

01; 12

© 1992

НЕКОТОРЫЕ МЕТОДЫ В НЕЛИНЕЙНОМ РЕГРЕССИОННОМ АНАЛИЗЕ

А.А. Г а л и ц к а с

В в е д е н и е

Большинство физических явлений описываются нелинейными относительно своих параметров математическими моделями. Оценка этих параметров по эмпирическим данным, или так называемый регрессионный анализ, относится к числу наиболее актуальных и также весьма сложных проблем. Несмотря на многочисленные работы по регрессионному анализу, до сих пор не имеется общей и эффективной методики вычисления параметров многомерных нелинейных моделей. Так как задача не имеет аналитического решения, используются поисковые итерационные вычислительные процедуры. Характерные трудности, возникающие при таком анализе, следующие: 1) отсутствие гарантии сходимости вычислений; 2) слишком медленная сходимость. Важно отметить, что даже в тех случаях, когда сходимость гарантируется теорией, практически она не всегда реализуется из-за слишком большого числа итераций, или из-за чрезмерного возрастания значений переменных в процессе вычислений. Методы и алгоритмы совершенствуются и развиваются, но общего метода, совершенного во всех отношениях, не существует.

В данной работе предлагаются вычислительные приемы общего характера, способствующие сходимости вычислительного процесса, и приводятся результаты численных экспериментов, подтверждающие формальные рассуждения.

О с н о в н ы е п о л о ж е н и я

Пусть исследуемое явление описывается некоторой моделью $Q(X, P)$, где $P = (P_1, \dots, P_M)$ — набор параметров, относительно которых она нелинейная. Задача заключается в определении этих параметров по измерительным данным $Y = (Y_1, \dots, Y_N)$, полученным на заданном множестве значений $X = (X_1, \dots, X_N)$ предиктной (управляемой) переменной X . Параметры вычисляются по критерию минимума среднеквадратической ошибки, т.е. задача сводится к минимизации целевой функции

$$\sigma(P) = \sum_{n=1}^N (Y_n - Q(X_n, P))^2 \quad (1)$$

по вектору P . Такого рода задачи решаются итерационными способами, т.е. параметры на текущей итерации вычисляются согласно алгоритму

$$P_{next} = P + \Delta P, \quad P = P_{next}. \quad (2)$$

Методика и критерии, по которым определяется поправка к параметрам ΔP , определяют и методику регрессионного анализа в целом [1, 2]. Существует несколько основных методик анализа. В данной работе за основу принимается широко применяемый метод Гаусса-Ньютона [1], согласно которому вектор коррекции вычисляется следующим образом:

$$\Delta P = \lambda (FF^T)^{-1} F \Delta Y, \quad (3)$$

где $\lambda < 1$ - коэффициент, определяющий длину шага коррекции, прямоугольная матрица

$$F = \left\{ \frac{\partial Q(X_n, P)}{\partial P_m} \right\}, \quad n=1, \dots, N, \quad m=1, \dots, M \quad (4)$$

составлена из производных модели по всем параметрам по всех точках измерения, F^T - матрица, транспонированная по отношению к F , а вектор невязки

$$\Delta Y = \{ Y_n - Q(X_n, P) \}, \quad n=1, \dots, N. \quad (5)$$

Ниже предлагаются проверенные на ЭВМ вычислительные приемы, оказавшиеся эффективными в нелинейном регрессионном анализе.

Р а н д о м и з а ц и я в ы б о р а т о ч е к н а б л ю д е н и я

Первая нами предлагаемая идея - это рандомизация выбора измерительных точек на каждой итерации. Точки анализа выбираются из полной их совокупности [1, ... N] случайным образом согласно равномерному закону распределения. Их число R должно быть не меньшем числа параметров M , но не может превышать полного числа измерений N . Нами принято среднее геометрическое этих пределов: $R = (MN)^{1/2}$. Таким образом, каждая итерация реализуется на новой, случайно выбранной и не полной совокупности измерительных точек. Такая процедура вычислений логически обоснована, так как любая, даже случайная выборка точек, определяет некоторую совокупность подлинных экспериментальных данных, относящихся к одной и той же исследуемой модели, поэтому и поправка к параметрам, вычисленная согласно (3), при каждой итерации является правомерной. Благодаря этому методическому приему объем вычислений на каждой итерации сокращается в $(N/M)^{1/2}$ раз, кроме этого, на вычислительный процесс мало влияет „канавы” и другие тупиковые ситуации, которые могут

встретится в анализируемой многомерной модели. При этом окрестность минимума, как показали эксперименты на компьютере, достигается за нескольких итераций. По заданному условию рандомизированный режим автоматически переключается на режим с полным набором измерительных точек, а вычисления завершаются по заданному критерию останова.

