

МЕТОД ДИСКРЕТНЫХ НАПРАВЛЕНИЙ В НЕЛИНЕЙНОЙ ЗАДАЧЕ ФИТИРОВАНИЯ СПЕКТРОВ НЕКОГЕРЕНТНОГО РАССЕЯНИЯ

Б.Г. Шпынев, А.Л. Воронов

DISCREET DIRECTION METHOD IN NONLINEAR PROBLEM OF INCOHERENT SCATTERING SPECTRUM FITTING

B.G. Shpynev, A.L. Voronov

Для определения параметров ионосферной плазмы по данным спектральных измерений, полученных на Иркутском радаре некогерентного рассеяния, разработан новый метод фитирования, основанный на разбиении функционала невязки симплексами n -го порядка. Поиск минимума многомерного функционала в данной постановке сводится к минимальному количеству вычислений модельных функций в каждой итерации. Алгоритм позволяет также упростить критерий останова поиска.

A new fitting technique based on functional partition by simplexes of the n^{th} order has been developed to determine ionospheric plasma parameters with spectral measurement data from the Irkutsk incoherent scatter radar. The search for functional minimum in this approach adds up to as few model function calculations in each iteration as possible. The algorithm also allows us to simplify the search stop criterion.

Введение

Определение параметров ионосферной плазмы по экспериментальным спектрам некогерентного рассеяния (НР) является некорректной обратной задачей, в которой искомые параметры входят нелинейным образом в выражение для теоретического спектра мощности [Faley, 1966]. Для задач такого типа используют численные методы решения – прямой расчет набора модельных спектров и выбор среди модельных спектров наиболее близкого к экспериментальному спектру. Такая процедура является стандартной в подобного рода задачах и называется фитированием, т. е. подгонкой модельной функции к экспериментальной. Спектры некогерентного рассеяния должны рассчитываться для разных значений электронной и ионной температур T_e , T_i , относительной концентрации ионов атомарного кислорода N_{O^+} и скорости дрейфа плазмы вдоль луча зрения радара V_d . На начальном этапе исследований методом НР [Hagen, Hsu, 1974] в алгоритме фитирования использовалась библиотека таких модельных спектров, и экспериментальный спектр сравнивался с заранее рассчитанными библиотечными спектрами. Очевидно, такая библиотека требовала большой оперативной памяти компьютера и на тот момент не могла обеспечить достаточного мелкого шага поиска. С увеличением быстродействия компьютеров стало возможным [Swartz, 1978] рассчитывать модельные спектры непосредственно перед сравнением их с экспериментальным и выбирать произвольный шаг поиска. Для проведения фитирования обычно использовался один из методов, описанных в работах [Bard, 1970; Dennis, et al., 1983], основанных на использовании частных производных спектра мощности по каждому из искомых параметров, а производные, если это было возможно, упрощались аналитически, что позволяло ускорить вычисления. Экспериментальный спектр мощности представляет собой дискретный набор из N отсчетов с известным шагом по частоте. Для возможности сравнения отсчеты модельного спектра вычисляются в тех же точках частотного диапазона.

Развитие технологий фитирования в методе НР в последние десятилетия проводилось в двух направ-

лениях. Первое – это улучшение качества статистических оценок при зондировании кодированными последовательностями импульсов на основе традиционных алгоритмов [Vallinkoski, 1988; Lehtinen, Huuskonen, 1996]. Используемые здесь классические методы минимизации до настоящего времени являются единственными алгоритмами, работающими в реальном масштабе времени. Второе направление – это оптимальный анализ [Holt, et al., 1992; Nikoukar, et al., 2008], при котором функционал невязки вычисляется не для отдельных точек пространства, а целиком для профилей температуры ионов и электронов, и далее производится минимизация этого функционала. Такой подход является более точным, поскольку позволяет устранить неоднозначности интерпретации спектров, вызванные одновременным изменением газового состава и температуры, которые приводят к одинаковым изменениям формы спектра [Erickson, 1998]. Однако время работы таких алгоритмов настолько велико, что в 90-х годах обработка часового эксперимента требовала более суток вычислений [Holt, et al., 1992], и даже при использовании современных вычислительных средств [Nikoukar, et al., 2008] работа этих алгоритмов в масштабе реального времени проблематична.

