

ОБ ОДНОМ ПОДХОДЕ К РЕШЕНИЮ НЕЛИНЕЙНЫХ ПАРАМЕТРИЧЕСКИХ ЗАДАЧ

Л. Л. ФРУМИН

Новосибирский государственный университет, Россия

e-mail: frumin@phys.nsu.ru

An approach to obtaining linear estimations of parameters of the function, which approximates an experimental data, is presented. The approach is based on the consideration of the equations, with this function as the solution. The examples of parametric problems, including problem of estimation of parameters of some higher transcendental functions, are presented.

1. Постановка задачи

Обработка экспериментальных данных имеет особое значение для практики естественных наук и часто приводит к обратным задачам определения параметров некоторой функции по экспериментально измеренным ее значениям.

Интерес к параметрическим задачам в последнее время во многом обусловлен прогрессом аппаратных цифровых методов и устройств, которые позволяют проводить быструю, массовую и эффективную обработку данных эксперимента уже на этапе сбора и первичной обработки информации.

Параметрические задачи обычно трактуют в следующей математической постановке.

Пусть $\lambda = \{\lambda_j\}, j = 1, 2, \dots, m$ — искомый вектор параметров определенной функции $y = f(\lambda, t)$, где t — сканируемая переменная. Имеется набор экспериментальных точек $t_i, i = 1, 2, \dots, N, N \geq m$, в которых с некоторой погрешностью, присущей данному эксперименту, измерены значения функции $y_i = f(\lambda, t_i)$. Требуется по данным эксперимента найти вектор параметров λ .

Традиционный и наиболее распространенный способ оценки параметров основан на методе наименьших квадратов и сводится к задаче минимизации функционала наименьших квадратов (НК) $M[f]$, который, вообще говоря, нелинейно зависит от вектора параметров. Численные реализации такого подхода, как правило, приводят к сравнительно громоздким итерационным численным методам минимизации [1–3]. Обширный объем вычислений в нелинейном методе НК обычно не допускает аппаратной реализации таких алгоритмов. Еще более серьезное ограничение нелинейного метода НК связано с тем, что случайная погрешность эксперимента в нелинейных задачах приводит к появлению побочных локальных минимумов функционала НК. При этом уравнения Эйлера для задачи минимизации уже не имеют единственного решения. Поиск глобального минимума функционала НК не только усложняет задачу, но и не гарантирует состоятельности получаемых оценок параметров.

Недостатки нелинейного метода НК стимулируют поиск и развитие других, в особенности линейных, подходов в задаче оценки параметров. Среди последних прежде всего отметим простой, но весьма распространенный подход, основанный на сведении задачи к линейной путем использования обратной функции. Этот подход обычно ограничен сравнительно узким классом задач, где вектор параметров входит в качестве линейных коэффициентов некоторого полинома. Причем этот полином является единственным аргументом измеряемой функции, монотонно зависящей от своего аргумента.

Следует отметить также метод моментов, в котором m моментов функции связывают с искомым вектором параметров и затем решают получившуюся систему алгебраических уравнений. И в этом случае имеются существенные ограничения, заметно сужающие область применимости метода. Искомые моменты далеко не всегда можно найти в явном виде, а система уравнений для моментов часто сама оказывается нелинейной. Заметим, наконец, что метод моментов не приводит к вариационной задаче поиска минимума функционала НК и получаемые с его помощью оценки трудно назвать оптимальными.

Настоящая работа посвящена новому способу решения параметрических задач, использующему некоторое операторное обобщение линейного метода НК и его разностную аппроксимацию. Отсутствие итераций и сложных алгоритмов принятия решения допускает аппаратную реализацию этого подхода, что может представлять значительный интерес для практики.

2. Метод решения

Идея предлагаемого метода оценки параметров основана на том, что для любой функции $f(\lambda, t)$ всегда существует (и не одно!) уравнение, решением которого она является. Если искомые параметры, от которых зависит эта функция, линейно входят в уравнение, то их можно найти с помощью линейного метода НК путем минимизации невязки конечно-разностной аппроксимации этого уравнения на множестве экспериментальных данных.

Прототипом этого метода послужил алгоритм барона де Прони 1795 года [4]. Де Прони заметил, что решением линейных разностных уравнений с постоянными коэффициентами является сумма экспонент. В алгоритме де Прони показатели экспонент отыскиваются непосредственно из разностного уравнения. В последнее время метод де Прони был обобщен на случай комплексных экспонент [5]. В современном варианте метода де Прони для учета избыточной информации используется метод наименьших квадратов.

