УДК 519.24

А. А. ПАВЛОВ, д-р техн. наук, проф. каф. АСОИУ НТУУ «КПИ»,
А. С. ШТАНЬКЕВИЧ, студент НТУУ «КПИ»

ВОССТАНОВЛЕНИЕ НЕИЗВЕСТНОЙ ЗАКОНОМЕРНОСТИ ПО ОГРАНИЧЕННОМУ НАБОРУ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ С ОШИБКОЙ

У статті запропоновано модифікацію методів відновлення невідомої закономірності за невеликим числом експериментальних даних для випадку наявності помилки вимірювань. Наведено приклад використання методу.

В статье предложена модификация методов восстановления неизвестной закономерности по небольшому числу экспериментальных данных для случая наличия ошибки измерений. Приведен пример использования метода.

In the article the modification of methods for recovery of an unknown law by small quantity of experimental data is suggested. The method usage example is stated.

Постановка задачи. В [1] были предложены и обоснованы методы решения задачи восстановления числовой скалярной функции действительных аргументов, однозначно задающей некоторую закономерность, по анализу наблюдаемых данных (вход–выход) ограниченного объема. Однако, в случае наличия ошибки в используемом наборе экспериментальных данных (например, погрешности измерений), методы требуют существенной модификации для эффективного решения задачи восстановления. Далее в статье будут сформулированы и рассмотрены требуемые изменения для предложенных ранее вероятностных методов (методы 2 и 3 из [1]).

Приведем основные определения. Существует некоторая закономерность, которая однозначно задается функцией $f\left(\overline{x}\right)$, не известной наблюдателю. Имеются результаты пассивного эксперимента $f\left(\overline{x}^i\right) = y_i$, $i = \overline{1,P}$ (P – небольшое число), содержащие ошибку измерений ε (случайная величина). По результатам пассивного эксперимента необходимо найти закономерность, то есть найти истинную функцию $f\left(\overline{x}\right)$, а не её аппроксимацию. Искомая функция $f\left(\overline{x}\right)$ представляется взвешенной суммой компонент $\psi_{i_j}(\overline{x})$ на подмножестве I_1 множества известных базовых функций $I = \left\{\psi_i(\overline{x}), i = \overline{1,L}\right\}$:

$$f\left(\overline{x}\right) = \sum_{j=1}^{K} a_{i_j} \psi_{i_j}(\overline{x}), \quad K < P < L \tag{1}$$

где $\psi_i(\bar{x})$ – известные базовые функции, a_i – неизвестные коэффициенты, $\bar{x} \in R_n$ – n-мерный вектор, K – число базовых функций в множестве I_1, $\psi_{i_j}(\bar{x}) \in I_1$ и заранее не известны.

Модифицированный метод 2 (вероятностный). Если P – число экспериментов – не на много меньше L – числа базовых функций в множестве I, то в предположении, что K (число базовых функций в множестве I_1) существенно меньше P, данный метод с большой вероятностью (легко определяемой как функция от K) дает точное решение по критерию из [1]. Для случая наличия ошибки в используемом для восстановления наборе экспериментальных данных был модифицирован второй этап этого метода.

Этап 1. Случайным образом конструируется последовательность базовых функций

$$\psi_{i_1}\left(\bar{x}\right), \psi_{i_2}\left(\bar{x}\right), \ldots, \psi_{i_L}\left(\bar{x}\right). \tag{2}$$

Функции $\psi_{i_{p+1}}\left(\bar{x}\right), \ldots, \psi_{i_L}\left(\bar{x}\right)$ исключаются из последовательности (2), получаем последовательность

$$\psi_{i_1}\left(\bar{x}\right), \ldots, \psi_{i_P}\left(\bar{x}\right). \tag{3}$$

В предположении, что базовые функции из множества I_1 находятся в последовательности (3), переходим к этапу 2.

Этап 2. Этот этап итерационный. На каждой итерации из последовательности (13) поочередно исключается каждая базовая функция $\psi_{i_p}\left(\bar{x}\right)$ и для оставшихся базовых функций решается задача линейного программирования (ЛП)

$$\min \sum_{l=1}^{P} y_l \tag{4}$$

$$-y_l \le f\left(\bar{x}^l\right) - \sum_{j=1}^{P-t} a_{i_j} \psi_{i_j}\left(\bar{x}^l\right) \le y_l \tag{5}$$

$$y_l \ge 0, \quad l = \overline{1, P}.$$

где переменными являются a_{i_j}, $j = \overline{1, P-t}$ и y_l, $l = \overline{1, P}$, а t – текущий порядковый номер итерации (начиная с 1). После решения задачи (4)-(5)

исключенная компонента возвращается в (3).