Описанный в настоящей работе алгоритм предназначен для классической обработки данных НР в реальном масштабе времени и должен обеспечить работу в условиях низкого отношения сигнал/шум (S/N). Причиной низкого отношения S/N для Иркутского радара НР являются два фактора. Во-первых, его частотный диапазон (~160 МГц) обладает относительно высоким уровнем естественных радиошумов. Во-вторых, для волноводно-щелевых антенн характерно наличие случайных искажений спектров в присутствии резких градиентов электронной концентрации. Суть этих искажений состоит в том, что для волноводно-щелевой антенны Иркутского радара НР изменение частоты принимаемого сигнала в полосе приема 10–15 кГц эквивалентно эффективному изменению углу наклона приемного луча на 0.05–0.08° (приблизительно 7–10 % от ширины луча). Это значит, что разные части спектра НР изменяются при разных углах наклона луча, и в присут-

ствии значительных градиентов электронной концентрации спектр имеет характерную асимметрию, зависящую от знака градиента, поэтому основным требованием к алгоритму фильтрования является его устойчивость к таким нестационарным шумам эксперимента.

Нелинейная задача о наименьших квадратах

Нелинейная задача о наименьших квадратах обычно возникает в приложениях, связанных с наилучшим приближением данных, где стараются наилучшим образом приблизить данные $\tilde{S}(\omega_i)$ с помощью модели $S(x_k, \omega_i)$, которая нелинейна по x . С математической точки зрения задача фильтрования есть поиск глобального минимума многомерного функционала невязки

$$\min_{x \in R^m} \Phi(x_1, x_2 \dots x_m) = \sum_{i=1}^N (\tilde{S}(\omega_i) - S(\omega_i, x_1, x_2 \dots x_m))^2, \quad (1)$$

где \tilde{S} – экспериментальный спектр, S – модельный спектр, x_k – набор параметров, по которым проводится минимизация, ω_i – сетка частот, m – число фильтруемых параметров.

Поскольку в эксперименте присутствуют шумы, для описания экспериментальных данных часто используют статистические методы, считая экспериментальную функцию случайным процессом с нормальным законом распределения [Bard, 1970; Dennis, et al., 1983; Lehtinen, Huuskonen, 1996]. На основе центральной предельной теоремы определяются вероятностная оценка среднего значения случайного процесса и его дисперсия. В таком подходе ищется минимум распределения:

$$\chi^2(x_1, x_2 \dots x_m) = \sum_{i=1}^N \frac{(\tilde{S}(\omega_i) - S(\omega_i, x_1, x_2 \dots x_m))^2}{\sigma_i^2}, \quad (2)$$

где σ_i – дисперсия оценки каждого отсчета. При отсутствии корреляции между ошибками определения разных параметров распределения доверительную область полученной оценки с вероятностью 68 % определяют из уравнения

$$\Delta\chi^2 = 1. \quad (3)$$

Иными словами, доверительной областью значений полученных параметров x_k с заданной вероятностью будет кривая на поверхности χ^2 , где значение невязки равно 2σ . Именно из статистических соображений в качестве меры для наилучшего приближения данных используют квадрат отклонения, для которого статистическая теория разработана лучше [Bard, 1970].

Основной задачей при обработке спектров некогерентного рассеяния является поиск электронной и ионной температур, от которых спектр сигнала зависит нелинейно, поэтому в первую очередь мы рассмотрим двухпараметрический функционал невязки $\Phi(T_e, T_i)$. На рис. 1 приведены примеры функционалов невязки для реальных спектров, измеренных на радаре НР ИСЗФ [Жеребцов и др., 2002; Shpynev, 2004]. В данных примерах глобальный минимум очевиден, хотя доверительная область, описываемая второй от минимума изолинией, имеет несимметричный вид, характеризующий наиболее вероятные ошибки изме-

