

ляется нежелательным для устройств с большим значением этой величины. Неэффективность системы в этом случае следует из того, что фотоны от всех частей АОЭ смешиваются в каждом из детекторов, и в результате относительный вклад от одного сегмента в приращение ФР становится незначительным и сравнимым с величиной «квантовых» пульсаций. Левая часть неравенства соответствует эффективному числу фотонов, приходящих на один сегмент в пространстве и во времени. Экспоненциальный множитель может интерпретироваться как коэффициент поглощения фотонов, зависящий от соотношения параметров турбулентной среды и самой системы, причем наиболее оптимальное их соотношение с энергетической точки зрения достигается при $t_0 \simeq \rho_0$.

В заключение отметим, что приведенное условие в некоторой степени зависит от типа алгоритма функционирования. Синтез и реализация алгоритмов, позволяющих проводить уже не последовательную, а параллельную (одновременную) подстройку сегментов АОЭ, позволит в будущем смягчить требования, предъявляемые к быстродействию АОЭ и энергетическим характеристикам сигнала.

ЛИТЕРАТУРА

1. Бакут П. А. и др. Методы обработки световых полей при наблюдении объектов через турбулентную среду.— Зарубеж. радиоэлектроника, 1977, № 1.
2. Бакут П. А., Троицкий И. Н., Харитонова О. И. Адаптивные методы улучшения качества оптических изображений.— Автометрия, 1979, № 4.
3. Muller R. A., Buffington A. Real-Time Correction of Atmospherically Degraded Telescope Images through Image Sharpening.— JOSA, 1974, vol. 64, p. 1200.
4. Бакут П. А. и др. Исследование оптимальных условий регистрации голограмм интенсивности и оптических изображений.— Квант. электроника, 1975, т. 2, № 8.
5. Троицкий И. Н. Выбор отношения сигнал/шум и размера считывающей апертуры в радиографии.— Дефектоскопия, 1974, № 2.
6. Lutomirski R., Yura H. Propagation of a Finite Optical Beam in an Inhomogeneous Medium.— Appl. Opt., 1971, vol. 10, p. 1652.
7. Fante R. L. Some Results on the Effect of Turbulence on Phase-Compensated Systems.— JOSA, 1976, vol. 66, p. 730.
8. Миддлтон Д. Введение в статистическую теорию связи. М.: Сов. радио, 1961.

Поступила в редакцию 4 апреля 1979 г.

УДК 519.853

Г. Д. РОДИОНОВ, М. Т. ХОДОНОВ, Б. М. ШТОКМАН, М. И. ШТОКМАН
(Новосибирск)

ЭФФЕКТИВНЫЙ АЛГОРИТМ НЕЛИНЕЙНОЙ ГЛОБАЛЬНОЙ МИНИМИЗАЦИИ ПРИ НАЛИЧИИ ОГРАНИЧЕНИЙ

1. **Введение.** Задача о нелинейной глобальной минимизации (НГМ) является одной из самых часто встречающихся в прикладной математике и программировании. Для определенности сформулируем задачу НГМ в следующем достаточно общем виде. Пусть имеется функция $f(\lambda)$, зависящая нелинейно от N -мерного вектора параметров λ . Необходимо найти минимумы этой функции $f^{(k)}$ ($k=0, 1, \dots$) и соответствующие стационарные векторы $\lambda^{(k)}$, среди которых должен быть глобальный минимум $f^{(0)}$:

$$f^{(k)} = f(\lambda^{(k)}), \quad k = 0, \dots, n_{\min} - 1. \quad (1)$$

Минимизация функции проводится при наложенных дополнительных условиях:

$$F_i(\lambda) = 0, \quad i = 1, \dots, m. \quad (2)$$

В частности, к этой задаче сводится решение систем нелинейных уравнений, включая переопределенные и несовместные системы, а также задачи подгонки кривых. В экспериментальной физике проблема НГМ возникает при параметрической обработке экспериментальных данных с целью нахождения значений физических параметров, в вычислительной физике к НГМ сводятся если не все, то большинство вариационных задач.