Когда получены значения показателя качества (4) при исключении каждой компоненты из (3), окончательно из последовательности (3) исключается та компонента, для которой это значение минимально. На этом текущая итерация завершена.

Описанная процедура повторяется. Критерием останова является резкий прирост минимального значения показателя качества (4) на текущей итерации по сравнению с предыдущей, после чего компоненты из (3) более не исключаются (установление граничного значения этого прироста в конкретных практических задачах является предметом статистического исследования метода). Оставшиеся в (3) базовые функции и являются истинными компонентами восстанавливаемой закономерности. Коэффициенты при них получены в последнем решении задачи ЛП (4)-(5) с текущим набором (3) на предпоследней итерации.

Этап 2 может быть ускорен, если в конце каждой итерации из последовательности (3) исключать не одну, а несколько компонент, для которых получены наименьшие значения показателя качества (4) в текущей итерации.

Введенные изменения метода следуют из особенностей решаемой задачи. При наличии ошибки ε минимальное значение показателя качества (4) на любой итерации этапа 2 статистически всегда больше нуля и увеличивается при последовательном удалении компонент из (3). При этом вследствие влияния ε при отбрасывании истинной компоненты получаемое значение показателя (4) не всегда будет превышать любое из значений показателя, полученных при отбрасывании фиктивных компонент, однако не будет минимальным в полученном наборе его значений на текущей итерации. Последнее утверждение, как и, в целом, эффективное применение модифицированного метода 2, имеет место при определенных границах ошибки ε, устанавливаемых статистическими исследованиями метода.

По результатам предварительных исследований эффективности метода были выработаны рекомендации относительно критерия останова итераций этапа 2 (табл. 1). Статистические исследования проводились для случаев, когда восстанавливаемая закономерность зависела от 3 до 5 входных величин (аргументы восстанавливаемой функции), количество базовых компонент в последовательности (3) – входной последовательности для этапа 2 – варьировалось от 15 до 100, а количество истинных компонент, которые требовалось восстановить, изменялось при этом от 2 до 5. При проведении экспериментов вычислялись точные оценки восстанавливаемых функций для случайных сгенерированных значений входных величин, после чего к полученным оценкам прибавлялась случайная величина ε (ошибка), равномерно распределенная в интервале $\left(-k_\varepsilon \cdot f; k_\varepsilon \cdot f\right)$, где f – точная оценка восстанавливаемой функции, k_ε – коэффициент зашумления. k_ε в сгенерированных примерах варьировался от 0,001 до 0,02, при более высоких его значениях неизвестная закономерность восстанавливалась статистически редко.

Таблица 1

Рекомендованные граничные значения коэффициента прироста минимального значения показателя (4) k_{opt} для останова этапа 2 модифицированного метода 2 в зависимости от числа компонент в последовательности (3)

Количество элементов в последовательности (3)	Мин. граничное значение k_{opt}
10	411,88
15	327,09
20	253,49
25	360,07
30	256,04
35	203,60
40	183,28
45	168,32
50	156,85
55	145,01
60	134,63
65	125,09
70	116,21
75	108,97
80	99,34
85	91,27
90	84,86
95	77,89
100	72,32

Метод 3 (вероятностный). Все утверждения в отношении данного метода, изложенные в [1], справедливы также и в случае наличия ошибки в используемом для восстановления наборе экспериментальных данных. Так, данный метод сводится к построению последовательности (2) и применению модифицированного метода 2 нужное количество раз для получения решения поставленной задачи с заданной вероятностью (детально описано в [1]).

Заключение. Изложенные модификации методов 2 и 3 могут эффективно использоваться в практических задачах для восстановления неизвестной закономерности по набору данных пассивного эксперимента, содержащих

ошибку измерений ε. По проведенным исследованиям можно сделать вывод, что предложенные методы особенно эффективны в случаях, когда ошибка ε не превышает 2% от истинного значения, что на практике может соответствовать условиям задачи восстановления точной закономерности при грубых требованиях к измерительным приборам.

