

Робастное согласование неточных данных на основе аппроксимации плотности вероятности рядами на основе полиномов Лагерра

В. А. Гаранин

*Высшая школа компьютерных технологий и информационных систем
Санкт-Петербургский политехнический университет Петра Великого
garanin_va@spbstu.ru*

К. К. Семенов

*Высшая школа компьютерных технологий и информационных систем
Санкт-Петербургский политехнический университет Петра Великого
semenov_kk@spbstu.ru*

Аннотация. Данная работа посвящена разработке робастного метода согласования результатов совместных измерений взаимосвязанных величин, выполняемых в многоканальных информационно избыточных измерительных и управляющих системах. Метод позволяет повысить точность получаемых результатов измерений, а также устойчивость к выбросам за счет аппроксимации закона распределения случайных погрешностей через разложение по полиномам Лагерра. Данное обстоятельство обеспечивает более высокую робастность оценок значений измеряемых величин, получаемых в ходе процедуры согласования, по сравнению с традиционными подходами к данной задаче.

Ключевые слова: согласование результатов измерений, робастная статистика, учет априорной информации, проекционная оценка плотности распределения вероятностей, полиномы Лагерра

I. ВВЕДЕНИЕ

Современные многоканальные измерительные и управляющие системы часто характеризуются намеренной информационной избыточностью измерительной задачи. Такая избыточность может быть структурной, подразумевающей дублирование датчиков и/или измерительных каналов, а также временной, связанной с передискретизацией медленно изменяющихся сигналов измерительной информации. Согласование неточных данных представляет собой совокупность методов математической обработки результатов, имеющих целью дополнить измерительную информацию априорными сведениями об объекте или процессе, на котором выполняются измерения [1, 2]. Данная группа методов успешно применяется для решения двух основных задач. Во-первых – повышения точности прямых измерений за счет согласованного пересчета показаний всех измерительных каналов, служащих для измерения значений взаимосвязанных величин x , что позволяет уменьшить дисперсию получаемых итоговых оценок x^* для x , обусловленную случайной погрешностью измерений. Во-вторых – увеличения надежности информационно-измерительных и управляющих систем: показания от дублирующих друг друга измерительных каналов сравниваются между собой для выявления выбросов, то есть аномальных

результатов, резко отличающихся от основной массы получаемого потока данных и вызванных неисправностью датчика или воздействием сильной помехи.

Существующие методы для согласования результатов измерений строятся следующим образом. На основе обоснованного предположения о виде закона распределения случайной погрешности выполняемых прямых измерений вычисляется оценка методом максимального правдоподобия (далее – ММП) как результат максимизации логарифмической функции правдоподобия L , в которую при необходимости в качестве условий включаются уравнения связей между измеряемыми величинами. При построении робастных ММП-оценок, предназначенных для снижения чувствительности получаемых результатов к тем результатам многократных измерений, что оказались значительно удалены от центра выборки, в качестве предполагаемого закона распределения случайной погрешности обычно выбирают распределение с заведомо тяжелыми хвостами либо так модифицируют функцию правдоподобия (L^*), чтобы сделать вклад в ее значение крайних членов вариационного ряда заведомо малым [3-5].

К недостаткам существующих методов согласования результатов совместных измерений, выполняемых в информационно-измерительных и управляющих системах, следует отнести их очевидную негибкость или, если речь идет об упомянутой модификации функции правдоподобия, предполагающей введение одного или двух свободных параметров, ограниченную гибкость [6-8]. Необходимость выбора закона распределения для получения ММП-оценки или вида выполняемой модификации L^* функции правдоподобия неизменно остается наиболее уязвимым местом задачи согласования результатов измерений, поскольку на практике почти всегда сводится к выбору приемлемого, но не оптимального решения из числа методов, зарекомендовавших себя на широком классе практических задач. Не оспаривая целесообразности подобного подхода, необходимо подчеркнуть следующее. При разработке измерительно-измерительных и управляющих систем и испытаниях их опытных образцов в лабораторных условиях, как правило, проводится серия детальных экспериментов, которые позволяют всесторонне изучить измерительную

Исследование выполнено за счет гранта Российского научного фонда № 25-29-20141, <https://rscf.ru/project/25-29-20141/> и средств гранта Санкт-Петербургского научного фонда № 25-29-20141.

