

УДК 621.391.26

© В. В. Манойлов, И. В. Заруцкий

АЛГОРИТМЫ ОЦЕНКИ НАЛИЧИЯ ДВУХ И БОЛЕЕ ПИКОВ В ВЫБОРКЕ ДАННЫХ ФРАГМЕНТА МАСС-СПЕКТРА ДЛЯ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ОБРАБОТКИ МАСС-СПЕКТРОВ В УСЛОВИЯХ НЕДОСТАТОЧНОГО РАЗРЕШЕНИЯ

Рассматриваются алгоритмы, позволяющие сделать оценку наличия двух и более пиков в выборке данных фрагмента масс-спектра. Необходимость разработки подобных алгоритмов возникает в связи с тем, что время выполнения программ оценки параметров наложившихся пиков существенно больше времени оценки параметров одиночных пиков, и есть смысл оперативно в режиме on-line выдать информацию о параметрах одиночных пиков, а затем в режиме off-line произвести разделение пиков и выдать окончательный результат.

Кл. сл.: методы обработки сигналов, масс-спектрометрия, оценка параметров пиков в мультиплетах

ПОСТАНОВКА ЗАДАЧИ

В настоящее время существует ряд различных методов оценки параметров отдельных масс-спектрометрических пиков, образующих мультиплет [1, 2, 3, 4]. Все методы разделения мультиплетов требуют гораздо больше машинного времени на обработку, чем оценка параметров в одиночном пике. Рассматриваются три типа алгоритмов, позволяющих ответить на вопрос, содержится ли во фрагменте масс-спектра одиночный пик или в нем два или более пика.

Первый тип алгоритмов основан на вычислении моментов третьего и четвертого порядков, позволяющих оценить эксцесс и асимметрию исходного сигнала.

Второй тип алгоритмов основан на вычислении сверток исходных данных с функциями, представляющими собой вторую и четвертую производную от функций, описывающих форму спектральных пиков. Если у обеих сверток в одной и той же точке имеются максимумы, то это значит, что на экспериментальной кривой в этом месте расположен пик, а в данной точке — его вершина. Использование подобных алгоритмов позволяет оценить наличие "мультиплетности" по количеству максимумов в сигналах после сверток, а также по отношению максимумов и минимумов в сигналах.

Третий тип алгоритмов основан на использовании разложения исходного сигнала на отдельные составляющие в приспособленном к обрабатываемым сигналам ортогональном базисе. В качестве базиса используется усредненный масс-спектр, в котором отсутствуют примеси — стандартный об-

разец. Для такого масс-спектра в результате его разложения в приспособленном ортогональном базисе мы получаем единственную линию (начальную компоненту), если выборка данных содержит одиночный пик. Критериями для принятия решения о наличии мультиплетности является сравнение норм и квадратов максимальных значений начальных компонент в векторах, полученных в результате выполнения преобразования соответственно стандартного образца и тестируемых сигналов.

Указанные алгоритмы могут быть применены в программном обеспечении как в отдельности, так и в комплексе алгоритмов нескольких типов в разных комбинациях.

ВЫЧИСЛЕНИЕ МОМЕНТОВ ВЫСШИХ ПОРЯДКОВ

Использование моментов третьего и четвертого порядков исходного сигнала позволяет производить оценку наложения пиков. Наложение пиков приводит к асимметричности кривой, описывающей форму спектрального сигнала относительно оси, проходящей через центр тяжести. В первом приближении центральный момент третьего порядка при наложении пиков для симметричной функции, описывающей форму пика, должен быть отличен от нуля.

Вычисление параметра центра:

$$t_0 = \frac{\sum_i I_i t_i}{\sum_i I_i},$$

I_i — значения отсчетов на пике, t_0 — положение центра, t_i — значения независимой переменной.

Центральный момент третьего порядка можно выразить через начальные моменты первых трех порядков:

$$\mu_3 = m_3 - 3m_1m_2 + 2m_1^3,$$

m_3 , m_2 , m_1 — начальные моменты соответственно 3, 2 и 1-го порядков.