Пример. Приведем пример восстановления неизвестной закономерности по результатам пассивного эксперимента, содержащим ошибку измерений ε, с применением модифицированного метода 2.

Пусть имеется закономерность $f(x_1, x_2, x_3, x_4)$. Её истинное выражение, неизвестное исследователю, состоит из двух компонент ($K = 2$) и имеет вид:

$$f(x_1, x_2, x_3, x_4) = -1.1569 \cdot \frac{x_3^2 x_4}{x_1^3 x_2} - 10.2517 \cdot x_2^2 x_3 \qquad (6)$$

Пусть также проведено 40 экспериментов ($P = 40$), в результате которых получены оценки неизвестной закономерности (табл. 2), содержащие некоторую ошибку измерений ε (в данном примере ошибка ε имеет равномерное распределение в интервале $(-k_\varepsilon \cdot f; k_\varepsilon \cdot f)$, где f – точное значение неизвестной функции, k_ε – коэффициент зашумления, $k_\varepsilon = 0.009$).

Таблица 2

Результаты пассивного эксперимента

№ эксп.	x_1	x_2	x_3	x_4	$f(x_1, x_2, x_3, x_4)$
1	+2.1165	-9.0054	+3.0700	-7.9078	-2537.15642422
2	+0.4087	+4.8609	-2.2114	+5.3952	+443.83786786
3	-8.0021	-5.0053	+8.5131	-6.8002	-2169.46500967
4	-2.5576	+9.7587	+6.8150	-8.5469	-6664.58925567
5	+4.7753	+1.4053	+2.5028	-9.9808	-49.81259220
6	+6.8859	+3.9424	-2.8453	+6.5617	+450.29283917
7	+0.8676	+0.4977	-0.3404	+2.7476	-0.26923936
8	-5.9096	-1.3885	+4.9655	-2.3072	-97.30253936
9	-6.1665	-1.6907	+7.7483	-7.5765	-224.90974295
10	-1.2545	+1.2902	-6.9635	-6.9882	-35.33509069
11	-8.3049	-0.4434	+9.5221	+7.1077	-21.96607079
12	-8.0023	-2.4945	-3.6066	+0.6226	+232.03867772
13	+2.5783	-0.9534	-6.9706	-4.5868	+49.16157063
14	+4.2356	-7.0788	-7.5737	+0.5863	+3886.83771643

Продолжение табл. 2

15	-8.0476	+5.4021	-1.0713	+6.6859	+320.37610068
16	-8.4204	+5.8121	+3.8760	-6.1894	-1338.71646619
17	-3.9062	-9.0721	-4.2190	+3.1127	+3528.93773065
18	-1.2620	+5.0623	-0.5006	+5.4115	+132.10075784
19	+7.5440	-8.9714	-3.2194	-6.4533	+2670.56500700
20	-4.1086	-3.3509	-4.5514	+1.9004	+527.54245571
21	-6.2152	-7.4419	+3.6616	+5.0244	-2063.86337238
22	+9.9281	-0.1586	+6.8698	+1.5877	-1.20490211
23	+6.8596	-2.7245	-2.7952	-2.7802	+211.48953980
24	-4.2538	-5.8509	-1.9844	-6.0298	+698.39079562
25	-8.1043	+2.9085	-1.9886	+7.1603	+173.81978567
26	-9.6955	-7.6986	-9.0158	-2.6109	+5431.12110435
27	+9.0544	-1.8311	-8.4082	-8.6804	+287.85624015
28	+4.5527	+2.4737	-3.3562	-3.0350	+209.38570907
29	+1.6332	+9.2973	-9.7020	+3.7074	+8554.91821163
30	+6.4841	-0.6785	-1.6540	-3.6276	+7.67650587
31	-9.5750	+4.8119	-4.6473	+2.7897	+1093.43641487
32	-5.1380	-6.4746	-9.8390	-4.1160	+4263.58569621
33	-9.6268	+9.7516	+4.9342	-2.8029	-4768.75138376
34	-0.0690	-9.8866	+4.8232	-8.5365	+66096.12022904
35	-2.5514	-6.8366	-9.3896	+4.0006	+4483.85794420
36	+1.1012	-3.4042	+6.2342	+4.2133	-703.13059664
37	+2.1838	-6.6772	+8.2155	-8.5150	-3774.22887091
38	-6.5666	-9.6278	+1.9927	+1.1854	-1896.35097643
39	+6.2236	+0.5106	+3.4386	-9.1188	-8.23001049
40	-8.0323	-7.2624	-9.0651	+4.7031	+4890.42795779

По данным табл. 2 ставится задача восстановления истинной закономерности $f(x_1,x_2,x_3,x_4)$.

