

СИЛЬНАЯ СОГЛАСОВАННОСТЬ В ЗАДАЧАХ ВОССТАНОВЛЕНИЯ ЗАВИСИМОСТЕЙ ПО ИНТЕРВАЛЬНЫМ ДАННЫМ

С.П. Шарый

Институт вычислительных технологий СО РАН, г. Новосибирск, Российская Федерация

E-mail: shary@ict.nsc.ru

Для задачи восстановления зависимостей по данным с интервальной неопределённостью вводится понятие сильной согласованности данных и параметров. Дается его содержательная интерпретация. Показывается, что получающаяся усиленная формулировка задачи сводится к исследованию непустоты и дальнейшему оцениванию так называемого допускового множества решений для интервальной системы уравнений, построенной по обрабатываемым данным.

Ключевые слова: задача восстановления зависимостей; согласование параметров и данных; сильное согласование; интервальная система уравнений; допусковое множество решений.

Задача восстановления зависимостей – это популярная и практически важная задача, в которой по эмпирическим данным требуется построить зависимость заданного вида между «входными» и «выходными» величинами. Далее в работе мы рассматриваем простейшую зависимость вида

$$b = a_1x_1 + a_2x_2 + \dots + a_nx_n, \quad (1)$$

в которой значения b являются линейной функцией от независимых переменных a_1, a_2, \dots, a_n . Необходимо определить неизвестные коэффициенты x_i , чтобы получившаяся зависимость «наилучшим образом» соответствовала заданному набору значений a_i и b , полученному в результате m измерений (наблюдений)

$$\begin{aligned} & a_1^{(1)}, a_2^{(1)}, \dots, a_n^{(1)}, b^{(1)}, \\ & a_1^{(2)}, a_2^{(2)}, \dots, a_n^{(2)}, b^{(2)}, \\ & \dots \\ & a_1^{(m)}, a_2^{(m)}, \dots, a_n^{(m)}, b^{(m)} \end{aligned} \quad (2)$$

(верхние индексы в скобках означают номер измерения). Нередко эту постановку называют также задачей оценивания параметров объекта или задачей идентификации.

Подставляя данные (2) в равенство (1), после переобозначения $a_{ij} := a_j^{(i)}$ и $b_i := b^{(i)}$ получаем систему уравнений относительно x_1, x_2, \dots, x_n :

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1, \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= b_2, \\ &\dots \quad \dots \quad \dots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n &= b_m, \end{aligned}$$

или, кратко,

$$Ax = b$$

с $m \times n$ -матрицей $A = (a_{ij})$ и m -вектором $b = (b_i)$. Её решение, обычное или в обобщённом смысле, принимается за оценку параметров x_1, x_2, \dots, x_n . Наглядная графическая иллюстрация задачи восстановления зависимости показана на рис. 1: требуется найти прямую, проходящую через начало координат, которая «наилучшим образом» приближает множество точек с координатами (2).

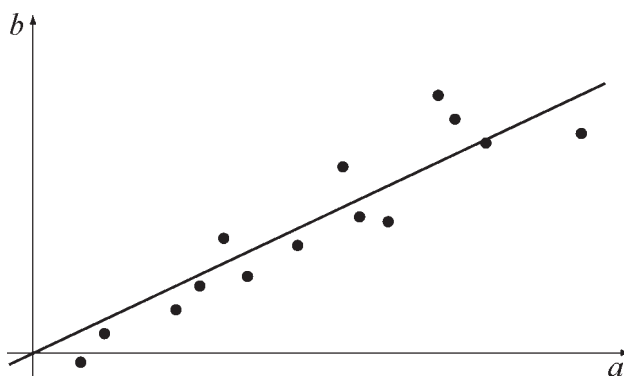


Рис. 1. Иллюстрация задачи восстановления линейной зависимости: точки – это данные измерений, по которым необходимо построить функцию заданного вида, «наиболее точно» приближающую их

Но в практических задачах восстановления зависимостей данные почти всегда неточны, поскольку на результаты измерений влияют внешние неконтролируемые факторы, сами измерительные приборы не являются абсолютно точными т. д. и т. п. Таким образом, реально мы должны иметь дело с той или иной *неопределённостью* – состоянием частичного знания об измеряемой величине, когда нам известно какое-то её значение, но оно приближённое, и имеется также некоторая информация (качественная и количественная) о погрешности этого значения.

Как описывать эти погрешности? Иными словами, какую «модель неопределённости» данных мы принимаем? Традиционный выбор – это теоретико-вероятностная модель ошибок, основы которой заложили на рубеже XVIII и XIX веков К.Ф. Гаусс и П.С. Лаплас. Согласно этому подходу ошибки измерений и наблюдений являются случайными величинами, адекватно описываемыми математическим аппаратом теории вероятностей, и нам (более-менее) известны характеристики этих случайных величин. Теоретико-вероятностная модель ошибок за прошедшие два века получила очень большое развитие и популярность, сделавшись основным инструментом обработки данных. Тем не менее, её приложение вызывает необходимость ответа на многие нетривиальные вопросы, и эти ответы подчас не вполне удовлетворительны.

