УДК 535.853.1

ВОССТАНОВЛЕНИЕ ОПТИЧЕСКИХ СПЕКТРОВ И ДРУГИХ НЕОТРИЦАТЕЛЬНЫХ ФУНКЦИЙ ПО МЕТОДУ СТАТИСТИЧЕСКОЙ РЕГУЛЯРИЗАЦИИ

В. Ф. Турчин и *Л. С. Туровцева*

В работе описан метод статистической регуляризации для восстановления оптических спектров с учетом конечной разрешающей силы прибора и решения других задач такого рода. В отличие от прежних вариантов метода статистической регуляризации здесь используется дополнительная априорная информация о неотрицательности восстанавливаемой функции. Разработан алгоритм с учетом этой дополнительной информации и создана программа для проведения расчетов. Выполнена серия математических экспериментов, иллюстрирующих эффективность описанного метода.

Рассмотрим задачу о восстановлении оптического спектра $\varphi(x)$ по результату его измерения с помощью прибора, характеризующегося функцией разрешения $R(x)$. Измеряемая функция, которую мы обозначим $f(x)$, связана с $\varphi(x)$ следующим образом:

$$f(x) = \int_a^b R(x - x')\,\varphi(x')\,dx' + \varepsilon(x), \tag{1}$$

где $\varepsilon(x)$ — случайная функция, изображающая ошибки измерения. В более общем случае, когда учитывается изменение функции разрешения с частотой, соотношение (1) имеет вид

$$f(x) = \int_a^b K(x, x')\,\varphi(x')\,dx' + \varepsilon(x). \tag{2}$$

Мы предполагаем, что математическое ожидание функции $\varepsilon(x)$ равно нулю, а так как функция $\varepsilon(x)$ нам, разумеется, не известна, при нестатистическом подходе задача восстановления $\varphi(x)$ сводится к решению уравнения

$$f(x) = \int_a^b K(x, x')\,\varphi(x')\,dx'. \tag{3}$$

Учитывая отброшенный член $\varepsilon(x)$, мы можем удовлетвориться и приближенным решением этого уравнения.

Однако уравнение (3) при гладком ядре $K(x, x')$ некорректно, и его решение невозможно без дополнительных (априорных) предположений о $\varphi(x)$. Тихонов [1] разработал метод регуляризации решения некорректных задач, состоящий в замене исходного уравнения (3) на «регуляризованное» уравнение, являющееся корректным и включающее параметр регуляризации α, который введен таким образом, что при $\alpha=0$ регуляризованное уравнение переходит в исходное. Решение регуляризованного уравнения является гладкой функцией, и при малом α его можно рассматривать как приближенное решение уравнения (3). Метод статистической регуляризации [2-4] состоит в том, что соотношение (2) с самого начала рассматривается с позиций математической статистики, и априорная информация о решении вносится в виде априорного распределения вероятности для функции $\varphi(x)$. При этом оказывается, что если априорная информация о $\varphi(x)$ состоит в предположении о гладкости этой функции, то для наиболее правдоподобной функции $\varphi(x)$ мы получаем уравнение, совпадающее с регуляризованным уравнением Тихонова. Далее, статистические соображения дают оценку погрешности восстановления и алгоритм выбора параметра регуляризации α.

В задаче о восстановлении оптического спектра (как и во многих других задачах физики) искомая функция $\varphi(x)$, кроме того, что она является гладкой, удовлетворяет еще одному требованию: она принимает только неотрицательные значения, $\varphi(x) \geqslant 0$. Ясно, что учет дополнительной априорной информации о неотрицательности $\varphi(x)$ должен улучшить результат восстановления. В настоящей работе мы опишем получающийся при этом алгоритм и результаты некоторых модельных расчетов.

