

Е. П. ЧУРАКОВ

(Рязань)

ИТЕРАТИВНЫЕ АЛГОРИТМЫ ОЦЕНИВАНИЯ ПАРАМЕТРОВ
СЛУЧАЙНЫХ ПРОЦЕССОВ (ПОЛЕЙ)

Введение и постановка задачи. При построении автоматизированных систем сбора и обработки данных значительный интерес представляет задача оценивания параметров случайных процессов (или полей). Один из распространенных ее вариантов сводится к следующему. Пусть имеется случайный процесс $v(\vec{A}, \vec{t})$, представленный аддитивной смесью

$$v(\vec{A}, \vec{t}) = f(\vec{A}, \vec{t}) + p(\vec{t}). \quad (1)$$

Здесь $f(\vec{A}, \vec{t})$ — детерминированная с точностью до r -мерного вектора \vec{A} составляющая процесса, причем $f(\cdot)$ — дифференцируемая по вектору \vec{A} произвольная нелинейная функция; \vec{t} — в общем случае вектор, характеризующий временную, пространственную или пространственно-временную природу процесса; \vec{A} — неизвестный или случайный с априори заданной плотностью вероятностей $\sigma(\vec{A})$ вектор; $p(\vec{t})$ — случайная составляющая, интерпретируемая как гауссов процесс с известной статистикой. В конечной области определения вектора \vec{t} выделяется ряд пронумерованных по некоторому правилу дискретных точек $\vec{t}_i (i = \overline{1, l})$ и регистрируются значения процесса (1) в этих точках, так что следствием является l -мерный вектор измерений

$$\vec{V} = \vec{\Phi}(\vec{A}) + \vec{P}. \quad (2)$$

В предположении $l > r$ и $\vec{P} \in N(0, \vec{K}_p)$ задача заключается в отыскании в некотором смысле наилучшей оценки $\hat{\vec{A}}$ вектора \vec{A} по апостериорным данным (2).

Сформулированная задача широко распространена в таких технических приложениях, как радиолокация, гидроакустика, идентификация, системы управления и т. п. Традиционно ее решение сводится к поиску или байесовой оценке при случайном векторе \vec{A} , или максимально правдоподобной при неизвестном векторе (например, [1]). Первая из оценок наиболее часто ищется из условия минимума среднего риска при квадратичной функции стоимости

$$R = M_{\vec{V}, \vec{A}} \{ \|\vec{A} - \hat{\vec{A}}(\vec{V})\|^2 \} = \min,$$

где $M_{\vec{V}, \vec{A}}$ — символ осреднения по пространствам векторов \vec{V} и \vec{A} ; $\|\dots\|$ — евклидова норма вектора, и оказывается равной апостериорному среднему [2, 3]

$$\hat{\vec{A}} = M\{\vec{A}|\vec{V}\} = \int_{\vec{A}} \vec{A} \omega(\vec{A}|\vec{V}) d\vec{A}. \quad (3)$$

Здесь $\omega(\vec{A}|\vec{V})$ — апостериорная плотность вектора \vec{A} , и интегрирование проводится по всему пространству этого вектора.

Оценка второго вида обычно находится из условия максимума функции правдоподобия $L(\vec{V}/\vec{A})$ и является корнем уравнения правдоподобия

$$\frac{d}{d\vec{A}} \vec{\Phi}^T(\vec{A}) \vec{K}_p^{-1} (\vec{V} - \vec{\Phi}(\vec{A})) \Big|_{\vec{A}=\hat{\vec{A}}} = 0, \quad (4)$$

записанного применительно к модели (2) и гауссовому вектору \vec{P} . Здесь и далее символ τ означает транспонирование, -1 — обращение матрицы.

Нелинейная зависимость векторов \vec{V} и \vec{A} приводит к тому, что на основании (3), (4) не удастся аналитически установить в явном виде зависимость оценки $\hat{\vec{A}}$ от измерений \vec{V} . Это обусловлено отсутствием методов многомерного интегрирования при нелинейной модели (2) в случае (3) и общих рецептов решения нелинейных уравнений в ситуации (4). Поэтому приходится прибегать или к помощи приближенных методов, или искать пути последовательных достижений оптимального решения. Один из возможных подходов, связанных с последним направлением, излагается ниже.