Ниже мы конспективно перечислим некоторые из проблем, возникающих при применении теоретико-вероятностных методов в статистике. Наш короткий обзор естественно дополняет работы [1, 2], где есть подробное обсуждение проблем и трудностей теоретико-вероятностной статистики.

Статистическая устойчивость. Прежде всего, мы должны принимать во внимание тот факт, что в основе самого понятия вероятности лежит так называемая частотная интерпретация, при которой вероятность того или иного события понимается как отношение числа благоприятных исходов к общему числу исходов рассматриваемого явления (эксперимента и т.п.), либо близкая к ней конструкция. Несмотря на то, что в современной теории вероятностей, построенной на аксиоматике А.Н. Колмогорова, математическая вероятность определяется как некоторая специальная мера на множестве событий, она формализует именно частотное понимание вероятности. Наконец, именно частотная интерпретация вероятности является основой всех приложений теории вероятностей к практике (см., к примеру, [3]). Существование подобной частоты, как объективной характеристики реальных явлений и процессов, является фундаментом самого существования теории вероятностей и залогом её успешного применения к моделированию окружающего нас мира.

Но важно осознавать, что эта модель не универсальна, она является определённой идеализацией, имеющей свою сферу применимости, весьма широкую, но всё-таки ограниченную. Многие явления окружающего нас мира, в отношении которых вполне применим общепринятый термин «случайные», не обладают свойством существования устойчивой частоты, так как при росте числа наблюдений эта относительная частота не устанавливается, а имеет тенденцию к постоянным колебаниям [4]. Для описания и анализа подобных явлений традиционная теория вероятностей непригодна.

Свойство существования относительной частоты событий называется *статистической устойчивостью* (статистической однородностью), и часто теорию вероятностей определяют как «математическую теорию статистически устойчивых явлений» (так делается, к примеру, в классической книге [5]).

Так или иначе, если нет статистической устойчивости, теоретико-вероятностные конструкции напрямую применять к решению задачи нельзя. В этом случае и традиционная математическая статистика, основанная на теории вероятностей, также не может служить подходящим инструментом для обработки данных.

Проблема малых выборок. Теоретико-вероятностные закономерности проявляются как тенденции, которые наиболее ярко видны для массовых явлений. При малом или небольшом количестве испытаний выводы теории вероятностей могут оказаться весьма далёкими от истинной картины явления. «Проблема малых выборок» — это вопрос о том, достаточен ли объём выборки (количество измерений и т. п.) для того, чтобы выводы, получаемые на основе теоретико-вероятностной модели ошибок, имели приемлемую практическую достоверность. Связанный с этим вопрос: какие методы следует применять для обработки выборок, являющихся «малыми», где теория вероятностей не способна адекватно описать поведение погрешностей?

Неизвестные вероятностные характеристики распределения. Каков конкретный вид распределения погрешностей? Каковы его числовые характеристики? Имеют ли данные корреляцию между собой? Или же они независимы? Многие классические результаты теоретико-вероятностной статистики требуют, как известно, независимости рассматриваемых случайных величин либо заданного уровня корреляции. Проверка этих условий на практике представляется почти невозможной.

«Робастность» модели обработки данных. Под этим требованием понимается адекватная устойчивость оценок, получаемых на основе тех или иных моделей, к малым возмущениям в данных, т. е. к вероятностным характеристикам распределений и их форме. Некоторые вероятностно-статистические методы не обладают этим свойством, давая ответы, чувствительность которых к возмущениям в данных неразумно велика.

Удобство вычислительных методов. Насколько удобны и практичны вычислительные технологии для решения соответствующих задач статистики? Некоторые традиционные методы теоретико-вероятностной статистики удовлетворяют этому условию. Например, широчайшее распространение метода наименьших квадратов в задачах обработки данных обусловлено, помимо ясного теоретико-вероятностного смысла, также его удобной вычислительной схемой: в линейном случае решение задачи наименьших квадратов сводится к решению системы линейных алгебраических уравнений. Но в более сложных ситуациях методы теоретико-вероятностной статистики технологической простотой уже не обладают (например, тот же самый метод наименьших квадратов в нелинейном случае). Это повышает «конкурентоспособность» альтернативных подходов к обработке данных.

В связи с поднятыми вопросами следует вспомнить многолетнюю дискуссию Ю.И. Алимова и В.Н. Тутубалина в 70–90-е годы XX века [6–9] по вопросам применимости и адекватности теоретико-вероятностных методов в статистике.