схему: установить характер случайных погрешностей, доминирующих в каждом измерительном канале, учесть выполняемые нелинейные преобразования (при их наличии) и, в конечном счете, сформировать достаточный объем априорной информации для построения оптимального метода согласования данных, учитывающего особенности конкретной измерительной ситуации. Как следствие, для достижения универсальности при решении задачи согласования в качестве L^* необходимо использовать одну общую гибкую модель с достаточным числом свободных параметров, значения которых должны оцениваться по результатам выполняемых упомянутых испытаний измерительно-измерительной или управляющей системы.

В настоящей работе рассмотрен подход к построению непараметрических согласованных ММП-оценок, отличающийся от ранее предложенных авторами [9-11] тем, что в L^* традиционные моменты случайной погрешности результатов измерений заменены на их робастные аналоги. Независимость от вида закона распределения достигается за счет использования его разложения в ряд с усечением до первых нескольких членов. Робастность обеспечена тем, что разложение осуществляется по модифицированным полиномам Лагерра, ортогональным на всей числовой оси с весовой функцией $e^{-|x|}$, что соответствует выбору в качестве опорного распределения симметричного экспоненциального распределения (Лапласа) вместо традиционно используемого нормального закона распределения. Как известно, ММП-оценкой для математического ожидания распределения Лапласа является выборочная медиана, обладающая предельной робастностью в сравнении с классическим выборочным средним. Данное свойство и обеспечивает замену в разложении плотности вероятностей случайной погрешности в ряд традиционных моментов на их робастные аналоги.

II. ПОСТАНОВКА ЗАДАЧИ СОГЛАСОВАНИЯ НЕТОЧНЫХ ДАННЫХ

Пусть задан вектор результатов совместных измерений $\mathbf{x} \in \mathbb{R}^N$, $\mathbf{x} = (x_1, x_2, \dots, x_N)^T$ взаимосвязанных величин, зависимости между которыми формализованы системой k уравнений

$$\mathbf{H}(\mathbf{x}, M) = \mathbf{0}, \quad (1)$$

где $\mathbf{H} = (h_1, h_2, \dots, h_k)^T$ – вектор-функция, M – матрица ее параметров, $\mathbf{x} = \mathbf{m} + \boldsymbol{\varepsilon}$, $\mathbf{m} \in \mathbb{R}^N$ – вектор действительных значений искомым величин, $\boldsymbol{\varepsilon} \in \mathbb{R}^N$ – вектор случайных погрешностей результатов измерений, $\mathbf{0}$ – нулевой вектор соответствующей размерности. В случае многократно выполняемых совместных измерений полученные реализации вектора \mathbf{x} образуют матрицу $X = (\mathbf{x}_1, \dots, \mathbf{x}_n)$, где n – количество выполненных повторных измерений. Будем считать, что случайные погрешности результатов измерений либо независимы, либо корреляцией между ними можно обоснованно пренебречь. Соответственно полагаем ковариационную матрицу $\Sigma_{\boldsymbol{\varepsilon}}$ вектора $\boldsymbol{\varepsilon}$ диагональной.

В соответствии с методом максимального правдоподобия согласование полученных результатов

измерений X за счет учета (1) сводится к поиску максимума логарифмической функции правдоподобия:

$$\mathbf{m}^* = \arg \max_{\mathbf{m}: \mathbf{H}(\mathbf{m}, M) = \mathbf{0}} \sum_{j=1}^N \sum_{i=1}^n \ln f_{\varepsilon_j}(x_{ji} - m_j, \boldsymbol{\Theta}_j), \quad (2)$$

где функции $f_{\varepsilon_j}(\varepsilon, \boldsymbol{\Theta}_j)$ – плотности распределения вероятности случайной погрешности $\varepsilon_{ji} = x_{ji} - m_j$ результата измерения величины m_j , $\boldsymbol{\Theta}_j$ – вектор параметров, значения которых оцениваются отдельно по значениям X .