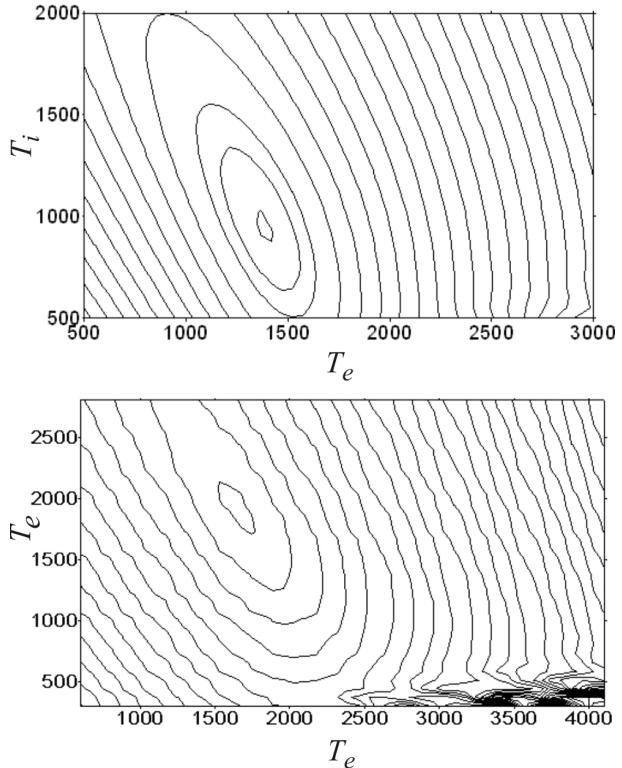


Рис. 1. Изолинии проведены с шагом, равным дисперсии σ .

рений. На функционале, изображенном в нижней части рис. 1, положение глобального минимума не соответствует физической реальности $T_e < T_i$, и требуется привлечение дополнительных физических условий для корректного решения данной задачи. Большая протяженность доверительной области на функционалах (рис. 1) является следствием упомянутых выше больших шумов эксперимента.

В литературе [Dennis, et al., 1983, Hamming, 1972] рассмотрен ряд методов поиска решения нелинейной задачи о наименьших квадратах. Наиболее распространенными являются методы Ньютона, Гаусса–Ньютона, метод секущих и их вариации [Dennis, et al., 1983]. Основа этих методов состоит в определении первых частных производных от $\Phi(x_1, x_2, \dots, x_m)$ по каждому из параметров, построение линейной модели $\Phi(x_1, x_2, \dots, x_m)$ в окрестности текущей точки и спуск в направлении глобального минимума функционала. Иногда для оценки вклада вторых производных необходимо пользоваться квадратичной моделью функционала, что требует вычисления вторых производных по каждому из параметров.

Данные методы обладают хорошей локальной сходимостью для задач, которые не являются слишком нелинейными или имеют в решении достаточно малые невязки. Напротив, для задач, которые слишком нелинейны или имеют в решении достаточно большие невязки, методы не обладают локальной сходимостью и не гарантируют глобальную сходимость. Необходимость вычисления производных по параметрам x_i является существенным недостатком данных методов, поскольку часто требует применения конечных разностей порядка выше первого, а иногда и вычисления вторых производных, что ведет к дополнительным $(m^2+3m)/2$ вычислениям

функционала $\Phi(x_1, x_2, \dots x_m)$ на каждой итерации. Поскольку модель $S(\omega_i, x_1, x_2, \dots x_m)$ обычно имеет довольно сложный вид [Шеффилд, 1978], такие затраты бывают непозволительными, и пользователям алгоритмов приходится прибегать к дополнительной информации для ускорения работы.

Еще одной общей проблемой для методов решения нелинейной задачи о наименьших квадратах является критерий останова алгоритма. Во-первых, ноль производной может означать как точку минимума, так и точку перегиба функционала, тогда для определения минимума нужен расчет вторых производных. Во-вторых, если вблизи глобального минимума $\Phi(x_1, x_2, \dots x_m)$ невязка остается достаточно большой, в классических методах тяжело выработать критерий останова, который удовлетворял бы требованиям точности, хотя такая ситуация часто встречается в реальных измерениях.

