

А. Л. Буляница

(Институт аналитического приборостроения РАН, Санкт-Петербург)

ОЦЕНКА ПАРАМЕТРОВ ПОЛОЖЕНИЯ СИГНАЛА ТИПА "ЛИНЕЙНЫЙ ТРЕНД" В УСЛОВИЯХ НЕСИММЕТРИЧНОЙ ДИСКРЕТНОЙ ПОМЕХИ

The method of estimation of tracks dip angle for linear signal with accidental error is described. Some statistic characteristics of these estimates are determined. The describes method, which called the method of first differences (MFD), use well-known method of least quadrates (MLQ) as alternative method. The MFD has some advantages in comparison with MLQ in cases unsimmetrial errors distributing with small dispersion or complex discrete error is demonstrated.

В настоящее время все большее распространение получают кинетические виды анализа различных органических и неорганических соединений, в частности, фотометрические в биохимии, биофизике и аналитической химии.

При создании новых кинетических методов стараются получить информативный сигнал в виде линейно изменяющегося во времени напряжения или тока, т.е. получить сигнал в виде случайного линейного кусочно-детерминированного тренда. При использовании в приборах АЦП происходит искажение информативного сигнала несимметричной дискретной помехой, обусловленной:

— дрейфом параметров аналоговых элементов, имеющим характер систематической погрешности в форме линейного тренда, приведенным к цифровому виду;

— случайной погрешностью ξ показаний АЦП.

Оценки параметров положения и рассеивания трендов являются важнейшими характеристиками качества получаемой аналитиками информации о веществе. Указанные оценки для микропроцессорных фотометров в условиях несимметричной дискретной помехи известны лишь для методов конечной точки, т.е. вне зависимости от времени [1].

Целесообразно рассмотреть оценки параметров положения сигнала типа "линейный тренд". В работе выводятся оценки на примере случаев, когда скорость химической реакции ограничена скоростью диффузии реагентов, хотя не вызывает трудностей и обобщение предлагаемого подхода и на другие случаи.

Выходной сигнал J_j микропроцессорного фотометра удовлетворяет уравнению

$$J_j = a_0 \sqrt{t_j}, \quad j = 1, N, \quad (1)$$

где t_j — текущее время j -го отсчета, а J_j выражается так:

$$J_j = \lg(A_b - A_t) - \lg(A_j - A_t).$$

Здесь A_b , A_t и A_j — выходные сигналы АЦП, равные величине бланка, темноты (или паразитного рассеяния) и текущей оптической плотности соответственно.

Систематическая погрешность, обусловленная дрейфом, может не иметь монотонного характера, но для интервалов времени, соизмеримых с временами проведения анализов, систематическая погрешность может быть аппроксимирована зависимостью

$$J_j = b\sqrt{t_j}, \quad j = 1, N.$$

Оценивание величины b является основной целью данной работы. Оценка величины тангенса угла наклона прямой (1) может быть произведена с помощью той же самой процедуры, а для учета величины возможного рассеяния значения a_0 следует оценить параметры рассеяния b .

Постановка задачи

Берем холстую пробу, т.е. пробу при условии отсутствия химического взаимодействия. Процесс описывается уравнением

$$J_j = b\sqrt{t_j} + \mu_j, \quad j = 1, N, \quad (2)$$

где μ_j — случайная погрешность определения J_j , обусловленная случайной погрешностью ξ_j показаний АЦП. В случае малости тангенса угла наклона b очевидно, что в течение всего процесса измерения значения A_j близки к A_b и возможно ограничиться только двумя членами ряда в разложении J_j в ряд Тейлора. При этом μ_j пропорционально ξ_j , т.е. зная статистические свойства μ_j , можно оценить приведенную погрешность прибора ξ_j . Таким образом, решение задачи определения тангенса угла наклона для холстой пробы тесно связано с выявлением статистических свойств случайной погрешности μ_j . По существу, появляется возможность одновременного оценивания и систематической и случайной погрешностей измерения.

