РОССИЙСКАЯ АКАДЕМИЯ НАУК

СИБИРСКОЕ ОТДЕЛЕНИЕ

АВТОМЕТРИЯ

2002, том 38, № 4

УДК 519.27

Ч. М. Гаджиев

(Стамбул, Турция)

ПОДХОД К ОТБРАКОВКЕ АНОМАЛЬНЫХ ИЗМЕРЕНИЙ, РОБАСТНЫЙ К СИСТЕМАТИЧЕСКИМ ПОГРЕШНОСТЯМ

Предлагается подход к отбраковке аномальных измерений, робастный к систематическим погрешностям. Для этого вводится невязка, включающая разности двух последовательных измерений и расчетных значений выходных координат модели системы. Исследованы статистические характеристики введенной невязки. В результате разработан оперативный метод отбраковки аномальных измерений, не требующий информации о величине и знаке систематической погрешности, а также о статистических характеристиках аномальных измерений.

Введение. Появление аномальных измерений на входе ЭВМ, ведущей статистическую обработку измерений, обусловлено машинными сбоями при предварительной обработке, сбоями при передаче данных по каналам связи и резким нарушением условий работы измерительных средств.

Существует множество процедур отбраковки аномальных измерений, использующих различного рода статистические данные об измеряемом процессе [1–5]. При классическом подходе к решению указанной задачи необходимо: предположить, что выборочные наблюдения производятся над случайной величиной, построить соответствующую статистику для обнаружения выбросов, чувствительную к резким изменениям, найти ее распределение при нулевой гипотезе (все наблюдения принадлежат одной и той же совокупности) и затем отвергнуть гипотезу, если окажется маловероятным, что вычисленная статистика имеет место в данной случайной выборке. В работах по обработке результатов измерений приводится большое количество таких статистик [1–3]. При таком подходе к решению задачи формирование упомянутой статистики снижает оперативность процедуры отбраковки, так как для этого необходимо использовать выборку большого объема. В то же время отбраковку аномальных измерений желательно производить в темпе поступления измерительной информации. Ряд таких алгоритмов можно найти в [4, 5]. Указанные алгоритмы базируются на знании математического описания измеряемого процесса, а сама процедура отбраковки сводится к простейшим операциям сравнения модуля разности между измерением и оценкой сигнала с некоторым заранее вычисленным порогом. Однако в этих алгоритмах систематические погрешности измерений и их влияние на процедуру принятия решения не учитываются. В реальных условиях решение задачи отбраковки аномальных измерений в большинстве случаев приходится проводить при наличии систематических погрешностей измерений, среднее значение которых не равно нулю. Появление систематической погрешности связано с наличием неучтенных постоянных или медленно меняющихся факторов, к которым можно отнести условия распространения радиоволн, изменение опорной частоты генераторов, ошибки геодезической привязки измерительного средства и т. д.

Присутствие систематических погрешностей снижает достоверность вышеупомянутых методов обнаружения аномальных измерений. При этом неучет систематической погрешности может привести к тому, что в результате обработки нормальные измерения могут отбраковываться как аномальные, а аномальные приниматься нормальными.

Таким образом, систематические составляющие погрешностей измерений могут привести к значительному снижению эффективности процесса отбраковки аномальных измерений, поэтому их учет имеет большое практическое значение. Далее предлагается новый подход к конструированию алгоритмов отбраковки аномальных измерений, робастный к систематическим погрешностям в измерениях.

Постановка задачи. В данной работе предложен метод отбраковки аномальных результатов в линейных моделях динамических измерений в условиях нормального распределения случайных ошибок измерений. Пусть измеряется выходная координата x_i линейной динамической системы, математическая модель которой в конечных разностях имеет вид

$$x_i = ax_{i-1} + by_{i-1}, \qquad (1)$$

где y_{i-1} – входное воздействие; a – известный параметр системы; b – коэффициент, характеризующий интенсивность входного воздействия.

При этом истинные значения параметров модели (1) связаны с их оценками, полученными при идентификации, в виде уравнений $a = \hat{a} + \Delta a$; $b = \hat{b} + \Delta b$, где $\hat{a}$ и $\hat{b}$ – оценки параметров рассматриваемой динамической системы; Δa и Δb – случайные составляющие с нулевыми математическими ожиданиями и конечными дисперсиями.