Оценивание случайных параметров. Пусть \vec{A} — случайный вектор, причем в соответствии с широким распространением в реальных ситуациях положим $\vec{A} \in N(\vec{m}_A, \vec{K}_A)$. Поиск оценки $\hat{\vec{A}}$ осуществим по схеме, подобной [4], обозначив через $\hat{\vec{A}}^{(n)}$ оценку на n -м шаге и заметив, что

$$\Delta \vec{A}^{(n)} = \vec{A} - \hat{\vec{A}}^{(n-1)} \in N(\vec{m}_A - \hat{\vec{A}}^{(n-1)}, \vec{K}_A).$$

Руководствуясь критерием

$$M_{\vec{V}, \Delta \vec{A}^{(n)}} \{ \|\Delta \vec{A}^{(n)} - \hat{\Delta \vec{A}}^{(n)}\| \} = \min \quad (5)$$

и известным [2] способом минимизации функционала (5), вычисляем оптимальную оценку приращения $\Delta \vec{A}^{(n)}$, что приводит к итеративному алгоритму

$$\hat{\vec{A}}^{(n+1)} = \hat{\vec{A}}^{(n)} + (\hat{\vec{D}}^{(n)})^{-1} [(\hat{\vec{I}}^{(n)})^T \vec{K}_p^{-1} (\vec{V} - \vec{\Phi}^{(n)}) + \vec{K}_A^{-1} (\vec{m}_A - \hat{\vec{A}}^{(n)})]. \quad (6)$$

Здесь

$$\hat{\vec{D}}^{(n)} = \vec{K}_A^{-1} + (\hat{\vec{I}}^{(n)})^T \vec{K}_p^{-1} \hat{\vec{I}}^{(n)}; \quad (7)$$

$$\hat{\vec{I}}^{(n)} = \frac{d}{d\vec{A}} \vec{\Phi}(\vec{A}) \Big|_{\vec{A}=\hat{\vec{A}}^{(n)}}; \quad (8)$$

$$\vec{\Phi}^{(n)} = \vec{\Phi}(\hat{\vec{A}}^{(n)}). \quad (9)$$

Асимптотические свойства алгоритма и сходимость. Предположим, что рекуррентная процедура (6) сходится. Установим предельное, т. е. при неограниченном увеличении числа итераций, значение вычисляемой в соответствии с (6) оценки. Обозначим

$$\lim_{n \rightarrow \infty} \hat{\vec{A}}^{(n+1)} = \lim_{n \rightarrow \infty} \hat{\vec{A}}^{(n)} = \hat{\vec{A}}.$$

Как следует из (6) и (8), при условии сходимости эта предельная оценка должна удовлетворять уравнению

$$\frac{d}{d\vec{A}} \vec{\Phi}^T(\vec{A}) \vec{K}_p^{-1} (\vec{V} - \vec{\Phi}(\vec{A})) + \vec{K}_A^{-1} (\vec{m}_A - \vec{A}) \Big|_{\vec{A}=\hat{\vec{A}}} = 0. \quad (10)$$

Нетрудно показать, что это уравнение определяет оптимальную по критерию максимума апостериорной плотности вероятностей [1] оценку вектора \vec{A} . Действительно, в соответствии с формулой Байеса имеем

$$w(\vec{A}/\vec{V}) = \sigma(\vec{A}) L(\vec{V}/\vec{A}) \frac{1}{w(\vec{V})}, \quad (11)$$

где $w(\vec{V})$ — безусловная плотность вероятностей вектора \vec{V} . Так как для модели (2) при гауссовом векторе \vec{P}

$$\ln L(\vec{V}/\vec{A}) = \text{const} - \frac{1}{2} (\vec{V} - \vec{\Phi}(\vec{A}))^T \vec{K}_p^{-1} (\vec{V} - \vec{\Phi}(\vec{A})),$$

то, продифференцировав (11) по вектору \vec{A} после предварительного логарифмирования, приходим к уравнению (10).

Установленное свойство позволяет придать алгоритму (6) иную форму записи:

$$\hat{A}^{(n+1)} = \hat{A}^{(n)} + (\vec{D}^{(n)})^{-1} \text{grad} \ln w(\hat{A}^{(n)}/\vec{V}), \quad (12)$$

где $\text{grad} \ln w(\cdot)$ — градиент логарифма апостериорной плотности вероятностей вектора \vec{A} , вычисленный в точке $\hat{A}^{(n)}$. Полученное соотношение представляет релаксационный итеративный алгоритм, многочисленные разновидности которого нашли весьма широкое распространение при решении разнообразных задач управления и смежных областей [5]. В ряду итеративных методов его можно рассматривать как модификацию известного метода Ньютона [6], характеризующуюся выбором оптимального шага перемещения.