Процедура оценивания тангенса угла наклона прямой

Хорошо известен путь получения оценки b тангенса угла наклона прямой с помощью метода наименьших квадратов (МНК). В случае нормального закона распределения погрешности измерения оценка МНК является несмещенной, состоятельной и оптимальной в смысле установившейся дисперсии ошибки. Вместе с тем, в [2] показано, что по мере отклонения закона распределения от нормального эффективность МНК резко падает.

Проведенные опыты позволяют утверждать, что распределение μ_j весьма далеко от нормального, и в большинстве случаев распределение μ_j носит асимметричный характер.

Основа предлагаемого метода — процедура построения первых разностей

$$X_j = J_{j+1} - J_j, \quad j = 1, N-1.$$

Из (2) видно, что

$$X_j = b(\sqrt{t_{j+1}} - \sqrt{t_j}) + \eta_j,$$

где $\eta_j = \mu_{j+1} - \mu_j$, $j = 1, N-1$.

Рассмотрим свойства случайной величины η_j .

Пусть функция распределения μ_j имеет вид

$$\mu_j = \begin{cases} \frac{P_k}{h}, & \mu_j = kh \\ \dots & \\ \frac{P_{-k}}{h}, & \mu_j = -kh \end{cases}$$

где h — интервал дискретности J_j (и μ_j).

Тогда, очевидно, $\eta_j \in [-2kh, 2kh]$. Вероятность состояния $\eta_j = mh$ равна

$$\begin{cases} \mathcal{P}_m = \sum_{l=-k+m}^k P_l P_{l-m} & m = 0, 2k. \\ \mathcal{P}_{-m} = \mathcal{P}_m \end{cases} \quad (3)$$

Исходя из (3) имеем

$$\begin{cases} M\{\eta\} = 0 \\ D\{\eta\} = 2h^2 \sum_{m=1}^{2k} m^2 \mathcal{P}_m \end{cases} \quad (4)$$

Из (2) и с учетом (4) получим

$$\sum_{j=1}^{N-1} X_j = b \sum_{j=1}^{N-1} (\sqrt{t_{j+1}} - \sqrt{t_j}) + \sum_{j=1}^{N-1} \eta_j,$$

а $\bar{\eta} = \frac{1}{N-1} \sum_{j=1}^{N-1} \eta_j$, где $\bar{\eta}$ — выборочное среднее (одна из оценок $M\{\eta\}$).

Благодаря (4) подобная оценка тангенса угла наклона (\hat{b}) является несмещенной. Тогда

$$\hat{\delta} = \frac{\sum_{j=1}^{N-1} X_j}{N-1} \cdot \frac{1}{\sum_{j=1}^{N-1} (\sqrt{t_{j+1}} - \sqrt{t_j})}$$

или, с учетом старых обозначений

$$\hat{b} = \frac{J_N - J_1}{\sqrt{t_N} - \sqrt{t_1}}.$$

Разброс величины \hat{b} связан со статистическими свойствами η_j ,

$$D \{ \bar{\eta} \} = \frac{D \{ \eta \}}{N-1}. \quad (5)$$

В работе [2] показано, что

$$D \{ M \{ \eta \} \} = E \{ \bar{\eta} \} D \{ \bar{\eta} \}, \quad (6)$$

где $E \{ \bar{\eta} \}$ — эффективность оценки центра распределения с помощью выборочного среднего

$$E \{ \bar{\eta} \} = 1,50 \varepsilon^{-0,4}.$$

Здесь ε — эксцесс распределения η_j .

Поскольку величина η_j дискретна, следует учитывать поправки Шенпарда к моментам распределения η_j . В [3] приведены эти поправки

$$\left\{ \begin{array}{l} \mu_4 = 2 h^4 \sum_{m=1}^{2k} m^4 \mathcal{P}_m, \\ \Delta D \{ \eta \} = \frac{1}{12} h^2, \\ \Delta \mu_4 = -\frac{1}{2} h^2 D \{ \eta \} + \frac{7}{400} h^4, \end{array} \right. \quad (7)$$

где μ_4 — 4-й статистический момент распределения величины η_j , а

$$\varepsilon = \frac{\mu_4}{(D \{ \eta \})^2}. \quad (8)$$

С учетом выражений (4) — (8) можно получить $D \{ M \{ \eta \} \}$. В работе [2] показан критерий учета выбросов для случайной величины η_j с эксцессом распределения ε

$$t = 1,55 + 0,8 \sqrt{\varepsilon - 1} \lg \left(\frac{N-1}{10} \right),$$

где N — объем выборки.