Опыт других авторов [Erickson, 1998] и проведенные нами исследования показали, что для глобальной сходимости алгоритмов решения нелинейной задачи о наименьших квадратах применительно к спектральным измерениям в методе НР одним из самых важных моментов является выбор начального приближения параметров задачи. В некоторых случаях, встречающихся на практике, функционал $\Phi(x_1, x_2, \dots x_m)$ имеет множественные локальные минимумы, а глобальный минимум иногда представляет собой желоб (набор решений задачи с одинаковой невязкой). Такая неоднозначность решения в задаче НР чаще всего вызвана наличием в плазме ионов разного сорта или неопределенностью в значении T_e/T_i , видимой как сильная вытянутость функционалов на рис. 1.

Для выбора единственного решения из возникающего множества (или корректировки данных) необходимо использовать дополнительную информацию. Так, в программе обработки данных радара Миллстон-Хилл [Erickson, 1998] использовалась информация о температурах с ближайших высот, и в функционал невязки вносится ограничение, изменяющее его в пользу наиболее вероятных температур. С учетом указанных проблем, возникающих при обработке экспериментальных данных в методе НР, нами был разработан метод фитирования спектров НР, хорошо сходящийся в различных шумовых условиях и позволяющий избежать некоторых проблем стандартных алгоритмов.

Метод дискретных направлений

В данном методе не проводятся вычисления производных по параметрам функционала $\Phi(x_1, x_2, \dots x_m)$ и используется фиксированный шаг при спуске к глобальному минимуму. Поскольку в реальных экспериментальных данных шумы не позволяют определять положение глобального минимума с произвольно заданной точностью, величину начального шага поиска неразумно делать меньше типичной ошибки измерений. Лишь при нахождении (приближенно) минимума $\Phi(x_1, x_2, \dots x_m)$ можно уменьшить шаг и определить минимум более точно. При последовательном проведении минимизации $\Phi(x_1, x_2, \dots x_m)$ на серии однотипных спектров шаг поиска для каждого последующего функционала невязки может быть определен из предшествующих расчетов.

Суть метода дискретных направлений (ДН) заключается в том, что вместо вычисления производных по каждому из параметров функционала $\Phi(x_1, x_2, \dots x_m)$ и построения его линейной модели использован поиск минимального значения функционала на окружностях с радиусами Δx_i , соответствующими шагу поиска по каждой из переменных.

Для наглядности рассмотрим работу метода на двухпараметрическом функционале $\Phi(x_1, x_2)$ с одинаковым шагом Δx . На каждой итерации будем вычислять значения функционала $\Phi(x_1, x_2)$ в дискретном наборе k точек плоскости (x_1, x_2) , расположенных симметрично вдоль окружности с радиусом Δx , причем отсчет углов будем производить от направления на точку предыдущей итерации или от нуля при первой итерации. Выбирая из k точек ту, в которой функционал $\Phi(x_1, x_2)$ минимален, переместимся в нее и продолжим процесс вычисления по данной схеме, пока значение функционала в центральной точке окажется меньше, чем где-либо на окружности.

При минимальном значении $k=3$, которое может быть использовано в данном алгоритме, плоскость параметров (x_1, x_2) разбивается на шестиугольники (рис. 2), количество необходимых вычислений $\Phi(x_1, x_2)$ равно $k-1$, т. е. двум (третья точка определяется на предыдущей итерации). Таким образом, по мере спуска от начального приближения $\Phi(x_1^0, x_2^0)$ к минимуму процесс поиска продвигается по ломаной, показанной на рисунке стрелками.

Только в трех случаях ($k=3; 4; 6$) точки плоскости (x_1, x_2) лежат в узлах равномерной сетки из правильных шестиугольников ($k=3$), четырехугольников ($k=4$) и треугольников ($k=6$). Для других значений $k>3$ периодическая сетка не существует.

Для выбора значения k , обеспечивающего максимальную скорость работы алгоритма, обратимся вновь к рис. 2. Пунктирная линия показывает возможные пути движения алгоритма при $k=6$. Как видно из рисунка, для спуска к минимуму, при $k=3$ алгоритму требуется восемь итераций и расчет 16 значений $\Phi(x_1, x_2)$. При $k=6$ алгоритму потребуется семь итераций и расчет 35 значений $\Phi(x_1, x_2)$. Быстродействие алгоритма с $k=3$ оказывается более чем в два раза выше. Можно показать, что быстродействие алгоритма при $k=4$ и при $k=3$ совпадает при движении по прямой вдоль одной из осей x_1 или x_2 , в остальных случаях быстродействие при $k=3$ будет выше, т. е. при $k=3$ данный алгоритм обладает максимальной скоростью.