Большое значение для итеративных алгоритмов имеют условия сходимости. Используя подход, подобный [4], можно показать, что для сходимости алгоритма (12) с вероятностью единица достаточно выполнения условия

$$\max_{\hat{A}} M_{\vec{V}} \left\{ \|\vec{E} + \vec{J} \left\{ (\vec{D}(\vec{A}))^{-1} \text{grad} \ln w(\hat{A}/\vec{V}), \vec{A} \right\}\| \right\} < 1. \quad (13)$$

Здесь \vec{E} — единичная матрица, \vec{J} — якобиан вектора $(\vec{D}(\vec{A}))^{-1} \text{grad} \ln w(\vec{A}/\vec{V})$, вычисленный в точке \vec{A} , т. е.

$$\vec{J}\{\cdot\} = \frac{d}{d\vec{A}} \left(\vec{D}(\vec{A}) \right)^{-1} \text{grad} \ln w(\vec{A}/\vec{V}) \Big|_{\vec{A}=\hat{A}}, \quad (14)$$

$\|\vec{B}\|$ — евклидова норма матрицы \vec{B} , определяемая как корень квадратный из наибольшего собственного числа матрицы $\vec{B}^T \vec{B}$ [7].

Оценивание неизвестных параметров. Условимся неизвестный вектор интерпретировать как случайный с некоррелированными компонентами, характеризующимися бесконечно большими дисперсиями [3].

Используя это обстоятельство, предположим, что в (6) матрица \vec{K}_A диагональна с бесконечно большими элементами главной диагонали. В результате находим итеративный алгоритм поиска оценки неизвестного вектора, подобный [4]:

$$\hat{A}^{(n+1)} = \hat{A}^{(n)} + (\vec{D}^{(n)})^{-1} (\vec{I}^{(n)})^T \vec{K}_p^{-1} (\vec{V} - \vec{\Phi}^{(n)}). \quad (15)$$

Помимо уже использованных символов, здесь обозначено

$$\vec{D}^{(n)} = (\vec{I}^{(n)}) \vec{K}_p^{-1} \vec{I}^{(n)}. \quad (16)$$

Так как при нулевой матрице \vec{K}_A^{-1} уравнение (10) вырождается в условие максимума правдоподобия, алгоритм (15) в случае сходимости приводит к максимально правдоподобной оценке и может быть представлен в форме

$$\hat{\vec{A}}^{(n+1)} = \hat{\vec{A}}^{(n)} + (\vec{D}^{(n)})^{-1} \text{grad ln } L(\vec{V}/\hat{\vec{A}}^{(n)}). \quad (17)$$

Здесь $\text{grad ln } L(\vec{V}/\hat{\vec{A}}^{(n)})$ — градиент логарифма функции правдоподобия, вычисленный в точке $\hat{\vec{A}}^{(n)}$. Условия, достаточные для сходимости алгоритма с вероятностью единица, принимают вид

$$\max_{\vec{A}, \vec{A}} M_{\vec{V}/\vec{A}} \left\{ \left\| \vec{E} + \vec{J} \left((\vec{D}(\vec{A}))^{-1} \text{grad ln } L(\vec{V}/\vec{A}), \vec{A} \right) \right\| \right\} < 1, \quad (18)$$

где $M_{\vec{V}/\vec{A}}$ — символ условного осреднения по вектору \vec{V} . Следовательно, для сходимости алгоритма достаточно, чтобы наибольшее значение условного математического ожидания евклидовой нормы матрицы $\vec{E} + \vec{J}$ было меньше единицы.

С некоторым ущербом для скорости сходимости правило (17) можно упростить, ограничив построение матрицы \vec{D} лишь начальной точкой, так что

$$\hat{\vec{A}}^{(n+1)} = \hat{\vec{A}}^{(n)} + (\vec{D}^{(0)})^{-1} \text{grad ln } L(\vec{V}/\hat{\vec{A}}^{(n)}). \quad (19)$$

Последующее упрощение алгоритма достигается переходом к фиксированной матрице \vec{I} , вследствие чего соотношение (15) приобретает вид

$$\hat{\vec{A}}^{(n+1)} = \hat{\vec{A}}^{(n)} + (\vec{D}^{(0)})^{-1} (\vec{I}^{(0)})^T \vec{K}_p^{-1} (\vec{V} - \vec{\Phi}^{(n)}). \quad (20)$$

Как показывает моделирование, возникающие при этом потери в точности оказываются разумно малыми.