Предельно допустимое отклонение центра распределения Δ_{\max} имеет вид

$$\Delta_{\max} = t \sqrt{D\{M\{\eta\}\}},$$

$$\sum_{j=1}^{N-1} \eta_j \approx (N-1) M\{\eta\},$$

тогда

$$|\Delta b|_{\max} = \frac{(N-1) \Delta_{\max}}{\sqrt{t_N} - \sqrt{t_1}}. \quad (9)$$

Очевидно, что при запоминании данных через равные интервалы времени приращение величины $\sqrt{t_j}$ или значение выражения $(\sqrt{t_{j+1}} - \sqrt{t_j})$ не будет постоянным в зависимости от номера шага j . Однако из всей последовательности отсчетов можно выделить подпоследовательность, которая будет равномерная в смысле приращения $(\sqrt{t_{j+1}} - \sqrt{t_j})$. Например, можно отбирать номера отсчетов по закону

$$k = j^2, \quad j = 1, l.$$

Тогда

$$X_k = b \Delta + \eta_k. \quad (10)$$

После процедуры упорядочивания элементов подпоследовательности можно воспользоваться оценками, предлагаемыми в работах [2] и [4]:

1) выборочное среднее \bar{X} ,

2) медиана $X_{\text{мед}}$,

3) центр сгиба $X_{\text{сг}} = \frac{1}{2} (X_{0,25} + X_{0,75})$,

4) центр размаха $X_{\text{раз}} = \frac{1}{2} (X_1 + X_l)$,

5) α — усеченное среднее $\bar{X}_\alpha = \frac{1}{l-2[l_\alpha]} \sum_{i=l_\alpha}^{l-[l_\alpha]} X_i$, где α — квантиль

распределения, $[l_\alpha]$ — целая часть (l_α) ,

6) оценка Гаствирта $X_r = 0,3 (X_{0,33} + X_{0,67}) + 0,4 X_{0,5}$,

7) оценка Тьюки

$$X_T = 0,25 (X_{0,25} + X_{0,75}) + 0,5 X_{0,5}. \quad (11)$$

Для оценок (11), кроме выборочного среднего, требуется создание упорядоченной последовательности. В случае, когда элементы выборки не удовлетворяют (10), невозможно однозначно сопоставить элементы b_i (оценки

Оценки тангенса угла наклона прямой для равномерной последовательности

Номер файла	Оценка								
	$b_{\text{МНК}}$	\bar{b}	$b_{\text{мед}}$	$b_{\text{сг}}$	$b_{\text{раз}}$	$b_{0,25}$	$b_{0,33}$	$b_{\text{г}}$	$b_{\text{г}}$
001	-0,171	-0,186	0,000	0,000	-1,000	0,125	0,000	0,000	0,000
002	-0,944	-0,807	0,000	-0,500	-1,000	-0,625	-1,000	-0,300	-0,250
004	0,087	0,062	0,000	0,500	0,000	0,000	0,000	0,000	0,250
005	0,030	0,062	0,000	0,000	0,500	0,000	0,000	0,000	0,000
007	-0,402	-0,311	0,000	0,000	-0,500	-0,375	-0,400	-0,300	0,000
010	-0,701	-0,559	0,000	-0,500	0,000	-0,750	-0,600	-0,300	-0,250
012	-1,057	-0,932	-1,000	-1,500	-1,000	-1,000	-0,800	-1,300	-1,250
102	-0,429	-0,311	0,000	-1,000	0,000	-0,500	-0,400	-0,600	-0,500
103	-0,070	-0,062	-1,000	-0,500	0,500	-0,250	-0,200	-0,700	-0,750
105	-0,834	-0,684	-1,000	-1,000	-0,500	-0,875	-0,800	-0,700	-1,000
108	-0,530	-0,373	0,000	-1,000	0,500	-0,375	0,000	-1,200	-0,500

величины b на i -м шаге) и X_i упорядоченной последовательности, также невозможно связать и значения погрешности определения b и X .