Выделение линейного этапа

Нелинейная в целом модель может содержать и параметры, относительно которых она линейная. В этом случае каждую итерацию целесообразно выполнить в два этапа. Первый – это выполнение вычислений только по тем параметрам, относительно которых модель линейная, фиксируя значения остальных параметров. Из теории известно [3], что минимум целевой функции в линейных случаях достигается сразу, при этом следует принять $\lambda = 1$. Во втором этапе текущей итерации, коррекция выполняется по полному набору параметров, а шаг λ вычисляется согласно принятой методике. Описанная процедура значительно ускоряет процесс сходимости, кроме этого, позволяет избежать чрезмерного возрастания переменных и превышения числовых пределов процессора.

Уместно отметить, что даже в тех случаях, когда физическая модель полностью нелинейная, практически всегда необходимо искусственно ввести два параметра, по которым модель линейная – это масштабный коэффициент C и параметр H , определяющий некоторый постоянный фон, на уровне которого проводятся измерения. Эти параметры не имеют физического смысла, но необходимы ввиду возможной неопределенности масштаба или сдвига „нуля“ при измерениях. Их учет может привести к большим ошибкам. Таким образом, анализируемую модель следует представить в виде

$$Q(X, P) = Cq(X, P) + H, \quad (6)$$

где $q(X, P)$ – настоящая физическая модель.

Вычисление шага

Как отмечалось выше, анализ нелинейной по параметрам модели выполняется согласно (3) и при $\lambda < 1$. В этом методе, который носит название демпфированного метода Гаусса–Ньютона, существенным является подходящий подбор или вычисление шага λ . Широко известен способ определения шага λ путем дробления (начиная с $\lambda = 1$) [1] до тех пор, пока коррекция ΔP не превышает некоторого заданного уровня. Но это – способ проб и ошибок, требующий много вычислений, и поэтому малоэффективен.

Здесь предлагается способ вычисления λ , учитывающий два существенных фактора: число параметров модели и степень приближения кривой регрессии к измерительным точкам. Некоторые формальные рассуждения, а также вычислительные эксперименты привели к следующей формуле для расчета шага:

$$\lambda = \exp(-M\eta), \quad (7)$$

где M — полное число параметров, а η — нормированное отклонение кривой регрессии от точек наблюдения:

$$\eta = b(p) / \sum_{n=1}^N \gamma_n^2. \quad (8)$$

В начале анализа при большом отклонении η вероятность расходимости вычислений большая, а вблизи минимума почти обеспечена сходимость. Исходя из этого положения, коэффициент в начале анализа мал, а по мере приближения к минимуму — увеличивается, ускоряя локальную сходимость.

Реализация

Описанные вычислительные приемы были реализованы и проверены на соответствующей программе. Исследовалась конкретная физическая модель, содержащая 7 параметров, по 3-ем из которых она нелинейная. На базе подлинных данных, полученных в результате экспериментального исследования, был проведен довольно полный и всесторонний анализ этой модели. Число измерительных точек принималось в пределах 100–250. Сходимость вычислений и определение оптимальных значений параметров достигалось всего за 5–20 итераций и занимало 1–3 минуты вычислений на компьютере IBM PC AT без сопроцессора при тактовой частоте 12 МГц. В качестве критерия останова использовано известное [4] правило, по которому сходимость считается достигнутой, если изменения вычисляемых параметров достигают заданного, достаточно малого значения.

Кроме описанного анализа, нами были проведены вычислительные эксперименты на широком спектре других моделей при различных условиях, которые доказали эффективность предлагаемых приемов. По сравнению с известными вычислительными процедурами, применяемыми в рамках метода Гаусса–Ньютона, предлагаемые процедуры обеспечивают сходимость вычислений в 3–5 раз быстрее того, с их помощью достигается сходимость в ряде случаев, когда известные способы оказываются бесплодными. Например, без осложнений нами была проанализирована модель, содержащая 12 параметров, по 8-ми из которых она была нелинейной.

Список литературы

- [1] Деннис Дж., Шнабель Р. Численные методы безусловной оптимизации и решения нелинейных уравнений. М.: Мир, 1988.

- [2] Дрейпер Н., Смит Г. Прикладной регрессионный анализ. М.: Финансы и статистика, 1987.
- [3] Федоров В.В. Теория оптимального эксперимента. М.: Наука, 1971.
- [4] Бард Й. Нелинейное оценивание параметров. М.: Статистика, 1979.

Институт физики
полупроводников
АН Литвы,
Вильнюс

Поступило в Редакцию
26 марта 1992 г.