

# Применение локальной регуляризации в подгонке контуров линий в молекулярных спектрах

ЩЕРБАКОВ А.П.

*Институт Оптики Атмосферы СО РАН*

Мицель А.А.

*ТУСУР, кафедра АСУ*

ПШЕНИЧНИКОВ А.М.

*ТУСУР, кафедра АСУ*

8 января 2004 г.

Results of the water vapor molecular spectrum processing are presented in this article. The results are presented as a peak list. The problem of forming the peak list is solved using the method of potential functions. This method is used for peak recognition in the absorption spectral curve. A comparison with the published results of R.A. Toth is made.

## 1. Введение

Молекулярные спектры являются уникальным источником информации о строении окружающих нас веществ и многих процессах, происходящих в природе. Получаемая спектроскопическая информация применяется в самых разнообразных областях науки и техники. Например, спектральные свойства водяного пара ответственны за парниковый эффект. Благодаря тому, что водяной пар (а так же другие парниковые газы —  $CO_2$ ,  $CH_4$ ) пропускает солнечное излучение в видимом диапазоне и поглощает большую часть энергии, переизлученной поверхностью Земли в инфракрасном (тепловом) диапазоне, на Земле сохраняется тепло и климат, пригодный для жизни. Следовательно, исследования спектров  $H_2O$ ,  $CO_2$ ,  $CH_4$  и других парниковых газов имеют важное значение для климатологии. Спектроскопическая информация об атмосферных газах используется при интерпретации результатов зондирования атмосферы Земли и результатов наблюдений в астрофизике. Из спектров молекул определяют фундаментальные характеристики молекул, их геометрию и энергетическую структуру, информацию о внутренних силах и процессах, что представляет интерес для теории строения атомов и молекул.

В настоящее время измерения проводятся на высокоточных и высокочувствительных Фурье— и лазерных спектрометрах при спектральном разрешении  $10^{-3}$ – $10^{-4}$   $cm^{-1}$  в широкой спектральной области: от микроволновой до видимой части спектра. Спектры содержат обычно несколько тысяч линий и представляют трудный объект для теоретического анализа, прежде всего в силу большого объема информации. Спектроскопическая информация о водяном паре и других атмосферных газах накапливается в специализированных банках и используется в дальнейшем для решения прикладных задач.

Анализ спектров обычно состоит из двух этапов. Первый из них — это математическая обработка спектральной кривой и подгонка параметров модельных контуров линий для определения центров, интенсивностей и полуширин линий. Результатом этого этапа является список линий (Peak list), содержащий значения этих спектральных характеристик. Второй этап — идентификация линий, то есть отнесение их к тем или иным квантовым переходам. Оба этапа являются достаточно трудоемкими и требуют, в зависимости от типа молекулы от нескольких недель до нескольких месяцев кропотливой работы высококвалифицированных специалистов. Как было замечено в [1], «время, необходимое для регистрации спектра, становится пренебрежимо малым, по сравнению со временем, затрачиваемым на его теоретическую обработку».

## 2. Анализ

### 2.1. Обнаружение линий в спектре

Задача поиска и определения линий в спектрах является типичной задачей теории распознавания образов, для решения которой использовался метод потенциальных функций [2–4]. Согласно этому методу, в зависимости от ширины линий, выбирается размер окна. При этом количество точек во входном векторе процедуры распознавания остается неизменным. Точки выбираются равномерно по ширине окна (рис. 1).

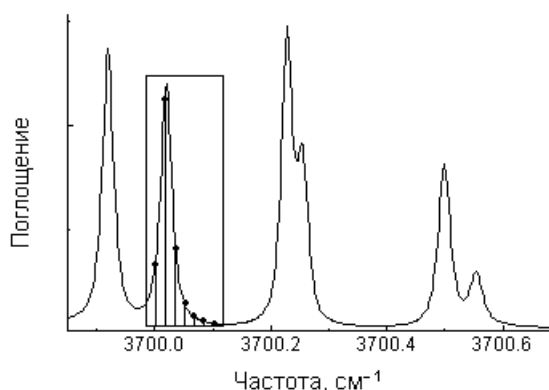


Рис. 1. Анализируемое окно (спектр H<sub>2</sub>O)

В качестве признакового пространства, согласно терминологии теории распознавания образов, был взят набор нормированных коэффициентов поглощения в точках спектра, находящихся в анализируемом окне. Пространство решений содержит 4 класса (см. рис. 2): «центр линии» (рис. 2в– 2е), «правый край» (рис. 2а), «левый край» (рис. 2б), «не край и не центр». В результате работы данной процедуры, изображение окна должно быть отнесено к одному из этих классов.

