НАУЧНОЕ ПРИБОРОСТРОЕНИЕ, 1999 том 9, № 1 с. 60–64

ОРИГИНАЛЬНЫЕ СТАТЬИ

УДК 519.245

ЗНАКОВЫЙ КРИТЕРИЙ ОПРЕДЕЛЕНИЯ РАЗЛАДКИ В ПОСЛЕДОВАТЕЛЬНОСТИ ИЗМЕРЕНИЙ

© А.Л. Буляница, Д.А. Бурылов

Институт аналитического приборостроения РАН, Санкт-Петербург.

Поступила в редакцию 3 февраля 1999г.

В работе предложен метод выявления разладки в последовательности измерений, базирующийся на анализе распределения вспомогательной случайной величины. Эта величина — число перемен знаков разностей текущего измеренного значения и соответствующей ему оценки в скользящем окне — формируется с учётом промежуточных результатов, получаемых в ходе выполнения алгоритма стохастической аппроксимации. Показана связь между мерой различия двух гистограмм — указанной случайной величины и эталонной (симметричной относительно нуля) — и параметрами разладки. Доказано, что в качестве меры можно выбрать начальный статистический момент третьего порядка или величину χ^2-критерия согласия.

ВВЕДЕНИЕ

При измерениях встречаются ситуации, когда существует априорная информация об анализируемом сигнале: в течение некоторого временного интервала известна его форма. Также предполагается, что на анализируемый сигнал накладывается аддитивная помеха с произвольным законом распределения. Кроме этого возможно изменение формы сигнала в произвольный момент времени.

Рассмотрим ситуацию, когда анализируемый сигнал представляет собой либо сигнал постоянного уровня, либо линейный тренд первого порядка. Информативными параметрами для них будут величина и тангенс угла наклона соответственно. Как было показано в работе [1], для оценки тангенса угла наклона можно перейти к первой разности исходного сигнала и оценивать сигнал постоянного уровня, при этом распределение помехи становится центрированным и симметричным. В этом случае для получения оценки тангенса угла наклона необходимо знать значения периода дискретизации и оценки получившегося сигнала постоянного уровня. В работе [2] для решения задачи оценивания сигнала постоянного уровня было предложено использовать следующую модификацию алгоритма стохастической аппроксимации:

$$x_{n+1} = x_n + \gamma_n sign(x_n - y_{n+1}),$$
$$\gamma_n = \beta(n+1)^{-1}, \qquad (1)$$

где x_n — оценка сигнала в момент времени n, y_{n+1} — наблюдаемое значение в момент времени $n+1$, $y_{n+1} = c^* + \varepsilon_{n+1}$, где c^* — истинное значение оцениваемого параметра, ε_{n+1} — величина помехи в $n+1$-ый момент времени, β — масштабный множитель, для которого необходимо выполнение условия $\beta \geq \beta_{\min}$ [3]. Было доказано, что в случае, если помеха является центрированной случайной величиной с симметричной плотностью распределения вероятностей, алгоритм (1) позволяет получить асимптотически несмещённую оценку анализируемого сигнала.

В случае изменения сигнала в течение времени оценивания, то есть, при наличии разладки в последовательности измерений, величина оценки может оказаться существенно смещенной, а сам этот факт остаться незамеченным.

Таким образом, нерешенной в рамках алгоритма (1) остаётся задача определения факта наличия или отсутствия разладки. Виды разладки могут быть различными, однако, в данной работе следует ограничиться двумя основными : скачок величины постоянного уровня в момент $t = t_0$, появление аддитивной систематической помехи в виде линейного тренда первого порядка — $\alpha(t - t_0)$. В случае возникновения разладки при оценивании линейного тренда первого порядка применение предлагаемого подхода приведёт к следующему: в результате вычисления первой разности измеряемого сигнала в момент разладки произойдёт скачкообразное изменение величины сигнала постоянного уровня.

Целью работы является поиск методики, позволяющей в реальном времени параллельно с определением оценки информативного параметра выявить разладку указанных двух типов. При этом, методика и количественный критерий разладки должны предполагать применение информации, полученной в ходе реализации алгоритма (1), и ограниченность используемых вычислительных ресурсов.

СПОСОБЫ ВЫЯВЛЕНИЯ РАЗЛАДКИ

Задача определения разладки непосредственно связана с решением задачи принятия гипотезы об однородности выборки. Эта задача, в частности, решалась на основе критерия Аббе, требующего, во-первых, нормальности закона распределения помехи, а во-вторых, значительного объема вычислений (выборочных средних и дисперсии, частичных дисперсий и т.д.). Кроме того, данный критерий не обладает устойчивостью к выбросам — робастностью.