Мы принимаем за основу метод статистической регуляризации с выбором наиболее вероятного параметра регуляризации α (см. [3], а также обзорную статью [4]). В этом варианте априорная информация о гладкости функции $\varphi(x)$ задается следующей плотностью вероятности:

$$P(\boldsymbol{\varphi}) = \mathrm{const} \int\limits_0^\infty C_\alpha \exp\left\{-\frac{\alpha}{2}(\boldsymbol{\varphi}, \Omega\boldsymbol{\varphi})\right\} d\alpha. \tag{4}$$

Здесь $\boldsymbol{\varphi} = \{\varphi_1, \varphi_2, \ldots, \varphi_n\}$ — вектор, компонентами которого являются значения функция $\varphi(x)$ в некоторых опорных точках $x_i \in [a, b]$; $\varphi_i = \varphi(x_i)$, $i = 1, 2, \ldots, n$; Ω — матрица такая, что квадратичная форма $(\boldsymbol{\varphi}, \Omega\boldsymbol{\varphi})$ представляет конечно-разностное приближение к интегралу

$$\int\limits_a^b \left[\frac{d^2\varphi(x)}{dx^2}\right]^2 dx,$$

переменная интегрирования α имеет смысл параметра регуляризации, а C_α — нормировочный множитель (зависящий от α), который определяется из соотношения

$$\int\limits_0^\infty C_\alpha \exp\left\{-\frac{\alpha}{2}(\boldsymbol{\varphi}, \Omega\boldsymbol{\varphi})\right\} d\boldsymbol{\varphi} = 1. \tag{5}$$

Пусть функция $f(x)$ измеряется в m точках x_j; обозначим ее значения через f_j, а через $\mathbf{f}$ — соответствующий m-мерный вектор. Обозначим через s_j среднеквадратичную ошибку измерения f_j

$$s_j^2 = \langle \varepsilon(x_j)^2 \rangle \tag{6}$$

и будем считать ошибки $\varepsilon_j = \varepsilon(x_j)$ независимыми и распределенными нормально. Заменяя интеграл в (2) на его конечно-разностное приближение, представим соотношение (2) в виде системы линейных уравнений

$$f_j = \sum_{i=1}^n k_{ji}\varphi_i + \varepsilon_j, \quad j = 1, 2, \ldots, m. \tag{7}$$

В работе [3] показано, что в этих условиях может быть вычислена апостериорная плотность вероятности $P(\alpha \mid \mathbf{f})$ для параметра α, найдено значение α_0, при котором эта функция принимает максимальное значение (оценка гладкости функции по методу наибольшего правдоподобия), и вместо интеграла (4) в качестве априорной вероятности для $\boldsymbol{\varphi}$ взята функция

$$P_{\alpha_0}(\boldsymbol{\varphi}) = C_{\alpha_0} \exp\left\{-\frac{\alpha_0}{2}(\boldsymbol{\varphi}, \Omega\boldsymbol{\varphi})\right\}. \tag{8}$$

Апостериорная плотность вероятности для φ определяется, опираясь на нормальный закон для ε_j и формулу Бейеса,

$$P(\varphi \mid \mathrm{f}) = C \exp\left\{-\frac{1}{2}(\varphi, [K^+WK + \alpha\Omega]\varphi) + (K^+W\mathrm{f}, \varphi)\right\}. \tag{9}$$

Здесь W — диагональная матрица ошибок с матричными элементами

$$W_{ij} = \frac{1}{s_i^2}\delta_{ij}\ (i, j = 1, 2, \ldots, m), \quad \delta_{ij} = \begin{cases} 0 & \text{при } i = j, \\ 1 & \text{при } i \neq j. \end{cases}$$

Восстановленная функция φ^0 и квадрат погрешности восстановления определяются как математическое ожидание и дисперсия компонент φ_i по ансамблю (9)

$$\varphi_i^0 = \int \varphi_i P(\varphi \mid \mathrm{f})\, d\varphi, \tag{10}$$

$$\sigma_i^2 = \int (\varphi_i - \varphi_i^0)^2 P(\varphi \mid \mathrm{f})\, d\varphi. \tag{11}$$

Эти интегралы легко берутся аналитически и для них получаются формулы, включающие лишь основные матричные операции [3].