Рассмотрим класс функций, описываемых операторным уравнением довольно общего вида, в которое вектор параметров λ входит в качестве вектора коэффициентов:

$$\sum_{j=1}^m \lambda_j P^j[f, t] = Q[f, t], \quad j = 1, 2, \dots, m. \quad (1)$$

Входящие в уравнение операторы P^j и Q зависят от сканируемой переменной t как от параметра и могут быть линейными или нелинейными алгебраическими, дифференциальными или интегральными, скалярными или векторными операторами.

Операторы P^j должны быть линейно независимы. Это условие необходимо для разрешимости задачи поиска параметров. Для частного случая линейных операторов оно является и достаточным.

Для всех операторов P^j и Q потребуем существование разностных аппроксимаций их термов $P^j[f, t_i]$ и $Q[f, t_i]$, которые мы будем обозначать как D_i^j и R_i соответственно. Эти

аппроксимации определены на некотором подмножестве Θ множества дискретных значений сканируемой переменной $\{t_i\}$. Подмножество Θ — область определения дискретной аппроксимации уравнения, зависит от вида операторов и схемы их дискретизации и может не совпадать со всем сеточным множеством, состоящим из N точек, но должно содержать, по крайней мере, $N_\Theta \geq m$ значений.

Кроме того, будем предполагать невырожденность симметричной $m \times m$ -матрицы A , компоненты которой даются суммой попарных произведений разностных аппроксимаций термов D :

$$A_{jk} = \sum_i D_i^j D_i^k. \quad (2)$$

Сумма здесь берется по дискретной области определения разностной аппроксимации уравнения Θ . Разностную аппроксимацию уравнения (1) можно теперь представить в виде

$$\sum_{j=1}^m \lambda_j D_i^j = R_i. \quad (3)$$

Для определения оценок вектора параметров λ потребуем минимума функционала НК (невязки) для уравнения (3):

$$M = \sum_i \left(\sum_{j=1}^m \lambda_j D_i^j - R_i \right)^2. \quad (4)$$

Минимизация функционала НК (4) требует обращения в нуль производных $\partial M / \partial \lambda_k$ и приводит к следующим уравнениям Эйлера для определения вектора λ :

$$\sum_i \sum_{j=1}^m \lambda_j D_i^j D_i^k = \sum_i D_i^k R_i, \quad k = 1, 2, \dots, m. \quad (5)$$

Решение системы уравнений (5) можно представить в матричном виде

$$\lambda = A^{-1} \mathbf{b}. \quad (6)$$

Здесь симметричная и невырожденная по предположению матрица A дается уравнением (2), а вектор \mathbf{b} описывается следующим выражением:

$$b_k = \sum_i D_i^k R_i. \quad (7)$$

Соотношение (6) и дает решение задачи нахождения оценок параметров.

Что касается точности полученных оценок параметров, то она зависит от ряда факторов, таких как уровень погрешности эксперимента и ее статистика, свойства операторов задачи и порядок их конечно-разностной аппроксимации. Вероятно, наиболее общий подход в исследовании точности состоит в численном моделировании задачи.

В том случае, когда операторы P представляют собой алгебраические (полиномиальные) выражения, решение (6) совпадает с методом де Прони [5]. Наше обобщение этого решения на случай дифференциальных или интегральных операторных уравнений существенно расширяет класс задач, где можно использовать простые линейные оценки метода НК. Чтобы убедиться в этом, изучим далее ряд примеров.

Пример 1. Рассмотрим задачу определения частоты Ω гармонических колебаний по экспериментально полученному ряду значений синусоидальной функции $f_i = f(t_i) = A \sin(\Omega t_i + \phi)$, $i = 1, 2, \dots, N$. Нелинейный метод НК для определения частоты потребовал бы решения 3-параметрической задачи поиска сразу трех параметров — Ω , A , ϕ . Однако во многих прикладных задачах амплитуда колебания A и фаза ϕ не представляют интереса для исследователя. Модуль или квадрат частоты Ω часто и есть та единственная величина, оценку которой требуется найти.