Развитием критерия Аббе является алгоритм кумулятивных сумм Пейджа и его различные робастные аналоги [4–7]. Использование этого семейства алгоритмов снимает некоторые из указанных выше недостатков (чувствительность к выбросам и большой объем вычислений). Однако, алгоритм принятия решения «однородность-неоднородность» достаточно сложен — требуется установить два пороговых значения, определяемых априорно заданными временем задержки и вероятностью ложных тревог. Последствия неправильного выбора данных порогов могут привести к существенным ошибкам в определении момента разладки.

Еще одна группа подходов предполагает на основе статистических величин (кумулятивных сумм, фишеровской информации) разрешить альтернативы «однородность-неоднородность выборки» либо «принадлежность измерений семейству 1 — семейству 2». Данные методы точны (при отсутствии выбросов), но громоздки в вычислительном плане и неэффективны в случае больших (более 100) объемов выборки [8, 9].

Использование в качестве основы для создаваемого критерия алгоритма кумулятивных сумм представляется более перспективным в силу его робастности и вычислительной простоты. Модификацию его следует вести с целью упрощения применяемых решающих правил.

Таким образом, методику определения разладки следует искать как рекурсивную модификацию робастного алгоритма кумулятивных сумм. При этом, эта модификация должна быть сопряжена с алгоритмом (1), то есть, использовать уже полученные в ходе его реализации оценки.

МЕТОДИКА ОПРЕДЕЛЕНИЯ РАЗЛАДКИ. ПОСТАНОВКА ЧИСЛЕННОГО ЭКСПЕРИМЕНТА

В ходе реализации алгоритма (1) на каждом шаге вычисляется величина $\phi_n = sign(x_n - y_{n+1})$. Введём величину $\nu_n = \sum_{i=n-L}^{n} \phi_i$ — сумму значений ϕ_n в скользящем окне ширины L. В процессе формирования оценок с помощью алгоритма (1) достаточно быстро возникает условие:

$$\beta / n << s, \qquad (2)$$

где s — некоторый характерный масштаб случайной помехи (стандартное отклонение, медиана абсолютной величины, размах для ограниченных распределений, интерквартильная ширина и т.п.). При выполнении условия (2) последовательность знаков поправок в скользящем окне ширины L, во-первых, определяется исключительно знаком помехи, во-вторых, из-за малости ошибки оценивания, знак помехи на следующем шаге не зависит от предыдущего значения помехи. Таким образом, внутри скользящего окна каждый знак поправки формируется независимо.

Тем самым, для значительной части выборки, распределение знаков в окне описывается биномиальным распределением Бернулли. То есть, искомая величина ν_n детерминированно связана со схемой достижения m успехов в L испытаниях и представляет собой разность числа успехов и неудач, то есть, $(2m - L)$. При этом, при отсутствии систематических погрешностей и при правильном выборе параметра алгоритма (1) — β, вероятностью успеха (p) можно считать вероятность положительной помехи, а неудачи ($q = 1 - p$) — неположительной (отрицательной) помехи. В случае симметрии помехи обе вероятности есть 1/2, и гистограмма знаков симметрична относительно нуля.

Гистограмма строится на основании рекурсивной процедуры: на каждом шаге предполагается суммирование единицы к содержимому одной из $(L + 1)$ ячеек. То есть, само ее построение не требует значительных вычислительных затрат.

Наличие разладки типа скачка величины h на шаге n_0 приводит к существенному изменению формирования ν_n. Симметрия схемы Бернулли нарушается, так как вероятность успеха становится не 1/2 или $F(0)$, а $F(\pm h)$, где F — функция распределения помехи. В случае значительного отношения h / s степень асимметрии относительно нуля гистограммы может стать существенной. Фактически, при достаточно больших h для участка длиной $(n - n_0)$ наиболее распространенные значения ν_i будут $\pm L$ или $\pm(L - 2)$. Учитывая малость величины β / n по сравнению с h, вероятности успеха и неудачи (p, q) будут практически неизменными при $n > n_0$.

В случае разладки типа наличия линейного тренда, появившейся на шаге n_0, схема Бернулли также потеряет симметричность относительно нуля. В этом случае вероятностью успеха станет

$p = F(\alpha T(n - n_0))$, где α — тангенс угла наклона тренда. T — интервал дискретизации. Тем самым, смещение центра симметрии гистограммы относительно нуля связано, помимо крутизны линейного тренда, еще и с частотой дискретизации. Очевидно, что для повышения чувствительности, при выявлении подобного типа разладки, требуется увеличение значения T.