### 2.2. Подгонка модельных контуров

Для определения параметров линий использовалась подгонка модельных контуров методом наименьших квадратов. В качестве начального приближения используются координаты

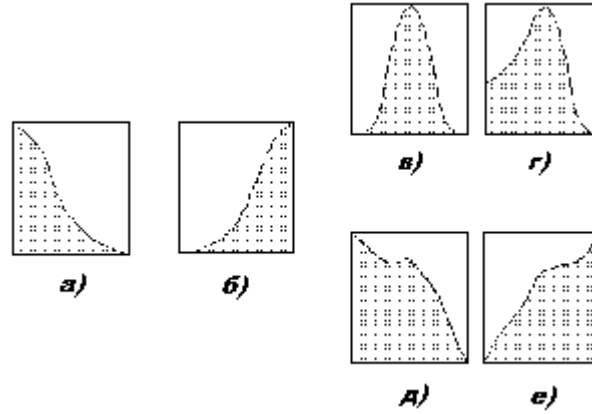


Рис. 2. Классы пространства решений данной задачи

ты центров линий, полученные поисковой процедурой, их амплитуды, и полуширины.

Надо заметить, что обратная задача вычисления параметров модельных контуров становится в ряде случаев некорректной, т.к. нарушается одно из условий устойчивости. В данной задаче малые возмущения во входных данных могут приводить к большим изменениям в решении. Неустойчивость может возникать, например, при обработке группы из 5–6 линий и более. В данной работе был использован так называемый «локальный» регуляризирующий алгоритм (АР). Суть локальной регуляризации состоит в том, что для каждой неизвестной компоненты искомого решения вводится свой параметр регуляризации, являющийся некоторой функцией, зависящей от отношения сигнал/шум либо от искомого решения. Введение локальной регуляризации позволяет повысить точность решения. Наибольшая эффективность локальной регуляризации достигается в тех случаях, когда компоненты решения существенно различаются друг от друга по величине.

### 2.3. Локальный АР

Рассмотрим функционал вида

$$J(\theta) = \sum_{i=0}^{N_{\text{exp}}} \frac{1}{\sigma_i^2} \{Y(x_i|\theta) - Y_i^{\text{exp}}\}^2 + \mu_0 \sum_{j=1}^{N_{\text{par}}} \left( \frac{\theta_j - \theta_j^*}{\theta_j^*} \right)^2 \quad (1)$$

где  $\theta$  – априорная оценка параметра  $\theta$ ,  $\mu_0$  – параметр регуляризации ( $\mu_0 > 0$ ),  $N_{\text{exp}}$  – число экспериментальных данных,  $N_{\text{par}}$  – число параметров модели,  $\sigma_i$  – ошибки измерений величин  $y_i$ .

После линейризации модели (1) относительно точки  $\theta$ , функционал примет вид:

$$J(\theta, \mu_0) = \sum_{i=1}^{N_{\text{exp}}} \frac{1}{\sigma_i^2} \left\{ z_i - \sum_{j=1}^{N_{\text{par}}} a_{ij} \Delta\theta_j \right\}^2 + \mu_0 \sum_{j=1}^{N_{\text{par}}} \left( \frac{\Delta\theta_j}{\theta_j^*} \right)^2, \quad (2)$$

где  $z_i = Y(x_i|\theta) - Y_i^{\text{exp}}$ ,  $a_{ij} = \left. \frac{\partial Y(x_i, \theta)}{\partial \theta_j} \right|_{\theta=\theta^*}$ ,  $\Delta\theta_j = \theta_j - \theta_j^*$ .

Параметр регуляризации  $\mu_0$  зависит от шума измерения. При стремлении шума измерения к нулю, параметр  $\mu_0$  также должен стремиться к нулю. В этом случае обеспечивается сходимость регуляризованного решения  $\theta$  к псевдорешению  $\theta^+$ , являющемуся точкой минимума функционала  $J(\theta, \mu_0)$ .