Функция $f(t, \Omega) = A \sin(\Omega t + \phi)$, очевидно, удовлетворяет следующему дифференциальному уравнению:

$$\frac{d^2 f}{dt^2} + \lambda f = 0, \quad (8)$$

где $\lambda = \Omega^2$ и есть параметр, подлежащий определению. Сравнивая это уравнение с общим уравнением (1), мы видим, что оператор P^1 представляет собой единичный оператор (умножение функции f на 1), а оператор Q есть дифференциальный оператор второй производной.

Предположим для простоты равномерность сетки $\{t_i = i\tau\}$, $i = 1, 2, \dots, N$, с шагом τ и рассмотрим на этой сетке некоторую конечно-разностную аппроксимацию уравнения (8). Эта аппроксимация определяется, в первую очередь, аппроксимацией дифференциального оператора второй производной, которую мы пока не будем конкретизировать и обозначим как f_i'' . Разностную аппроксимацию уравнения (8) запишем в виде

$$f_i'' + \lambda f_i = 0. \quad (9)$$

Далее запишем функционал наименьших квадратов, представляющий собой сумму квадратов отклонений (невязки) разностного уравнения (9) по области определения разностной аппроксимации:

$$M = \sum_i (f_i'' + \lambda f_i)^2. \quad (10)$$

Минимизируя M по параметру λ , получим следующее уравнение Эйлера:

$$\sum_i f_i f_i'' + \lambda \sum_i f_i^2 = 0. \quad (11)$$

Решение уравнения (11) нетрудно представить в виде

$$\lambda = -\frac{\sum_i f_i f_i''}{\sum_i f_i^2}. \quad (12)$$

Последнее выражение показывает, что оценка квадрата частоты определяется конечно-разностной аппроксимацией второй производной. Точность получаемых оценок зависит от выбранного способа численного дифференцирования. При численном дифференцировании функции с шумами погрешность результата может сильно возрасти.

Задача численного дифференцирования давно и хорошо изучена. Разработано несколько подходов решения этой задачи на основе интерполяционных формул или путем предварительного сглаживания или фильтрации высокочастотной компоненты шума. Известно, что при данном уровне шумов существует оптимальная реализация численного дифференцирования, дающая минимальную погрешность решения [6].

3. Численное моделирование

Для проверки подхода проводилось численное моделирование задачи определения частоты на основе выражения (12). Моделирование осуществлялось по так называемой схеме “замкнутого цикла”. При этом для заданных частоты, фазы и амплитуды строилась синусоидальная функция. Затем к ней добавлялся гауссов случайный шум с постоянной дисперсией, имитирующий ошибку эксперимента. Полученные “данные эксперимента” сглаживались с помощью прямоугольного фильтра шириной p точек. Для оценки частоты использовались выражение (12) и простейшая центрированная разностная аппроксимация второй производной второго порядка точности $-f_i'' = (f_{i+1} - 2f_i + f_{i-1})/\tau^2$.

В ходе численного моделирования исследовался вопрос выбора оптимального размера прямоугольного фильтра. Для заданного размера фильтра p определялось стандартное отклонение s для ошибки оценки частоты $\Omega = \sqrt{\lambda}$. В численном моделировании использовался один период синусоидальной функции. Расчетная сетка содержала 250 узлов.

Результат такого моделирования в виде графика зависимости стандартного отклонения ошибки оценки частоты от ширины фильтра для 50 % уровня шумов эксперимента представлен на рисунке. Из графика видно, что существует набор оптимальных ширин фильтра в интервале $p \in [40 - 100]$. Оценка стандартного отклонения ошибки определения частоты проводилась по выборке из 400 независимых статистических реализаций “случайного” шума.

Расчеты показали, что численная реализация алгоритма для задачи оценки частоты на основе выражения (12) и подходящего алгоритма численного дифференцирования не представляет особых трудностей. Алгоритм содержит только простейшие арифметические операции, и его можно использовать для построения аппаратных способов оценки частоты с помощью простого программного автомата.

Еще раз подчеркнем, что полученная оценка (квадрата) частоты Ω не требует одновременной оценки амплитуды A и фазы ϕ , что заметно отличает предложенный подход от нелинейного метода НК.

Возможность исключить лишние параметры из рассмотрения связана с тем, что мы исследуем операторное, дифференциальное, а не алгебраическое уравнение и “лишние” параметры являются просто его константами интегрирования.