В целом, при неудовлетворённости теоретико-вероятностным описанием погрешностей часто удобнее работать с неопределённостями и неточностями в данных с помощью методов интервального анализа. При этом вместо вероятностных распределений заданными считаются интервальные оценки результатов измерений величин, т. е. их принадлежности некоторым интервалам. В частности, в рассматриваемой нами задаче оценивания параметров линейной зависимости мы считаем, что

$$a_{ij} \in \mathbf{a}_{ij} = [\inf a_{ij}, \sup a_{ij}] \quad \text{и} \quad b_i \in \mathbf{b}_i = [\inf b_i, \sup b_i]$$

(через « \inf » и « \sup » здесь и далее обозначены нижний и верхний концы интервала).

Пионером нового подхода к обработке данных выступил Л.В. Канторович [10]. В дальнейшем в развитие теории восстановления зависимостей с интервальными неопределённостями в данных у нас в стране значительный вклад внесли М.Л. Лидов [11], С.И. Спивак [12, 13], А.П. Вошинин [1, 14], Н.М. Оскорбин [15], С.И. Жилин [16–18], Б.Т. Поляк [19] и другие исследователи. За рубежом первой публикацией по теме стала работа Ф. Швеппе [20]. Далее значительные результаты в новом направлении были получены в работах Дж.П. Нортонна,

М. Миланезе, Дж. Бельфорте, Л. Пронцато, Э. Вальтера и других (см. монографию [21] и коллективный обзор [22]). Этому же вопросу посвящены работы автора настоящей статьи [23–26], развивающие так называемый метод максимума согласования для восстановления линейных зависимостей по интервальным данным.

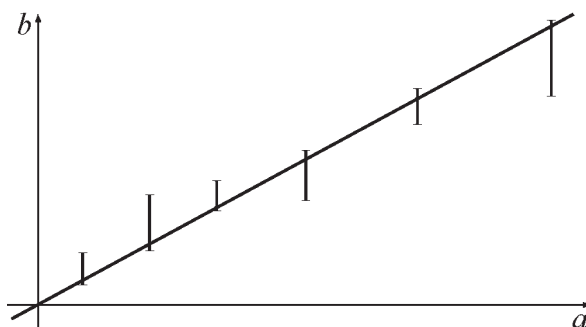


Рис. 2. Иллюстрация согласования параметров линейной модели и интервальных данных измерений, когда входные значения являются известными точно

Итак, нам необходимо найти или как-то оценить коэффициенты x_j , $j=1,2,\dots,m$, для которых линейная функция (1) «наилучшим образом» приближала бы интервальные данные. При этом идеальным является, конечно, случай, когда график восстанавливаемой зависимости проходит через все точки наблюдений, т. е. когда приближение данных в самом деле полное и имеет почти тот же смысл, что и в задаче интерполирования.

Отметим, что в постановке Л.В. Канторовича [10] и его последователей задача восстановления зависимостей по неточным данным имела не самый общий случай: неопределённости во входных данных предполагались отсутствующими, т. е. $a_{ij} = a_{ij}$. Тогда (см. рис. 2)

$$\inf b_i \leq \sum_{j=1}^n a_{ij} x_j \leq \sup b_i,$$

$i=1,2,\dots,m$. Согласование параметров и данных следует понимать как прохождение регрессионной прямой через все коридоры неопределённости выходных данных. Этот случай, тем не менее, практически очень важен и именно его тщательное решение способствовало широкому распространению новых подходов на практике. При этом с математической точки зрения получаем систему линейных неравенств, которую можно решать, к примеру, методами линейного программирования.

В общем случае, когда интервальную неопределённость имеют как входные, так и выходные данные, естественным представляется следующее

Определение 1. Будем говорить, что набор параметров x_1, x_2, \dots, x_n линейной зависимости (1) согласуется с интервальными экспериментальными данными $a_{i1}, a_{i2}, \dots, a_{in}, b_i$, $i=1, 2, \dots, m$, если для каждого наблюдения i в пределах измеренных интервалов найдутся такие представители $a_{i1} \in a_{i1}, a_{i2} \in a_{i2}, \dots, a_{in} \in a_{in}$ и $b_i \in b_i$, что имеет место равенство

$$b_i = a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n.$$

В соответствии с этим определением данные каждого замера входов и выхода представляют собой в пространстве \mathbb{R}^{n+1} как бы большие точки, «раздувшиеся» до брусков (прямоугольных параллелепипедов с гранями, параллельными координатным осям), а прохождение графика конструируемой зависимости через такую точку понимается как её пересечение с этим бруском (см. рис. 3).