Оценку "наложенности" пиков можно сделать, вычислив безразмерную величину

$$k = \frac{\mu_3}{\sqrt{\mu_2^3}}. \quad (1)$$

При $k > \text{const}_3$ пики считаются наложившимися. Вычисление оценок начальных моментов производится по приближенным формулам:

$$\mu_k = \frac{1}{N} \sum_{i=1}^N (t_i - \mu_1)^k, \quad (2)$$

$k=1,2,3,4$ — порядки моментов.

Кроме характеристики асимметрии для оценки наложения пиков можно также использовать коэффициент эксцесса, который вычисляется с помощью центрального момента четвертого порядка по следующей формуле:

$$\gamma = \frac{\mu_4}{\mu_2^2} - 3. \quad (3)$$

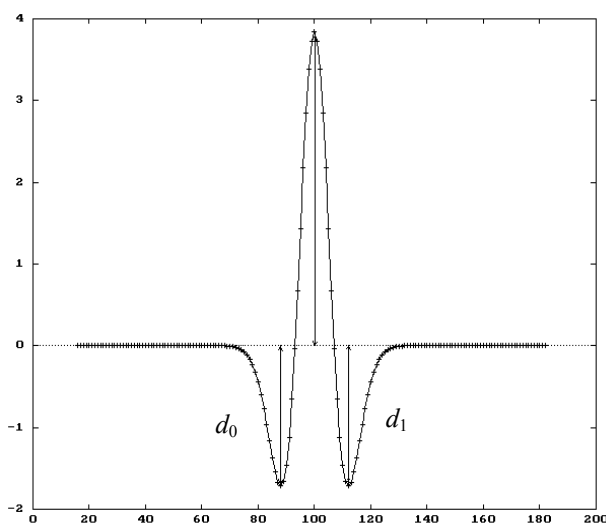


Рис. 1. Признак наложения линий по свертке со второй производной базовой (аппаратной) функции (одиночный пик)

При $\gamma > \text{const}_4$ пики считаются наложившимися.

ВЫЧИСЛЕНИЕ СВЕРТОК С ПРОИЗВОДНЫМИ

$$F_n(t_1) = \int_{-\infty}^{\infty} S_n \left(\frac{t}{\mu_0} - t_1 \right) f \left(\frac{t}{\mu} \right) dt, \quad (4)$$

где $S_n \left(\frac{t}{\mu_0} \right) = \frac{\partial^{(n)}}{\partial \left(\frac{t}{\mu_0} \right)} f \left(\frac{t}{\mu_0} \right)$ — весовые функции,

являющиеся производными от функции $f \left(\frac{t}{\mu_0} \right)$, описывающей стандартную форму пика единичной амплитуды и полушириной μ_0 .

Для свертки со второй производной при $\frac{A^2}{\sigma^2} > 10$, где A — амплитуда пика, а σ^2 — дисперсия шумов для составного пика (совокупность наложившихся пиков), должно выполняться неравенство: пик составной, если

$$R = |d_0 - d_1| > \text{const}_1, \quad (5)$$

где d_0 и d_1 — минимумы второй производной свертки (см. рис. 1). Для свертки с четвертой производной признак наложения (рис. 2):