Достижимая точность оценок \mathbf{m}^* зависит в первую очередь от точности определения функций $f_{\varepsilon_j}(\varepsilon, \boldsymbol{\Theta}_j)$. На практике обычно полагают, что имеет место нормальный закон распределения. В работах [9-11] показано, что для достижения непараметричности можно заменить в (2) функции $f_{\varepsilon_j}(\varepsilon, \boldsymbol{\Theta}_j)$ на их аппроксимации $f_{\varepsilon_j}^*(\varepsilon, \boldsymbol{\Psi}_j)$, в качестве которых использованы разложения в ряд, коэффициенты которых $\boldsymbol{\Psi}_j$ зависят исключительно от моментов величин ε_j . Таким образом, получение оценок $\boldsymbol{\Psi}_j^*$ оказывается эквивалентным оценке моментов, для чего, вообще говоря, не требуется знать вида закона распределения f_{ε_j} . Данное обстоятельство упрощает задачу, поскольку освобождает от необходимости построения параметрических методов оценки $\boldsymbol{\Theta}_j$. Выбор аппроксимации $f_{\varepsilon_j}^*$ для f_{ε_j} определяет робастность итоговых результатов.

III. ОТ КЛАССИЧЕСКИХ РАЗЛОЖЕНИЙ РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ К ЕГО РОБАСТНОМУ АНАЛОГУ

Пусть ε – некоторая непрерывная случайная величина, чьей плотностью является $f_{\varepsilon}(\varepsilon)$. Рассмотрим для нее два приближения: через классический ряд Грама-Шарлье по полиномам Эрмита $He_k(y)$ с нормальным распределением в качестве опорного [10], и через разложение в ряд по полиномам Лагерра $L_k(y)$ с упомянутым выше распределением Лапласа в качестве опорного [12], который обеспечивает более робастную аппроксимацию плотности.

1) Аппроксимация $f_{\varepsilon}^*(\varepsilon)$ для плотности вероятностей $f_{\varepsilon}(\varepsilon)$ рядом по полиномам Эрмита со стандартным нормальным распределением $\varphi(t) = \frac{1}{\sqrt{2\pi}} \cdot e^{-t^2/2}$ в качестве опорного имеет вид:

$$f_{\varepsilon}^*(\varepsilon) = \frac{1}{\sigma^*} \cdot \varphi(\varepsilon/\sigma^*) \cdot \sum_{k=0}^{z-1} c_k \cdot He_k(\varepsilon/\sigma^*), \quad (3)$$

где σ^{*2} – оценка дисперсии, осуществляемая по выборке значений ε , z – количество удерживаемых членов ряда. Коэффициенты c_k разложения при полиномах He_k определяются соотношениями

$$c_k = \frac{1}{k!} \cdot \frac{\mu_k^*}{\sigma^{*k}},$$

где μ_k^* – оценка k -го центрального момента случайной величины ε .

В таком случае логарифмическая функция правдоподобия для ε оказывается равна:

$$\ln f_{\varepsilon}^*(\varepsilon) = -\frac{1}{2} \cdot \left(\frac{\varepsilon}{\sigma^*}\right)^2 - \ln(\sigma^* \cdot \sqrt{2\pi}) + \ln \sum_{k=0}^{z-1} c_k \cdot He_k\left(\frac{\varepsilon}{\sigma^*}\right).$$

Если в (2) все плотности имеют вид (3), то (2) сводится к соотношению

$$\mathbf{m}^* = \arg \min_{\mathbf{H}(\mathbf{m}, M)=0} \left(\frac{1}{2} \cdot \sum_{j=1}^N \sum_{i=1}^n t_{ji}^2 - \sum_{j=1}^N \sum_{i=1}^n \ln Q(t_{ji}, z) \right),$$

$$\text{где } t_{ji} = \frac{x_{ji} - m_j}{\sigma_j^*}, Q(t_{ji}, z) = \sum_{k=0}^{z-1} c_k \cdot He_k(t_{ji}).$$

При $z=1$ и $c_0=1$ получаем $Q(t, z) \equiv 1 \forall t$, что приводит к тому, что представленная оценка \mathbf{m}^* сводится к методу наименьших квадратов. Данное обстоятельство является следствием того, что в качестве опорного выбрано нормальное распределение.