Введем функцию $\zeta(\varphi)$, обращающуюся в нуль в тех случаях, когда $\varphi_i < 0$ хотя бы при одном $i = 1, \ldots, n$. Наиболее последовательный способ учета дополнительной информации о неотрицательности $\varphi(x)$ — домножить априорную вероятность (4) на $\zeta(\varphi)$. При этом, однако, интеграл по φ, через который выражается апостериорное распределение $P(\alpha|\mathrm{f})$ для параметра регуляризации α, уже не берется аналитически, что создает трудности для нахождения наиболее вероятного значения α_0. Поэтому поступим следующим образом. Сначала найдем α_0 без учета неотрицательности $\varphi(x)$, а затем в качестве априорной вероятности возьмем функцию (8), домноженную на $\xi(\varphi)$

$$P'(\varphi) = C'\zeta(\varphi) \exp\left\{-\frac{\alpha_0}{2}(\varphi, \Omega\varphi)\right\}. \tag{12}$$

Константа C' определяется из условия нормировки и не совпадает с C_{α_0}.

Множитель $\zeta(\varphi)$, введенный в априорную вероятность $P'(\varphi)$, переходит в апостериорную вероятность

$$P'(\varphi \mid \mathrm{f}) = C''\zeta(\varphi) P(\varphi \mid \mathrm{f}) \tag{13}$$

и оказывается под знаком интеграла в формулах (10) и (11). Здесь снова возникает та же трудность: новые интегралы не берутся аналитически.

Чтобы избежать взятия интеграла (10), будем вычислять не математическое ожидание вектора φ, по апостериорному распределению, а тот вектор φ, при котором апостериорная плотность вероятности $P'(\varphi|\mathrm{f})$ принимает максимальное значение; обозначим его φ^m и примем в качестве восстановленной функции. При отсутствии множителя $\zeta(\varphi)$ апостериорное распределение для φ является нормальным, поэтому φ^0 и φ^m совпадают. При наличии $\zeta(\varphi)$ они не совпадают, и с точки зрения теоретической φ^0 предпочтительнее, чем φ^m. Однако и φ^m имеет ясный смысл и в нашем случае отличается от φ^0 незначительно.

Хуже обстоит дело с вычислением ошибки восстановления. Однако мы можем воспользоваться ошибкой, вычисляемой по формуле (11) без учета неотрицательности функции $\varphi(x)$, как оценкой сверху, ибо введение обрезающей функции $\zeta(\varphi)$ может только уменьшить этот интеграл.

Точку φ^m можно определить как точку гипероктанта $\varphi_i > 0$, $i = 1, 2, \ldots, n$, в которой функционал

$$Z(\varphi) = \frac{1}{2}(\varphi, [K^+WK + \alpha\Omega]\varphi) - (K^+W\mathrm{f}, \varphi)$$

принимает минимальное значение. Легко показать, что если φ^0 имеет отрицательные компоненты, то φ^m лежит на одной из граней гипероктанта, т. е. в области, где хотя бы одно из φ_i равно нулю. Грань Г гипероктанта можно задать множеством $\nu(\Gamma)$ тех i, при которых $\varphi_i > 0$. Через $\bar{\nu}(\Gamma)$ обозначим дополнительное множество, через $R(\Gamma)$ — подпространство, образованное ортами множества $\nu(\Gamma)$. Грань Γ^m, содержащая φ^m, обладает следующими свойствами:

$$\frac{\partial}{\partial \varphi_i} Z(\varphi) = 0 \quad \text{для} \quad i \in \nu(\Gamma), \tag{14а}$$

$$\frac{\partial}{\partial \varphi_i} Z(\varphi) \geqslant 0 \quad \text{для} \quad i \in \bar{\nu}(\Gamma). \tag{14б}$$

Если грань Γ^m найдена, нахождение φ^m не представляет труда — достаточно решить линейную систему уравнений (14а) относительно φ_i, $i \in \nu(\Gamma^m)$. Для остальных граней Г осуществляется один из двух случаев.