$$R1 = |d_0 - d_1| > \text{const}_2. \quad (6)$$

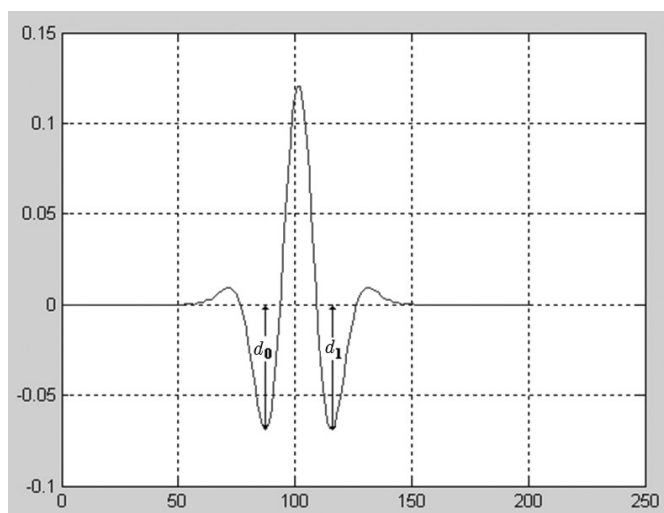


Рис. 2. Признак наложения линий по свертке с четвертой производной базовой (аппаратной) функции (одиночный пик)

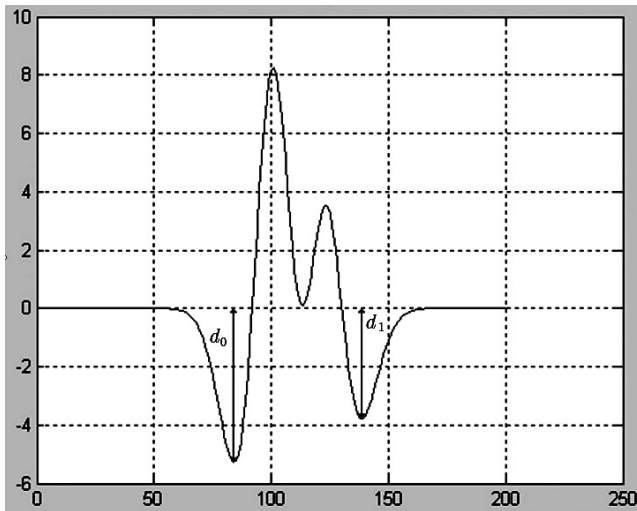


Рис. 3. Признаки наложения по свертке со второй производной (наложившиеся пики): минимумы различны, максимумов больше одного

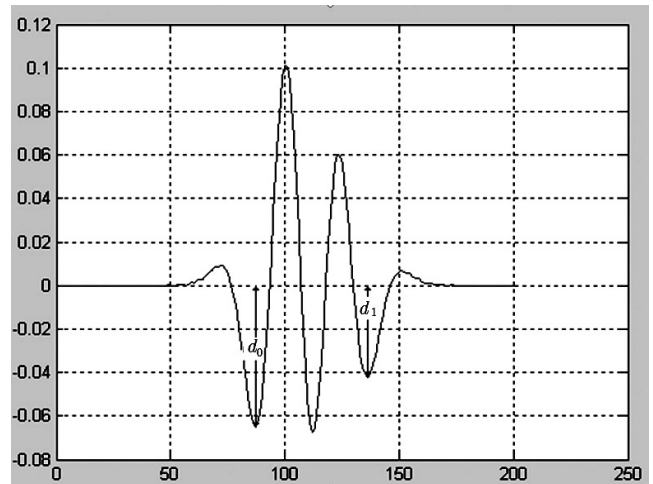


Рис. 4. Признаки наложения по свертке с четвертой производной (наложившиеся пики): минимумы различны, максимумов больше одного

Кроме того, пик составной, если в свертке со второй производной больше одного максимума, а для свертки с четвертой производной — больше трех максимумов.

На рис. 1–4 показаны значения минимума d_0 и минимума d_1 для гауссовых пиков.

Значения const_1 и const_2 вычисляются из функций, описывающих форму пика, и отношения сигнала к шуму.

ИСПОЛЬЗОВАНИЕ ОРТОГОНАЛЬНОГО ПРЕОБРАЗОВАНИЯ В СИСТЕМЕ БАЗИСНЫХ ФУНКЦИЙ (СБФ)

Сущность алгоритмов заключается в выполнении следующей последовательности операций [5, 6, 7].