Вполне естественно поэтому, что решению проблемы НГМ посвящено большое число работ (см., например, [1—7]). Если при этом не принимать во внимание эффективность разрабатываемых алгоритмов, то задача НГМ становится тривиальной и решается с необходимой точностью простым перебором возможностей, так что главным вопросом остается снижение затрат машинного времени. В случае большого числа параметров N количество возможностей и соответствующие затраты машинного времени T при простом переборе являются астрономическими, согласно оценке

$$T \sim T_0 n^N, \quad (3)$$

где n — число градаций по каждому параметру; T_0 — время, необходимое для однократного вычисления функции f .

Описываемый алгоритм позволяет найти для большинства задач если не все, то многие локальные минимумы, а в ряде случаев и глобальный минимум при разумном времени счета, для которого справедлива оценка

$$T \approx T_0(N + \log_2 n)[k_1 + k_2 N]. \quad (4)$$

Здесь k_1, k_2 — константы, зависящие от задачи (ср. с приложением).

Мы не ставили своей целью вывести строгие условия, в которых наш алгоритм находит все локальные минимумы или в которых среди найденных локальных минимумов будет глобальный. Исследование условий существования экстремумов в задачах невыпуклого программирования с ограничениями, а также вопросов сходимости алгоритмов можно найти в работах [1, 2, 8, 9]. Эффективность и устойчивость нашего метода были подтверждены опытом его применения для многих задач.

Ниже в разделах 2 и 3 будут изложены алгоритмы локальной и глобальной минимизаций, а в приложении рассмотрен пример многопараметрической задачи.

2. Алгоритм локальной минимизации. Под локальной минимизацией будет пониматься такая процедура, которая по исходному вектору параметров λ_0 находит достаточно близкий стационарный вектор λ_{oi} . Для осуществления нелинейной локальной минимизации (НЛМ) известен ряд алгоритмов, в частности метод скорейшего спуска и метод Ньютона (см., например, [1, 2, 5]). В этих методах вводится функция скалярного аргумента l

$$\varphi(l) = f(\lambda_0 + l\Lambda), \quad (5)$$

где вектор Λ линейно связан с градиентом функции f в точке λ_0 :

$$\Lambda_i = \sum_{j=1}^N S_{ij} G_j, \quad G \equiv \partial f / \partial \lambda |_{\lambda=\lambda_0}. \quad (6)$$

Далее находится минимум функции φ по l в окрестности точки $l = 0$, что дает значение следующего приближения для вектора λ_0 . Затем описанная процедура итерируется. Обычно проводят такой поиск минимума вдоль направления Λ (5) в N -мерном параллелепипеде:

$$F_i(\lambda) = \theta((\lambda_i - a_i)(b_i - \lambda_i)), \quad (7)$$

т. е. выбирают функции $F_i(2)$ в виде ступенчатых функций, где a_i, b_i — границы изменения параметра λ_i .

В методе скорейшего спуска $S_{ij} = \delta_{ij}$, в методе Ньютона S_{ij} определяются через обращенную матрицу вторых производных функции f . Метод Ньютона недостаточно эффективен, так как требует вычисления $N(N+3)/2$ частных производных функции f на одну итерацию по сравнению с N производными в методе скорейшего спуска. Кроме того, метод Ньютона может быть неустойчивым при удалении от минимума. Метод скорейшего спуска, судя по нашему опыту, обладает серьезным недостатком. Когда частные производные f по различным параметрам сильно отличаются по величине (это может быть либо свойством функции f , либо результатом неудачного выбора начальной точки), то осуществляется только минимизация по параметрам с большими производными, а прочие параметры меняются мало. Это приводит к увеличению числа итераций и потере точности.

Модифицируем метод скорейшего спуска следующим образом. Сначала вычисляем корректирующую матрицу S (6) в виде

$$S_{ij} = \delta_{ij} |b_i - a_i| / G_i. \quad (8)$$

Затем с этой (неизменной) матрицей проводим M итераций поиска минимума функции φ (5); далее вновь вычисляем корректирующую матрицу S (8) и всю процедуру повторяем.

Матрица S осуществляет периодическую адаптацию к виду функции f в окрестности минимума. При выборе матрицы S в виде (8) краевые ограничения (7) по всем переменным разрешают одинаковое (по порядку величины) число шагов вдоль направления Λ (5) вне зависимости от величин производных функции f и отрезков изменения параметров $|b_i - a_i|$. Число M подбирается опытным путем в зависимости от вида минимизируемой функции. Обычно хорошие результаты получаются при $M = 3 \div 9$.