В таблице приведены все виды оценок (11) для реальных данных. Эти результаты показывают возможность объективизации оценки значения b при использовании сразу нескольких оценок или создания какой-либо комплексной оценки на базе (11). Оценки выражены в "угловых квантах" (интервал дискретности h , деленный на масштаб по оси абсцисс Δ , $c^{1/2}$). Также в таблице представлены оценки значения b , полученные по МНК. С учетом (9) и (10) можно записать

$$|\Delta b|_{\max} = \frac{\Delta_{\max}}{\Delta}.$$

В ходе рассуждения выявляется очевидное преимущество предлагаемого метода (в дальнейшем — метода первых разностей или МПР) над МНК — он значительно проще. Так, число вычислений для построения только выборочного среднего многократно меньше, чем в МНК. Для процедуры упорядочивания требуется не более $l(l-1)$ перестановок элементов, но так как $l \leq \sqrt{N} - 1$, то число операций не превосходит $N - 1$. Таким образом, для построения оценок (11) требуется существенно меньше операций, чем в процедуре МНК (если N не слишком мало).

Численный эксперимент

В ходе численного эксперимента требовалось:

- 1) сравнить качество оценок МНК и предлагаемого МПР;
- 2) выявить области, где качество оценок МНК ниже, чем МПР, и сопоставить характеристики этих областей с условиями эксперимента.

Строилась последовательность отсчетов

$$X_j = a_0 \sqrt{j}, \quad j = 1, N,$$

где a_0 задавался в пределах от 10^{-6} до 10^1 .

В соответствии с [5] генерировались 3 вида случайных величин ξ :

- 1) $\xi \sim R(0,1)$ — равномерное распределение,
- 2) $\xi \sim N(0,1)$ — нормальное распределение,
- 3) $\xi_1 \xi_2 \sim N(0,1)$; $\xi = \text{sign}(\xi_1 - \xi_2)$ — сигнатурное распределение.

Все эти случайные величины имеют симметричное распределение, но закон распределения помехи 3-го вида далек от нормального

$$J_j = a_0 (\sqrt{j} + k \xi_j).$$

Здесь k — масштаб помехи.

Далее вычислялись оценки $a_{\text{МНК}}$, $a_{\text{МПР}}$ и, соответственно, погрешности определения

$$\left\{ \begin{array}{l} \delta_{\text{МНК}} = \frac{|a_{\text{МНК}} - a_0|}{a_0}, \\ \delta_{\text{МПР}} = \frac{|a_{\text{МПР}} - a_0|}{a_0}. \end{array} \right.$$

У фотометров, используемых для кинетического анализа, величина дрейфа a_0 лежит в пределах от 0,1 до 1 квант/с. Для 12-разрядного АЦП наибольшая величина кванта на рабочем участке в фотометрических единицах не превосходит $2,2 \cdot 10^{-3}$, наименьшая — $1,1 \cdot 10^{-4}$ Е.О.П. соответственно. На рис. 1 представлена зависимость относительной погрешности определения тангенса угла наклона δ от a_0 .

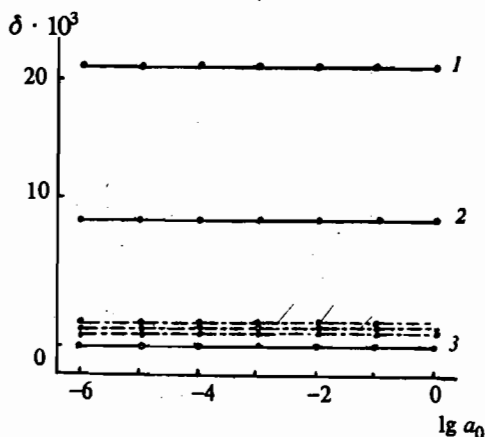


Рис. 1. Зависимость качества оценки δ от величины a_0 .