1. Выполняем синтез целевого оператора H ортогонального преобразования $Y = HX$ в приспособленном базисе. В качестве исходных данных для такого синтеза берется вектор цифровых значений функции, описывающей форму пика, например гауссова, или форму пика по экспериментальным данным.

2. Организуем скользящее окно шириной N , где N — длина вектора исходного сигнала, по которому был найден оператор преобразования H .

3. Для каждого вектора исходного сигнала в окне (F_1) находим вектор преобразованного сигнала в приспособленном базисе $Y = H F_1^T$. После спектрального преобразования в приспособленном базисе спектр сигнала в ортогональном базисе, построенном на основе формы самого сигнала, представляет собой одиночную линию, аналогично то-

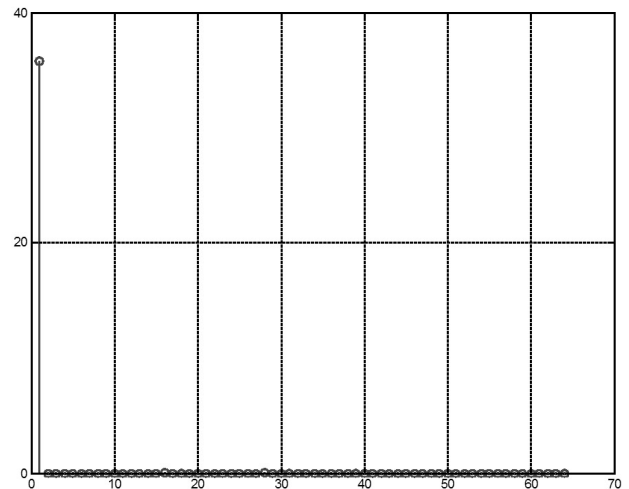


Рис. 5. Спектр одиночного пика в приспособленном базисе для $N = 64$

му как спектр функции $y = \cos(\omega t)$ в традиционном гармоническом базисе представляет одиночную линию, например $Y(1)$ на рис. 5.

4. Сравниваем с порогом величину $Y(1)$ для преобразованных данных каждого текущего окна. При превышении величины $Y(1)$ порога считаем, что пик обнаружен.

На рис. 5 представлен спектр сигнала одиночного пика с единичной амплитудой. На рис. 6 представлены два наложившихся пика масс-спектра. На рис. 7 представлен спектр сигналов двух наложившихся пиков масс-спектра, представленных на рис. 6.

Преимуществом данного алгоритма является

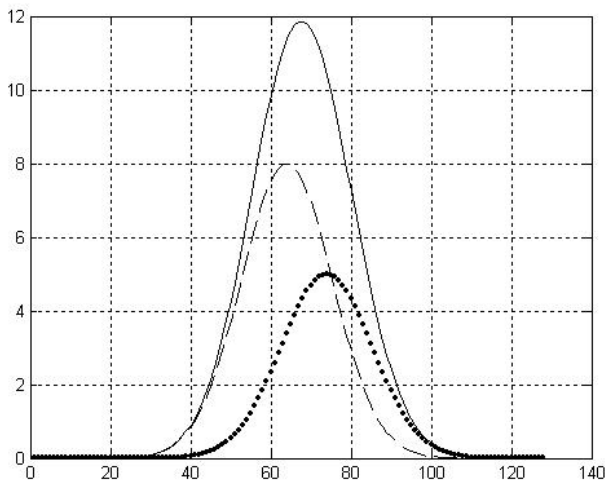


Рис. 6. Наложившиеся пики гауссовой формы. Сплошная линия — суммарный сигнал, пунктирная линия — сигнал F_1 , линия из точек — F_2

