

УДК 535.338.1+519.6+519.853.6

О методе производных разделения большого числа перекрывающихся компонент при наличии шума

В. С. Сизиков, А. В. Лавров

Санкт-Петербургский университет ИТМО,

Санкт-Петербург 197101. E-mail: sizikov2000@mail.ru, lavrov@itmo.ru

Аннотация. Рассмотрена эффективность метода производных для разделения (сепарации) большого числа перекрывающихся спектральных линий-компонент в условиях зашумленности. Линии моделируются гауссианами или лоренцианами. Выведены аналитические формулы для производных 1-4 порядков, а также для амплитуд и полуширин линий. Найденные число и параметры линий уточняются путем минимизации функционала невязки с ограничениями на решение. Путем моделирования спектра из 7 компонент с использованием 1-4 производных показано, что количество и значения параметров линий хорошо восстанавливаются при отсутствии шумов и при использовании аналитических выражений для производных. При наличии шумов (даже слабых) производные определяются (численно) с большими погрешностями. В этом случае эффективным является использование сглаживающих сплайнов. Однако если использовать сплайн лишь один раз для сглаживания спектра и вычисления 1-й производной, то погрешности резко возрастают с ростом порядка производной от 2-го до 4-го. Поэтому в работе предложено сглаживать сплайнами перед расчетом каждой производной. В этом случае значения производных получаются мало отличающимися от производных при отсутствии помех, а число линий и их параметры определяются достаточно точно.

Ключевые слова: разделение (сепарация) перекрывающихся спектральных линий, метод производных, гауссианы и лоренцианы, число линий, параметры линий, сглаживающие сплайны.

On the derivative method for separating large number of overlapping components in the presence of noise

V. S. Sizikov, A. V. Lavrov

Saint-Petersburg ITMO University, Saint-Petersburg 197101.

Abstract. The efficiency of the derivatives method for separation of an increased number of overlapping spectral lines (components) is considered, in addition, under noisy conditions. The lines are modeled by Gaussians or Lorentzians. Analytical formulas are derived for derivatives of 1-4 orders, as well as for amplitudes and half-widths of lines. Then, the number and the parameters of the lines are refined by minimizing the discrepancy functional. By modeling a spectrum of 7 components using 1-4 derivatives, it is shown that the number and parameters of the lines are well restored in the noise absence and when using analytical expressions for derivatives. And in the presence of noise (even weak), the derivatives are determined (numerically) with large errors. In this case, the use of smoothing splines is very effective. However, if the spline is used only once to smooth the spectrum and calculate the 1st derivative, then the errors will increase sharply with increasing order of the derivative from 2 to 4. Therefore, in this paper, it is proposed to use a spline before calculating each

derivative. In this case, the derivatives are slightly different from the derivatives in the absence of noises, and the number of lines and their parameters are determined quite accurately.

Keywords: separation of overlapping spectral lines, derivative method, Gaussian and Lorentzian, number of lines, line parameters, smoothing splines.

MSC 2010: 41A15, 65D07, 65Z05

Введение

Одной из актуальных задач спектроскопии является задача о разделении (сепарации) непрерывных перекрывающихся спектральных линий (см. [12, 9, 10, 1, 7, 2] и др.). Необходимо определить количество линий N , входящих в суммарный измеренный спектр Z , и параметры линий-компонент z . Данная задача, строго говоря, имеет множество решений. Поэтому математически её решают обычно путем моделирования линий гауссианами или лоренцианами (у каждой линии по 3 искомым параметра, см. дальше). А количество линий N и параметры линий, как правило, определяют методом производных (методом дифференцирования), а именно, находят производные от суммарного спектра Z до 4-го порядка и по значениям производных оценивают N и параметры линий. Затем параметры линий уточняют путем минимизации функционала невязки [2, 3, 4, 5].