Таким образом, наличие факта разладки возможно установить путём сравнения образцовой гистограммы, полученной с помощью схемы Бернулли с параметрами ($p = q = 1/2$), и гистограммы, полученной на основе текущих измерений. Выбор меры различия гистограмм и порога, позволяющего принять решение о наличии или отсутствии факта разладки, базируется на результатах математического эксперимента.

Для проверки выдвинутых предположений был проведён следующий численный эксперимент. На исходный сигнал постоянного уровня накладывается помеха с равномерным законом распределения и следующими параметрами $m_x = 0$, размах $(b - a) = 1$. Значение β выбрано оптимальным образом, исходя из требований, определённых в [3]. Величина ν_n определяется в скользящем окне ширины $L = 16$. Численный эксперимент был поставлен для выборки объёмом $N = 2000$, разладка в виде скачка величины сигнала постоянного уровня или линейного тренда первого порядка наступала в моменты $n_0 = 0{,}1N$, $0{,}3N$, $0{,}5N$, $0{,}7N$, $0{,}9\,N$. Результаты численного эксперимента представлены на рис. 1 ... 3 и в таблицах 1 ... 4.

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ ЭКСПЕРИМЕНТА

В работах [10, 11] правила идентификации сигнала постоянного уровня и отсутствия тренда первого порядка установлены эмпирически, без теоретического обоснования, практически на основании текущего значения величины ν_n. В работе [10] предложено правило «1 из 5», примененное в условиях значительных уклонов. Так признаком данного события является наличие не менее одной перемены знака в последовательности из 5-ти измерений . То есть, исключаются события «0 успехов» и «5 успехов», суммарная вероятность которых составляет 1/16 или около 6%. В работе [11] сформулировано правило «3 из 10», применимое, по мнению автора, при условиях малости ошибки оценивания по сравнению с масштабом помехи. То есть, приведенные выше рассуждения относительно применимости схемы Бернулли обоснованы. Признаком указанной ситуации является наличие не менее 3-х перемен знака в 10-ти измерениях. В принятой нами терминологии 0, 1, 2 или 8, 9, 10 успехов в 10-ти испытаниях считается практически невозможным для симметричной схемы с $p = q = 1/2$. На самом деле, исключаются события, суммарная вероятность которых есть 112/1024 или около 10% . Естественно, что наличие гистограммы распределения ν_n значительно более информативно по сравнению с одиночными измерениями и, следовательно, позволяет делать статистически более достоверные выводы.

В качестве искомой меры различия гистограмм предложено выбрать либо нормированный на-

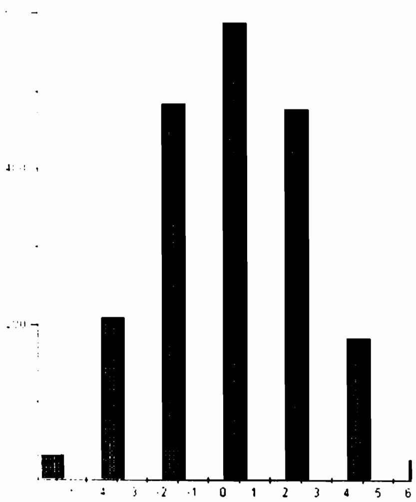

Рис. 1. Гистограмма распределения величины ν_n при отсутствии разладки.

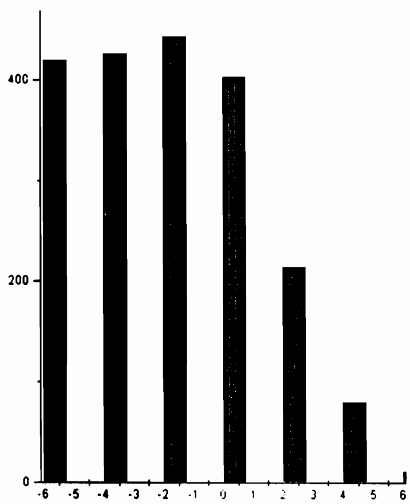

Рис. 2. Гистограмма распределения величины ν_n. Разладка — изменение величины сигнала постоянного уровня.