Согласно этапу 1 модифицированного метода 2, не зная истинных компонент и соответствующих коэффициентов функции $f(x_1,x_2,x_3,x_4)$, исследователь сформировал множество из 45 компонент ($L=45$), возможных, по его мнению, составляющих искомой закономерности, в которое вошли также и компоненты из (6), и упорядочил случайным образом, как показано в табл. 3.

Таблица 3

Компоненты последовательности (2)

№	Комп.	№	Комп.	№	Комп.	№	Комп.	№	Комп.
1	$x_1^2x_3x_4$	10	$\frac{x_3^2x_4^2}{x_1}$	19	$\frac{x_4}{x_1^2}$	28	$\frac{x_4}{x_1}$	37	$\frac{x_2}{x_4}$
2	$\frac{x_1^2x_4}{x_2x_3}$	11	$\frac{x_2^3x_4}{x_1}$	20	$\frac{1}{x_1x_3^2x_4^2}$	29	$\frac{x_1^2x_2}{x_3^3}$	38	x_3x_4
3	$\frac{x_2x_4^2}{x_1}$	12	$\frac{x_1x_2^4x_4}{x_3^2}$	21	$\frac{x_1^2}{x_4}$	30	$\frac{x_1^3x_4^2}{x_3}$	39	$\frac{1}{x_3}$
4	$\frac{x_1}{x_2x_3}$	13	$\frac{x_3x_4}{x_1x_2}$	22	$x_1x_4^2$	31	$\frac{x_4}{x_2x_3}$	40	$\frac{x_3^2x_4}{x_2}$
5	$\frac{x_1}{x_3x_4}$	14	$x_2^2x_3$	23	$\frac{1}{x_2x_3x_4^2}$	32	$\frac{x_2}{x_1x_4}$	41	$\frac{x_2}{x_1x_4^2}$
6	x_2	15	$\frac{1}{x_1^2x_2x_3x_4^3}$	24	$\frac{x_2x_3}{x_4^2}$	33	$\frac{x_2^2x_3^2x_4^2}{x_1^2}$	42	$1(const)$
7	$\frac{x_1^2x_3}{x_4}$	16	$\frac{x_2x_4}{x_3^3}$	25	$\frac{1}{x_1}$	34	$x_2^3x_4^5$	43	$x_2x_3x_4^2$
8	$\frac{1}{x_1^4x_3^2}$	17	$\frac{1}{x_1x_3}$	26	$\frac{1}{x_3^3}$	35	$\frac{x_3^2x_4}{x_1^3x_2}$	44	$\frac{x_3}{x_1^2x_2^2x_4^2}$
9	$\frac{x_3^2}{x_1x_2}$	18	$\frac{x_4}{x_1x_3^2}$	27	$\frac{x_2x_4}{x_3}$	36	$\frac{x_4}{x_1}$	45	$\frac{x_1x_4^3}{x_2^2}$

Таким образом, исследователь построил последовательность (2). Далее он получает последовательность (3), оставив первые 40 компонент из таблицы 3. Так как в этом случае P не намного меньше L и намного больше K (на практике это обычно предположение исследователя), истинные компоненты неизвестной функции также попали (поскольку велика вероятность этого события) в последовательность (3).

Согласно модифицированному этапу 2 на каждой итерации из (3) последовательно отбрасывается каждая компонента и для оставшихся компонент решается задача (4)-(5). Окончательно исключается та компонента, при отбрасывании которой значение показателя качества (4) было минимально. В табл. 4 приведены результаты решения задачи (4)-(5) в первой итерации этапа (на этой итерации в (3) входит 40 компонент).