Уравнение нормальных измерений записывается в виде

$$z_i = x_i + \xi_i + \lambda_i, \qquad (2)$$

где λ_i – значение систематической погрешности; аддитивный шум ξ_i имеет нормальный закон распределения со следующими статистическими характеристиками:

$$E[\xi_i] = 0, \quad E[\xi_i \xi_j] = D_{\xi_i}\delta_{ij}, \quad E[\xi_i \Delta a_j] = 0, \quad E[\xi_i \Delta b_j] = 0, \qquad (3)$$

здесь E – оператор статистического усреднения, δ_{ij} – символ Кронекера:

$$\delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

Принимаем, что

$$E[\lambda_i] = E[\lambda_{i-1}], \qquad (4)$$

так как систематическая погрешность при повторных измерениях остается постоянной или изменяется медленно.

В измерениях z_i выходной координаты системы содержатся редкие единичные сбои, моменты появления которых случайны, статистические характеристики неизвестны. Систематические погрешности λ_i также неизвестны.

При вышеизложенных условиях с целью повышения эффективности обработки измерительных потоков требуется оперативный метод отбраковки аномальных измерений, робастный к систематическим погрешностям в измерениях.

Алгоритм решения. Учитывая малость значений приращений оценок параметров и начальных значений состояния модели системы x_0, решение уравнения (1) приближенно запишем в следующем виде:

$$x_i \approx x_{p_i} + \frac{\partial x_i}{\partial(\Delta a)}\Delta a + \frac{\partial x_i}{\partial(\Delta b)}\Delta b + \frac{\partial x_i}{\partial(\Delta x(0))}\Delta x(0),$$

где x_{p_i} – решение уравнения (1), полученное на основе оценок параметров.

Поскольку исследуемая система имеет линейную модель, то можно предположить, что достаточное количество информации о статистической природе расчетного значения выходной координаты модели системы будет содержаться в описании его первыми двумя моментами:

$$E[x_i] = \bar{x}_{p_i}, \qquad E[(x_i - \bar{x}_{p_i})^2] = D_{x_i}, \qquad (5)$$

где принимаем

$$\bar{x}_{p_i} = \hat{a}\bar{x}_{p_{i-1}} + \hat{b}\hat{y}_{i-1}. \qquad (6)$$

Выражения для определения D_{x_i} можно найти в [5].

В силу центральной предельной теоремы теории вероятности, учитывая множество компонент, влияющих на значение дисперсии D_{x_i}, распределение истинного значения выходной координаты x_i в момент времени i принимаем нормальным $N(\bar{x}_{p_i}, D_{x_i})$. Таким образом, оценка $\bar{x}_{p_i}$, полученная в соответствии с выражением (6) путем прямой экстраполяции начальных условий (в качестве начальных условий используются оценки параметров и состояния системы на момент прекращения процесса идентификации с соответствующими дисперсиями), представляет собой расчетное значение измеряемой координаты.

Как известно, оценка не совпадает с истинным значением и вследствие этого существует ошибка $\varepsilon_i = x_i - \bar{x}_{p_i}$.

Обозначим

$$\Delta\bar{x}_{p_i} = \bar{x}_{p_i} - \bar{x}_{p_{i-1}}, \qquad \Delta z_i = z_i - z_{i-1} \qquad (7)$$

и для отбраковки аномальных измерений используем статистику

$$u_i = \frac{\Delta z_i - \Delta \bar{x}_{p_i}}{D_{\Delta_i}},$$

где D_{Δ_i} – дисперсия разности $\Delta z_i - \Delta \bar{x}_{p_i}$. Покажем, что величина $\Delta z_i - \Delta \bar{x}_{p_i}$ в случае нормальных измерений распределена по гауссовскому закону с нулевым математическим ожиданием. Для этого с учетом (7) разность $\Delta z_i - \Delta \bar{x}_{p_i}$ представим в следующем виде:

$$\Delta z_i - \Delta \bar{x}_{p_i} = z_i - z_{i-1} + \bar{x}_{p_{i-1}} - \bar{x}_{p_i}. \qquad (8)$$

Как известно [6], случайные величины, полученные в результате любых линейных преобразований нормально распределенных случайных величин, распределены нормально. На основе этого можно констатировать, что величина $\Delta z_i - \Delta \bar{x}_{p_i}$ распределена по нормальному (гауссовскому) закону.