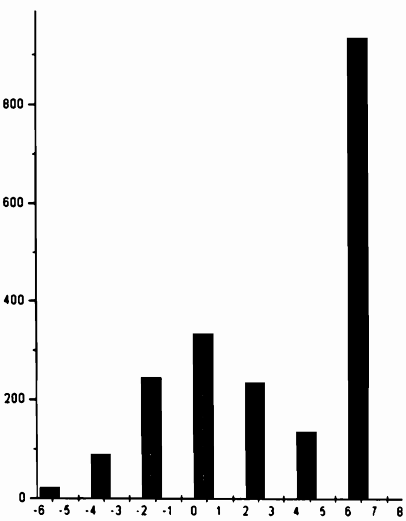

Рис. 3. Гистограмма распределения величины ν_n. Разладка - наложение линейного тренда первого порядка.

Таблица 1. Значения третьего статистического момента в зависимости от момента разладки для случая изменения величины сигнала постоянного уровня, $\beta = \beta_{опт}$

	n_0	0	0,1N	0,3N	0,5N	0,7N	0,9N
Величина	0,5(b–a)	–0,03	–2,67	–3,53	-2,75	-2,04	-1,60
	1(b–a)		–6,47	–5,88	–3,20	–2,20	–1,69
	2(b–a)		–29,69	–6,08	–3,22	–2,21	–1,66

Таблица 2. Значения третьего статистического момента в зависимости от момента разладки для случая наложения линейного тренда первого порядка $y_n = c^* + \varepsilon_n + \alpha T(n - n_0)$, $\beta = \beta_{опт}$

	n_0	0	0,1N	0,3N	0,5N	0,7N	0,9N
Параметр	α = 5,0e–4	2,44	2,29	1,84	1,31	0,66	–0,37
	α = 1,0e–3	5,15	4,87	3,30	2,27	1,40	0,09
	α = 5,0e–3	31,3	18,19	5,65	3,19	2,30	1,43

Таблица 3. Значения, полученные по χ^2 критерию в зависимости от момента разладки для случая изменения величины сигнала постоянного уровня, $\beta = \beta_{опт}$

	n_0	0	0,1N	0,3N	0,5N	0,7N	0,9N
Величина	0,5(b–a)	86	$3{,}07\cdot10^{4}$	$4{,}35\cdot10^{5}$	$8{,}44\cdot10^{5}$	$1{,}63\cdot10^{6}$	$6{,}75\cdot10^{5}$
	1(b–a)	107	$6{,}01\cdot10^{6}$	$4{,}75\cdot10^{7}$	$3{,}35\cdot10^{7}$	$1{,}25\cdot10^{7}$	$1{,}25\cdot10^{6}$
	2(b–a)	110	$1{,}01\cdot10^{8}$	$6{,}64\cdot10^{7}$	$3{,}35\cdot10^{7}$	$1{,}18\cdot10^{7}$	$1{,}19\cdot10^{6}$

Таблица 4. Значения, полученные по χ^2 критерию в зависимости от момента разладки для случая наложения линейного тренда первого порядка $y_n = c^* + \varepsilon_n + \alpha T(n - n_0)$, $\beta = \beta_{опт}$

	n_0	0	0,1N	0,3N	0,5N	0,7N	0,9N
Параметр	α = 5,0e–4	$2{,}16\cdot10^{6}$	$1{,}92\cdot10^{6}$	$4{,}01\cdot10^{5}$	$3{,}10\cdot10^{4}$	$9{,}31\cdot10^{2}$	$1{,}23\cdot10^{2}$
	α = 1,0e–3	$4{,}50\cdot10^{7}$	$4{,}25\cdot10^{7}$	$2{,}42\cdot10^{7}$	$7{,}50\cdot10^{6}$	$2{,}32\cdot10^{5}$	$1{,}65\cdot10^{2}$
	α = 5,0e–3	$1{,}20\cdot10^{8}$	$1{,}03\cdot10^{8}$	$6{,}11\cdot10^{7}$	$2{,}94\cdot10^{7}$	$9{,}06\cdot10^{6}$	$3{,}66\cdot10^{5}$

чальный статистический момент третьего порядка, либо "Хи-квадрат" критерий согласия.

Обе выбранных меры чувствительны как к вариации центра симметрии, так и к асимметрии гистограммы. Расчет мер несложен, если ширина окна не слишком велика (порядка 10–20).

Как показали данные эксперимента, при не очень значительном завышении β по сравнению с величиной $\beta_{\min}$ полученная гистограмма имеет параметры близкие к параметрам распределения Бернулли с $p = q = 1/2$ и $n = L$.