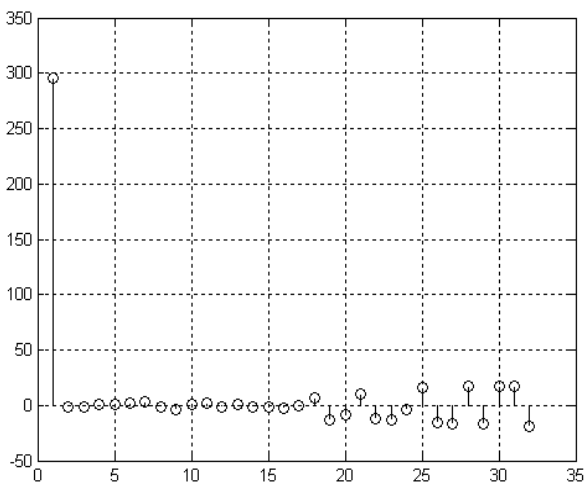


Рис. 7. Сигнал суммы $F_1 + F_2$ пиков гауссовой формы после спектрального преобразования в приспособленном базисе

его способность автоматически принимать решение о том, что обнаруженный пик состоит из двух или более наложившихся пиков. Для выполнения этой операции необходимо оценить близость полученного в результате преобразований вектора $\mathbf{Y}$ с аналогичным вектором $\mathbf{Y}_{\text{эт}}$, который характеризует одиночный пик. Например, в случае наложения двух пиков, представленных на рис. 6, преобразованный сигнал содержит дополнительные составляющие, кроме основной линии $Y(1)$, как показано на рис. 7. Оценка близости двух векторов производится путем сравнения с порогом разности норм и (или) сравнения с порогом разности квад-

ратов основных линий или по другим критериям, которые подробно описаны в [5, 6, 7]. Например, вектор $\mathbf{Y}$, полученный в результате преобразования исходного сигнала в приспособленном базисе, принадлежит классу одиночных пиков, если

$$P_1 = \|\mathbf{Y} - \mathbf{Y}_{\text{эт}}\| \leq \delta, \quad (7)$$

$$P_2 = |Y(1)^2 - Y_{\text{эт}}(1)^2| \leq \varepsilon. \quad (8)$$

В противном случае $\mathbf{Y}$ принадлежит классу сигналов "наложившиеся пики". Параметры порогов δ и ε выбираются в процессе обучения с использованием информации о функции, описывающей форму пика.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ АЛГОРИТМОВ НА МОДЕЛЯХ СИГНАЛОВ, СОДЕРЖАЩИХ НАЛОЖИВШИЕСЯ ПИКИ

Для определения возможностей указанных алгоритмов с помощью вычислительного эксперимента определялись параметры наложения пиков, которые вычислялись по приведенным выше формулам для сигналов в дублетах с соотношением амплитуд отдельных пиков 5 : 1 в отсутствие шума и с шумом (соотношение сигнал/шум S/N) при различных степенях наложения M для пиков гауссовой формы. Степень наложения M определяется как отношение полуширины пика к расстоянию между пиками в дублете; $M = \infty$ соответствует одиночному пику, $S/N = \infty$ соответствует сигналам без шума. Данные экспериментов без шума представлены в табл. 1. В табл. 2 представлены данные при различных степенях наложения при $S/N = 100$.

Анализ результатов, приведенных в табл. 1 и 2, показывает, что при отсутствии шума все рассмотренные алгоритмы позволяют оценивать наличие мультиплетности в обрабатываемых данных. Вычисление момента 3-го порядка и нормы вектора данных, полученного в результате преобразования с помощью ортогональных СБФ, позволяет оценить степень наложения как при наличии, так и при отсутствии шума.

ЗАКЛЮЧЕНИЕ

1. Рассмотренные алгоритмы каждого типа могут быть применены в программном обеспечении как по отдельности, так и в комплексе алгоритмов, состоящем из нескольких типов в разных комбинациях.