Поиск минимума вдоль направления Λ (5) проводится с помощью одной из модификаций метода деления отрезка пополам, что дает оценку времени для локальной минимизации (ср. (4))

$$T_2 = kT_0(N + \log_2 n), \quad (9)$$

где $k = \text{const}$ (обычно $k \approx 20 \div 40$).

3. Алгоритм глобальной минимизации. Рассмотрим наш подход к осуществлению глобальной минимизации, т. е. к нахождению и перебору локальных минимумов с целью поиска глобального минимума. Мы заложили в алгоритм НГМ следующие принципы. Переход с одного локального минимума на другой проводится путем фиксации значения некоторого параметра, в то время как по всем остальным $N-1$ параметрам осуществляется локальная минимизация. Значение зафиксированного параметра изменяется с шагом, который должен быть намного больше шага по этой переменной в процессе НЛМ, но не слишком велик, чтобы не пропустить минимум и чтобы сопутствующая НЛМ не потребовала много времени. Таким образом, наша процедура НГМ есть не что иное, как проверка устойчивости локального минимума по отношению к немалому изменению каждого из параметров (по очереди). Переход на другой локальный минимум происходит в момент потери этой устойчивости. Качественно алгоритм глобальной минимизации показан на рис. 1.

Количественно алгоритм НГМ можно описать следующим образом. Введем функцию одной переменной f_i , которую назовем профилем f по переменной λ_i :

$$f_i(\alpha) = \min^{(L)} f(\lambda) |_{\lambda_i = \alpha}, \quad (10)$$

где $\min^{(L)}$ — оператор локальной минимизации, распространяющийся на

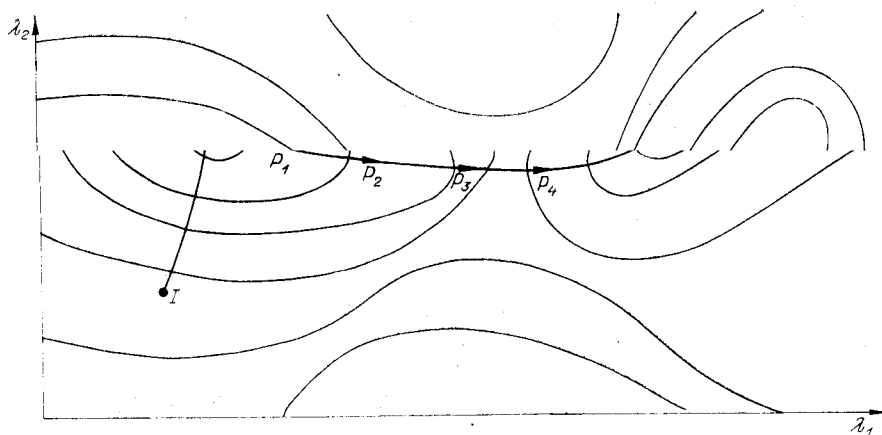


Рис. 1. Пояснение алгоритма глобальной минимизации:

топографическая карта рельефа минимизируемой функции для случая двух переменных λ_1, λ_2 и начальная точка I , от которой процедура локальной минимизации приводит к ближайшему минимуму M_1 . Далее переменная λ_1 исключается из минимизации и по ней начинается движение с шагом d_1 , в то время как по переменной λ_2 осуществляется минимизация, которая приводит к соответствующим точкам P_1, P_2, P_3 . В точке P_4 устойчивость минимума M_1 теряется и процедура локальной минимизации сразу же находит второй минимум M_2 .

все переменные, за исключением λ_i , которая фиксируется равной α . Поиск глобального минимума заключается в поиске минимумов N скалярных функций $f_i(\alpha)$, например, методом перебора, т. е. изменением переменной α с некоторым шагом d_i для $i = 1, \dots, N$.

Рассмотрим вопрос об оценке неопределенностей (ошибок) найденных параметров глобального минимума. Наш алгоритм обладает тем преимуществом, что в процессе глобальной минимизации исследуется поведение функции f при немалых отклонениях каждой из компонент вектора параметров λ от глобального минимума $\lambda^{(0)}$. Поэтому можно дать оценки ошибок параметров, не прибегая к обычной процедуре линеаризации вблизи минимума. Отметим, что это не влечет никаких дополнительных затрат времени счета.