Многие численные результаты [9, 10, 7, 2, 3, 4, 5], демонстрируют хорошее восстановление спектров методом производных. Однако это имеет место лишь в отсутствии шумов и при использовании аналитических выражений для производных. При наличии же шумов (даже слабых) производные определяются (численно) с большими погрешностями [2, 4, 5] и особенно при большом (порядка 10) количестве линий N . В работах [2, 4, 5] показано, что весьма эффективным при наличии шумов является использование сглаживающих сплайнов (о сплайнах см. [6, с. 215–224]). Однако в этих работах сплайн использован лишь один раз, а именно, для сглаживания экспериментального зашумленного суммарного спектра \tilde{Z} и вычисления 1-й производной \tilde{Z}' , а для устойчивого вычисления производных более высоких порядков сглаживающие сплайны не используются. В данной работе предлагается использовать сглаживающий сплайн перед расчетом каждой производной. Кроме того, акцент в работе сделан на большом (порядка 10) количестве линий N , как, например, в работе [9]. При этом не используется дополнительная информация о структуре вещества и его спектре, как в работе [8], и не применяется метод Фурье само-деконволюции [12], так как он искусственно занижает ширины линий (до 5 раз), что уже отмечалось в работах [2, 3, 4, 5].

1. Постановка задачи

Пусть в спектре присутствуют N непрерывных линий с профилями интенсивности $z_j(\lambda)$, $j = \overline{1, N}$, где j – номер линии, а λ – длина волны. Считаем, что линии налагаются друг на друга в той или иной степени. Полагаем, что измеряется за-

шумленный суммарный спектр-профиль $\tilde{Z}(\lambda)$, равный сумме N линий-компонент:

$$\tilde{Z}(\lambda) = \sum_{j=1}^N z_j(\lambda) + \delta z, \quad a \leq \lambda \leq b, \quad (1.1)$$

где $[a, b]$ – пределы суммарного спектра, а δz – шум.

Ставится **задача**: на основе измеренного зашумленного суммарного спектра $\tilde{Z}(\lambda)$ определить количество N линий-компонент и их профили $z_j(\lambda)$. Такая задача в её общей постановке имеет множество решений, поскольку задана лишь одна функция $\tilde{Z}(\lambda)$, а нужно определить N функций $z_j(\lambda)$. Поэтому для корректности задачи допускается моделирование профилей линий некоторыми функциями (гауссианами, лоренцианами и т.д.) и использование дополнительной (априорной) информации о линиях (начальные приближения параметров линий, диапазоны изменения параметров и т.д.).

2. Моделирование линий гауссианами, их дифференцирование и оценка параметров линий по производным

Представим каждую линию $z_j(\lambda)$ гауссианом:

$$z_j(\lambda) = A_j \exp\left(-\frac{(\lambda - \bar{\lambda}_j)^2}{2\sigma_j^2}\right), \quad a \leq \lambda \leq b, \quad (2.1)$$

где A_j – амплитуда линии, $\bar{\lambda}_j$ – координата максимума, σ_j^2 – дисперсия, при этом $\tau_j = \sqrt{2 \ln 2} \sigma_j = 1.1773 \sigma_j$ – полуширина гауссиана по уровню 0.5 [4]. На рис. 1 представлен пример спектра из $N = 7$ спектральных линий в виде гауссиан.

В работе [11] впервые предложен метод производных для оценки параметров линий. В этой работе была использована производная лишь 1-го порядка, а в дальнейшем в работах [9, 10, 7, 2, 3, 4, 5] и др. использованы производные до 4-го порядка. Воспользуемся методом производных.

Производные порядков $l = 1, 2, 3, 4$ от $z_j(\lambda)$ по λ равны (ср. [7, 4])

$$z_j'(\lambda) = -\frac{A_j}{\sigma_j} \exp\left(-\frac{(\lambda - \bar{\lambda}_j)^2}{2\sigma_j^2}\right) \left(\frac{\lambda - \bar{\lambda}_j}{\sigma_j}\right), \quad (2.2)$$

$$z_j''(\lambda) = -\frac{A_j}{\sigma_j^2} \exp\left(-\frac{(\lambda - \bar{\lambda}_j)^2}{2\sigma_j^2}\right) \left(-\frac{(\lambda - \bar{\lambda}_j)^2}{\sigma_j^2} + 1\right), \quad (2.3)$$