Недостаток глобальной регуляризации проявляется в том, что найденные параметры  $\theta_j$  могут иметь существенно различную погрешность, в зависимости от их значимости в модели. Для устранения этого недостатка используется локальная регуляризация [5], согласно которой, вместо (2) используется функционал вида

$$J(\theta, \mu) = \sum_{i=1}^{N \text{ exp}} \frac{1}{\sigma_i^2} \left\{ z_i - \sum_{j=1}^{N \text{ par}} a_{ij} \Delta \theta_j \right\}^2 + \sum_{j=1}^{N \text{ par}} \mu_j^2 \left( \frac{\Delta \theta_j}{\theta_j^*} \right)^2. \quad (3)$$

Введем функционал  $\Gamma[\mu, \theta] = \sum_{j=1}^{N \text{ par}} (\mu_0 - \mu_j)^2$ , и составим новый сглаживающий функционал

$$\Phi[\mu, \theta] = J[\mu, \theta] + \gamma \Gamma[\mu]. \quad (4)$$

Недостатком обычной локальной регуляризации (например, [5]) является отсутствие учета информативности выборки  $\{XY\}^{\text{exp}}$  для оценки той или иной из компонент вектора  $\theta$ . Для учета этой информации введем функционал

$$\Gamma[\mu, \theta] = \sum_{j=1}^M (\mu_0 \eta(\varepsilon_j) - \mu_j)^2, \quad (5)$$

где  $\varepsilon_j$  — оценка восстановления компонент решения  $\theta_j$ ,  $\eta(x)$  — монотонно возрастающая и ограниченная функция, такая что

$$\begin{aligned} \eta(0) &= 0, \\ \eta(+\infty) &= 1. \end{aligned} \quad (6)$$

Функцию  $\eta(x)$  можно взять, например,  $\arctg(x)$ . Минимизация функционала (4), с учетом (5) приведет к следующей системе уравнений:

$$\sum_{j=1}^{N \text{ par}} \tilde{a}_{kj} \Delta \theta_j + \mu_k^2 \frac{\Delta \theta_k}{(\theta_k^*)^2} = \tilde{z}_k, \quad k = \overline{1, N \text{ par}}, \quad (7)$$

$$\mu_k = \frac{\gamma \mu_0 \cdot \eta(\varepsilon_k)}{\left( \frac{\Delta \theta_k}{\theta_k^*} \right)^2 + \gamma}, \quad k = \overline{1, N \text{ par}} \quad (8)$$

где  $\tilde{a}_{kj} = \sum_{i=1}^{N \text{ exp}} \frac{a_{ij} a_{ik}}{\sigma_i^2}$ ,  $\tilde{z}_k = \sum_{i=1}^{N \text{ exp}} \frac{z_i a_{ik}}{\sigma_i^2}$ , и  $\theta_k^*$  — начальное приближение.

Приведенный алгоритм обеспечивает сходимость регуляризованного решения к псевдорешению  $\theta^+$ . Сходимость следует из неравенства  $\mu_k \leq \mu_0$  (обеспечиваемого соотношениями (6) и (8) и из сходимости глобального регуляризованного решения задачи (2) к псевдорешению  $\theta^+$ , при  $\mu_0 \rightarrow 0$ .

На первом шаге делается попытка итерации с глобальной регуляризацией, для получения оценок  $\varepsilon_j$  восстановления компонент параметра  $\theta_j$ . Затем, применяя (8), вычисляются  $\mu_k$ , которые затем подставляются в (7). После этого снова используется (8), и итерационный процесс повторяется до стабилизации значений  $\theta_k$  и  $\mu_k$ .

### 3. Результаты

В рамках данной работы система ASA [6], была применена для анализа Фурье-спектра поглощения  $H_2O$  в области 1800–2500  $\text{см}^{-1}$ , зарегистрированного К.Кепплером [7] при следующих условиях: длина оптического пути 288.5 м, давление газа 25.81 mbar, температура 297К, спектральное разрешение  $10^{-3} \text{ см}^{-1}$ . Ранее этот спектр обрабатывался С.Н.Михайленко, результаты представлены в [7]. В этот интервал попадают переходы на первое возбужденное колебательное состояние молекулы воды. Эта полоса является наиболее сильной полосой в КВ спектре водяного пара и ее спектроскопические характеристики явились предметом многих исследований (см., например, [7–9]).