1 — нормальная помеха, 2 — равномерная, 3 — сигнатурная. Здесь и на рис. 2, 3: штриховая линия — МНК, сплошная — МПР.

Помимо введенных выше 3-х видов случайной помехи вводится сложная помеха вида

$$\left\{ \begin{array}{l} \mathcal{F}_i = \alpha_i N(C_i, \beta_i \sigma), \\ \xi = \sum_{i=1}^m \mathcal{F}_i, \end{array} \right. \quad (12)$$

где α_i — весовые коэффициенты, C_i — центры распределения \mathcal{F}_i , β_i — относительный коэффициент ширины \mathcal{F}_i , σ — масштаб дисперсии.

Представление Тьюки [6] для плотности распределения результатов эксперимента с учетом выбросов является частным случаем (12). Так

$$\xi = (1 - \varepsilon) N(0, \sigma) + \varepsilon N(0, 9\sigma),$$

где ε от 0 до 0,2. При $\sigma \rightarrow 0$ $N(m, \sigma) \rightarrow \delta(\xi - m)$, где δ — дельта-функция Дирака, т. е. (12) в этом случае будет представлять плотность распределения случайной величины ξ_0 , которая принимает значения C_i с вероятностью, пропорциональной α_i , $i = 1, m$. Уменьшение масштаба σ приближает распределение величины ξ к распределению дискретной величины ξ_0 .

Во всем исследуемом диапазоне абсолютных значений a_0 качество обеих оценок практически не зависит от величины a_0 . Таким образом, можно изме-

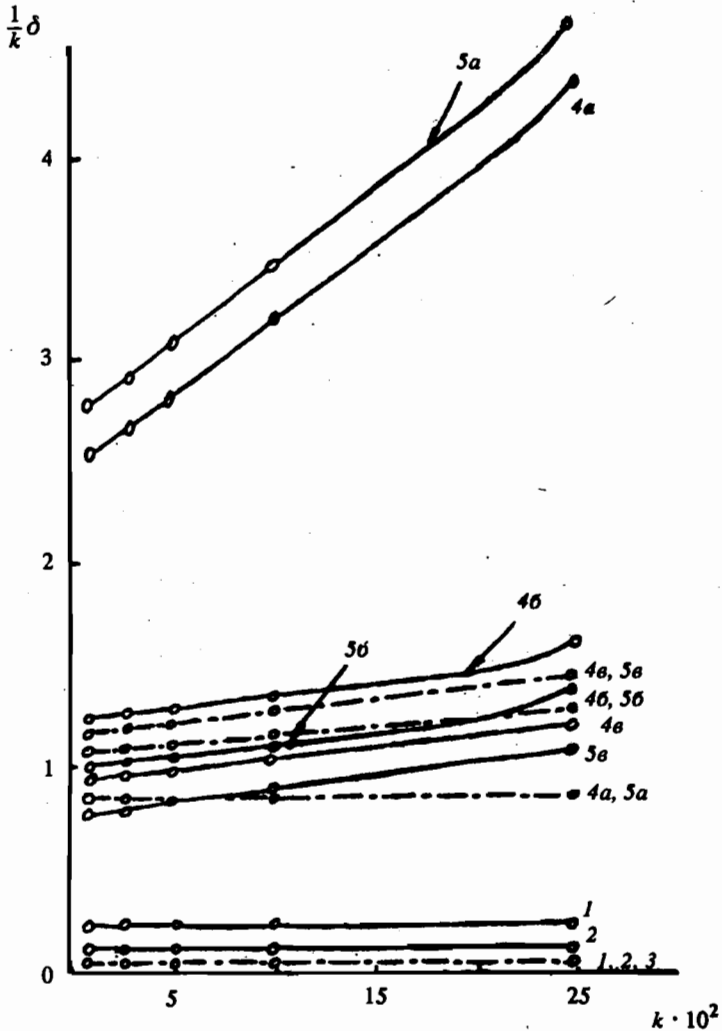


Рис. 2. Зависимость качества оценки δ от масштаба помехи k .
 Помехи: 1 — нормальная; 2 — равномерная; 3 — сигнатурная;
 4 — $N(0, \sigma) + 3N(1, \sigma)$; 5 — $3N(0, \sigma) + N(1, \sigma)$; для 4 и 5:
 а — $\sigma = 1$, б — $\sigma = 0,1$, в — $\sigma = 0,01$.