$$z_j'''(\lambda) = \frac{A_j}{\sigma_j^3} \exp\left(-\frac{(\lambda - \bar{\lambda}_j)^2}{2\sigma_j^2}\right) \left(-\frac{(\lambda - \bar{\lambda}_j)^2}{\sigma_j^2} + 3\right) \left(\frac{\lambda - \bar{\lambda}_j}{\sigma_j}\right), \quad (2.4)$$

$$z_j^{IV}(\lambda) = \frac{A_j}{\sigma_j^4} \exp\left(-\frac{(\lambda - \bar{\lambda}_j)^2}{2\sigma_j^2}\right) \left(\frac{(\lambda - \bar{\lambda}_j)^4}{\sigma_j^4} - 6\frac{(\lambda - \bar{\lambda}_j)^2}{2\sigma_j^2} + 3\right). \quad (2.5)$$

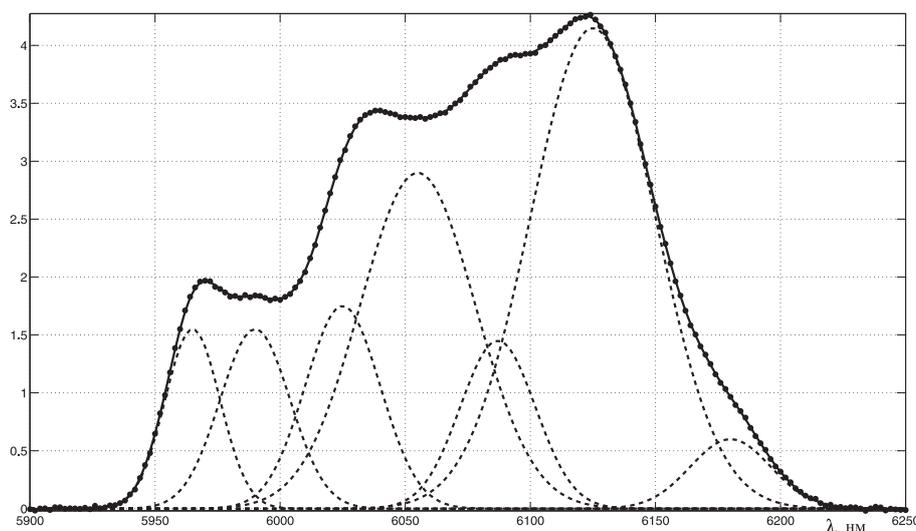


Рис. 1. Семь линий $z_j(\lambda)$ (пунктирные кривые), суммарный спектр $\tilde{Z}(\lambda)$ с шумом (точки) и сглаживающий сплайн (непрерывная кривая, проходящая гладко через точки)

Из условий экстремума и формул (2.2)–(2.5) имеем:

$$z'_j(\bar{\lambda}_j) = 0, \quad z''_j(\bar{\lambda}_j) = -\frac{A_j}{\sigma_j^2} < 0, \quad z'''_j(\bar{\lambda}_j) = 0, \quad z_j^{IV}(\bar{\lambda}_j) = 3 \frac{A_j}{\sigma_j^4} > 0. \quad (2.6)$$

Из соотношений (2.6) следуют важные **выводы**:

- 1) нулевые значения 1-й и 3-й производных при условии, что $z''_j(\bar{\lambda}_j) < 0$, $z_j^{IV}(\bar{\lambda}_j) > 0$, указывают на координату $\bar{\lambda}_j$ максимума j -й линии;
- 2) отрицательный минимум 2-й производной и положительный максимум 4-й производной также указывают на j -ю линию.