С учетом (2) и (8) определим математическое ожидание разности $\Delta z_i - \Delta \bar{x}_{p_i}$:

$$E[\Delta z_i - \Delta \bar{x}_{p_i}] = E[z_i - z_{i-1} - \bar{x}_{p_i} + \bar{x}_{p_{i-1}}] =$$

$$= E[x_i - \bar{x}_{p_i} + \bar{x}_{p_{i-1}} - x_{i-1}] + \{E[\xi_i] - E[\xi_{i-1}]\} + \{E[\lambda_i] - E[\lambda_{i-1}]\}.$$

Первый член в правой части этого уравнения равен нулю вследствие выражения (5). Второй и третий члены равны нулю согласно (3) и (4) соответственно. Следовательно, имеем $E[\Delta z_i - \Delta \bar{x}_{p_i}] = 0$.

Покажем, что при принятых допущениях (3), (4) и отсутствии аномальных измерений дисперсия разности $\Delta z_i - \Delta \bar{x}_{p_i}$ определяется выражением

$$D_{\Delta_i} = 2D_{\xi_i} + D_{x_i} + D_{x_{i-1}} - 2D_{x_{i,i-1}}. \qquad (9)$$

Учитывая независимости измерений и расчетных значений выходной координаты модели, дисперсию разности $\Delta z_i - \Delta \bar{x}_{p_i}$ в общем виде можем представить как

$$D[\Delta z_i - \Delta \bar{x}_{p_i}] = D[x_i + \xi_i + \lambda_i - x_{i-1} - \xi_{i-1} - \lambda_{i-1} - \bar{x}_{p_i} + \bar{x}_{p_{i-1}}] =$$

$$= D[\xi_i - \xi_{i-1}] + D[(x_i - \bar{x}_{p_i}) - (x_{i-1} - \bar{x}_{p_{i-1}})] = 2D_{\xi_i} + D[\varepsilon_i - \varepsilon_{i-1}], \qquad (10)$$

где $\varepsilon_{i-1} = x_{i-1} - \bar{x}_{p_{i-1}}$.

Получение дисперсии величины $\varepsilon_i - \varepsilon_{i-1}$ представляет сложную задачу, так как ошибки в двух последовательных расчетных значениях выходной координаты модели являются коррелированными из-за использования при их вычислении одной и той же модели, а также одних и тех же начальных условий. В результате между указанными ошибками появляются перекрестные ковариации.

Принимая обозначение $\mu_i = \varepsilon_i - \varepsilon_{i-1}$ и применяя полученные в [7] результаты для решения данной задачи, запишем:

$$D[\varepsilon_i - \varepsilon_{i-1}] = E[\mu_i^2] = E[\varepsilon_i^2 - \varepsilon_i\varepsilon_{i-1} - \varepsilon_{i-1}\varepsilon_i + \varepsilon_{i-1}^2] =$$

$$= D_{x_i} + D_{x_{i-1}} - D_{x_{i,i-1}} - D_{x_{i-1,i}}, \qquad (11)$$

где $D_{x_{i,i-1}}$ и $D_{x_{i-1,i}}$ – перекрестные ковариации между ошибками двух последовательных расчетных значений выходной координаты модели, причем

$$D_{x_{i,i-1}} = D_{x_{i-1,i}}. \qquad (12)$$

Подставляя выражение (11) в (10), с учетом (12) получим доказываемое выражение (9).

На основе этого можем констатировать, что предложенная для обнаружения аномальных измерений статистика u_i при нормальных измерениях имеет нормированное нормальное распределение:

$$u_i = \frac{\Delta z_i - \Delta\bar{x}_{p_i}}{D_{\Delta_i}} \sim N(0,1). \qquad (13)$$

Измерение считается аномальным, если значение статистики u_i превышает при выбранном уровне значимости квантиль стандартного нормального распределения, т. е. $u_i > u_\beta$, где u_β – квантиль стандартного нормального распределения; $\alpha = 1 - \beta$ – выбранный уровень значимости.

Предложенный алгоритм обнаружения аномальных измерений не требует никакой информации о величине и знаке систематической погрешности, а также о статистических характеристиках аномальных измерений. Указанный алгоритм достаточно прост и легко реализуется на ЦВМ.

Следует отметить, что при использовании данного подхода выявленное аномальное измерение из дальнейшей обработки исключается. Вследствие того что в контролируемую статистику (13) входит разность двух последовательных измерений, возникает трудность при формировании разности Δz_{i+1} после отбраковки i-го измерения.