Ошибки найденных параметров глобального минимума могут быть следствием либо конечности шага при минимизации, либо случайной ошибки в минимизируемой функции. Последняя ситуация типична для обработки результатов измерений, когда минимизируется норма отклонения экспериментальных данных от теоретических предсказаний.

Для определенности рассмотрим минимизацию величины $f(\lambda) = \chi^2$ — взвешенного среднеквадратичного отклонения значений J независимых гауссовских случайных величин x_m ($m = 1, \dots, J$) от их средних значений $\bar{x}_m(\lambda)$, зависящих от вектора параметров λ (см., например, [10]):

$$\chi^2(\lambda) = \sum_{m=1}^J \left[\frac{x_m - \bar{x}_m(\lambda)}{\sigma_m} \right]^2, \quad (11)$$

где σ_m — дисперсии величин x_m . Введем также (по аналогии с соотношением (10)) N скалярных функций $\chi_i^2(\lambda_i)$ — профилей исходной функции $\chi^2(\lambda)$ (11).

Если дисперсии параметров $\sigma(\lambda_i)$ достаточно малы, то применима обычная процедура линеаризации, приводящая к следующей оценке:

$$\sigma(\lambda_i) = 2 \langle d^2\chi_i^2/d\lambda_i^2 \rangle^{-1}, \quad (12)$$

здесь угловые скобки означают статистическое усреднение. Выражение для второй производной χ_i^2 имеет вид

$$d^2\chi_i^2/d\lambda_i^2 = \sum_{m=1}^J \frac{1}{\sigma_m^2} [(d\bar{x}/d\lambda_i)^2 + d^2\bar{x}_m/d\lambda_i^2 (\bar{x}_m - x_m)], \quad (13)$$

где полные производные по λ_i вычислены на кривой локальной минимизации, задаваемой оператором $\min^{(L)}$ (10). Первый член в (13) в квадратных скобках представляет собой когерентную сумму (все слагаемые входят с одинаковым знаком), в то время как второй член — сумму знакопеременных слагаемых и в среднем равен нулю. Поэтому ясно, что при большом числе слагаемых J усреднение по соседним точкам эквивалентно статистическому усреднению, а вклад второго члена по порядку величины будет, как правило, в \sqrt{J} раз меньше вклада первого члена в квадратных скобках в (13). Такое приближение, которое можно назвать «приближением хаотических знаков», дает возможность снять операцию усреднения в (12) и выразить дисперсию ошибок через вычисленное значение второй производной $d^2\chi_i^2/d\lambda_i^2$ (для данной выборки величин x_m):

$$\sigma^2(\lambda_i) \approx 2 (d^2\chi_i^2/d\lambda_i^2)^{-1}. \quad (14)$$

Поскольку профили $\chi_i(\lambda_i)$ вычисляются в процессе глобальной минимизации, то наш алгоритм позволяет найти линейные оценки дисперсий взаимно-коррелированных величин λ_i согласно (14), не тратя дополнительного времени счета и не прибегая к обычному (см., например, [41]) вычислению и обращению информационной матрицы Фишера. Формула (14) позволяет также написать выражение для функций $\chi_i^2(\lambda_i)$ вблизи глобального минимума:

$$\Delta\chi_i^2(\lambda_i) \approx \frac{\Delta\lambda_i^2}{\sigma^2(\lambda_i)}, \quad (15)$$

где введены отклонения от глобального минимума $\Delta\lambda_i = \lambda_i - \lambda_i^{(0)}$, $\Delta\chi_i^2(\lambda_i) = \chi_i^2(\lambda_i) - \chi_i^2(\lambda_i^{(0)})$. Из этого выражения видно, что при удалении данного параметра λ_i от глобального минимума на k стандартных отклонений, т. е. на величину $\Delta\lambda_i = k\sigma(\lambda_i)$, соответствующая функция χ_i^2 возрастает на величину $\Delta\chi_i^2 = k^2$.