2. Алгоритм, основанный на вычислении момента 3-го порядка, и алгоритм вычисления нормы вектора, полученного в результате разложения

Табл. 1. Параметры наложения, вычисленные по предлагаемым формулам, при различной близости одиночных пиков в отсутствие шума

№ п/п	Алгоритм	Формула	Параметры наложения при различной близости (M) пиков при отсутствии шума			
			$M = \infty$	$M = 20\,000$	$M = 2000$	$M = 500$
1	Момент 3	1	3.5E-015	1.04E-004	0.0010	0.0042
2	Момент 4	3	1.7523	1.7523	1.7523	1.7523
3	Свертка с 2-й производной	5	3.5E-015	7.78E-005	1.18E-004	0.0034
4	Свертка с 4-й производной	6	0	2.65E-007	1.15E-006	3.23E-006
5	СБФ	7	0.0041	0.0043	0.0170	0.0670
6	СБФ	8	6.5E-004	7.1E-004	7.91E-004	1.69E-004

Табл. 2. Параметры наложения, вычисленные по предлагаемым формулам, в присутствии шума

№ п/п	Алгоритм	Формула	Средние значения параметров наложения при различной близости пиков (M) при $S/N = 100$			
			$M = 20\,000$	$M = 2000$	$M = 500$	$M = 10$
1	Момент 3	1	0.001	0.0014	0.0040	0.21
2	Момент 4	3	1.7523	1.7523	1.7523	1.7636
3	Свертка с 2-й производной	5	0.0430	0.0234	0.0261	0.07
4	Свертка с 4-й производной	6	1E-005	1E-005	1.3E-005	1.7E-005
5	СБФ	7	0.2	0.22	0.24	3.31
6	СБФ	8	0.02	0.04	0.06	0.08

с помощью ортогональных СБФ, позволяют оценить не только сам факт наличия "мультиплетности", но и степень наложения как при наличии, так и при отсутствии шума.

СПИСОК ЛИТЕРАТУРЫ

1. Сирвидас С.И., Заруцкий И.В., Ларионов А.М., Манойлов В.В. Использование метода сверток с производными базовых функций для обнаружения и разделения пиков в экспериментальных данных // DSPA'99. Доклады, т. 1. С. 105.
2. Сирвидас С.И., Заруцкий И.В., Ларионов А.М., Манойлов В.В. Обнаружение, разделение и оценка параметров масс-спектрометрических пиков методом свертки экспериментальных данных с производными гауссовых функций // Научное приборостроение. 1999. Т. 9, № 4. С. 84.
3. Разников В.В., Разникова М.О. Информационно-аналитическая масс-спектрометрия. М.: Наука, 1991. 248 с.
4. Гуревич А.Л., Могильницкий А.М., Русинов Л.А. и др. Автоматизация обработки масс-спектрометрической информации. М.: Энергия, 1978. 182 с.
5. Солодовников А.И., Сиваковский А.М. Основы теории и методы спектральной обработки ин-

- формации. Учебн. пособие. Л.: Изд-во Ленингр. ун-та, 1986. 272 с.
6. Абденби А. Методы классификации сигналов на основе приспособляемых спектральных ортогональных преобразований. Дис. ... канд. техн. наук. СПбГЭТУ, 2005. 120 с.
7. Манойлов В.В. Развитие методов обработки информации в масс-спектрометрии для изотопного и элементного анализа. Дис. ... д-ра техн. наук. СПб.: ИАП РАН, 2008. 263 с.

*Институт аналитического приборостроения РАН,
Санкт-Петербург*

Материал поступил в редакцию 3.07.2009.

ALGORITHMS FOR ESTIMATE OF PRESENCE TWO OR MORE PEAKS IN THE SAMPLE OF FRAGMENT MASS-SPECTRA FOR SOFTWARE DATA PROCESSING IN DEFICIENT RESOLUTION

V. V. Manoylov, I. V. Zarutsky

Institute for Analytical Instrumentation RAS, Saint-Petersburg

The results of investigation of algorithms for separate mass spectra data with singlet peaks and mass spectra data with multiplets are discussed. The necessity of these algorithms is due to the fact that time of data processing of multiplet peaks is much longer than that of single peaks, and it's reasonable to present on-line information on single peak parameters and then separate peaks off-line and give the final result.

Keywords: methods for data processing, mass-spectrometry, evaluation of peak parameters in multiplets