2) Аппроксимация $f_{\varepsilon}^{**}(\varepsilon)$ для плотности вероятностей $f_{\varepsilon}(\varepsilon)$ рядом по полиномам Лагерра со стандартным распределением Лапласа $g(t) = \frac{1}{2} \cdot e^{-|t|}$ в качестве опорного имеет вид

$$f_{\varepsilon}^{**}(\varepsilon) = \frac{1}{b^*} \cdot g\left(\frac{\varepsilon}{b^*}\right) \cdot \sum_{k=0}^{z-1} a_k \cdot L_k(|\varepsilon/b^*|), \quad (4)$$

где $b^* = \text{MAD}^*/\ln 2$, MAD^* – оценка медианы абсолютного отклонения значений ε от медианы ε (median of absolute deviation from median, MAD). Коэффициенты a_k разложения при полиномах L_k определяются соотношениями:

$$a_k = \sum_{r=0}^k C_k^r \cdot \frac{(-1)^r}{r!} \cdot \frac{v_r^*}{b^{*r}}, \quad v_r^* = \frac{1}{n} \cdot \sum_{i=1}^n |\varepsilon_i - m|^r, \quad (5)$$

где $\boldsymbol{\varepsilon}^T = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)$ – выборка значений случайной величины ε ; m – математическое ожидание ε ; $C_k^r = \frac{k!}{r!(k-r)!}$ – биномиальный коэффициент.

Логарифмическая функция правдоподобия для ε оказывается равна:

$$\ln f_{\varepsilon}^{**}(\varepsilon) = -\left|\frac{\varepsilon}{b^*}\right| - \ln(2 \cdot b^*) + \ln \sum_{k=0}^{z-1} a_k \cdot L_k\left(\left|\frac{\varepsilon}{b^*}\right|\right).$$

Если в (2) все плотности имеют вид (4), то (2) сводится к соотношению

$$\mathbf{m}^* = \arg \min_{\mathbf{H}(\mathbf{m}, M)=0} \left(\sum_{j=1}^N \sum_{i=1}^n |t_{ji}| - \sum_{j=1}^N \sum_{i=1}^n \ln P(|t_{ji}|, z) \right),$$

$$\text{где } t_{ji} = \frac{x_{ji} - m_j}{b_j^*}, P(t_{ji}, z) = \sum_{k=0}^{z-1} a_k \cdot L_k(t_{ji}).$$

По аналогии с разложением в ряд Грама-Шарлье при $z=1$ и $a_0=1$ имеем $P(t, z) \equiv 1 \forall t$, что приводит к тому, что представленная оценка \mathbf{m}^* сводится к l_1 -минимизации в нормах Минковского, известному как метод наименьших модулей.

Видно, что минимизируемые функционалы в выражениях для \mathbf{m}^* в обоих случаях имеют схожую форму:

$$J(\mathbf{m}) = \sum_{j=1}^N \sum_{i=1}^n \rho_0(t_{ji}) - \sum_{j=1}^N \sum_{i=1}^n \ln S(t_{ji}, z),$$

где ρ_0 – базовая функция невязки, а значения S – поправка, призванная отразить расхождения между данными и реперным законом. Для рассмотренных случаев:

$$\rho_0(t) = \begin{cases} t^2/2, \varepsilon \in N, \\ |t|, \varepsilon \in Lap, \end{cases} \quad S(t, z) = \begin{cases} Q(t, z), \varepsilon \in N, \\ P(|t|, z), \varepsilon \in Lap, \end{cases}$$

где N – нормальный закон распределения, Lap – закон распределения Лапласа.

При $z \rightarrow +\infty$ оба ряда (3) и (4) сходятся к истинной плотности $f_{\varepsilon}(\varepsilon)$ и решение задачи согласования сводится к отысканию минимума нормы $\|\mathbf{x} - \mathbf{m}\|$, не противоречащего условию $\mathbf{H}(\mathbf{m}, M) = \mathbf{0}$.

IV. ММП-ОЦЕНИВАНИЕ НА ОСНОВЕ АППРОКСИМАЦИИ ПЛОТНОСТИ ВЕРОЯТНОСТИ РАЗЛОЖЕНИЕМ В РЯД ПО ПОЛИНОМАМ ЛАГЕРРА

Найдем робастную ММП-оценку, заменив точную плотность $f_{\varepsilon}(\varepsilon)$ ее аппроксимацией $f_{\varepsilon}^{**}(\varepsilon)$ из соотношения (4), ограничившись первыми четырьмя ($z=4$) полиномами Лагерра соответственно. Более высокий порядок разложения в ряд нецелесообразен в связи с ростом статистической погрешности при оценке моментов (классических или робастных) соответствующих старших порядков.