Можно также дать согласованное определение ошибок параметров в нелинейном случае, когда эти ошибки, вообще говоря, не малы. Согласованность означает, что в пределе малых ошибок их нелинейные оценки будут приближаться к линейным. Таким определением является следующее. Нелинейной оценкой ошибки параметра λ_i на уровне $\pm q$ стандартных отклонений будем называть величину $\pm \Delta^{(\pm q)}\lambda_i$, такую, что

$$\Delta\chi_i^2(\lambda_i^{(0)} \pm \Delta^{(\pm q)}\lambda_i) = q^2. \quad (16)$$

Согласованность следует из того, что при стремлении ошибки $\Delta\lambda$ к нулю становится применимым соотношение (15), из которого следует $\Delta^{(\pm q)}\lambda_i = q\sigma(\lambda_i)$. Но в нелинейном случае соотношение (16) лучше передает ожидаемые ошибки параметров. Эта ситуация иллюстрируется рис. 2. На практике обычно задается $q = 2 \div 3$.

Для снижения времени счета целесообразно ограничить величину отклонения параметров от предполагаемого глобального минимума. При-

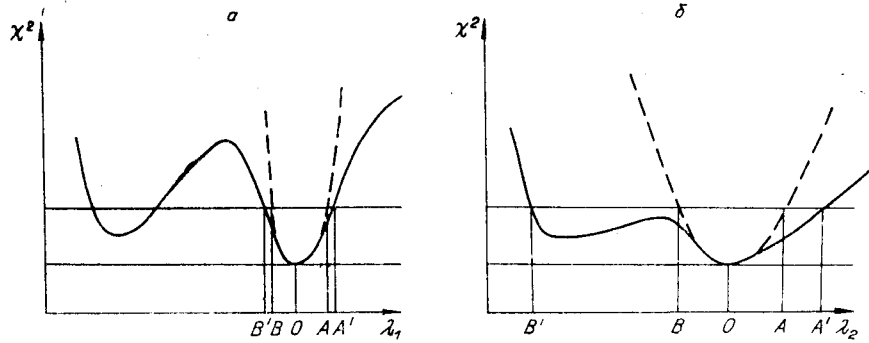


Рис. 2. Соотношение между линейными и нелинейными оценками ошибок параметров, показанное на примере профиля χ^2 , соответствующего рельефу на рис. 1:

расстояние между тонкими горизонтальными линиями равно единице, что соответствует одному стандартному отклонению; линейные оценки определяются по аппроксимирующей параболе (15) (штриховая линия) и даются отрезками OA и OB , равными между собой; нелинейные оценки — OA' и OB' ; a — случай хорошей применимости линейного приближения, b — существенно нелинейный случай (видно, что нелинейные оценки могут быть значительно больше линейных и лучше передают точность вычисления параметров).

мая во внимание (15), (16), это разумно сделать следующим образом:

$$-\Delta^{(-Q)}\lambda_i \leq \lambda_i - \lambda_i^{(0)} \leq \Delta^{(Q)}\lambda_i. \quad (17)$$

Здесь Q есть некоторая константа, зависящая от задачи. Практически хорошие результаты получаются при $Q = 5 \div 7$. Во многих случаях рационально задать $q = Q$.

Рассмотрим теперь вопрос о выборе шага. Ясно, что шаг d_i изменения переменной λ_i должен быть мал по сравнению с величиной $\Delta^{(Q)}\lambda_i$, но для экономии времени счета не слишком мал по сравнению с $\Delta^{(q)}\lambda_i$. Поэтому в алгоритм НГМ введен автоматический выбор шага, удерживающий его в пределах, определенных соотношениями (ср. (15), (16)):

$$Q^2/N_2 < \Delta\chi_i^2(\lambda_i^{(0)} \pm d_i) < Q^2/N_1, \quad (18)$$

где $N_1, N_2 = \text{const}$, причем $N_2 > N_1 > 1$. Аналогичный подход к автоматическому выбору шага был предложен в [12]. На практике хорошие результаты получаются при $N_1 = 4 \div 9, N_2 = 9 \div 25$. При этом введение автоматического выбора шага обычно снижает время счета на 20–40%.

