

УДК 621.38

© В. В. Манойлов, Л. В. Новиков

ОЦЕНКА ПАРАМЕТРОВ МАСС-СПЕКТРОМЕТРИЧЕСКОГО ПИКА В ДУБЛЕТЕ

Предлагается метод оценки положения и интенсивности пика в дублете, скрытого главным пиком, интенсивность которого превышает искомым в 10 и более раз в присутствии шума. Положение главного пика при этом задано точно, его форма известна с некоторой точностью, а амплитуда неизвестна. Метод позволяет оценить параметры скрытого пика при отношении сигнал/шум от десяти и выше и расстоянии между пиками от половины ширины на половине высоты и более.

Кл. сл.: обработка сигналов, метод независимых компонент, собственные векторы матриц, оценка параметров масс-спектрометрических пиков

ВВЕДЕНИЕ

В ряде практических приложений элементного, изотопного и других методов анализа требуется разделить компоненты дублета — смеси двух наложившихся пиков, амплитуда одного из которых в десятки раз превосходит второй. Оценка параметров таких пиков: в мультиплетах, и в частности в дублетах, основана на математических методах восстановления сигналов, искаженных аппаратной функцией. Среди этих методов, применяемых в приборостроении, наиболее известными являются метод Тихонова [1], метод максимального правдоподобия [2, 3], метод деконволюции [4, 5], метод на основе фильтров Винера [6], метод сингулярных разложений (SVD) [7], метод наименьших квадратов [8], метод сверток [9, 10], метод независимых компонент (ICA) [11] и др. С целью подавления шума во всех этих операциях в последнее время стали использовать избыточное вейвлет-преобразование с последующей пороговой обработкой [12]. В настоящей работе на основе идеи метода ICA предлагается метод разделения пиков известной формы в присутствии шумов.

ТЕОРИЯ

Предположим, что для разделения пиков в дублете требуется по суммарному пику и положению главного пика найти положение и амплитуду второго (маленького) пика. Как правило, форма главного и искомого пиков одна и та же, но они несколько смещены друг относительно друга. Амплитуда главного пика неизвестна. Модель такого сигнала можно представить в виде

$$\mathbf{x}_1 = C\mathbf{s}_1 + \mathbf{s}_2, \quad (1)$$

где C — произвольная положительная константа, определяющая амплитуду главного пика, $C \geq 10$, и может быть значительно больше 10; $\mathbf{s}_1, \mathbf{s}_2$ — векторы-столбцы отсчетов размерности N главного (\mathbf{s}_1) и искомого (\mathbf{s}_2) пиков. Эти пики положительны, а суммарный вектор \mathbf{x}_1 образует вектор исходных данных, как правило искаженный аддитивным шумом. Пусть известна (с некоторой точностью) форма первого пика \mathbf{s}_1 , что при известном его положении позволяет сформировать вектор-столбец модели $\mathbf{s}_0 \cong \mathbf{s}_1$. Требуется, используя эту модель и исходный сигнал \mathbf{x}_1 , выделить из дублета компоненту \mathbf{s}_2 . Решение, т. е. оценку $\hat{\mathbf{s}}_2$, будем искать в виде

$$\hat{\mathbf{s}}_2 = a\mathbf{x}_1 + b\mathbf{s}_0. \quad (2)$$

С учетом выражения (2), полагая что это — точное решение, образуем скалярные произведения исходных данных:

$$\begin{aligned} (\mathbf{s}_2, \mathbf{x}_1) &= a(\mathbf{x}_1^T \cdot \mathbf{x}_1) + b(\mathbf{s}_0^T \cdot \mathbf{x}_1) = k_{11}a + k_{12}b, \\ (\mathbf{s}_2, \mathbf{s}_0) &= a(\mathbf{x}_1^T \cdot \mathbf{s}_0) + b(\mathbf{s}_0^T \cdot \mathbf{s}_0) = k_{21}a + k_{22}b, \end{aligned} \quad (3)$$

которые удобно представить в матричном виде как:

$$\begin{pmatrix} \mathbf{s}_2^T \cdot \mathbf{x}_1 \\ \mathbf{s}_2^T \cdot \mathbf{s}_0 \end{pmatrix} = \begin{pmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \mathbf{K} \begin{pmatrix} a \\ b \end{pmatrix}. \quad (4)$$

Построим двухстолбцовую матрицу $\mathbf{x} = (\mathbf{x}_1 \ \mathbf{s}_0)$.