Первые четыре полинома Лагерра соответственно равны

$$\begin{aligned} L_0(t) &= 1, \\ L_1(t) &= 1 - t, \\ L_2(t) &= 1 - 2 \cdot t + t^2/2, \\ L_3(t) &= 1 - 3 \cdot t + 3 \cdot t^2/2 - t^3/6. \end{aligned}$$

Итого получаем, что

$$\frac{2 \cdot b^*}{e^{-|t|}} \cdot f_{\varepsilon}^{**}(t) \approx$$

$$1 + a_1 \cdot L_1(|t|) + a_2 \cdot L_2(|t|) + a_3 \cdot L_3(|t|).$$

Здесь $t = \frac{\varepsilon}{b^*} = \varepsilon \cdot \ln 2 / \text{MAD}^*$. Поскольку ε имеет смысл случайной погрешности, то $\varepsilon = x - m$, где x – случайная величина с тем же симметричным распределением, что и у ε , но с математическим ожиданием, равным m , а не ноль. Поскольку коэффициенты a_k по сути дела представляют собой также оценки, то для простоты будем считать их известными фиксированными константами. Выполним при указанных условиях оценку значения m .

Пусть для x получена выборка значений x_1, \dots, x_n . Тогда логарифмическая функция правдоподобия равна

$$L(m) = \sum_{i=1}^n \ln f_{\varepsilon}(x_i - m) =$$

$$= -n \cdot \ln(2 \cdot b^*) - \frac{1}{b^*} \cdot \sum_{i=1}^n |x_i - m| + \sum_{i=1}^n \ln P^*(|t_i|),$$

где $t_i = (x_i - m)/b^*$.

Получаем, что

$$\begin{aligned} P^*(|t|) &= 1 + a_1 \cdot (1 + |t|) + a_2 \cdot \left(1 - 2 \cdot |t| + \frac{t^2}{2}\right) + \\ &+ a_3 \cdot \left(1 - 3 \cdot |t| + \frac{3 \cdot t^2}{2} - \frac{|t|^3}{6}\right) = \\ &= A - B \cdot |t| + C \cdot t^2 - D \cdot |t|^3, \end{aligned}$$

где $A = 1 + a_1 + a_2 + a_3$; $B = a_1 + 2 \cdot a_2 + 3 \cdot a_3$; $C = a_2/2 + 3 \cdot a_3/2$; $D = a_3/6$.

Продифференцируем $L(m)$ по m , чтобы отыскать максимум правдоподобия. Так как

$$s_i = \frac{\partial}{\partial m} |x_i - m| = -\text{sign}(x_i - m) = \begin{cases} +1, & x_i > m, \\ -1, & x_i < m, \\ 0, & x_i = m \end{cases}$$

и $\frac{\partial t_i}{\partial m} = -\frac{1}{b^*}$, то

$$\frac{dL(m)}{dm} = \frac{1}{b^*} \cdot \sum_{i=1}^n s_i - \frac{1}{b^*} \cdot \sum_{i=1}^n \frac{\partial P^*(|t_i|)/\partial m}{P(|t_i|)},$$

где $\frac{\partial P^*(|t_i|)}{\partial m} = \frac{\partial P^*(|t_i|)}{\partial t_i} \cdot \frac{\partial t_i}{\partial m} = -\frac{1}{b^*} \cdot \frac{\partial P^*(|t_i|)}{\partial t_i}$.

Найдем $\partial P^*(|t_i|)/\partial t_i$, заменив $|t_i|$ на $t_i \cdot s_i$, а $|t_i|^3$ на $t_i^2 \cdot |t_i|$:

$$\frac{\partial P^*(|t_i|)}{\partial t_i} = -B \cdot s_i + 2 \cdot C \cdot t_i - 3 \cdot D \cdot t_i \cdot |t_i|.$$

Итого получаем, что

$$\frac{dL(m)}{dm} = \frac{1}{b^*} \cdot \sum_{i=1}^n s_i + \frac{1}{b^*} \cdot \sum_{i=1}^n \frac{s_i \cdot (B - 2 \cdot C \cdot |t_i| + 3 \cdot D \cdot t_i^2)}{P(|t_i|)}.$$

Приравняем $dL(m)/dm$ к нулю:

$$\frac{1}{b^*} \cdot \sum_{i=1}^n s_i \cdot \left[1 + \frac{B - 2 \cdot C \cdot |t_i| + 3 \cdot D \cdot t_i^2}{A - B \cdot |t_i| + C \cdot t_i^2 - D \cdot |t_i|^3} \right] = 0.$$

Получаем уравнение

$$\sum_{i=1}^n s_i \cdot \left[1 + \frac{B - 2 \cdot C \cdot |t_i| + 3 \cdot D \cdot t_i^2}{A - B \cdot |t_i| + C \cdot t_i^2 - D \cdot |t_i|^3} \right] = 0, \quad (6)$$

решение которого относительно m при условии, что $t_i = (x_i - m)/b^*$, дает ММП-оценку m^* для параметра m для одномерной случайной величины.