В заключение рассмотрим вопрос о нормировке величины χ^2 . При правильной нормировке дисперсий σ_m в (11) величина χ^2 имеет среднее значение, равное числу степеней свободы $\approx J$, и дисперсию $\sqrt{2J}$ [10]. Однако положение минимумов $\chi^2(\lambda)$ не изменится, если все величины σ_m умножить на постоянный фактор, что часто и делается для практического удобства. Примером может быть случай $\sigma_m = \text{const}$, когда просто полагают $\sigma_m = 1$ (метод наименьших квадратов). При этом формулы (14), (16) уже не применимы непосредственно, тем не менее по ним можно рассчитать правильный ответ, если умножить вычисленную величину χ^2 на корректирующий коэффициент

$$C = J/\chi_0^2, \quad (19)$$

где χ_0^2 — значение χ^2 в глобальном минимуме. Такой коэффициент обеспечивает правильную нормировку величины χ^2 вне зависимости от общей нормировки дисперсий σ_m . Описанная процедура применима при $J \gg 1$, когда можно пренебречь дисперсией величины χ^2 по сравнению с ее средним значением.

При реализации алгоритма НГМ применяем следующую процедуру нормировки. Если величина χ^2 отклоняется от ожидаемого значения J

не более, чем на две своих дисперсии, т. е. на $\sqrt{8J}$, то считаем такое отклонение незначительным и не корректируем величины χ^2 (т. е. полагаем $C = 1$); в противном случае при вычислении погрешностей параметров величина χ^2 умножается на фактор (49). Пример использования описанного алгоритма НГМ приведен в приложении.

ПРИЛОЖЕНИЕ

В качестве проверочной задачи осуществляем минимизацию функции $\chi^2(\lambda)$, зависящей от десяти параметров λ_i , $i = 1 \div 10$:

$$\chi^2 = \sum_t (F_E(t) - F_T(t, \lambda))^2,$$

где сумма распространяется по t от 0,23 до 4,52 с шагом 0,07; $F_E(t)$ — значения функции, найденные из эксперимента [13] (рис. 3), а $F_T(t, \lambda)$ — теоретические значения этой функции:

$$F_T(t, \lambda) = \lambda_1 \int_0^t \left\{ e^{-\frac{t-t'}{\lambda_{10}}} + \frac{1}{\lambda_{10}\lambda_5} \operatorname{Im} [(1 + \lambda_{10}\lambda_9 - i\lambda_{10}\lambda_5) e^{(i\lambda_5 - \lambda_9)(t-t')}] \right\} f(t') dt', \quad (20)$$

где

$$f(t) = 2\lambda_3 u_A(t) - u_A^2(t), \quad u_A(t) = \lambda_2 \{ (1 + \lambda_4) e^{-\lambda_6 t} - e^{-\lambda_7 t} - \lambda_4 e^{-\lambda_8 t} \}. \quad (21)$$

Время счета на ЭВМ М-4030 составляло 2 ч 20 мин, при этом были определены $k_1 = 36$, $k_2 = 9$ (см. (4)). Алгоритм обнаружил девять минимумов. График теоретической функции $F_T(t, \lambda)$, соответствующий глобальному минимуму χ^2 , приведен на рис. 3 вместе с экспериментальной зависимостью $F_E(t)$.