Используя выражение (3) легко убедиться в том, что элементы матрицы \mathbf{K} образованы произведением $\mathbf{x}^T \cdot \mathbf{x}$. Введем далее вектор $\mathbf{e}_1 = \begin{pmatrix} e_{11} \\ e_{21} \end{pmatrix}$ и по-

ложим, что $\begin{pmatrix} a \\ b \end{pmatrix} = d \begin{pmatrix} e_{11} \\ e_{21} \end{pmatrix}$, где d — некоторая константа. Имея ввиду (1) и (2), положим $a = de_{11} = 1$.

Следовательно, $b = de_{21} = e_{21}/e_{11}$ и $\begin{pmatrix} a \\ b \end{pmatrix} = \frac{1}{e_{11}} \begin{pmatrix} e_{11} \\ e_{21} \end{pmatrix}$.

Используя последнее выражение, покажем, что вектор \mathbf{e}_1 является собственным вектором матрицы \mathbf{K} , для которого должно выполняться равенство $\mathbf{K} \mathbf{e}_1 = \lambda_1 \mathbf{e}_1$, где λ_1 — собственное число. Действительно, продолжая равенства (4), имеем

$$\mathbf{K} \begin{pmatrix} a \\ b \end{pmatrix} = \frac{1}{e_{11}} \mathbf{K} \begin{pmatrix} e_{11} \\ e_{21} \end{pmatrix} = \frac{1}{e_{11}} \lambda_1 \begin{pmatrix} e_{11} \\ e_{21} \end{pmatrix}. \quad (5)$$

Следовательно, для оценки $\hat{\mathbf{s}}_2$ из (2) будем иметь

$$\hat{\mathbf{s}}_2 = \mathbf{x}_1 + \frac{e_{21}}{e_{11}} \mathbf{s}_0. \quad (6)$$

Чтобы определить ошибку оценки $\hat{\mathbf{s}}_2$ и, следовательно, справедливость предположения (2), получим выражение для отношения e_{21}/e_{11} . Из выражения (5) с учетом (4) будем иметь

$$k_{21} e_{11} + (k_{22} - \lambda_1) e_{21} = 0$$

и

$$\frac{e_{21}}{e_{11}} = -\frac{k_{21}}{k_{22} - \lambda_1} = -\frac{C(\mathbf{s}_1^T \cdot \mathbf{s}_0) + (\mathbf{s}_2^T \cdot \mathbf{s}_0)}{(\mathbf{s}_2^T \cdot \mathbf{s}_0) - \lambda_1}. \quad (7)$$

Вычисление λ_1 показывает, что $\lambda_1 < (\mathbf{s}_0^T \cdot \mathbf{s}_0)$, поэтому

$$\frac{e_{21}}{e_{11}} = -C_0, \text{ причем } C_0 > C.$$

Подставляя выражение (1) в (6), с учетом последнего вывода можно убедиться в том, что компенсация главного пика при вычислении оценки $\hat{\mathbf{s}}_2$ происходит с избытком, т. е. в $\hat{\mathbf{s}}_2$ на месте главного пика будет всегда присутствовать отрицательная составляющая.

С целью устранения этой компоненты вводится итерационная процедура

$$\hat{\mathbf{s}}_2^j = \mathbf{x}_1 + A_0^{j-1} \mathbf{s}_0, \quad j = 1, 2, \dots, j_{\max}.$$

Величина A_0^j изменяется от $A_0^0 = \frac{e_{21}}{e_{11}}$ с шагом $-\Delta A$ до $A_0^{j_{\max}}$, при которой норма отрицательной составляющей оценки $\hat{\mathbf{s}}_{2-}$ достигает минимума, а именно $\min_{A_0} \{ \|\hat{\mathbf{s}}_{2-}\| / \|\hat{\mathbf{s}}_2\| \}$. Величину шага удобно выбирать из условия $\Delta A_0^j = \alpha \|\hat{\mathbf{s}}_{2-}^{j-1}\| / \|\mathbf{x}_1\|$ при $\alpha = 0.1, \dots, 0.5$, чтобы обеспечить плавное достижение искомого минимума независимо от амплитуды главного пика.