Более того, из выражений для $z''_j(\bar{\lambda}_j)$ и $z_j^{IV}(\bar{\lambda}_j)$ в (2.6) можно определить A_j и σ_j , решая систему двух уравнений:

$$\left. \begin{aligned} \frac{A_j}{\sigma_j^2} &= -z''_j(\bar{\lambda}_j), \\ \frac{A_j}{\sigma_j^4} &= \frac{1}{3} z_j^{IV}(\bar{\lambda}_j), \end{aligned} \right\}$$

откуда получаем:

$$A_j = 3 \frac{[z''_j(\bar{\lambda}_j)]^2}{z_j^{IV}(\bar{\lambda}_j)}, \quad \sigma_j = \sqrt{3 \frac{-z''_j(\bar{\lambda}_j)}{z_j^{IV}(\bar{\lambda}_j)}}, \quad j = \overline{1, N}. \quad (2.7)$$

Однако реальный (не модельный) измеренный спектр обычно зашумлен и его численное (не по аналитическим формулам (2.2)–(2.5)) дифференцирование выполняется со значительными погрешностями (см. [2, 4]). Можно отфильтровать

шумы, например, путем аппроксимации спектра сглаживающим сплайном [6, с. 223] и дифференцировать сплайн, как это сделано в работах [2, 4]. Однако, как сказано выше, в этих работах сплайн использован лишь для устойчивого вычисления 1-й производной. В данной работе сплайн использован для устойчивого вычисления каждой производной (см. далее).

3. Численная иллюстрация

Рассмотрим пример типа примера из статьи [9, Fig. 3, левый верхний рис.]. В суммарный спектр (1.1) входят $N = 7$ линий-компонент в виде гауссиан (2.1).

Параметры линий:

амплитуды (в условных единицах) $A_j = 1.55, 1.55, 1.75, 2.9, 1.45, 4.15, 0.6$;
 центральные длины волн $\bar{\lambda}_j$, нм = 5965, 5990, 6025, 6055, 6087, 6125, 6180;
 СКО σ_j , нм = 11, 14, 15, 23, 15, 25, 16.

Параметры примера:

$\lambda_{\min} = a = 5900$ нм, $\lambda_{\max} = b = 6250$ нм, шаг дискретизации $h = 2$ нм, число узлов дискретизации $n = (b - a)/h + 1 = 176$.

На рис. 1 представлены 7 линий-компонент в виде гауссиан (2.1) (пунктир).

На рис. 2 представлены производные порядков $l = 1, 2, 3, 4$ от $Z(\lambda)$ по λ в отсутствие шума, вычисленные численно (не по аналитическим формулам (2.2)–(2.5)) с помощью m-функции `diff.m` системы MatLab.

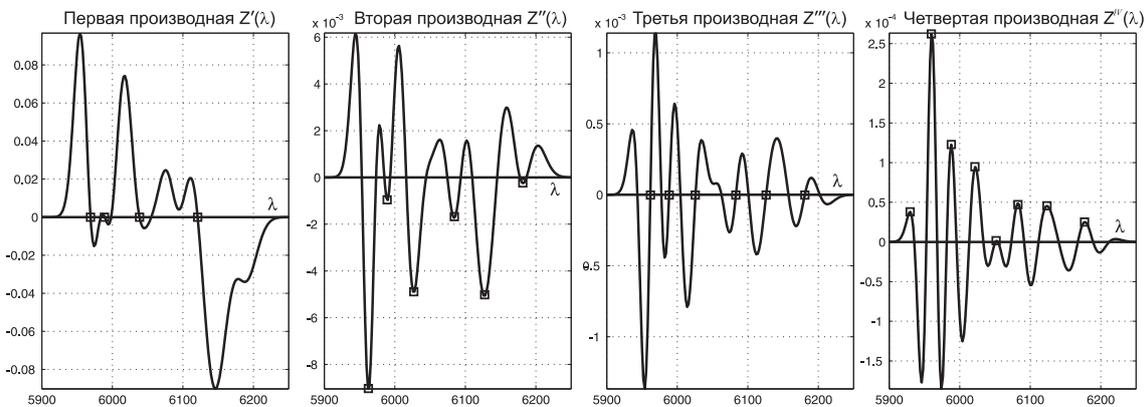


Рис. 2. Производные 1–4 порядков без шума (λ в нм)

Для определения количества линий N по производным использованы следующие критерии (согласно выше приведенным выводам): отрицательный экстремум 2-й производной и положительный экстремум 4-й производной соответствуют максимуму j -й линии, а нулевое значение 1-й и 3-й производных (при условии: $z_j''(\bar{\lambda}_j) < 0$, $z_j^{IV}(\bar{\lambda}_j) > 0$) также указывают на j -ю линию.