Для решения указанной задачи предлагается следующий подход. При $u_i > u_\beta$ принимаем $u_i = u_\beta$, отсюда

$$\Delta z_i = u_\beta D_{\Delta_i} + \Delta\bar{x}_{p_i}.$$

С учетом $\Delta z_i = z_i - z_{i-1}$ получим

$$z_i = z_{i-1} + u_\beta D_{\Delta_i} + \Delta\bar{x}_{p_i}, \qquad (14)$$

т. е. выявленное аномальное измерение z_i из дальнейшей обработки исключается и в разработанном алгоритме отбраковки при формировании разности

$\Delta z_{i+1} = z_{i+1} - z_i$ в качестве z_i используется его вычисленное с помощью (14) значение.

При формировании разности $\Delta z_i = z_i - z_{i-1}$ должна быть уверенность в том, что измерение z_{i-1} не является аномальным (z_{i-1} – либо нормальное измерение, либо вычисленное с помощью (14) граничное значение). В противном случае использование данного подхода может привести к неверным результатам.

В реальных условиях эксплуатации системы предложенный алгоритм обнаружения и отбраковки аномальных измерений сводится к следующей последовательности вычислений.

1. Определяется допустимое значение (квантиль) стандартного нормального распределения u_β соответствующей выбранной доверительной вероятности β.

2. В момент времени i по поступившим последовательным измерениям z_{i-1} и z_i (должна быть гарантирована нормальность измерения z_{i-1}) и расчетным значениям выходной координаты модели $\Delta\bar{x}_{p_i}$ и $\Delta\bar{x}_{p_{i-1}}$ с помощью (8) вычисляется значение $\Delta z_i - \Delta\bar{x}_{p_i}$.

3. На основе (9) и (13) определяются соответственно дисперсия разности $\Delta z_i - \Delta\bar{x}_{p_i}$ и значение статистики u_i.

4. Проверяется выполнение неравенства $u_i > u_\beta$. Возможны два случая: а) при невыполнении указанного неравенства измерение z_i считается нормальным и поступает в дальнейшую обработку и последовательность предложенных контрольных вычислений повторяется, начиная с п. 2, для следующего момента времени $i+1$; б) в случае выполнения неравенства $u_i > u_\beta$ с помощью (14) определяется допустимое граничное значение измерения z_i и последовательность вычислений повторяется, начиная с п. 2, для следующего момента времени $i+1$. При этом в качестве предыдущего измерения используется значение z_i, найденное на основе (14).

Заключение. Предложен новый подход к отбраковке аномальных измерений, робастный к систематическим погрешностям в измерениях. Данный подход с целью обнаружения аномальных измерений использует разности двух последовательных измерений и расчетных значений выходных координат модели системы. В результате чего систематические погрешности измерений взаимно исключают друг друга. Основным достоинством такого подхода является то, что при обнаружении аномальных измерений с учетом их смещений не требуется никакой информации о величине и знаке этих смещений, а также о статистических характеристиках аномальных измерений.

Предложенный алгоритм отбраковки не требует больших вычислительных затрат и может широко использоваться в различных технических отраслях при обработке измерительной информации, например, в радиоинерциальных системах навигации, автоматизированных системах управления технологическими процессами, устройствах обработки данных в радиолокационных системах, при управлении летательными аппаратами и др.

СПИСОК ЛИТЕРАТУРЫ

1. **Летные** испытания ракет и космических аппаратов /Под ред. Е. И. Кринецкого. М.: Машиностроение, 1979.

2. **Стогов Г. В., Макшанов А. В., Мусаев А. А.** Устойчивые методы обработки результатов измерений (обзор) // Зарубежная радиоэлектрон. 1982. № 9. С. 3.

3. **Фомин А. Ф., Новоселов О. Н., Плющев А. В.** Отбраковка аномальных результатов измерений. М.: Энергоатомиздат, 1985.

4. **Зелененький В. П.** Применение методов теории статистических решений при исключении аномальных измерений // Изв. АН СССР. Техническая кибернетика. 1969. № 2. С. 139.

5. **Гаджиев Ч. М.** Отбраковка аномальных измерений с учетом погрешностей математической модели измеряемого процесса // Метрология. 1998. № 2. С. 3.

6. **Пугачев В. С.** Теория вероятностей и математическая статистика. М.: Наука, 1979.

7. **Brumback B. D., Srinath M. D.** Chi-square test for fault detection in Kalman filters // IEEE Trans. Automat. Contr. 1987. **AC-32**, N 6. P. 552.

Стамбульский технический университет,
E-mail: cingiz@itu.edu.tr

Поступила в редакцию
26 сентября 2001 г.