ОПИСАНИЕ АЛГОРИТМА

Алгоритм вычисления оценки $\hat{\mathbf{s}}_2$ включает выполнение следующих операций:

- формирование вектора-столбца исходных данных \mathbf{x}_1 ;
- формирование вектора-столбца модели \mathbf{s}_0 ;
- формирование матрицы $\mathbf{x} = (\mathbf{x}_1 \ \mathbf{s}_0)$;
- вычисление матрицы

$$\mathbf{K} = (\mathbf{x}^T \cdot \mathbf{x}) = \begin{pmatrix} \mathbf{x}_1^T \cdot \mathbf{x}_1 & \mathbf{s}_0^T \cdot \mathbf{x}_1 \\ \mathbf{x}_1^T \cdot \mathbf{s}_0 & \mathbf{s}_0^T \cdot \mathbf{s}_0 \end{pmatrix};$$

- вычисление собственных векторов $\mathbf{e} = (\mathbf{e}_1, \mathbf{e}_2) = \begin{pmatrix} e_{11} & e_{12} \\ e_{21} & e_{22} \end{pmatrix}$ (и собственных значений $\lambda = (\lambda_1, \lambda_2)$) матрицы \mathbf{K} ;

- начало итерационного цикла $j = 1$ от $A_0^0 = e_{21}/e_{11}$ до $A_0^{j_{\max}}$;

- вычисление оценки искомого вектора $\hat{\mathbf{s}}_2^j = \mathbf{x}_1 + A_0^{j-1} \mathbf{s}_0$;

- подавление шума в оценке $\hat{\mathbf{s}}_2^j$;

- вычисление на каждой итерации функционала $\varepsilon(j) = \|\hat{\mathbf{s}}_{2-}^j\| / \|\hat{\mathbf{s}}_2^j\|$;

- вычисление минимума ε по нескольким итерациям. В качестве условия минимума принимается выполнение неравенства

$$(\varepsilon(j) - \varepsilon(j-2)) / 2 < P,$$

где P — порог, величина которого определяется уровнем шума (например, $P = \sigma / N + 5 \cdot 10^{-2}$), σ — среднеквадратические отклонения (СКО) шума;

- если условие минимума не выполняется — переход к п. (д), иначе — к п. (л);

- оценка положения и амплитуды пика $\hat{\mathbf{s}}_2$ на итерации $j-2$.

МОДЕЛИРОВАНИЕ

Проверка качества разделения компонент дублета по предлагаемому алгоритму выполнена в среде МАТЛАБ на двух моделях исходного сигнала $x_1 = Cs_1 + s_2$, где векторы s_1 и s_2 построены

а) по формуле асимметричной гауссианы (задний фронт в 2 раза круче переднего);

б) в виде пика с плоской вершиной.

Модель пиков асимметричной гауссианы строилась по формуле:

$$s_1(t) = A_1 \exp\left\{-\frac{(t-t_1)^2}{2\mu_k^2}\right\}, \quad (8)$$

где A_1 — амплитуда, t_1 — положение, μ_k — среднеквадратическая ширина пика, причем $\mu_k = \mu_1$ при $t < t_1$, $\mu_k = \mu$ при $t > t_1$ и $\mu_1 > \mu$.

Из отсчетов функции $s_1(t)$ формируется вектор s_1 , причем компонента этого вектора $s_{1,i} = s_1(i\Delta t)$, где интервал Δt выбирается из условия 50–60 отсчетов на пик. По формуле (2) формируется также вектор s_2 с параметрами A_2, t_2, μ_{2k} и строится модель исходных данных $x_1 = Cs_1 + s_2$ с добавлением белого шума с дисперсией σ^2 при отношении сигнал/шум, определяемом как A_2/σ .

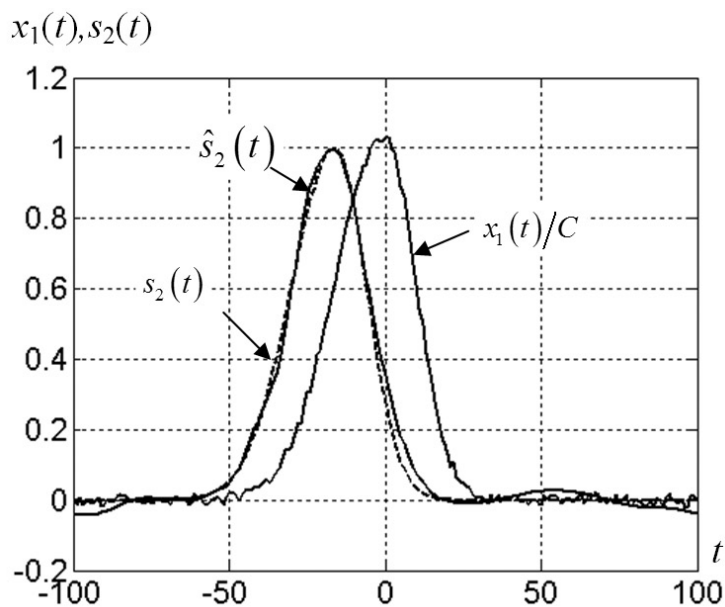


Рис. 1. Бигауссова модель дублета.
 $x_1(t)$ — исходные данные; $s_2(t)$ — искомый пик (штриховая линия); $\hat{s}_2(t)$ — оценка пика (сплошная линия)