Если из интервальных данных задачи организовать $m \times n$ -матрицу $A = (a_{ij})$ и m -вектор $b = (b_i)$, то множество параметров, согласующихся с данными в смысле первого определения – это множество, определяемое как

$$\{x \in \mathbb{R}^n \mid \text{существуют такие } A \in A \text{ и } b \in b, \text{ что } Ax = b\}.$$

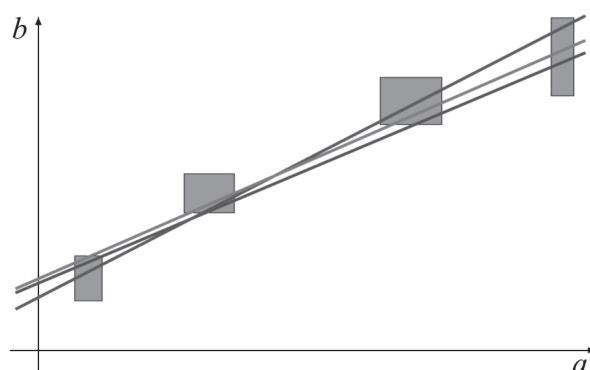


Рис. 3. Иллюстрация согласования параметров линейной модели и интервальных данных измерений

С помощью формального языка логики предикатов его можно переписать также в виде

$$\mathcal{E}_{uni}(A, \mathbf{b}) = \left\{ x \in \mathbb{R}^n \mid (\exists A \in \mathcal{A})(\exists b \in \mathbf{b})(Ax = b) \right\},$$

и в интервальном анализе это множество называется *объединённым множеством решений* интервальной линейной системы уравнений $Ax = \mathbf{b}$ (нижний индекс «uni» в обозначении множества – от английского термина *united solution set*).

Но раздувшаяся точка-брус приобретает уже дополнительную структуру, которой не было у исходных бесконечно малых точек. Как следствие, в условиях неточности данных, когда каждое измерение-наблюдение вместо точки представляет собой целое множество возможных значений рассматриваемой величины, само понятие «прохождения через точки наблюдений» должно быть переосмыслено. Дело в том, что теперь наличие у множеств неопределённости наблюдений «тонкой структуры» вызывает необходимость различать те или иные случаи прохождения графика конструируемой функции через эти множества.

Прежде всего, нужно различать входные переменные и выходные. Входы и выходы системы (соответствующие независимым переменным функции и её значениям) отличаются друг от друга по целевому назначению, а их измерения могут выполняться отличным друг от друга способом или даже в разное время. Как следствие, различные грани бруса неопределённости замера имеют разный смысл (на рис. 3 это вертикальные и горизонтальные стороны прямоугольников), а задача восстановления зависимостей по неточным данным приобретает иной контекст. Становится важным, как именно график восстанавливаемой зависимости проходит через брус неопределённости. Если процесс измерения значений входа и выхода разорван во времени и разделён на этапы, когда выходы измеряются *после* фиксации значений входов, то более адекватно другое понимание «согласования» параметров и данных, при котором ограничение на выходе должно выполняться *равномерно* при любых значениях входов. Иными словами, действительное значение b_i на выходе принадлежит \mathbf{b}_i вне зависимости от того, каковы входные значения $a_{i1}, a_{i2}, \dots, a_{in}$ в пределах соответствующих интервалов $\mathbf{a}_{i1}, \mathbf{a}_{i2}, \dots, \mathbf{a}_{in}$.

Формально эта ситуация описывается другим определением:

Определение 2. Будем говорить, что набор параметров x_1, x_2, \dots, x_n линейной зависимости (1) *сильно согласуется* с интервальными экспериментальными данными $\mathbf{a}_{i1}, \mathbf{a}_{i2}, \dots, \mathbf{a}_{in}, \mathbf{b}_i, i = 1, 2, \dots, m$, если для каждого наблюдения i для любых значений $a_{i1} \in \mathbf{a}_{i1}, a_{i2} \in \mathbf{a}_{i2}, \dots, a_{in} \in \mathbf{a}_{in}$ в пределах измеренных интервалов \mathbf{b}_i на выходе найдётся такое $b_i \in \mathbf{b}_i$, что выполняется равенство

$$b_i = a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n.$$

Множество параметров, сильно согласующихся с данными в смысле второго определения, можно описать как

$$\mathcal{E}_{tol}(A, \mathbf{b}) = \left\{ x \in \mathbb{R}^n \mid \text{для любой } A \in \mathcal{A} \text{ найдётся такое } b \in \mathbf{b}, \text{ что } Ax = b \right\}.$$

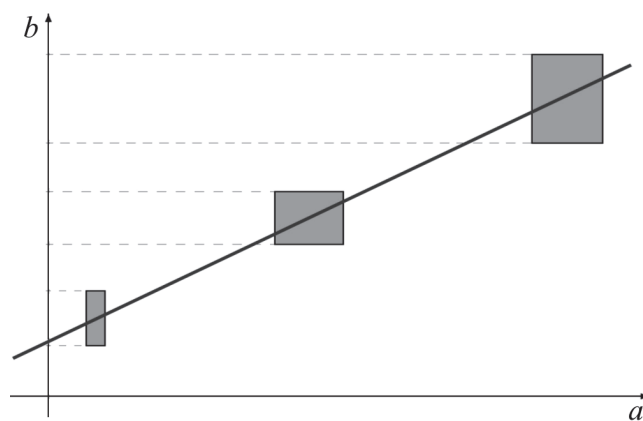


Рис. 4. Иллюстрация сильного согласования параметров линейной модели и интервальных данных измерений