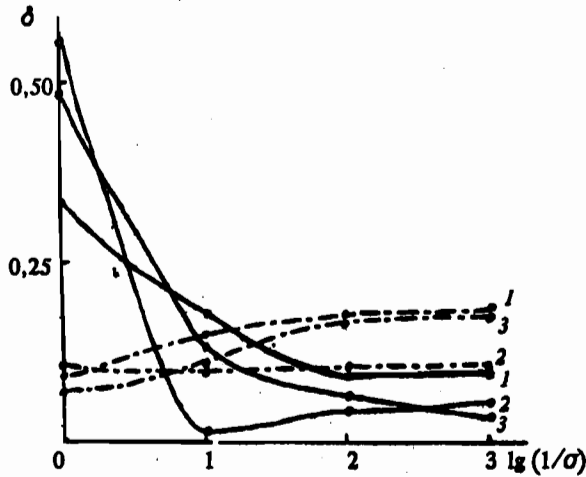


Рис. 3. Зависимость качества оценки δ от масштаба дисперсии σ (при масштабе помехи $k=0,10$).

Помехи: 1— $2N(0, \sigma) + N(1, \sigma)$; 2— $25N(0, \sigma) + N(1, \sigma)$; 3— $N(0, 2\sigma) + 3N(1, \sigma)$.

нить масштаб и рассмотреть задачу в безразмерном виде, в качестве масштаба используя квант дискретности h . Практически, распределение можно считать дискретным, если для любого $\mathcal{F}_i = \alpha_i N(C_i, \beta_i, \sigma)$ выполняется условие: $3\sqrt{\beta_i \sigma} \leq 0,5$, т.е. \mathcal{F}_i локализована вблизи C_i и округляется до уровня C_i квантов. Тогда $\beta_i \sigma \leq 1/36$, или $\sigma \leq [36\beta_i]^{-1}$. Случай $m=2$, $C_i = i-1$, $\max \beta_i = 25$ ниже рассмотрен более подробно.

На рис. 2 представлены зависимости относительной погрешности δ от масштаба помехи k . Качество оценок для $\sigma = 0,01$ и $\sigma = 0,001$ очень близко, потому случай $\sigma = 0,001$ на рис. 2 не представлен.

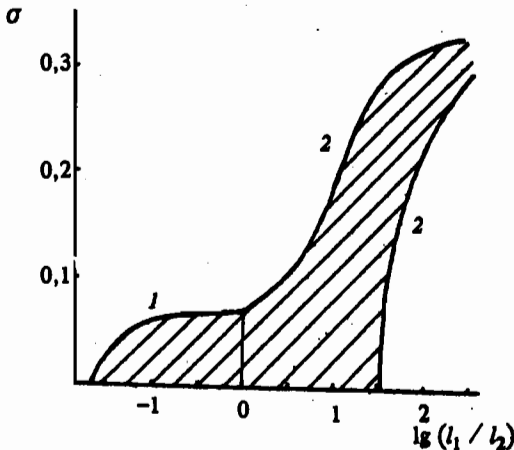


Рис. 4. Область использования МПР (цифрами 1, 2 обозначены участки).

На рис. 3 приведена зависимость относительной погрешности δ при $k=0,10$ от масштаба дисперсии σ .

На рис. 4 изображена область, где качество оценок МПР выше, чем качество оценок МНК. Рассматривается в качестве границы области предельное значение масштаба дисперсии σ в зависимости от степени несимметричности помехи

$$\xi = l_1 N(0, \sigma) + l_2 N(1, \sigma),$$

где 1-й участок $l_1 = 1$ и 2-й участок $l_2 = 1$.