1. Решение системы (14а) содержит отрицательные компоненты φ_i, иначе говоря, точка подпространства $R(\Gamma)$, в которой достигается минимум $Z(\varphi)$, лежит вне грани Г.

2. Решение системы (14а) принадлежит грани Г, но не выполнено условие (14б).

Алгоритм нахождения φ^m состоит в том, что некоторая опорная точка φ^p перемещается с грани на грань, пока не попадает на грань Γ^m и обращается в φ^m. Когда задана грань Г и опорная точка $\varphi^p \in \Gamma$, правила перемещения определяют новую грань Γ' и опорную точку на ней $\varphi^{p'} \in \Gamma'$. Начальной опорной точкой может служить, например, точка $\varphi^p = 0$ или точка, полученная из φ^0 заменой всех отрицательных φ_i^0 на нуль. Начальной гранью Г может служить наименьшая грань, содержащая опорную точку. Правила перемещения таковы. Решаем систему (14а) для Г; решение обозначим через φ^r. Если имеет место случай 2, то положим $\varphi^{p'} = \varphi^r$, а грань Γ' получим «расширением» грани Г: множество $\nu(\Gamma')$ образуется из $\nu(\Gamma)$ включением того i, для которого $-\partial Z/\partial\varphi_i$ максимально. Из того, что $\Gamma' \supset \Gamma$ и $\varphi^{p'} \in \Gamma$, следует $\varphi^{p'} \in \Gamma'$. Если имеет место случай 1, поступаем следующим образом. Соединяем точки φ^p и φ^r отрезком прямой

$$\varphi^d(\xi) = \varphi^p(1 - \xi) + \varphi^r \xi.$$

Поскольку φ^p не имеет отрицательных компонент, а φ^r имеет, существует такое $i \in \nu(\Gamma)$, что при движении по этому отрезку от φ^p к φ^r компонента φ_i^d меняет знак первой. Соответствующее значение параметра ξ, которое мы обозначим через ξ_0, нетрудно найти. Грань Γ' определим как «сужение» грани Г: множество $\nu(\Gamma')$ образуется исключением i из $\nu(\Gamma)$. В качестве опорной точки $\varphi^{p'}$ возьмем $\varphi^d(\xi_0)$. Из того, что $\varphi_i^d(\xi_0) = 0$, следует: $\varphi^{p'} \subset \Gamma$. Наконец, если не имеет места ни случай 1, ни случай 2, то $\Gamma = \Gamma^m$, $\varphi^r = \varphi^m$ и выполнение алгоритма завершается.

Нетрудно убедиться в сходимости описанного метода. Случай 1 не может, очевидно, повторяться более чем n раз подряд. В случае же 2 опорная точка φ^p однозначно определяется гранью Г. Так как при каждом шаге значение функционала $Z(\varphi^p)$ убывает, пройдя такую грань, мы не можем попасть на нее снова. Из ограниченности числа граней следует сходимость метода. На практике она оказывается очень быстрой.

На рис. 1 ÷ 5 приводятся результаты модельных расчетов по методу статистической регуляризации с использованием дополнительной информации о неотрицательности искомой функции. Расчеты проводились так. Была выбрана некая неотрицательная функция φ, которую мы будем называть «истинной» функцией; на всех рисунках она показана сплошной жирной линией. Затем эта функция свертывалась с «функцией разрешения» треугольной формы (полуширина ее варьировалась) и к результату добавлялась нормально распределенная случайная ошибка («ошибка эксперимента»). Полученная функция рассматривалась как левая часть $\mathbf{f}$ и восстанавливалась функция φ^0 без учета и φ^m с учетом информации о