На формальном языке оно определяется как

$$\mathcal{E}_{tol}(A, \mathbf{b}) = \left\{ x \in \mathbb{R}^n \mid (\forall A \in \mathbf{A})(\exists \mathbf{b} \in \mathbf{b})(Ax = \mathbf{b}) \right\},$$

и в интервальном анализе это множество называется *допусковым множеством решений* интервальной линейной системы уравнений $Ax = \mathbf{b}$ [27–29]. Соответствующий английский термин *tolerable solution set* (отсюда и индекс «tol» в обозначении множества), а точки из этого множества часто называют «tolerance solutions» [28]. Нетрудно понять, что

$$\mathcal{E}_{tol}(A, \mathbf{b}) \subseteq \mathcal{E}_{uni}(A, \mathbf{b}),$$

т.е. допусковое множество решений интервальной системы уравнений всегда является подмножеством объединённого.

Как и в традиционном неинтервальном случае, иногда не существует набора параметров, согласующихся с данными, т. е. линии, проходящей через все брусы неопределённости замеров в нужном нам смысле, сильном или обычном. В этом случае оценкой параметров конструируемой зависимости следует взять точку, которая обеспечивает «наименьшее несогласование» параметров и данных, аналогично тому, как это сделано в [23–26].

Итак, множество параметров модели, удовлетворяющих условию сильного согласования, является *допусковым множеством решений* для интервальной системы уравнений, построенной по данным наблюдений. Допусковое множество решений для интервальных систем линейных алгебраических уравнений сравнительно хорошо изучено [27–31]. Известно, что оно всегда является выпуклым полиэдральным множеством. Существуют практические методы для распознавания пустоты или непустоты допускового множества решений, а также для его внутреннего и внешнего оценивания. В частности, можно порекомендовать читателю свободно распространяемую программу [32].

Отметим, что ранее возможность использования допускового множества решений в задаче восстановления зависимостей отмечалась в [33]. Но никаких соображений в пользу такого выбора в этой работе не представлено. Между тем в дополнение к уже сказанному о содержательных причинах применения сильного согласования и допускового множества решений можно добавить следующее. Уникальными особенностями допускового множества решений для интервальных линейных систем уравнений, коренным образом отличающими его от других множеств решений и оправдывающими введение на его основе отдельного понятия «сильной согласованности», является то, что это

- наиболее «устойчивое» из множеств решений,
- множество решений, имеющее полиномиальную сложность распознавания.

Таким образом, основанные на использовании допускового множества решений подходы к оцениванию параметров и восстановлению зависимостей будут обладать вычислительной эффективностью и, как следствие, смогут завоевать популярность у специалистов, решающих прикладные задачи. Для сравнения напомним, что распознавание и оценивание объединённого множества решений интервальных линейных систем уравнений является NP-трудной задачей. По этой

причине сильная согласованность данных и параметров приводит (по крайней мере, в линейном случае) к более практичной вычислительной технологии решения задачи восстановления зависимостей, чем обычная согласованность.