Обсуждение результатов и выводы

Основными результатами численного эксперимента являются:

1. Построение зависимости точности оценок от масштаба помехи k . Для простых помех (кроме помех 3-го типа) оценка МНК значительно точнее. Кроме того, при больших k (более 0,25) происходит резкое ухудшение качества оценки МПР, особенно для сложных помех.

2. Учет зависимости качества оценок δ от масштаба дисперсии σ . В случае больших σ и помехи вида (12) качество оценки МНК низкое, а оценка МПР — просто неприемлемое. При уменьшении σ , т. е. дискретизации помехи, качество оценок МНК ухудшается (тем сильнее, чем больше асимметричность помехи), а качество оценок МПР возрастает. Эти результаты иллюстрируются рис. 2 и 3.

Вместе с тем, резкая асимметричность помехи (см. рис. 4), по-видимому, приводит к полному доминированию одного из двух слагаемых, т. е. практически приводит к нормальному распределению, где качество оценок МНК значительно выше. Участки 1 и 2 на рис. 4 несимметричны, так как в одном случае доминирует $N(1, \sigma)$, а в другом — $N(0, \sigma)$.

Кроме преимуществ, изложенных выше, МПР обладает еще одним: эта процедура позволяет построить функцию распределения вероятности случайной величины η . Обратив систему (3), можно получить плотность распределения вероятности для μ_j и для ξ_j , тем самым оценив приведенную случайную погрешность прибора. Таким образом, возможна одновременная компенсация и случайной и систематической погрешностей прибора и уточнение значения a_0 в (1). Обращение (3) возможно в явном виде при условии $P_k = P_{-k}$. В иных случаях решение получается приближенными итерационными методами.

Все аспекты решения задачи обращения системы (3), а именно: составление функционала качества, выбор метода итерации, построение приближенных оценок значений P_i в случае преобладания какой-либо величины P_j , построение первого приближения итерационной процедуры, проверка алгоритмов на модельных примерах — остаются за рамками данной работы.

Обсуждение подобной задачи производится единственно с целью показа роли МПР для оценки тангенса угла наклона тренда в контексте полного решения задачи кинетического анализа, т. е. получения оценки концентрации исследуемого вещества.

Существенным недостатком МПР является его низкое качество работы в условиях больших шумов. Вместе с тем, в качественных приборах соотношение шум/сигнал вряд ли может быть даже соизмеримо с величинами порядка 0,10—0,25.

Предлагаемая процедура МПР для оценки тангенса угла наклона прямой обладает рядом преимуществ по сравнению с таковой для МНК:

1) она дает лучшее качество оценки в условиях дискретных помех и помех с малой дисперсией;

2) она обеспечивает экономию вычислительных ресурсов по сравнению с МНК, что важно при использовании микропроцессорной техники в компакт-

ных сенсорных устройствах, так как позволяет использовать ОЗУ с малым объемом;

3) в некоторых случаях она дает возможность построения еще 6-ти робастных оценок, с помощью которых возможно производить уточнение результата в случае выбросов.

ЛИТЕРАТУРА

1. Курочкин В. Е., Полек А. М., Рахманкулов Р. В. и др. // Техническое оснащение медицинских лабораторных исследований. М., 1988. С. 51.

2. Новицкий П. В., Зограф И. А. Оценка погрешностей результатов измерений. Л., 1991. 248 с.

3. Кендалл М. Дж., Стьюарт А. Теория распределений: Пер. с англ. М., 1966. 587 с.

4. Устойчивые методы оценки данных / Под ред. Лонера Р. Л., Уилкинсона Г. Н. М., 1984. 314 с.

5. Голенко Д. И. Моделирование и статистический анализ псевдослучайных чисел на ЭВМ. М., 1965. 227 с.

6. Хампель Ф., Рончетти Э., Рассеу П., Штаэль В. Робастность в статистике. Подход на основе функций влияния: Пер. с англ. М., 1989. 512 с.

Рукопись поступила 24.11.92