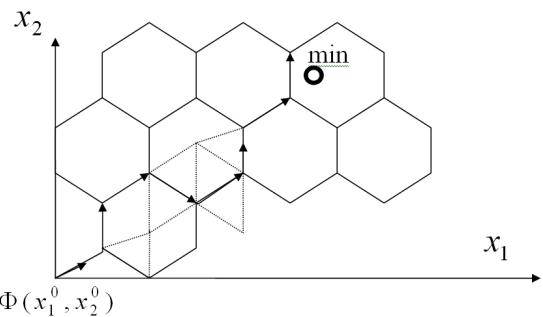


Рис. 2. Схема спуска в алгоритме дискретных направлений.

Поскольку на каждой итерации при $k=3$ существует три возможных направления движения по плоскости (x_1, x_2) , для каждого из этих направлений будет рассчитываться два значения функционала $\Phi(x_1, x_2)$, третье значение известно из предыдущей итерации, и в любой итерации, кроме первой, оно больше текущего значения и может не рассматриваться.

Таким образом, в двумерном случае точка текущей итерации лежит в центре равнобедренного треугольника, а новые узлы являются его вершинами, которые можно определить, воспользовавшись тем, что равнобедренный треугольник является симплексом размерности 2. Пользуясь правилом построения симплекса размерности 2, поместим центр равнобедренного треугольника в начало координат, как это изображено на рис. 3, и положим длину его стороны $\Delta x=1$.

Координаты вершин будут равны $D(0, \sqrt{3}/3)$, $B(-1/2, -\sqrt{3}/6)$ и $C(1/2 - \sqrt{3}/6)$.

По мере движения алгоритма к точке с наименьшей невязкой центр треугольника перемещается в новую точку:

1) при движении вверх вычисляются значения $\Phi(x_1, x_2)$ в точках $(x_1 - 1/2, x_2 + \sqrt{3}/6)$ и $(x_1 + 1/2, x_2 + \sqrt{3}/6)$;

2) при движении влево (вверх, вниз) – в точках $(x_1, x_2 \pm \sqrt{3}/3)$ и $(x_1 - 1/2, x_2 \mp \sqrt{3}/6)$;

3) при движении вправо (вверх, вниз) – в точках $(x_1, x_2 \pm \sqrt{3}/3)$ и $(x_1 + 1/2, x_2 \mp \sqrt{3}/6)$.

По аналогии можно построить правило расчета новых вершин для функционала типа $\Phi(x_1, x_2, x_3)$. В этом случае новые точки будут вершинами симплекса размерности 3 – тетраэдра (рис. 4). Для построения общего правила определения координат узлов сетки, составленной из вершин симплексов, обратимся к рис. 5 и 6, которые описывают двумерный и трехмерный случай соответственно.

В двумерном случае (рис. 5) координаты новых узлов A^1, B^1, C^1 определяются как отражение узлов A, B, C на предыдущей итерации относительно прямой OE , проведенной через середину вектора $\vec{BB^1}$, проведенного из узла на предшествующей итерации в текущий узел. По условию построения, точки A, B^1, C являются вершинами равнобедренного треугольника, а точка B делит высоту MB^1 треугольника AB^1C в соотношении $1/2$. Из рисунка видно, что для нахождения координат новых узлов необходимо к координатам узлов на предыдущей итерации добавить вектор $\vec{BB^1}$, умноженный на два.

В трехмерном случае (рис. 6) новые узлы находятся путем отражения старых узлов относительно плоскости, проходящей через середину вектора $\vec{BB^1}$ и перпендикулярной ему. Преобразования координат узлов будут представлять собой параллельный перенос на вектор $\vec{BB^1}$ с коэффициентом $5/3$, равным отношению высоты тетраэдра к расстоянию от вершины до центра.