Подгонка параметров проводилась с использованием контура Фойгта:

$$f_F(\nu) = S \frac{a}{\pi} \int_{-\infty}^{\infty} \frac{\exp(-x^2) \cdot dx}{a^2 + [w - x]^2}, \quad (9)$$

где

$$a = \sqrt{\ln 2} \gamma_L / \gamma_d; \quad w = \sqrt{\ln 2} |\nu - \nu_0| / \gamma_d, \quad (10)$$

$\nu_0$  — частота центра линии поглощения,  $\gamma_L$  и  $\gamma_d$  — значения ее лоренцевской и доплеровской полуширин.

На первом этапе работы система ASA обнаружила 1728 линий с порядками пропускания в пиках от  $10^{-6}$  до  $10^{-3}$ . На втором этапе проводилась автоматическое определение параметров линий. Поскольку спектр содержал значительную часть достаточно интенсивных линий со значениями пропускания около нуля («насыщенные» линии), из этих линий были успешно подогнаны 1398 (без регуляризирующего алгоритма), т.е. 80.9%. Обработка оставшейся части спектра проводилась с использованием алгоритма локальной регуляризации.

На рис. 3 приведены типичные результаты подгонки для изолированной линии с центром 1989.931  $\text{см}^{-1}$ .

В верхней части рисунка изображены две кривые, одна из которых получена экспериментально, а вторая — подогнанная по методу наименьших квадратов к контуру Фойгта. Невязка на последней итерации процедуры подгонки для этой линии была равна  $2.7 \cdot 10^{-9}$ . В нижней части рисунка показана разность измеренного и вычисленного контуров данной линии. Необходимо отметить, что автоматическая подгонка дает результаты хорошо совпадающие с данными, полученными в [8]. Отклонение между рассчитанными и измеренными коэффициентами поглощения не превышает 1%.

На табл. 1 представлены результаты обработки спектра водяного пара. В двух последних столбцах для сравнения даны величины центров линий и их интенсивности, соответствующие переходу (010)–(000) изотопа воды  $H_2^{16}O$ , полученные Р.А. Тосом и опубликованные в [8]. Системой ASA положения центров линий были вычислены с дисперсией 0.001  $\text{см}^{-1}$ , что составляет  $5 \cdot 10^{-5}\%$  от их величин. Дисперсии интенсивностей спектральных линий в среднем составляют 10%.

Результаты сравнения показывают, что система ASA вполне удовлетворительно обрабатывает реальные спектры в автоматическом режиме, что позволяет значительно уменьшить время обработки спектров.

Авторы выражают благодарность Быкову А.Д. за консультации и ценные замечания, Михайленко С.Н. за предоставленный спектр водяного пара и полезное обсуждение результатов работы.

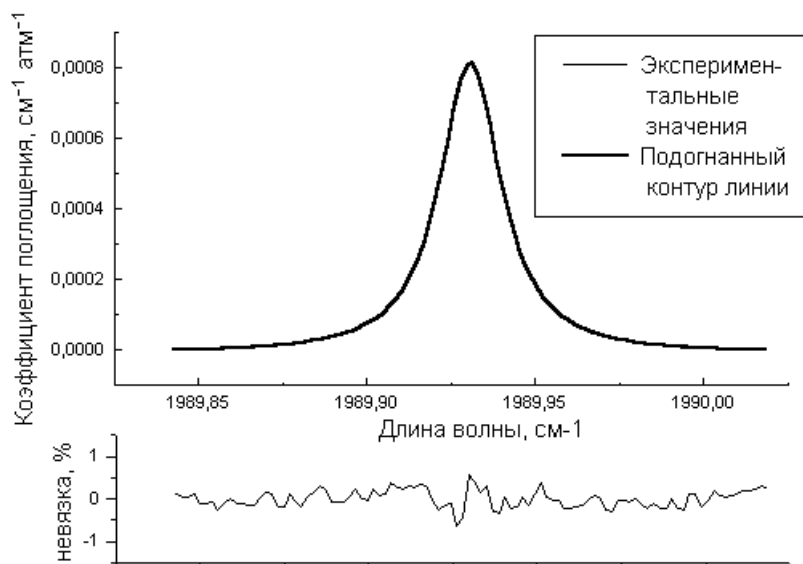


Рис. 3. Типичный результат подгонки, линия водяного пара с центром  $\nu_0 = 1989.931 \text{ см}^{-1}$