Литература

1. Вошинин, А.П. Интервальный анализ данных: развитие и перспективы / А.П. Вошинин // Заводская лаборатория. Диагностика материалов. – 2002. – Т. 68, №1. – С. 118–126.
2. Вошинин, А.П. Задачи анализа с неопределёнными данными – интервальность и/или случайность? / А.П. Вошинин // Труды Международной конференции по вычислительной математике. Рабочие совещания. Совещание «Интервальная математика и методы распространения ограничений» ИМРО-2004. – Издательство ИВМиМГ СО РАН: Новосибирск, 2004. – С. 147–158. http://www.nsc.ru/interval/Conferences/IMRO_04/Voschinin.pdf
3. Тутубалин, В.Н. Теория вероятностей: Краткий курс и научно-методические замечания / В.Н. Тутубалин. – Москва: Изд-во МГУ, 1972. – 230 с.
4. Горбань, И.И. Феномен статистической устойчивости / И.И. Горбань. – Киев: Наукова думка, 2014. – 444 с.
5. Крамер, Г. Математические методы статистики / Г. Крамер. – Москва: Мир, 1975. – 648 с.
6. Алимов, Ю.И. Альтернатива методу математической статистики / Ю.И. Алимов. – Москва: Знание, 1980. – 64 с.
7. Алимов, Ю.И. Является ли вероятность «нормальной» физической величиной? / Ю.И. Алимов, Ю.А. Кравцов // Успехи физических наук. – 1992. – Т. 162, №7. – С. 149–182.
8. Тутубалин, В.Н. Границы применимости (вероятностно-статистические методы и их возможности) / В.Н. Тутубалин. – Москва: Знание, 1977. – 64 с.
9. Тутубалин, В.Н. Вероятность, компьютеры и обработка результатов эксперимента / В.Н. Тутубалин // Успехи физических наук. – 1993. – Т. 163, №7. – С. 93–109.
10. Канторович, Л.В. О некоторых новых подходах к вычислительным методам и обработке наблюдений / Л.В. Канторович // Сибирский матем. журнал. – 1962. – Т. 3, №5. – С. 701–709.
11. Лидов, М.Л. Минимаксные методы оценивания / М.Л. Лидов // Препринты ИПМ им. М. В. Келдыша. – 2010. – № 071. – 87 с.
12. Спивак, С.И. Применение метода выравнивания по П.Л. Чебышёву при построении кинетической модели сложной химической реакции / С.И. Спивак, В.И. Тимошенко, М.Г. Слинько // Доклады Академии Наук. – 1970. – Т. 192, № 3. – С. 580–582.
13. Оценка погрешности и значимости измерений для линейных моделей / С.И. Спивак, О.Г. Кантор, Д.С. Юнусова и др. // Информатика и её применения. – 2015. – Т. 9, вып. 1. – С. 87–97.
14. Вошинин, А.П. Метод анализа данных при интервальной нестатистической ошибке / А.П. Вошинин, А.Ф. Бочков, Г.Р. Сотиров // Заводская лаборатория. Диагностика материалов. – 1990. – Т. 56, №7. – С. 76–81.
15. Оскорбин, Н.М. Построение и анализ эмпирических зависимостей методом центра неопределённости / Н.М. Оскорбин, А.В. Максимов, С.И. Жилин // Известия Алтайского государственного университета. – 1998. – № 1. – С. 37–40.
16. Жилин, С.И. Нестатистические модели и методы построения и анализа зависимостей: дис. ... канд. физ.-мат. наук / С.И. Жилин. – Барнаул: АлтГУ, 2004. – 119 с.
17. Zhilin, S.I. On fitting empirical data under interval error / S.I. Zhilin // Reliable Computing. – 2005. – Vol. 11, no. 5. – P. 433–442.
18. Zhilin, S.I. Simple method for outlier detection in fitting experimental data under interval error / S.I. Zhilin // Chemometrics and Intellectual Laboratory Systems. – 2007. – Vol. 88, no. 1. – P. 60–68.
19. Поляк, Б.Т. Оценивание параметров в линейных многомерных системах с интервальной неопределённостью / Б.Т. Поляк, С.А. Назин // Проблемы управления и информатики. – 2006. – № 1-2. – С. 103–115.
20. Schweppe, F.C. Recursive state estimation: unknown but bounded errors and system inputs / F.C. Schweppe // IEEE Trans. Autom. Control, AC-13. – 1968. – no. 1. – P. 22–28.
21. Прикладной интервальный анализ / Л. Жолен, М. Кифер, О. Дидри, Э. Вальтер. – Москва-Ижевск: Издательство «РХД», 2007. – 468 с.

22. Bounding approaches to system identification / Milanese, M., Norton, J., Piet-Lahanier, H., Walter, E., eds. – New York: Plenum Press. 1996. – 567 p.
23. Шарый, С.П. Разрешимость интервальных линейных уравнений и анализ данных с неопределённостями / С.П. Шарый // Автоматика и телемеханика. – 2012. – № 2. – С. 111–125.
24. Шарый, С.П. Распознавание разрешимости интервальных уравнений и его приложения к анализу данных / С.П. Шарый, И.А. Шарая // Вычисл. технологии. – 2013. – Т. 18, №3. – С. 80–109.
25. Shary, S.P. Maximum consistency method for data fitting under interval uncertainty / S.P. Shary // Journal of Global Optimization. – 2016. – Vol. 66, Issue 1. – P. 111–126.
26. Kreinovich, V. Interval methods for data fitting under uncertainty: a probabilistic treatment / V. Kreinovich, S.P. Shary // Reliable Computing. – 2016. – Vol. 23. – P. 105–140.
27. Shary, S.P. Solving the linear interval tolerance problem / S.P. Shary // Mathematics and Computers in Simulation. – 1995. – Vol. 39. – P. 53–85.
28. Шарый, С.П. Решение интервальной линейной задачи о допусках / С.П. Шарый // Автоматика и телемеханика. – 2004. – № 10. – С. 147–162.
29. Шарый, С.П. Конечномерный интервальный анализ / С.П. Шарый. – Новосибирск: XYZ, 2016. – 606 с.
30. Rohn, J. A handbook of results on interval linear problems / J. Rohn. – Prague: Czech Academy of Sciences, 2005. – 80 p. <http://www.nsc.ru/interval/Library/Surveys/ILinProblems.pdf>
31. Шарая, И.А. Строение допустимого множества решений интервальной линейной системы / И.А. Шарая // Вычисл. технологии. – 2005. – Т. 10, № 5. – С. 103–119. <http://www.nsc.ru/interval/sharaya/Papers/ct05.pdf>
32. <http://www.nsc.ru/interval/Programing/MCodes/tolsolvty.m>
33. Gutowski, M.W. Interval experimental data fitting / M.W. Gutowski // Focus on Numerical Analysis: сб. науч. тр. – New York: Nova Science Publishers, 2006. – P. 27–70.