Выбор подходящего уравнения для искомой функции позволяет включить или исключить из рассмотрения некоторые из параметров, понижая тем самым сложность или размерность задачи. Например, предположим, что нас интересует одновременное определение

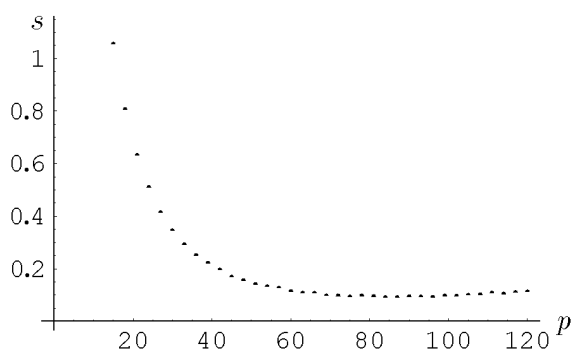


Рис. 1. Зависимость относительного стандартного отклонения s оценки частоты для уравнения (12) от ширины прямоугольного фильтра p .

частоты и амплитуды синусоиды.

Рассмотрим следующее нелинейное уравнение первого порядка, которому удовлетворяет синусоидальная функция

$$f^2 + \left(\frac{df}{dt}\right)^2 / \Omega^2 = A^2. \quad (13)$$

Вводя вектор параметров λ с компонентами $\lambda_1 = A^2$ и $\lambda_2 = 1/\Omega^2$, представим уравнение (13) в виде

$$\lambda_1 - \lambda_2 \left(\frac{df}{dt}\right)^2 = f^2. \quad (14)$$

Сравнивая уравнение (14) и уравнение (1), мы видим, что $P^1 = 1$, $P^2 = \left(\frac{df}{dt}\right)^2$, $Q = f^2$, и решение задачи оценки параметров дается уравнением (6).

Приведенный пример демонстрирует гибкость и эффективность нашего обобщения метода НК. В то же время он показывает, что этот подход требует известной изобретательности при поиске подходящего уравнения в рутинной, по сути, задаче оценки параметров.

Пример 2. Другой интересный пример представляет популярная в приложениях задача восстановления параметров положения центра λ_2 и квадрата ширины λ_1 функции Гаусса

$$f(t) = A \exp \left[-\frac{(t - \lambda_2)^2}{2\lambda_1} \right]. \quad (15)$$

Дифференциальное уравнение, которому удовлетворяет эта функция, можно получить, дифференцируя функцию (15) по параметру t . В результате мы приходим к уравнению

$$\lambda_1 \frac{df}{dt} - \lambda_2 f = -t f. \quad (16)$$

Обозначим f'_i разностную аппроксимацию первой производной и запишем дискретный аналог уравнения (16) в виде

$$\lambda_1 f'_i - \lambda_2 f_i = -t_i f_i, \quad (17)$$

а функционал наименьших квадратов — в виде

$$\sum_i [\lambda_1 f'_i + (t_i - \lambda_2) f_i]^2. \quad (18)$$

Минимизация функционала (18) приводит к следующим выражениям для параметров:

$$\lambda_1 = \frac{\sum_i t_i f_i^2 \sum_k f'_k f_k - \sum_i t_i f_i f'_i \sum_k f_k^2}{\left(\sum_i f_i f'_i\right)^2 - \sum_i f_i'^2 \sum_k f_k'^2}, \quad (19)$$

$$\lambda_2 = \frac{\sum_i t_i f_i^2 \sum_k f_k'^2 - \sum_i t_i f_i f'_i \sum_k f_k f'_k}{\left(\sum_i f_i f'_i\right)^2 - \sum_i f_i'^2 \sum_k f_k'^2}. \quad (20)$$

Пример 3. Особый интерес для практических приложений имеет оценка параметров специальных функций, которые нередко встречаются в практике обработки эксперимента. Параметрические задачи, приводящие к специальным функциям, таким как функции Бесселя, гипергеометрические функции и другим, часто остаются без ответа в силу относительной сложности их свойств. Ниже будет показано, что предложенный метод решения параметрических задач позволяет оценивать параметры специальных функций с помощью простейших арифметических операций, которые могут быть реализованы аппаратно.

Рассмотрим в качестве примера произвольную цилиндрическую функцию $f(t) = Z_\nu(\kappa t)$ порядка ν . Требуется по множеству экспериментально полученных значений функции определить параметр $\lambda = \kappa^2$.