На рис. 1 показан исходный сигнал $x_1(t)$ на сетке $t = -100\Delta t \div 100\Delta t$ при $C=10, \mu=10, A_1=1, t_1=0, \mu_1=\sqrt{2}\mu; A_2=1, t_2=15, \mu_2=\sqrt{2}\mu$ и отношении сигнал/шум 10. Из рисунка видно, что главный пик полностью скрывает искомый пик $s_2(t)$. Это обстоятельство, а также наличие шума усложняют решение задачи разделения пиков.

Построим модель главного пика дублета $s_0(t)$, используя формулу (8) с параметрами $A_0=1, t_0=0, \mu_0=\sqrt{2}\mu$, сформируем матрицу x и вычислим матрицу K . Вычислим далее собственные векторы (и собственные значения) этой матрицы, используя команду *eig* и выполним операции в соответствии с пп. (д), ..., (л) предлагаемого алгоритма.

Результат выделения искомого пика из дублета для приведенных выше исходных данных показан на рис. 1. Число итераций при этом оказалось равным 16. Итерационный процесс, который характеризуется зависимостью $\varepsilon(j)$, проиллюстрирован на рис. 2.

По результатам моделирования вычислялись среднеквадратические отклонения (СКО) и смещения оценок параметров искомого пика по M реализациям шума ($M=20$):

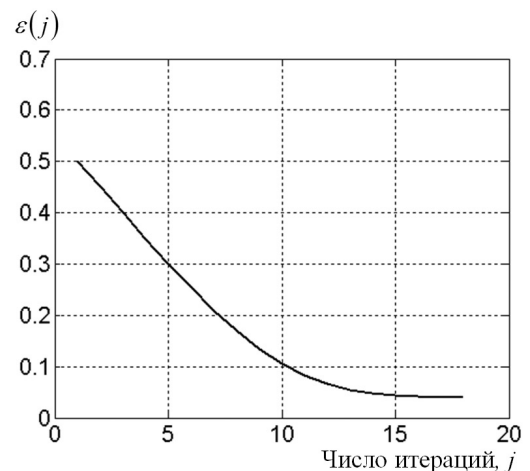


Рис. 2. Кривая сходимости итерационного процесса $\varepsilon(j) = \|\hat{s}_2^j\| / \|\hat{s}_2^j\|$ для бигауссовой модели при отношении сигнал/шум 10

СКО положения

$$\left\{ \sum_{m=1}^M (\hat{t}_{2,m} - \bar{t}_2)^2 / (M-1) \right\}^{1/2};$$

смещения положения $\bar{t}_2 - t_2$,

где

$$\bar{t}_2 = \frac{1}{M} \sum_{m=1}^M \hat{t}_{2,m};$$

СКО амплитуды

$$\frac{1}{A_2} \left\{ \sum_{m=1}^M (\hat{A}_{2,m} - \bar{A}_2)^2 / (M-1) \right\}^{1/2};$$

смещения амплитуды $\bar{A}_2 - A_2$,

где

$$\bar{A}_2 = \frac{1}{M} \sum_{m=1}^M \hat{A}_{2,m};$$

СКО восстановления формы

$$\frac{1}{M} \sum_{m=1}^M \|\hat{s}_2 - s_2\|_m,$$

где норма вычисляется при каждой реализации шума.

Алгоритм был апробирован при отношении сигнал/шум от 10 и более, при $C=10$ и 100. Результаты сведены в табл. 1.