Данные, представленные в таблицах 1 и 2 свидетельствуют, что для признания факта наличия разладки для величины третьего момента можно установить порог, равный по модулю единице.

В таблицах 3 и 4 представлены значения для «Хи-квадрат» критерия, полученные в ходе численного эксперимента. Во всех случаях гипотеза о равенстве гистограмм принимается с крайне низким уровнем значимости. Даже при отсутствии разладки, как правило, уровень значимости гипотезы не превосходит 1–5 %, но в случаях наличия факта разладки гипотеза равенства неприемлема практически со 100%-ным уровнем значимости. Известно, что на начальном участке условие $\beta / n < s$ не выполнено. Таким образом, схема Бернулли неприемлема, поскольку знак поправки может определяться начальной погрешностью оценивания. То есть, условие независимости испытаний нарушается (вероятность знака поправки на следующем шаге существенным образом зависит от величины ошибки оценивания на предыдущем шаге). Кроме того, на самых первых шагах, может быть выполнено обратное условие $\beta / n >> s$, и тогда будет осуществлено практически детерминированное чередование знаков. Этот эффект приведет к "утяжелению" центра гистограммы по сравнению с эталоном, соответствующим симметричной схеме Бернулли.

Таким образом, применение критерия согласия из-за его высокой чувствительности и заложенной методической ошибки, вызванной учетом начальных измерений, не даст статистически значимых результатов. Тем не менее, отличия полученных значений будут столь велики, что имеется возможность дискриминации ситуаций наличия-отсутствия разладки и подбора порога, обеспечивающего малые вероятности ошибок классификации 1-го и 2-го рода.

ВЫВОДЫ

Предложенная методика позволяет решать задачу определения разладки сигнала постоянного уровня или линейного тренда первого порядка. Она базируется на сравнении образцовой гистограммы и гистограммы, определяемой суммой знаков поправки за последние L шагов реализации алгоритма (1). Преимуществом данного подхода является то, что фактически он является дополнением к алгоритму (1) и позволяет определять разладку без особых вычислительных затрат. Следует также отметить, что для более точного определения разладки необходимо правильно выбрать параметр β алгоритма (1) и задать ширину скользящего окна L.

ЛИТЕРАТУРА

1. *Буляница А.Л.* // Научное приборостроение. 1993. Т.3, №2. С. 69–78.
2. *Цыпкин Я.З., Поляк Б.Т.* //Динамика систем. Математические методы теории колебаний. 1977. Вып. 12. Горький: Издательство Горьковского ун-та. С.22–46.
3. *Бедельбаева А.А.* // Автоматика и телемеханика. 1978. № 1. С.87–95.
4. *Page E.S.* //Biometrika. 1954. V.41, №1. P. 141–154.
5. *Page E.S.* // Biometrika. 1955. V.42, №4. P.523–527.
6. *Никифоров И.В.* Последовательное обнаружение изменения свойств временных рядов. М. Наука. 1983. 254с.
7. *Воробейчиков С.Э* // Автоматика и телемеханика. 1998. №3. С. 50–56.
8. *Бродский Б.Е., Дарховский Б.С.* // Автоматика и телемеханика. 1993. № 1. С.62–67.
9. *Курочкин В.Е., Макарова Д.Е., Манойлов В.В., Сирвидас С.И.* // Тезисы докладов I Международный симпозиум "Электроника в медицине. Мониторинг, диагностика, терапия". СПб. Вестник аритмологии. 1998. № 8. С.146.
10. *Ивницкий Д.М., Ситдыков Р.А., Курочкин В.Е., Рейфман Л.С.* // Журнал аналитической химии. 1991. Т. 46, №6. С.1239–1244.
11. *Поляк Б.Т.* // Автоматика и телемеханика. 1990. №7. С.98–107.

SIGN CRITERION FOR DETECTION OF A DISSENSION IN A MEASUREMENT SEQUENCE

A.L. Bulianitsa, D.A. Burylov

Institute for Analytical Instrumentation RAS, St.Petersburg

In this work, a method to reveal dissension in a measurement sequence based on the analysis of the distribution of an auxiliary random value is offered. This value — the number of sign changes of the difference between the current measurement and the respective estimate in a moving window — is formed in view of intermediate results obtained during execution of the stochastic approximation algorithm. The relation between the measure of the difference between two histograms of the specified random value and reference one (symmetric with respect to zero) and the dissension parameters is shown. It is proved that as such a measure it is possible to choose the initial statistical moment of the third order or the value of the χ^2 goodness-of-fit criterion.