На кривых рис. 2 квадратиками (\square) отмечены точки, удовлетворяющие сформулированным критериям. На основе этих критериев по рис. 2 определено: по 1-й производной критериям удовлетворяют 4 точки, т.е. $N = 4$; по 2-й производной

$N = 6$; по 3-й производной $N = 6$; по 4-й производной $N = 8$. Видим, что точность определения количества линий N по производным растет с повышением порядка производной l – неожиданный результат. Поэтому необходимо в методе производных доходить до $l = 4$, а не ограничиваться значением $l = 2$, как в ряде публикаций. В данном примере заключаем, что $N = 7$ (если проигнорировать нереальную точку при $\lambda \approx 5930$ на кривой $Z^{IV}(\lambda)$).

Далее, на суммарный спектр $Z(\lambda)$ наложен шум (точки на рис. 1). Через зашумленные значения $\tilde{Z}(\lambda)$ проведен сглаживающий кубический сплайн (непрерывная линия на рис. 1), по которому получена 1-я производная $\tilde{Z}'(\lambda)$ путем численного дифференцирования (рис. 3).

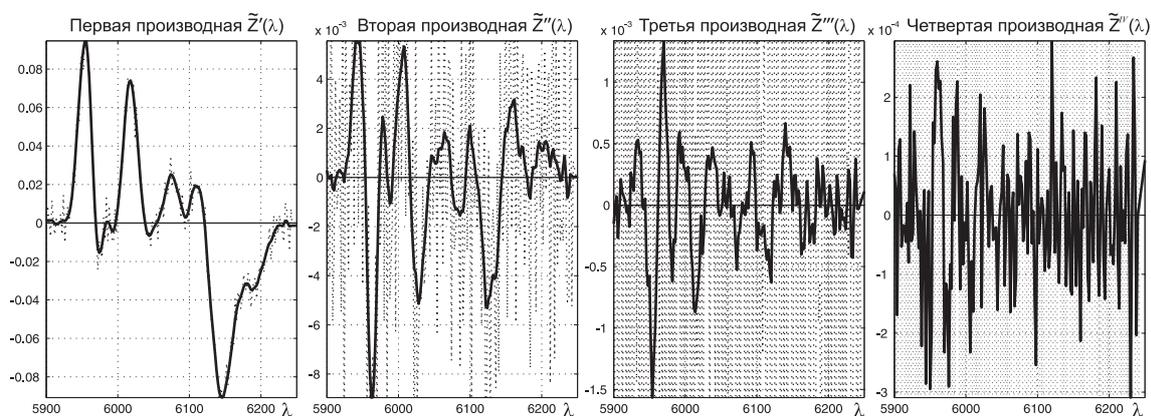


Рис. 3. Производные 1–4 порядков с шумом без сглаживания (пунктир) и с однократным применением сглаживающего сплайна (непрерывные линии)

В предыдущих публикациях [2, 3, 4, 5] для получения 2-й производной $\tilde{Z}''(\lambda)$ мы дифференцировали численно 1-ю производную $\tilde{Z}'(\lambda)$, а сплайн не использовался (как для получения 2-й, так и 3-й и 4-й производных). В данном примере такой подход дал большие искажения 2-й, 3-й и 4-й производных (рис. 3) в виде так называемой знакопеременной "пилой" большой амплитуды (ср. [6, с. 182]).

Чтобы повысить точность численного дифференцирования, мы в данном примере для вычисления производных каждый раз используем сплайн-сглаживание. Это значительно улучшило результаты, как видно из рис. 4, где сплайны 2-й, 3-й и 4-й производных практически такие же, как на рис. 2 без шума.