Поступила в редакцию 14 октября 2016 г.

*Bulletin of the South Ural State University
Series "Mathematics. Mechanics. Physics"
2017, vol. 9, no. 1, pp. 39–48*

DOI: 10.14529/mmph170105

STRONG COMPATABILITY IN DATA FITTING PROBLEMS WITH INTERVAL DATA

S.P. Shary

Institute of Computational Technologies of SB RAS, Novosibirsk, Russian Federation

E-mail: shary@ict.nsc.ru

The data fitting problem is a popular and practically important problem in which a functional dependency between “input” and “output” variables is to be constructed from the given empirical data. Real-life data are almost always inaccurate, and we have to deal with the measurement uncertainty. Traditionally, when processing the measurement results, models of probability theory are used, which are not always adequate to the situations under study. An alternative way to describe data inaccuracy is to use methods of interval analysis, based on specifying interval bounds of the measurement results.

Data fitting problems under interval uncertainty are being solved for about half a century. Most studies in this field rely on the concept of compatibility between parameters and measurement data in which any measurement result is a kind of a large point “inflated” to a box (rectangular parallelepiped with facets parallel to the coordinate axes). That the graph of the constructed function passes through such a “point” means a nonempty intersection of the graph with the box. However, in some problems, this natural concept turns out to be unsatisfactory.

In this work, for the data fitting under interval uncertainty, we introduce the concept of strong compatibility between data and parameters. It is adequate to the situations when measurements of input and output variables are broken in time, and we strive to uniformly take into account the interval results of output measurements. The paper gives a practical interpretation of the new concept. It is shown that the modified formulation of the problem reduces to recognition and further estimation of the so-called tolerable solution set to interval systems of equations constructed from the processed data.

Keywords: data fitting problem; compatibility between data and parameters; strong compatibility; interval linear equation system; tolerable solution set.

References

1. Voshchinin, A.P. *Industrial laboratory. Materials diagnostics*, 2002, Vol. 68, no. 1, pp. 118–126. (in Russ.).
2. Voshchinin, A.P. Zadachi analiza s neopredelyennymi dannymi – interval'nost' i/ili sluchaynost'? (Analysis problems with uncertain data – intervality and/or randomness?). *Trudy Mezhdunarodnoy konferentsii po vychislitel'noy matematike. Rabochie soveshchaniya. Soveshchanie "Interval'naya matematika i metody rasprostraneniya ogranicheniy" IMRO-2004* (Proceedings of the International Conference on Computational Mathematics. Working meetings. Meeting "Interval mathematics and methods of constraint propagation" IMRO-2004). Novosibirsk, IVMiMG SO RAN Publ., 2004, pp. 147–158. http://www.nsc.ru/interval/Conferences/IMRO_04/Voschinin.pdf
3. Tutubalin, V.N. *Teoriya veroyatnostey: Kratkii kurs i nauchno-metodicheskie zamechaniya* (Probability Theory: A Short Course, scientific and methodological notes). Moskva, MGU Publ., 1972. 230 p. (in Russ.).
4. Gorban', I.I. *Fenomen statisticheskoy ustoychivosti* (Phenomenon of statistical stability). Kiev: Naukova dumka, 2014, 444 p.
5. Kramer, G. *Matematicheskie metody statistiki* (Mathematical Methods of Statistics). Moskva, Mir Publ., 1975. 648 p.
6. Alimov, Yu.I. *Al'ternativa metodu matematicheskoy statistiki* (Alternative to the method of mathematical statistics). Moskva, Znanie Publ., 1980, 64 p. (In Russ.).
7. Alimov, Yu.I., Kravtsov, Yu.A. Is probability a 'normal' physical quantity? *Sov. Phys. Usp.*, 1992, Vol. 35, no. 7., pp.606–622 DOI: 10.1070/PU1992v035n07ABEH002250
8. Tutubalin, V.N. *Granitsy primenimosti (veroyatnostno-statisticheskie metody i ikh vozmozhnosti)* (Limits of applicability (probability and statistical methods and their possibilities)). Moscow, Znanie Publ., 1977, 64 p.
9. Tutubalin, V.N. Probability, computers, and processing the experimental data. *Phys. Usp.*, 1993, Vol. 36, no. 7, pp. 628–641. DOI: 10.1070/PU1993v036n07ABEH002294
10. Kantorovich, L.V. *Sibirskiy matematicheskiy zhurnal*, 1962, Vol. 3, no. 5, pp. 701–709. (in Russ.).
11. Lidov, M.L. Minimax methods for estimation. *Keldysh Institute preprints*, 2010, no. 071, 87 pp. (in Russ.).
12. Spivak, S.I., Timoshenko, V.I., Slin'ko, M.G. *Doklady Akademii Nauk*. 1970. Vol. 192, no. 3, pp. 580–582. (in Russ.).
13. Spivak, S.I., Kantor, O.G., Yunusova, D.S., Kuznetsov, S.I., Kolesov, S.V. Evaluation of measurement accuracy and significance for linear models. *Inform. Primen.*, 2015, Vol. 9, Issue 1, pp. 87–97. DOI: 10.14357/19922264150108
14. Voshchinin, A.P., Bochkov, A.F., Sotirov, G.R. *Zavodskaya laboratoriya. Diagnostika materialov*, 1990, Vol. 56, no. 7, pp. 76–81. (in Russ.).
15. Oskorbin, N.M., Maksimov, A.V., Zhilin, S.I. *Izvestiya Altayskogo gosudarstvennogo universiteta*, 1998, no. 1, pp. 37–40. (in Russ.).
16. Zhilin, S.I. *Nestatisticheskie modeli i metody postroeniya i analiza zavisimostey. Dissertatsiya kand. fiz.-mat. nauk* (Non-statistical models and methods for construction and analysis of dependencies. Cand. phys. and math. sci. diss.). Barnaul, AltGU Publ., 2004, 119 p.
17. Zhilin, S.I. On fitting empirical data under interval error. *Reliable Computing*, 2005, Vol. 11, no. 5, pp. 433–442. DOI: 10.1007/s11155-005-0050-3