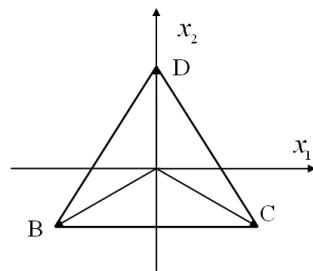


Рис. 3. Относительные координаты вершин равнобедренного треугольника на плоскости.

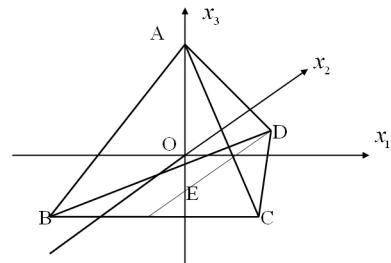


Рис. 4. Относительные координаты вершин тетраэдра в трехмерном пространстве.

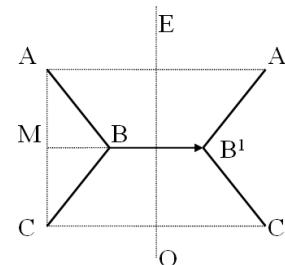


Рис. 5. Схема построения узлов сетки на плоскости.

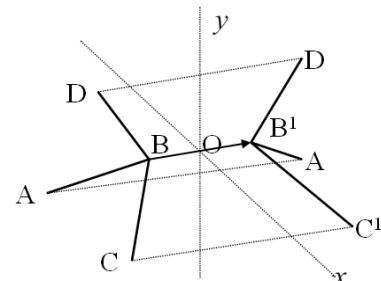


Рис. 6. Схема построения узлов сетки в объеме.

Данный метод можно обобщить на случай симплекса произвольной размерности порядка n , где $n>1$. При этом правильный $(n+1)$ -угольник (симплекс) строится из симплекса размерности $n-1$ по следующему алгоритму:

1) рассчитывается длина высоты h_n нового симплекса по формуле $h_n = \sqrt{1-h_{n-1}}$, где h_{n-1} – высота симплекса размерности $n-1$;

2) симплекс размерности $n-1$ сдвигается на $-\frac{h_n}{n}$ вдоль новой переменной x_n ;

3) добавляется новая точка с координатами $(0, 0, 0, \dots, 1 - \frac{h_n}{n})$.

После построения симплекса размерности n определяются координаты вершин симплекса и ищется вершина с наименьшей невязкой. Строится вектор \vec{K} из текущей точки в направлении точки наименьшей невязки, и координаты векторов $A_i(x_1, x_2, \dots, x_n)$ ($i = 1, \dots, n-1$) узлов в новой итерации определяются по формуле

$$A_i(x_1, x_2, \dots, x_n) = \bar{A}_i(x_1, x_2, \dots, x_n) + \frac{n+2}{n} \vec{K}, \quad (4)$$

где $\bar{A}_i(x_1, x_2, \dots, x_n)$ – текущие координаты вершин симплекса.

Алгоритм ДН останавливается, если значение функционала в текущей точке (в центре симплекса) меньше, чем в какой-либо из вершин.

Поскольку сетка узлов является равномерной по всем переменным, при использовании метода на практике необходимо проводить соответствующее масштабирование, чтобы скорость движения по каждой из переменных соответствовала реально получаемой точности и диапазону изменения параметра.

Тестирование

Для двумерного функционала невязки метод ДН сравнивался с классическим алгоритмом градиентного спуска на одинаковом массиве теоретических и экспериментальных спектров, соответствующих разным ионосферным условиям и разным отношениям сигнал/шум. При аппроксимации теоретических спектров мощности более высокую точность давал классический алгоритм с приблизительно равной скоростью расчета. При обработке экспериментальных спектров алгоритмы давали одинаковые в пределах точности измерений результаты при $S/N > 1$, а скорость работы алгоритма ДН была в 1.6–2.1 раза выше. При $0.1 < S/N < 1$ новый алгоритм работал в среднем в 2–3 раза быстрее, причем классический алгоритм в некоторых случаях выходил из области анализа.