На рис. 2 точки, отмеченные квадратиками, дают нам значения $\bar{\lambda}_j$ – координат максимумов (центральных длин волн) линий. По ним можно получить значения 2-й и 4-й производных $Z_j''(\bar{\lambda}_j)$ и $Z_j^{IV}(\bar{\lambda}_j)$. Аналогично можно снять значения $\bar{\lambda}_j$, а также производных $\tilde{Z}_j''(\bar{\lambda}_j)$ и $\tilde{Z}_j^{IV}(\bar{\lambda}_j)$ с кривых на рис. 4. Далее, используя формулу (2.7) и заменяя в ней z на Z и \tilde{Z} , можно вычислить параметры линий A_j и σ_j .

Однако параметры линий $\bar{\lambda}_j$, A_j и σ_j были оценены по результатам дифференцирования не непосредственно линий z , а суммарного спектра Z или \tilde{Z} , поэтому необходимо уточнить значения параметров линий. Для этого в работах [2, 3, 4, 5] использована разработанная модификация метода координатного спуска

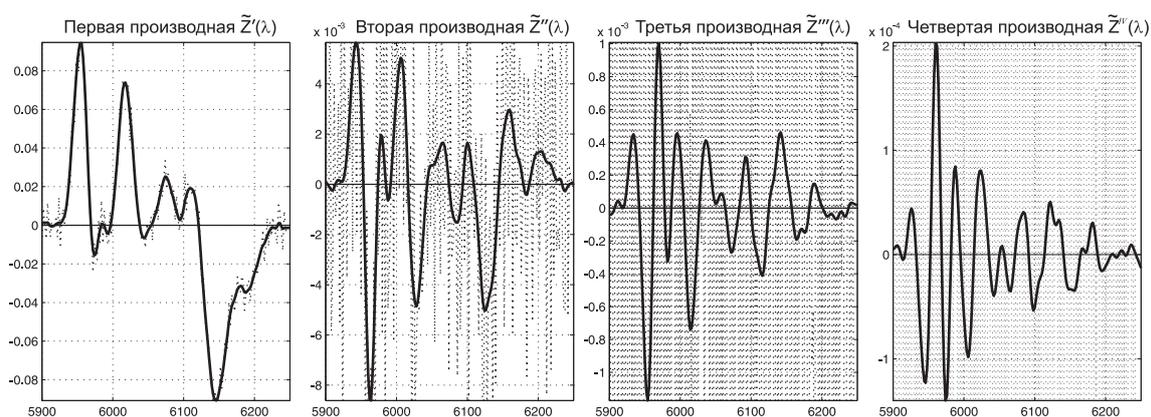


Рис. 4. Производные 1–4 порядков с шумом без сглаживания (пунктир) и с четырехкратным применением сглаживающего сплайна (непрерывные линии)

(КС) и (для сравнения) симплекс-метод Нелдера–Мида, сводящиеся к минимизации функционала невязки с ограничениями на решение. Однако в цели данной работы не входили оценка и уточнение параметров линий. Цель работы была показать большие возможности использования сглаживающих сплайнов при аппроксимации сильно неустойчивых, зашумленных функций (производных на рис. 3 и 4) и рис. 4 это подтвердил.

Заключение

Рассмотрен метод производных (дифференцирования спектра) для разделения большого числа перекрывающихся спектральных линий при наличии шума. Линии промоделированы гауссианами. Выведены аналитические формулы для производных 1–4 порядков от спектра по длине волны, а также для амплитуд и полуширин линий. Путем моделирования спектра из 7 линий с использованием 1–4 производных показано, что число и параметры линий восстанавливаются достаточно надежно в отсутствии шумов. При наличии же шумов производные определяются (численно) с большими погрешностями. В этом случае весьма эффективным является использование сглаживающих сплайнов. Однако если использовать сплайн лишь один раз для сглаживания спектра и вычисления 1-й производной, то погрешности будут резко возрастать с ростом порядка производной от 2-го до 4-го. В данной работе предложено использовать сплайн перед расчетом каждой производной. В этом случае значения производных почти не отличаются от производных в отсутствие помех, а число линий и их параметры определяются достаточно точно.