Выражение (6) можно трактовать как знаковый критерий: каждый элемент x_i выборки значений случайной величины x вносит в указанную сумму слагаемое вида ± 1 с весом, равным $1 + (\partial P^*(|t_i|)/\partial t_i)/P^*(|t_i|)$, значение которого указывает на отклонение фактической формы плотности распределения x от распределения Лапласа.

Проверим предельный случай. Если действительный закон распределения соответствует распределению Лапласа, то тогда $a_1 = a_2 = a_3 = 0$. Как следствие, $A=1$, $B=C=D=0$. Тогда уравнение (6) сводится к

$$\sum_{i=1}^n s_i = \sum_{i=1}^n \text{sign}(x_i - m) = 0,$$

откуда получаем, что m^* должно совпадать с выборочной медианой, как это и должно быть для распределения Лапласа.

V. МЕТОД РОБАСТНОГО СОГЛАСОВАНИЯ ДАННЫХ

Найдем решение задачи условной оптимизации (2) по методу множителей Лагранжа, заменив действительные плотности распределения случайных погрешностей $f_{\varepsilon_j}(\varepsilon_{ji}, \Theta_j)$ их приближениями $f_{\varepsilon_j}^{**}(\varepsilon_{ji})$, полученными разложением плотности в ряд по полиномам Лагерра. Как и прежде, полагаем, что корреляция между случайными погрешностями результатов измерений

искомых величин отсутствует или ей можно пренебречь. Пусть, как и выше, N – число величин, результаты измерения которых подлежат согласованию, n – число выполненных многократных измерений.

Введем вектор множителей Лагранжа $\lambda \in \mathbb{R}^k$, $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_k)^T$. Логарифмическая функция правдоподобия для лагранжиана задачи (2) имеет вид:

$$\mathcal{L}(\mathbf{m}, \lambda) = \sum_{j=1}^N \sum_{i=1}^n \ln f_{\varepsilon_j}(x_{ji} - m_j) - \lambda^T \cdot \mathbf{H}(\mathbf{m}, M).$$

Запишем условие стационарности:

$$\frac{\partial \mathcal{L}}{\partial m_j} = \frac{\partial}{\partial m_j} \sum_{j=1}^N \sum_{i=1}^n \ln f_{\varepsilon_j}(x_{ji} - m_j) - \sum_{p=1}^k \lambda_p \cdot \frac{\partial h_p(\mathbf{m}, M)}{\partial m_j} = 0.$$

Обозначим для краткости

$$\psi_j(x_{ji} - m_j) = -\frac{\partial}{\partial m_j} \ln f_{\varepsilon_j}(x_{ji} - m_j) = \frac{f'_{\varepsilon_j}(x_{ji} - m_j)}{f_{\varepsilon_j}(x_{ji} - m_j)}.$$

Следовательно, получаем, что для всех $j = 1, 2, \dots, N$

$$\sum_{j=1}^N \sum_{i=1}^n \psi_j(x_{ji} - m_j) + \sum_{p=1}^k \lambda_p \cdot \frac{\partial h_p(\mathbf{m}, M)}{\partial m_j} = 0.$$

Дополнив данные соотношения уравнениями $\mathbf{H}(\mathbf{m}, M) = \mathbf{0}$, получим полную систему уравнений относительно \mathbf{m} и λ , которая и определяет решения задачи согласования неточных данных:

$$\begin{cases} \sum_{j=1}^N \sum_{i=1}^n \psi_j(x_{ji} - m_j) + \sum_{p=1}^k \lambda_p \cdot \frac{\partial h_p(\mathbf{m}, M)}{\partial m_j} = 0, & j = 1, \dots, N, \\ \mathbf{H}(\mathbf{m}, M) = \mathbf{0}. \end{cases}$$