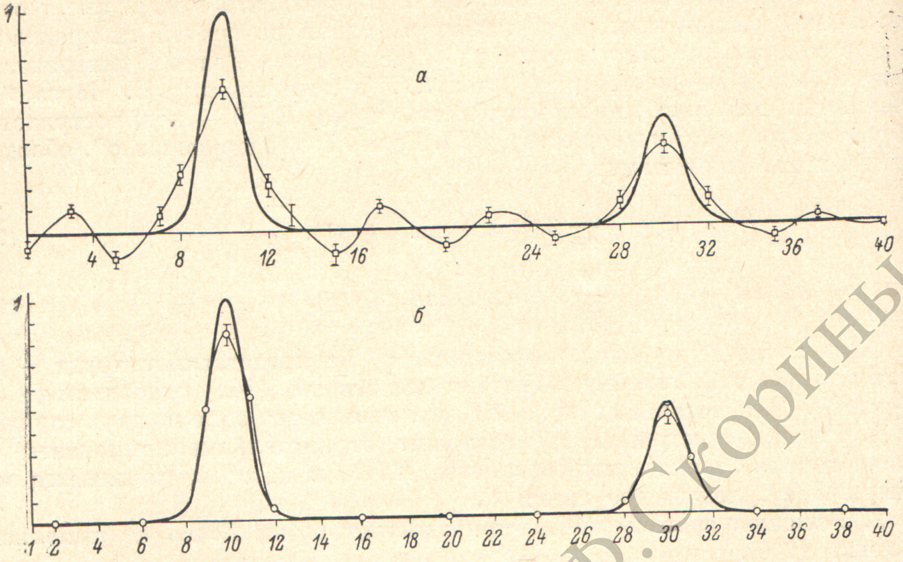

Рис. 1. Восстановление функции $\varphi(x)$ (жирная кривая без точек) в модельном расчете.

Полуширина функции разрешения $\Delta=5$. Среднеквадратичная величина «экспериментальной» ошибки $s_j=0.03\,f_j$. *а* — восстановление без условия неотрицательности $\varphi(x)$, *б* — с условием неотрицательности.

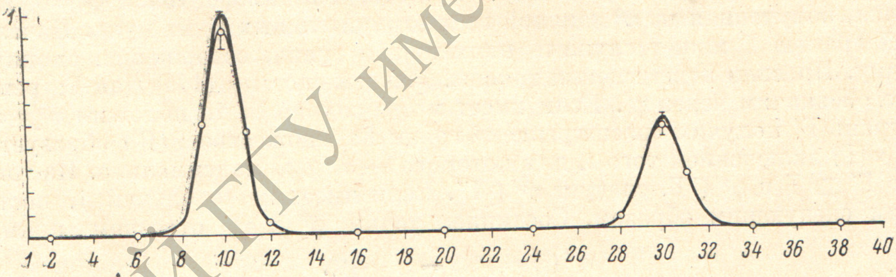

Рис. 2.
То же, что на рис. 1, *б*, но с поправкой в α_0.

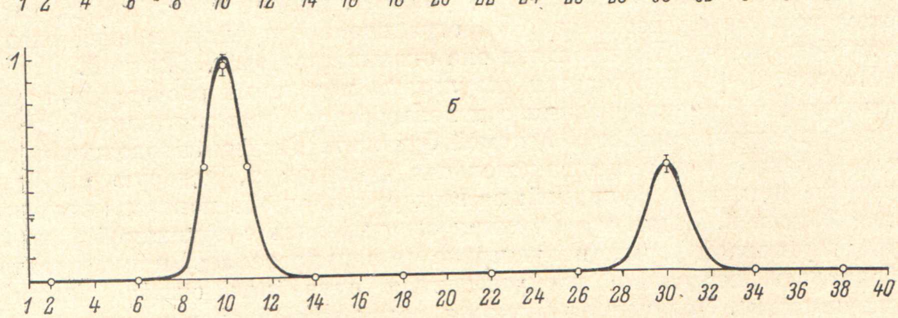

Рис. 3. Восстановление по алгоритму без учета неотрицательности (*а*) и с учетом неотрицательности (*б*) при $\Delta=3$, $s_j=0.03\,f_j$.