Экспериментальное исследование алгоритмов. С целью изучения свойств алгоритмов (15), (20) моделировалась частная задача оценивания параметров случайного поля, приближенное решение которой дано в [8]. В этой задаче

$$f(\vec{A}, \vec{t}_i) = u \left| \frac{\sin(a_1 - t_{1i})}{a_1 - t_{1i}} \right| \left| \frac{\sin(a_2 - t_{2i})}{a_2 - t_{2i}} \right|,$$

причем компоненты t_{1i}, t_{2i} задаются таблицей из [8]. При моделировании компоненты вектора P назначались некоррелированными с дисперсией σ^2 .

n/σ	Алгоритмы	$a_1, a_2=1,57$						$a_1, a_2=2,74$					
		\bar{a}_1	σ_1^2	k_1	\bar{a}_2	σ_2^2	k_2	\bar{a}_1	σ_1^2	k_1	\bar{a}_2	σ_2^2	k_2
5	I	1,52	0,06	2,9	1,56	0,1	2,7	2,72	0,06	2,8	2,74	0,07	2,7
	II	1,5	0,07	2,0	1,57	0,11	2,0	2,7	0,08	2	2,73	0,08	2
7	I	1,57	0,03	2,6	1,56	0,04	2,3	2,76	0,01	2,5	2,77	0,05	2,5
	II	1,57	0,03	2	1,57	0,04	1,9	2,74	0,03	2	2,77	0,05	2
9	I	1,55	0,02	2,3	1,55	0,02	2,4	2,75	0,01	2,5	2,75	0,02	2,4
	II	1,55	0,02	2	1,55	0,02	2	2,75	0,018	2	2,75	0,03	2

Большое значение при реализации итеративных алгоритмов имеет правило выбора начальной точки, так как оно в значительной степени определяет как сходимость алгоритма вообще, так и характер (локальный или глобальный) достигаемого экстремума в частности. Однозначных рекомендаций по ее выбору в настоящее время нет. При проведении моделирования поступали следующим образом.

Обычно область возможных значений вектора \vec{A} исследователю, пусть грубо, известна. Проквантуем эту область, выделив в ней N дискретных точек, с некоторым шагом заполняющих всю область. Пусть в j -й точке вектор \vec{A} принимает значение $\vec{A}_j (j = \overline{1, N})$. Предположим, что вектор \vec{A} , соответствующий апостериорным данным \vec{V} , совпадает с \vec{A}_j . Правомочность этой гипотезы условимся характеризовать соответствующей функцией правдоподобия или некоторым ее однозначным преобразованием, например величиной

$$l_j = -(\vec{V} - \vec{\Phi}(\vec{A}_j))^T \vec{K}_p^{-1} (\vec{V} - \vec{\Phi}(\vec{A}_j)) \quad (j = \overline{1, N}). \quad (21)$$

Из N величин (21) выделим наибольшую. Пусть

$$l_s = \sup_j \{l_j, j = \overline{1, N}\}. \quad (22)$$

Тогда в качестве начальной точки $\hat{A}^{(0)}$ в процедурах (15), (20) назначается величина $\vec{A}_{(s)}$ из (21). Заметим, что сходный прием используется при многоканальном оценивании параметров по методу максимального правдоподобия [9, 10]. При моделировании принималось

$$\vec{A} \in \Gamma \{(a_1, a_2) : 0 \leq a_1 \leq 2\pi, 0 \leq a_2 \leq 2\pi\}; N = 25,$$

причем анализируемые точки $\vec{A}_j (j = \overline{1, 25})$ равномерно размещались в области Γ . В результате моделирования были найдены осредненные по пятидесяти экспериментам значения оценок \hat{a}_1, \hat{a}_2 и соответствующие дисперсии, т. е. величины

$$\bar{a}_q = \frac{1}{m} \sum_{j=1}^m \hat{a}_{qj};$$

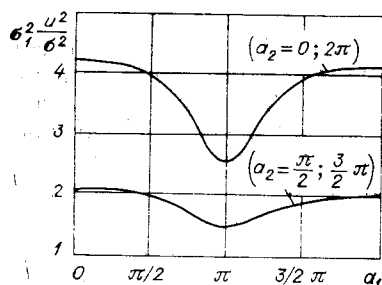
$$\sigma_q^2 = \frac{m}{m-1} \left(\frac{1}{m} \sum_{j=1}^m \hat{a}_{qj}^2 - \bar{a}_q \right) \quad (q = 1, 2; m = 50),$$