Заменим действительные плотности распределения случайных погрешностей $f_{\varepsilon_j}(\varepsilon_{ji}, \Theta_j)$ их приближениями $f_{\varepsilon_j}^{**}(\varepsilon_{ji})$, полученными разложением плотности в ряд по полиномам Лагерра для $z=4$. Тогда в приведенной выше системе уравнений функции $\psi_j(x_{ji} - m_j)$ заменяются на аппроксимацию:

$$\psi_j^{**}(x_{ji} - m_j) = \frac{\text{sign}(x_{ji} - m_j)}{b_j^*} \cdot \left[1 + \frac{B_j - 2 \cdot C_j \cdot |t_{ji}| + 3 \cdot D_j \cdot t_{ji}^2}{A_j - B_j \cdot |t_{ji}| + C_j \cdot t_{ji}^2 - D_j \cdot |t_{ji}|^3} \right],$$

где $t_{ji} = (x_{ji} - m_j)/b_j^*$, оценки параметров масштаба $b_j^* = \text{MAD}_j^*/\ln 2$ определяются по выборке (x_{j1}, \dots, x_{jn}) ; коэффициенты A_j, B_j, C_j и D_j предпочтительно также вычисляются на массиве экспериментальных данных, полученным, как отмечалось выше, по результатам предварительных испытаний тестового образца информационно-измерительной или управляющей системы – при решении системы, таким образом, участвуют их оценки A_j^*, B_j^*, C_j^* и D_j^* .

Решаемая система уравнений в таком случае приобретает следующий вид:

$$\begin{cases} \sum_{j=1}^N \sum_{i=1}^n \frac{\text{sign}(x_{ji} - m_j)}{b_j^*} \cdot \left[1 + \frac{B_j^* - 2 \cdot C_j^* \cdot |t_{ji}| + 3 \cdot D_j^* \cdot t_{ji}^2}{A_j^* - B_j^* \cdot |t_{ji}| + C_j^* \cdot t_{ji}^2 - D_j^* \cdot |t_{ji}|^3} \right] + \\ \sum_{p=1}^k \lambda_p \cdot \frac{\partial h_p(\mathbf{m}, M)}{\partial m_j} = 0, & j = 1, \dots, N, \\ \mathbf{H}(\mathbf{m}, M) = \mathbf{0}. \end{cases} \quad (7)$$

Решение системы \mathbf{m}^* является вектором согласованных оценок совместно измеряемых взаимосвязанных через $\mathbf{H}(\mathbf{x}, \mathbf{A}) = \mathbf{0}$ величин \mathbf{x} .

Поскольку все коэффициенты A_j , B_j , C_j и D_j связаны с величиной m_j через соотношения (5), то на самом деле имеем зависимости $A_j(m_j)$, $B_j(m_j)$, $C_j(m_j)$ и $D_j(m_j)$, которые могут быть учтены в составе решаемой системы уравнений:

$$\left\{ \begin{array}{l} \sum_{p=1}^k \lambda_p \cdot \frac{\partial h_p(\mathbf{m}, M)}{\partial m_j} + \sum_{j=1}^N \sum_{i=1}^n \frac{\text{sign}(x_{ji} - m_j)}{b_j^*} \cdot \\ \left[1 + \frac{B_j(m_j) - 2 \cdot C_j(m_j) \cdot |t_{ji}| + 3 \cdot D_j(m_j) \cdot t_{ji}^2}{A_j(m_j) - B_j(m_j) \cdot |t_{ji}| + C_j(m_j) \cdot t_{ji}^2 - D_j(m_j) \cdot |t_{ji}^3|} \right] = 0, \\ \mathbf{H}(\mathbf{m}, M) = \mathbf{0}, \end{array} \right. \quad (8)$$

где $A_j(m_j) = 1 + a_{1j}(m_j) + a_{2j}(m_j) + a_{3j}(m_j);$
 $B_j(m_j) = a_{1j}(m_j) + 2 \cdot a_{2j}(m_j) + 3 \cdot a_{3j}(m_j);$
 $C_j(m_j) = a_{2j}(m_j)/2 + 3 \cdot a_{3j}(m_j)/2;$
 $D_j(m_j) = a_{3j}(m_j)/6,$

$$a_{kj}(m_j) = \sum_{r=0}^k C_k^r \cdot \frac{(-1)^r}{r!} \cdot \frac{v_r^*(m_j)}{b_j^{*r}},$$