18. Zhilin, S.I. Simple method for outlier detection in fitting experimental data under interval error. *Chemometrics and Intellectual Laboratory Systems*, 2007, Vol. 88, no. 1, pp. 60–68.
19. Polyak, B.T., Nazin, S.A. Estimation of Parameters in Linear Multidimensional Systems under Interval Uncertainty. *Journal of Automation and Information Sciences*, 2006, Vol. 38, Issue 2, pp. 19–33. DOI: 10.1615/J Automat Inf Scien.v38.i2.20
20. Schweppe, F.C. Recursive state estimation: unknown but bounded errors and system inputs. *IEEE Trans. Autom. Control*, 1968, Vol. 13, Issue 1, pp. 22–28. DOI: 10.1109/TAC.1968.1098790
21. Jaulin, L., Kieffer M., Didrit, O., Walter, E. *Prikladnoi interval'nyi analiz* (Applied Interval Analysis). Moscow-Izhevsk, RKhD Publ., 2007. 468 p.
22. Milanese, M., Norton, J., Piet-Lahanier, H., Walter, E., eds. *Bounding approaches to system identification*. New York, Plenum Press, 1996, 567 p. DOI: 10.1007/978-1-4757-9545-5
23. Shary, S.P. Solvability of interval linear equations and data analysis under uncertainty. *Automation and Remote Control*, 2012, Vol. 73, no. 2, pp. 310–322 DOI: 10.1134/S0005117912020099
24. Shary, S.P., Sharaya, I.A. Recognizing solvability of interval equations and its application to data analysis. *Journal of Computational Technologies*, 2013, Vol. 18, no. 3, pp. 80–109. (in Russ.).
25. Shary, S.P. Maximum consistency method for data fitting under interval uncertainty. *Journal of Global Optimization*, 2016, Vol. 66, Issue 1, pp. 111–126. DOI: 10.1007/s10898-015-0340-1
26. Kreinovich, V., Shary, S.P. Interval methods for data fitting under uncertainty: a probabilistic treatment. *Reliable Computing*, 2016, Vol. 23, pp. 105–140.
27. Shary, S.P. Solving the linear interval tolerance problem. *Mathematics and Computers in Simulation*, 1995, Vol. 39, pp. 53–85. DOI: 10.1016/0378-4754(95)00135-K
28. Shary, S.P. An interval linear tolerance problem. *Automation and Remote Control*, 2004, Vol. 65, no. 10, pp.1653–1666. DOI: 10.1023/B:AURC.0000044274.25098.da
29. Shary, S.P. *Konechnomernyi interval'nyi analiz* (The finite-dimensional interval analysis). Novosibirsk, XYZ Publ., 2016. 606 p. (in Russ.).
30. Rohn, J. *A handbook of results on interval linear problems*. Prague, Czech Academy of Sciences, 2005, 80 p. <http://www.nsc.ru/interval/Library/Surveys/ILinProblems.pdf>
31. Sharaya, I.A. Structure of the tolerable solution set of an interval linear system. *Journal of Computational Technologies*, 2005, Vol. 10, no. 5, pp. 103–119. (in Russ.).
32. <http://www.nsc.ru/interval/Programing/MCodes/tolsolvty.m>
33. Gutowski, M.W. *Interval experimental data fitting. Focus on Numerical Analysis*. New York, Nova Science Publishers, 2006, pp. 27–70.

Received October 14, 2016