где \hat{a}_{qj} — оценка величины a_q в j -м эксперименте. Одновременно исследовалась скорость сходимости алгоритмов, в качестве характеристики которой принималось среднее число итераций, необходимое для завершения процедуры. При этом процесс вычисления оценки считался законченным на некотором шаге, если приращение оценки на последующих шагах оказывалось менее 10^{-2} величины $\hat{A}^{(0)}$. Таким образом, мерой скорости сходимости являлась величина

$$k = \frac{1}{m} \sum_s sm_s,$$

где m_s — число экспериментов, в которых итеративный алгоритм заканчивался за s шагов. Результаты эксперимента систематизированы в таблице, в которой цифрами I и II сопровождаются характеристики алгоритмов (15) и (20) соответственно.

Представлялось интересным сопоставить полученные результаты с предельно достижимыми характеристиками, следующими из неравен-



ства Рао — Крамера. С этой целью была составлена матрица Фишера $C = |c_{ij}|$ ($i, j = 1, 2$), элементы которой в исследуемом случае равны

$$c_{ij} = \sigma^{-2} \frac{\partial}{\partial a_i} \vec{\Phi}^* (\vec{A}) \frac{\partial}{\partial a_j} \vec{\Phi} (\vec{A}),$$

и вычислена корреляционная матрица ошибок оценивания $\vec{K}_{\Delta \vec{A}} = \vec{C}^{-1}$. На осно-

вании этой матрицы построено семейство кривых (см. рисунок), характеризующее предельно достижимые дисперсии ошибок оценивания величины a_1 в функции вектора \vec{A} . Из сопоставления этих кривых с результатом моделирования следует, что полученные экспериментальные данные довольно близки к предельным показателям.

ВЫВОДЫ

Обсуждаемые алгоритмы позволяют находить оптимальные с позиций максимума апостериорной плотности вероятностей и максимума правдоподобия оценки. Их реализация, ориентированная на широкое использование ЦВМ в современных автоматизированных системах сбора и обработки данных, не вызывает затруднений. Выбор начальной точки, являющийся традиционно болезненным местом большинства итеративных процедур, осуществляется с учетом рекомендаций, сопровождающих экспериментальное исследование алгоритмов. Возможные упрощения расчетных схем достигаются на пути фиксирования матриц \vec{D} , \vec{D} , \vec{I} их начальными значениями. Для алгоритмов характерна высокая скорость сходимости.

В заключение автор благодарит В. П. Кузнецова и В. П. Красина за помощь при экспериментальном исследовании алгоритмов.

ЛИТЕРАТУРА

1. Г. Ван Трис. Теория обнаружения, оценок и модуляции. М., «Советское радио», 1972.
2. Д. Миддлтон. Введение в статистическую теорию связи. Т. II. М., «Советское радио», 1962.
3. В. С. Пугачев. Теория случайных функций и ее применение к задачам автоматического управления. М., Физматгиз, 1962.
4. Е. П. Чураков. Построение дискретных преобразующих систем при наличии многих информационных каналов. — «Автоматика и телемеханика», 1974, № 3.
5. Я. З. Цыпкин. Адаптация и обучение в автоматических системах. М., «Наука», 1968.
6. Н. С. Бахвалов. Численные методы. Т. I. М., «Наука», 1973.
7. Ф. Р. Гантмахер. Теория матриц. М., «Наука», 1966.
8. В. П. Кузнецов, Е. П. Чураков. Об одном способе оценки векторного параметра при нелинейных измерениях многими устройствами в условиях гауссовых шумов. — «Автоматика», 1973, № 4.
9. С. Е. Фалькович. Оценка параметров сигналов. М., «Советское радио», 1970.
10. Л. П. Сысоев. Оценки параметров, обнаружение и различение сигналов. М., «Наука», 1969.

Поступила в редакцию 5 апреля 1974 г.;
окончательный вариант — 28 сентября 1974 г.