$$v_r^*(m_j) = \frac{1}{n} \cdot \sum_{u=1}^n |x_{ju} - m_j|^r.$$

Разработанный метод согласования неточных данных имеет ряд преимуществ как по сравнению с классическими алгоритмами согласования, так и по отношению к непараметрическим подходам, основанным на приближениях плотности распределения вероятностей случайных погрешностей. Примененная в настоящей работе схема аппроксимации на основе полиномов Лагерра позволяет учесть факторы, влияющие на распределение погрешностей каждого измерительного канала многоканальной системы, причем глубина учета регулируется выбором значения z количества членов ряда. Естественным ограничением для z выступает объем имеющейся в наличии выборки, по которой оцениваются необходимые коэффициенты разложения: устойчивое оценивание большего числа свободных параметров (как и моментов старших степеней) требует кратного увеличения размера выборки. Предварительные расчеты показывают более высокую эффективность представленного подхода в сравнении с другими схемами, основанными на иных разложениях плотностей распределения вероятностей случайных погрешностей в ряды – в частности в ряд Грама-Шарлье.

Повышение робастности результатов согласования по отношению к статистическим выбросам достигается выбором в качестве опорного закона распределения при

разложении в ряд не традиционно принимаемого нормального распределения, а распределения Лапласа, обладающего тяжелыми хвостами, что в свою очередь приводит при применении ММП к эффективности более робастных оценок для искомого параметров распределения. Представленные соотношения (7) и (8) позволяют выполнять отбраковку недостоверных результатов измерений прямо в рамках процедуры их согласования с информацией о взаимосвязях между измеряемыми величинами.

ЗАКЛЮЧЕНИЕ

В настоящей работе предложен новый метод робастного согласования результатов совместных измерений, выполняемых в информационно-измерительных и управляющих системах, в основе которого лежит аппроксимация априорно неизвестной плотности распределения случайных погрешностей усеченным рядом на основе полиномов Лагерра. Показано, что данная задача сводится к решению систем уравнений (7) и (8).

СПИСОК ЛИТЕРАТУРЫ

- [1] A. Bakhtouchi, "Data reconciliation and fusion methods: A survey," Applied Computing and Informatics, vol. 18(3/4), p. 182-194, 2022.
- [2] S. Narasimhan, C. Jordache, "Data Reconciliation and Gross Error Detection: An Intelligent Use of Process Data," Gulf Professional Publishing, Houston, TX, USA, 1999.
- [3] C.E. Llanos, M.C. Sanchez, R.A. Maronna, "Robust estimators for data reconciliation," Industrial & Engineering Chemistry Research, vol. 54(18), p. 5096-5105, 2015.
- [4] D. Wang, J.A. Romagnoli, "A framework for robust data reconciliation based on a generalized objective function," Industrial & engineering chemistry research, vol. 42(13), p. 3075-3084, 2003.
- [5] S. Jin, X. Li, Z. Huang, M. Liu, "A new target function for robust data reconciliation," Industrial & engineering chemistry research, vol. 51(30), p. 10220-10224, 2012.
- [6] S. Xie, C. Yang, X. Yuan, X. Wang, Y. Xie, "A novel robust data reconciliation method for industrial processes," Control Engineering Practice, vol. 83, p. 203-212, 2019.
- [7] A.S. Da Cunha, F.C. Peixoto, D.M. Prata, "Robust data reconciliation in chemical reactors," Computers & Chemical Engineering, vol. 145, paper 107170, 2021.
- [8] M. Alhaj-Dibo, D. Maquin, J. Ragot, "Data reconciliation: A robust approach using a contaminated distribution," Control Engineering Practice, vol. 16(2), p. 159-170, 2008.
- [9] V.A. Garanin, K.K. Semenov, "Increasing measurement accuracy by nonparametric data reconciliation," Measurement, vol. 238, paper 115235, 2024.
- [10] V.A. Garanin, K.K. Semenov, "Semi-nonparametric approach for measured data reconciliation based on the Gram-Charlier series expansion," Measurement: Sensors, vol. 18, paper 100351, 2021.
- [11] Гаранин В.А., Семенов К.К. Непараметрические методы решения задачи согласования данных // Записки научных семинаров Санкт-Петербургского отделения математического института им. В.А. Стеклова РАН. 2024. Т. 540. С. 351-405.
- [12] H. Mustapha, R. Dimitrakopoulos, "Generalized Laguerre expansions of multivariate probability densities with moments," Computers & mathematics with applications, vol. 60(7), p. 2178-2189, 2010.