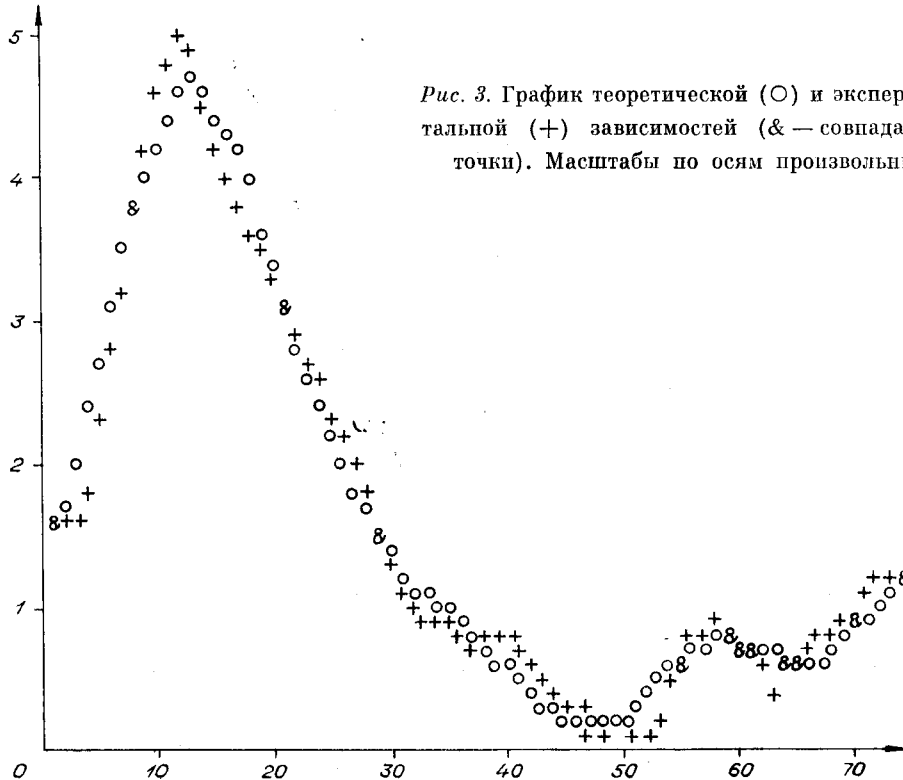


Рис. 3. График теоретической (○) и экспериментальной (+) зависимостей (& — совпадающие точки). Масштабы по осям произвольные.

ЛИТЕРАТУРА

1. **Numerical Methods for Non-Linear Optimization**/Ed. by F. A. Lootsma. L.—N.—Y.: Acad. Press, 1972.
2. **Численные методы условной оптимизации**/Под ред. Ф. Гилла, У. Мюррея. М.: Мир, 1977.
3. **Фиакко А., Мак-Кормик Г.** Нелинейное программирование. Методы последовательной безусловной минимизации. М.: Мир, 1972.
4. **Jehida K., Fujii Y.** An Interval Arithmetic Method for Global Optimization.—Computing, 1979, vol. 23, p. 1—23.
5. **Hesse R. A.** A Suboptimal Method for Global Solution of the Non-Linear Programming Problem.—Report N COD-1449-16. St. Louis, Missouri: Washington University, 1971.
6. **Zwart P. B.** Non-Linear Programming: Global use of the Lagrangian.—J. Optim. Theory and Appl., 1968, т. 18, с. 49.
10. **Корн Г., Корн Т.** Справочник по математике. М.: Наука, 1968, т. 1, с. 10.
11. **Линник Ю. В.** Метод наименьших квадратов и основы теории обработки наблюдений. М.: Физматгиз, 1962.
12. **Brent R. P.** Algorithms for Minimization Without Derivatives. Prentice-Hall, Englewood-Cliffs, 1973.
13. **Hill B. C., Schubert E. D., Nokes M. A., Michelson R. P.** Laser Interferometer Measurement of Changes in Crayfish Axon Diameter Concurrent with Action Potential.—Science, 1977, vol. 196, p. 426—428.

Поступила в редакцию 26 февраля 1980 г.

УДК 62.506 : 518.3

Б. Г. БЛАНК
(Ленинград)

АДАПТИВНЫЕ АЛГОРИТМЫ КУСОЧНО-ЛИНЕЙНОЙ АППРОКСИМАЦИИ НЕ СТРОГО МОНОТОННЫХ ФУНКЦИЙ РАСПРЕДЕЛЕНИЯ

В известных методах аппроксимации функций распределения [1—3] и других относительно последних обычно предполагается строгая монотонность, которая необходима для обеспечения требуемой точности при ограничениях на число участков аппроксимации. Однако строгая монотонность не всегда имеет место, в особенности при распределениях, более сложных, чем унимодальные. Применение в указанных случаях стандартных методов аппроксимации может привести к значительному числу интервалов разбиения области определения функции распределения либо, как упоминалось выше, к значительной погрешности при ограничениях на число участков аппроксимации. Последнее обстоятельство связано с ограничениями на объем используемой памяти ЭВМ, что в ряде случаев представляется весьма существенным.

В настоящей работе рассматривается задача наилучшей в отношении среднеквадратической погрешности кусочно-линейной аппроксимации функций распределения, главным образом не строго монотонных, при ограничениях на число участков аппроксимации либо на величину допустимой погрешности.

К рассматриваемой задаче может приводить также задача выделения информативных участков экспериментальных кривых в случаях, когда