Модели пиков с плоской вершиной строятся из двух пиков гауссовой формы с $\mu = 4\Delta t$, расстояние между которыми такое, что образуется плоская вершина. Ширина таких пиков выбиралась равной $30\Delta t$. На рис. 3 показан исходный сигнал $x_1(t)$ из пиков с плоской вершиной на сетке $t = -100\Delta t \div 100\Delta t$ при $C=10$; $A_1=1$, $t_1=0$, $A_2=1$, $t_2=10\Delta t$, ширине пиков $30\Delta t$. Фронты пиков — гауссовой формы при $\mu = 4\Delta t$ и отношении сигнал/шум 10. Из рисунка видно, что главный пик полностью скрывает искомым пик $s_2(t)$. Применение предлагаемой вычислительной процедуры позволило решить задачу оценки параметров скрытого в дублете пика $s_2(t)$, численные результаты которой приведены в табл. 2. Результат выделения искомого пика из дублета для приведенных выше исходных данных проиллюстрирован на рис. 3.

Табл. 1. Погрешность оценки параметров пика бигауссовой формы в дублете при $\mu = 10\Delta t$, $A_1=1$, $t_1=0$, $\mu_1 = \sqrt{2}\mu$; $A_2=1$, $t_2=15\Delta t$, $\mu_2 = \sqrt{2}\mu$

Превышение главного пика	Отношение сигнал/шум	Погрешность оценки положения Δt		Погрешность оценки амплитуды в %		СКО восстановления
		СКО	Смещение	СКО	Смещение	
C=10	10	1.4	-1.65	7.2	-0.5	0.13
	50	0.65	-1.8	3.2	-1	0.05
	Без шума	0	-2	0	-1	0.004
C=100	10	1.7	-1.4	9	0	0.15
	50	0.65	-1.8	3.2	-1	0.05
	Без шума	0	-2	0	-1	0.004

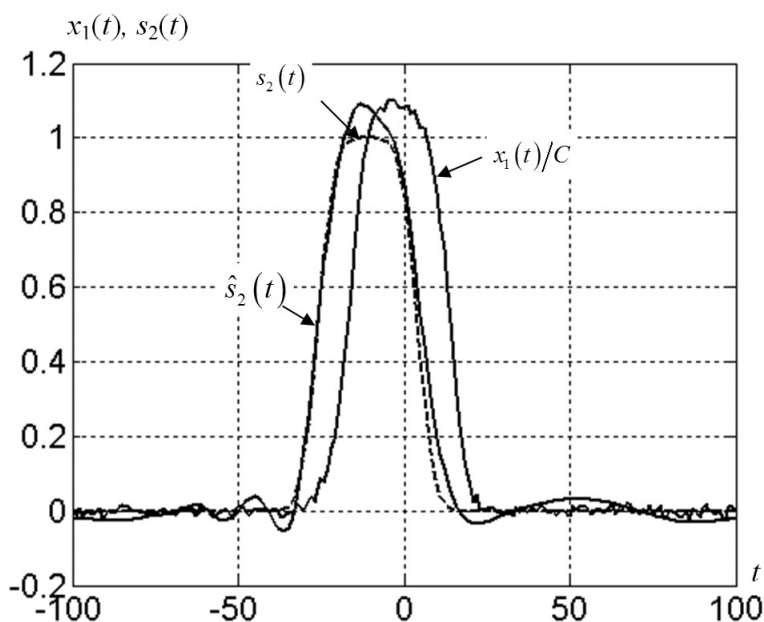


Рис. 3. Модель дублета с плоской вершиной: $x_1(t)$ — исходные данные; $s_2(t)$ — искомый пик (штриховая линия); $\hat{s}_2(t)$ — оценка пика (сплошная линия)

Табл. 2. Погрешность оценки параметров пика плоской формы в дублете при $A_1 = 1$, $t_1 = 0$, $A_2 = 1$, $t_2 = 10\Delta t$.
Ширина пиков — $30\Delta t$, фронты пиков — гауссовой формы при $\mu = 4\Delta t$

Превышение главного пика	Отношение сигнал/шум	Погрешность оценки положения Δt		Погрешность оценки амплитуды в %		СКО восстанав- ления
		СКО	Смещение	СКО	Смещение	
C=10	10	2.1	0.55	6	6	0.118
	50	1	0	3.2	1.2	0.035
	Без шума	0	0	0	-1	0.015
C=100	10	2.05	0.6	7	7	0.13
	50	1.1	0	3.5	1.2	0.04
	Без шума	0	0	0	1.2	0.02

При моделировании изменялись положения главного и искомого пиков от $5\Delta t$ до $35\Delta t$ для бигаусса и от $10\Delta t$ до $35\Delta t$ для пика с плоской вершиной, т. е. от полного наложения до практически полного разделения. При максимальном разделении главного и исходного пиков для обеих моделей погрешности оценки положения и отклонения формы практически не изменились. Погрешность оценки амплитуды при этом уменьшилась на порядок.