неотрицательности функции. Восстановленная функция показана тонкой линией с кружочками; показана также теоретическая (среднеквадратичная) ошибка восстановления. Полуширина Δ функции разрешения и

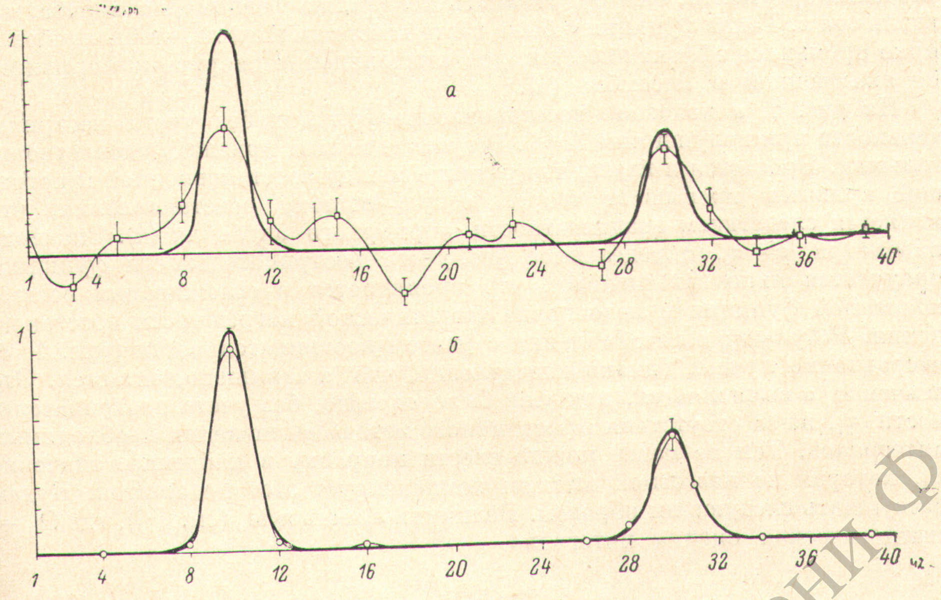

Рис. 4.
То же, что на рис. 3 при $\Delta=8$.

среднеквадратичная величина «экспериментальной» ошибки s_j указаны в подписях под рисунками.

Исходная функция φ имеет два острых максимума разной высоты, а в тех точках, которые не лежат вблизи точек максимума, обращается

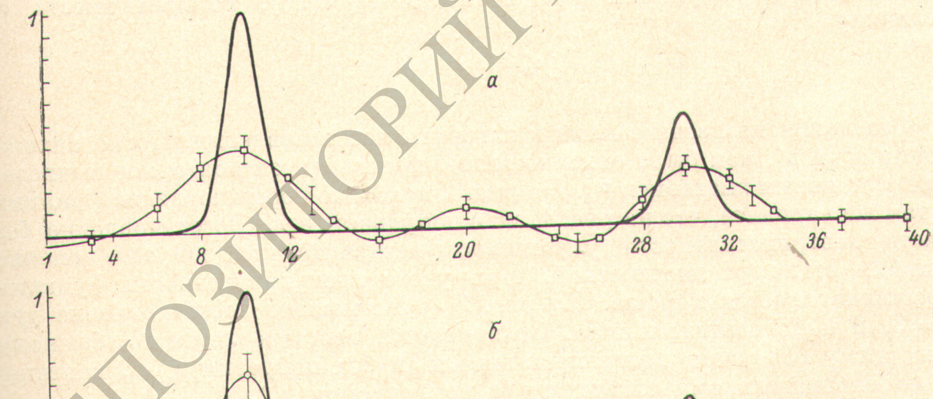

Рис. 5.
То же, что на рис. 3 при $\Delta=15$.

в нуль. Такая функция не принадлежит к ансамблю (4); точнее, вероятность получить такую функцию из ансамбля (4) исчезающе мала. Действительно, в этом ансамбле функции с заданной нормой второй производной равновероятны. Ясно, что если задать норму второй производной, а в остальном генерировать функции случайно, то они не станут концентрироваться около двух точек и обращаться в нуль в остальных областях значений аргумента; они будут иметь тенденцию равномерно колебаться вдоль оси x. Такая функция φ была взята нами с той целью, чтобы наиболее ярко проявился эффект внесения дополнительной априорной информации, не включенной в ансамбль (4).