Таблица 1. Фрагмент списка линий  $H_2O$  в области  $1977.6\text{--}1987.3 \text{ см}^{-1}$ , сравнение с Тосом [8]

Центр линии $\text{см}^{-1}$	Ошибка подгонки центра	Интенсивность $\text{см}^{-2}\text{атм}^{-1}$	Полуширина, $\text{см}^{-1}$	Центр линии*, $\text{см}^{-1}$	Интенсивность*, $\text{см}^{-2}\text{атм}^{-1}$
1981,3264	1,436E-04	1,703890E-04	9,221433E-03	1981,32806	1,91E-4
1982,1951	1,346E-04	4,453651E-05	6,008514E-03	1982,19772	4,45E-5
1983,0275	5,527E-05	9,324617E-05	9,720858E-03	1983,02876	1,01E-4
1986,6559	7,206E-06	3,164760E-05	5,254588E-03	1986,65831	3,23E-5
1987,3391	1,217E-04	7,431256E-03	1,352246E-02	1987,34013	7,07E-3

## Список литературы

- [1] JENSEN P. The Giessen Program for processing spectral data from BRUKER IFS 120 HR interferometer using an IBM compatible personal computer. // Justus Leigig University Giessen. January 1990.
- [2] АЙЗЕРМАН М.А., БРАВЕРМАН Э.И., РОЗОНОЭР Л.И. Метод потенциальных функций в задачах обучения машин., М.: Наука, 1970
- [3] PSHENICHNIKOV A.M., SHCHERBAKOV A.P. Computer-aided system for automatic peak searching and contour fitting in molecular spectra // SPIE Proceedings., 2000. Vol. 4143, p. 59–67.
- [4] ВYКOV A.D., NAUMENKO O.V., PSHENICHNIKOV A.M., SINITSA L.N., SHCHERBAKOV A.P. An Expert System for Identification of Lines in Vibrational-Rotational Spectra // Optics and Spectroscopy. Vol. 94, No. 4, 2003. p. 528–538.
- [5] ВОСКОВОЙНИКОВ Ю.Е., МУХИНА И.Н. Локальный регуляризирующий алгоритм восстановления контрастных сигналов и изображений // Автометрия., 2000., № 3., С. 45–53.
- [6] ПШЕНИЧНИКОВ А.М. Автоматизированная система получения банка линий из молекулярных спектров поглощения атмосферных газов (ASA) // Информационные системы: Труды постоянно действующей научно-технической школы-семинара студентов, аспирантов и молодых специалистов УИИформационные системы мониторинга окружающей средыФ (сентябре-октябре 2002 г.), Томск: Томский гос. ун-т систем управления и радиоэлектрон., 2002., С. 49–55.
- [7] МИХАЙЛЕНКО S.N., TYUTEREV V.L.G, KEPPLER K.A., WINNEWISSER B.P., WINNEWISSER M., MELLAN G., KLEE S., K. NARAHARI RAO. The 22 band of water: Analysis of new FTS measurements and high-Ka transitions and energy levels // J. Mol. Spectrosc. 1997., V. 184., p. 330–349.
- [8] TOTN R.A. Air-and N2-Droadening Parameters of Water Vapor: 600 to 2271  $cm^{-1}$  // J. of Molec.Spectrosc. Vol. 201., 2000., P. 218–243.
- [9] CLAVEAU C., HENRY A., HURTMANS D., VALENTIN A. Narrowing and broadening parameters of H2O lines perturbed by He, Ne, Ar, Kr and nitrogen in the spectral range 1850-2140  $cm^{-1}$  // J. of Quantitative Spectrosc. & Radiative Transfer., Vol 1194., 2000.