Цилиндрическая функция удовлетворяет дифференциальному уравнению Бесселя

$$t^2 \frac{d^2 f}{dt^2} + t \frac{df}{dt} + (\lambda t^2 - \nu) f = 0. \quad (21)$$

Выбирая подходящую разностную схему и обозначая разностную аппроксимацию первой и второй производных функции как f'_i и f''_i , оценку параметра λ , минимизирующую норму невязки разностной аппроксимации уравнения (21), можно получить по следующей формуле:

$$\lambda = - \frac{\left(\sum_i t_i^4 f''_i f_i + t_i^3 f'_i f_i - \nu t_i^2 f_i^2 \right)}{\sum_i t_i^4 f_i^2}. \quad (22)$$

Заметим также, что дифференциальное уравнение Бесселя (21) можно использовать и для двухпараметрической задачи одновременного поиска параметров λ и ν .

Аналогичное рассмотрение можно провести для гипергеометрической функции и целого ряда других специальных функций.

Заключительные замечания

В заключение коснемся вопроса о статистических свойствах получаемых оценок. Эти свойства определяются, в первую очередь, точностью разностных аппроксимаций операторов задачи. В случае дифференциальных операторов, которые в основном и рассматривались выше, их разностные аппроксимации приводят к смещенным оценкам искомых параметров. Это можно показать, воспользовавшись дифференциальным приближением Шокина — Яненко [7]. Действительно, пусть для некоторого дифференциального оператора задачи P используется разностная аппроксимация D порядка k . Разлагая эту аппроксимацию в ряд и переходя к дифференциальному оператору, в первом дифференциальном приближении мы получим добавку к исходному оператору P порядка k . Даже если погрешность эксперимента будет равна нулю, полученная нами оценка будет смещенной ввиду неточности разностной аппроксимации дифференциального оператора.

Проведенное схематическое рассмотрение показывает, что по крайней мере для дифференциальных уравнений получаемые оценки оказываются смещенными. Это смещение определяется шагом дискретизации и уменьшается с увеличением числа экспериментальных точек и уменьшением шага дискретной сетки. Вместе с тем, точность получаемых оценок, безусловно, зависит от качества аппроксимации. Более высокий порядок аппроксимации, разумеется, повышает точность оценок, однако при этом может уменьшиться

объем статистической выборки, так как чем выше порядок аппроксимации, тем больше точек на границах перестают участвовать в оценке параметров. К примеру, первый порядок аппроксимации приводит к потере минимум одной граничной точки, где дискретная аппроксимация не определена. Второй порядок может приводить уже к потере не менее двух граничных точек и т. д.

Несмотря на смещенность получаемых оценок, в целом предложенный подход представляет собой хорошее дополнение к существующим методам оценки параметров и может найти широкое применение в прикладных задачах автоматизации сбора и обработки данных.

Автор благодарит А. Г. Игнатенко, обратившего внимание на первые приложения представленного подхода и отметившего его сходство с алгоритмом де Прони, а также В. В. Пая и Л. Б. Чубарова за полезные обсуждения, ценные советы и интерес к этой работе.

Список литературы

- [1] ЛОУСОН Ч., ХЕНТОН Р. Численное решение задач метода наименьших квадратов. М.: Наука, 1986.
- [2] ОРТЕГА Д., РЕЙНБОЛДТ В. Итерационные методы решения систем уравнений со многими неизвестными. М.: Мир, 1975
- [3] ВАСИЛЬЕВ Ф. П. Методы решения экстремальных задач. М.: Наука, 1981
- [4] DE PRONY (GASPAR RISCHE). Essay expérimental et analitique: sur les lois de la dilatabilité de fluides élastiques et sur celles de la force expansive de la vapeur de l'eau et de la vapeur de l'alkool, à différentes températures // J. E. Polytech. 1795. Vol. 1, No. 2. P. 24–76.
- [5] МАРПЛ мл. С. Л. Цифровой спектральный анализ и его приложения. М.: Мир, 1990 (гл. 11).
- [6] БАХВАЛОВ Н. С., ЖИДКОВ Н. П., КОБЕЛЬКОВ Г. М. Численные методы. М., СПб.: Физматлит, 2001.
- [7] ШОКИН Ю. И. Метод дифференциального приближения. Новосибирск: Наука, 1979.

*Поступила в редакцию 18 апреля 2002 г.,
в переработанном виде — 7 мая 2002 г.*