Список цитируемых источников

1. Михайленко В. И., Михальчук В. В. Методы разложения спектров с неразрешенной структурой // Журнал прикл. спектроскопии. – 1987. – Т. 46, № 4. – С. 535–543.
Mikhailenko V. I., Mikhal'chuk V. V. Method of expanding spectra with unresolved structure. J. Appl. Spectrosc. 46, no.4, 327-335 (1987).

2. *Сизиков В. С., Лавров А. В.* Исследование погрешностей некоторых методов разделения перекрывающихся спектральных линий в условиях воздействия помех // Научно-техн. вестник ИТМО. – 2017. – Т. 17, №5. – С. 879–889.
Sizikov V. S., Lavrov A. V. Study of errors of some methods for separating overlapped spectral lines under noise effect. Sci. Techn. J. ITMO 17, no.5, 879–889 (2017).
3. *Сизиков В. С., Лавров А. В.* Сепарация непрерывных линий, взаимно перекрывающихся и сглаженных аппаратной функцией // Опт. и спектроск. – 2017. – Т. 123, № 5. – С. 678–688.
Sizikov V. S., Lavrov A. V. Separation of continuous lines mutually overlapping and smoothed by the instrumental function. Opt. Spectrosc. 123, no.5, 682–691 (2017).
4. *Сизиков В. С., Лавров А. В.* Сравнение различных методов разделения непрерывных перекрывающихся спектральных линий. // Опт. и спектроск. – 2018. – Т. 124, №6. – С. 723–731.
Sizikov V. S., Lavrov A. V. A comparison of different methods of separation of continuous overlapping spectral lines // Opt. Spectrosc. 124, no. 6, 753–762 (2018).
5. *Сизиков В. С., Лавров А. В.* Современные устойчивые математические и программные методы восстановления искаженных спектров. Обзорная статья // Научно-техн. вестник ИТМО. – 2018. – Т. 18, № 6. – С. 911–931.
Sizikov V. S., Lavrov A. V. Modern stable mathematical and software-based methods for distorted spectra restoration. Review article. Sci. Techn. J. ITMO 18, no. 6, 911–931 (2018).
6. *Сизиков В. С.* Математические методы обработки результатов измерений. – СПб.: Политехника, 2001. – 240 с.
Sizikov V. S. Mathematical methods for processing the results of measurements. – St. Petersburg: Polytechnika Publ., 2001. – 240 p. (in Russian)
7. *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, № 2. – С. 71–75.
Sirvidas S. I., Zarutsky I. V., Larionov A. M., Manoilov V. V. The convolution of a signal with derivatives of gaussians as an approach suitable for detection, separation and estimation of mass spectrometer peaks. Sci. Instrum. 9, no. 2, 71–75 (1999).
8. *Усольцев Д. А., Ситникова В. Е., Носенко Т. Н., Олехнович Р. О., Успенская М. В.* Сравнение методик расчета вторичной структуры белков на основе деконволюции инфракрасных спектров // Научно-техн. вестник ИТМО. – 2019. – Т. 19, № 4. – С. 586–593.
Usoltsev D. A., Sitnikova V. E., Nosenko T. N., Olekhovich R. O., Uspenskaya M. V. Comparison of protein secondary structure calculation methods based on infrared spectra deconvolution. Sci. Techn. J. ITMO 19, no. 4, 586–593 (2019).
9. *Byler D. M., Susi H.* Examination of the secondary structure of proteins by deconvolved FTIR spectra. // Biopolymers. – 1986. – Vol. 25. – P. 469–487.
10. *Cameron D. G., Moffatt D. J.* A generalized approach to derivative spectroscopy // Appl. Spectrosc. – 1987. – Vol. 41, no. 4. – P. 539–544.

11. *Giese A. T., French C. S.* The analysis of overlapping spectral absorption bands by derivative spectrophotometry // *Appl. Spectrosc.* – 1955. – Vol. 9, no. 2. – P. 78–96.
12. *Kauppinen J. K., Moffatt D. J., Mantsch H. H., Cameron D. G.* Fourier self-deconvolution: a method for resolving intrinsically overlapped bands. // *Appl. Spectrosc.* – 1981. – Vol. 35, no. 3. – P. 271–276.

Получена 01.06.2019