Алгоритм дискретных направлений тестировался также на функционалах с размерностью 3 и 4. Для функционала $\Phi(T_e, T_i, N_{O^+})$ он также дает хорошую устойчивость и скорость работы и используется в настоящее время при регулярной обработке спектров НР. При размерности 4 для функционала типа $\Phi(T_e, T_i, N_{O^+}, V_d)$ точность фитирования ухудшается ввиду упомянутых выше искажений спектра, делающих задачу неоднозначной. Поэтому определение скорости дрейфа V_d проводится нами по доплеровскому сдвигу частоты вне основного цикла фитирования. Возможно, в задачах, где функционал невязки не является сильно нелинейным, алгоритм ДН будет устойчив и для больших размерностей функционала невязки, однако такое тестирование в данной работе не проводилось и не предполагалось.

Выводы

Предложенный метод фитирования показал себя более эффективным по сравнению с традиционными в нелинейной обратной задаче в присутствии шумов. Он требует только n вычислений функционала невязки на каждой итерации (в сравнении с 2^n вычислений в методах, использующих первые частные производ-

ные). Последовательное применение метода с разным значением шага сетки позволяет обеспечить необходимую точность, а возможность работы с большим шагом позволяет достаточно быстро проводить поиск глобального минимума на всей поверхности для функционалов, являющихся слишком нелинейными и имеющих большую невязку. Важным преимуществом нового алгоритма является простота и однозначность критерия останова, что обычно представляет определенную проблему в классических алгоритмах.

Описанный метод фитирования реализован в рамках комплекса вторичной обработки спектров мощности радара НР ИСЗФ. Высокая скорость работы и гибкость алгоритма в выборе шага сетки позволяет проводить повторные или последовательные итерации поиска решения задачи минимизации (1) и эффективно корректировать точность фитирования. Для программирования алгоритм удобен, так как обладает простотой и единообразием построения симплексов высших порядков.

Работа выполнена при поддержке гранта РФФИ №08-05-00618-а.

СПИСОК ЛИТЕРАТУРЫ

Жеребцов Г.А., Заворин А.В., Медведев А.В. и др. Иркутский радар некогерентного рассеяния // Радиотехника и электроника. 2002. Т. 47, № 11. С. 1339–1345.

Шеффилд Дж. Рассеяние электромагнитного излучения в плазме, М: Атомиздат, 1978. 280 с.

Bard Y. Nonlinear Parameter Estimation. New York Academic Press, 1970. N 4. P. 753–758.

Dennis J.E., Jr., Schnabel R.B. Numerical Methods for Unconstrained Optimization and Nonlinear Equations, Prentice-Hall, Inc. 1983. 440 p.

Erickson P.J. Observation of light ions in the midlatitude and equatorial topside ionosphere, Ph. D. Thesis, Graduate School of Cornell University, 1998.

Faley D.T. A theory of incoherent scattering of radio waves by a plasma // J. Geophys. Res. 1966. V. 71, N. 17. P. 4091–4098.

Hagen J.B., Hsu P.Y. The structure of the protonosphere above Arecibo // J. Geophys. Res. 1974. V. 79 P. 4269–4275.

Holt, Rhoda, Tetenbaum, Van Eyken. Optimal analysis of incoherent scatter data // Rad. Sci. 1992. V. 27. P. 435.

Hamming R.W. Numerical Methods for Scientists and Engineers. New York: McGraw- Hill, 1972. 721 p.

Lehtinen M.S., Huuskonen A. General incoherent scatter analysis and GUISDAP // Journal of Atmospheric and Terrestrial Physics. 1996. V. 58, N 1–4. P. 435–452.

Nikoukar R., Kamalabadi F., Kudeki E., Sulzer M. An efficient near-optimal approach to incoherent scatter radar parameter estimation // Rad. Sci. 2008. V. 43. P. RS2007. doi:10.1029/2007RS003724.

Swartz W.E. Analytic partial derivatives for least-squares fitting incoherent scatter data // Rad. Sci. 1978. V. 13. P. 581–589.

Shpynev B.G. Incoherent scatter Faraday rotation measurements on a radar with single linear polarization // Rad. Sci. 2004. V. 39, N. 3. P. RS3001. doi: 10.1029/2001RS002523.

Vallinkoski. Statistics of incoherent scatter multiparameter fits // JATP. 1988. V. 50. P. 839.