На рис. 1, *а* показано восстановление функции без учета неотрицательности при полуширине $\Delta=5$. Результат нельзя признать удовлетворительным. Бросается в глаза, что разница между восстановленной и истинной кривыми превышает теоретическую ошибку восстановления, что является следствием неадекватности априорного ансамбля (4). Значение параметра $\alpha=\alpha_0$, определенное без учета неотрицательности функции, получается явно завышенным, что проявляется в сглаживании восстановленной функции и недооценке теоретической погрешности восстановления. Результат восстановления с учетом неотрицательности (рис. 1, *б*) несравненно лучше, однако впечатление, что α завышено, остается. Это значение α вычисляется, как уже было сказано, без учета неотрицательности φ_i, из-за отсутствия эффективного метода вычисления необходимых интегралов. Тем не менее можно ввести поправку в найденное значение α_0, которая до известной степени компенсирует этот недостаток. Будем рассуждать следующим образом. Величина α_0 есть n/Ω_0 (см. [3]), где Ω_0 — статистическая оценка интеграла

$$\int\left(\frac{d^2\varphi}{dx^2}\right)^2 dx \tag{15}$$

в предположении, что функция $\varphi(x)$ равномерно распределена во всем интервале x, соответствующем точкам i от 1 до n. Но мы знаем, что в действительности она отлична от нуля только на множестве точек $i\in\nu(\Gamma^m)$. Обозначим через n' число этих точек и представим себе, что функция $\varphi(x)$ сжата по оси x, чтобы уместиться в интервале длиной n'. Очевидно, интеграл (15) возрастает при этом в $(n/n')^3$ раз. Поэтому чтобы правильно учесть характер функции $\varphi(x)$ в точках $i\in\nu(\Gamma^m)$, возьмем вместо α_0 величину

$$\alpha_0'=\alpha_0(n'/n)^3.$$

Эта поправка введена нами в алгоритм восстановления. После нахождения φ^m и n', как было описано выше, производится повторное вычисление φ^m с новым значением α_0. На рис. 2 показано восстановление функции в том же варианте, что на рис. 1, но с поправкой в α. Восстановление без учета неотрицательности улучшилось незначительно, но зато оценка ошибки стала реалистической. При учете неотрицательности функция восстанавливается хорошо. На рис. 3 $\div$ 5 показано восстановление при $\Delta=3$, 8 и 15 соответственно. Сравнивается восстановление по обычному алгоритму без учета неотрицательности (и без поправки в α_0, конечно) и по окончательному алгоритму (с поправкой в α_0).

Отметим, что при $\Delta=8$ (рис. 4, *б*) полуширина функции разрешения вчетверо больше полуширины максимумов исходной функции, а «экспериментальная» ошибка не слишком мала (3% от «измеренной» функции); тем не менее восстановление получается вполне удовлетворительным. На рис. 5 показано восстановление φ при $\Delta=15$. Здесь уже и учет неотрицательности не спасет положения: восстановленная функция содержит ложный максимум (правда, в пределах теоретической ошибки).

Заметим, что одного лишь ограничения неотрицательными функциями недостаточно для эффективной регуляризации. Расчеты, проведенные нами, показали, что при слишком маленьком α начинается «раскачка», несмотря на неотрицательность функции. Именно комбинация априорной гладкости и неотрицательности дает хороший результат.

Программы на Алголе и Фортране, осуществляющие описанный алгоритм, будут опубликованы отдельно.

Литература

[1] А. Н. Тихонов. ДАН СССР, *151*, 501, 1963.
[2] В. Ф. Турчин. Ж. вычисл. мат. и матем. физ., *7*, 1270, 1967.
[3] В. Ф. Турчин, В. З. Нозик. Изв. АН СССР, сер. «Физика атм. и океана», *5*, 29, 1969.
[4] В. Ф. Турчин, В. П. Козлов, М. С. Малкевич. Усп. физ. наук, *102*, 345, 1970.

Поступило в Редакцию 20 июня 1972 г.