Таблица 4

Первая итерация этапа 2 модифицированного метода 2

№ отбр. компоненты	Значение показ. (4)	№ отбр. компоненты	Значение показ. (4)
1	+6.15798846	21	+19.97473174
2	+11.49255949	22	+10.61960516
3	+52.31254785	23	+15.81905781
4	+12.81325862	24	+19.48833838
5	+29.81379867	25	+8.51000588
6	+2.47518204	26	+24.30393908
7	+52.81285628	27	+21.02555301
8	+43.62772649	28	+15.68329164
9	+18.35329664	29	+24.27761766
10	+28.43070566	30	+25.13257786
11	+32.44892636	31	+8.90352016
12	+22.04431326	32	+23.23289859
13	+24.74367155	33	+57.26741220
14	+3394.77183228	34	+37.10437398
15	+9.23731262	35	+50.13540479
16	+23.90057386	36	+13.14950502
17	+24.35497304	37	+22.88837808
18	+1.78271230	38	+6.42942292
19	+0.48067162	39	+30.76551714
20	+27.97114710	40	+21.53112191

На первой итерации минимальное значение показателя качества (4) получено при отбрасывании компоненты с номером 19, поэтому эта компонента окончательно исключается из последовательности (3) и не участвует в дальнейших итерациях.

Аналогично выполняются следующие итерации этапа 2. Итерации прекращаем, если коэффициент прироста k_{opt} (отношение минимального значения показателя (4) на текущей итерации к этому же значению на предыдущей) составляет не менее 183,28 согласно таблице 1. В табл. 5 приведены результаты всех итераций этапа 2 (в столбцах таблицы: порядко-

вый номер итерации; номер компоненты, окончательно исключаемой из (3); минимальное значение показателя качества (4); коэффициент прироста k_{opt}).

Таблица 5

Результаты итераций

№ итерации	№ отбр. комп.	Мин. значение показателя качества (4)	k_{opt}
1	19	0.48067162	-
2	6	2.54545776	5.2956
3	18	5.42053786	2.1295
4	15	19.20954712	3.5438
5	1	30.77506318	1.6021
6	36	37.08658353	1.2051
7	25	44.49533939	1.1998
8	38	46.86268556	1.0532
9	31	50.34920905	1.0744
10	22	52.37363879	1.0402
11	12	59.34195627	1.1331
12	4	84.13350281	1.4178
13	23	96.69474388	1.1493
14	9	118.18589310	1.2223
15	13	119.65539208	1.0124
16	5	129.30322947	1.0806
17	3	134.44644208	1.0398
18	40	138.02184801	1.0266
19	27	144.22469929	1.0449
20	21	149.00304027	1.0331
21	28	158.09121343	1.0610
22	33	162.60399689	1.0285
23	2	168.23328838	1.0346
24	39	175.25844053	1.0418
25	30	182.07790461	1.0389
26	11	193.13531058	1.0607
27	24	198.51934639	1.0279
28	32	213.66718473	1.0763
29	7	221.85644211	1.0383

Продолжение табл. 5

30	37	230.60620624	1.0394
31	20	246.51191538	1.0690
32	29	252.76363441	1.0254
33	8	254.63230084	1.0074
34	17	256.08642727	1.0057
35	26	256.14460655	1.0002
36	16	256.45106368	1.0012
37	10	257.77143262	1.0051
38	34	285.91669153	1.1092
39	14	69117.12363753	241.7387

На последней итерации компонента с номером 14 не исключается, так как $k_{opt} \approx 242$ ($k_{opt} \geq 183,28$). Таким образом, получили, что закономерность $f(x_1, x_2, x_3, x_4)$ состоит из компонент с номерами 14 ($x_2^2 x_3$) и 35 ($\frac{x_3^2 x_4}{x_1^3 x_2}$). При этом значения коэффициентов при этих компонентах в восстанавливаемой закономерности получены в оптимальном решении задачи ЛП (4)-(5) на предпоследней итерации при исключении компоненты с номером 34 (приведены в табл. 6).

Таблица 6

Найденные коэффициенты

№ комп.	Истинные коэффициенты	Найденные коэффициенты	Модуль ошибки
14	-10.25166122	-10.22518428	0.026477
35	-1.15685002	-1.15950170	0.00265

Закономерность $f(x_1, x_2, x_3, x_4)$ восстановлена. Истинные коэффициенты найдены с высокой точностью.

Список литературы: ***1.*** *Павлов А. А., Штанькевич А. С.* Восстановление неизвестной закономерности по результатам пассивного эксперимента с ограниченным набором данных.– Вестник НТУ «ХПИ».– Харьков: НТУ «ХПИ».– 2009.– № 4.– С. 160-168.

Поступила в редколлегию 10.03.09