ЗАКЛЮЧЕНИЕ

Рассмотренный метод может быть, в частности, полезен при оценке чистоты материалов, анализируемых на масс-спектрометре. Метод позволяет сделать вывод о том, существует или нет загрязнение исследуемого объекта, а если существует, то на какой массе находится сторонний объект, вызвавший загрязнение. В случае отсутствия загрязнения на месте искомого пика наблюдается только шум.

Метод оценки параметров одиночных пиков в дублетах представляет собой нелинейный итерационный процесс. Предлагаемый алгоритм быстро сходится, и он не требует аналитического описания формы пика: она может быть любой, в том числе построенной на основе экспериментальных данных.

Авторы благодарят Тимченко Н.А. за активное участие в обсуждении результатов работы.

СПИСОК ЛИТЕРАТУРЫ

1. Тихонов А.Н., Арсенин В.Я. Методы решения некорректных задач. Наука, 1986. 288 с.
2. Разников В.В., Разникова М.О. Информационно-аналитическая масс-спектрометрия. М.: Наука, 1991. 248 с.
3. Gelfgat V.I., Kosarev E.L., Podolyak E.R. Program for signal recovery from noisy using the maximum likelihood principle // Comput. Phys. Commun. 1993. V. 74. P. 335.
4. Погуляй А.В., Васильев Ю.В., Туймедов Г.М., Мазунов В.А. Повышение разрешения спектрометра методами деконволюции // Химия и компьютерное моделирование. Бутлеровские сообщения. 1999. № 2. Приложение: Материалы VI Всероссийской конференции "Структура и динамика молекулярных систем".
5. Погуляй А.В., Аблазимов Р.Р., Туймедов Г.М., Мазунов В.А. Сравнительный анализ методов деконволюции // Химия и компьютерное моделирование. Бутлеровские сообщения. 2001. № 6. Приложение: Материалы VIII Всероссийской конференции "Структура и динамика молекулярных систем".
6. Бардин Б.В., Белов В.Д., Новиков Л.В., Чижов Ю.В. Использование оптимального фильтра Винера для деконволюции электронных спектров // Научное приборостроение. 1999. Т. 9, № 1. С. 53–59.
7. Shrager R.I. Chemical transition measured by spectra resolved using singular value decomposition // Chemom. Intell. Lab. Syst. 1986. V. 1, N 1. P. 59–70.
8. Bardin B.V., Belov V.D., Mamro N.V. et al. A Computerized control and data handling system of a multifunctional electron spectrometer and software for processing complex electron spectra // Instruments and Experimental Techniques. 1999. V. 42, N 2. P. 205–211.
9. 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 шумов функций // Научное приборостроение. 1999. Т. 9. № 2. С. 71–75.
10. Манойлов В.В., Заруцкий И.В. Возможности алгоритма сверток с производными для оценки параметров масс-спектров, содержащих наложившиеся пики // Научное приборостроение. 2009. Т. 19, № 4. С. 103–109.
11. Hувärinen Aapo. Survey on independent component analysis // Neural Computing Surveys. 1999. N 2. P. 94–128. URL: (<http://www.icsi.berkeley.edu>).
12. Князева Т.Н., Новиков Л.В., Орешко Н.И. Удаление нестационарного шума из экспериментальных данных // Научное приборостроение. 2008. Т. 18, № 2. С. 61–65.

*Институт аналитического приборостроения РАН,
г. Санкт-Петербург*

Контакты: Манойлов Владимир Владимирович,
manoilov_vv@mail.ru

Материал поступил в редакцию 26.04.2012.

PARAMETER ESTIMATION FOR THE MASS SPECTROMETRIC PEAK IN THE DOUBLET

V. V. Manoilov, L. V. Novikov

Institute for Analytical Instrumentation of RAS, Saint-Petersburg

We propose a method of estimating the position and the intensity of the peak in the doublet, mostly hidden peak, whose intensity is greater than the required 10 or more times in the presence of noise. The position of the main peak at exactly this set, its shape is known with some accuracy, but the amplitude is unknown. The method allows to estimate the parameters of the hidden peak at a signal / noise ratio of ten and above, and the distance between the peaks of the half-width at half height, and more.

Keywords: signal processing, method of independent components, eigenvectors of matrices, estimation of parameters of mass spectrometric peaks