Вычислительные технологии, 2024, том 29, № 4, с. 71–94. © ФИЦ ИВТ, 2024
Computational Technologies, 2024, vol. 29, no. 4, pp. 71–94. © FRC ICT, 2024
ISSN 1560-7534
eISSN 2313-691X

ВЫЧИСЛИТЕЛЬНЫЕ ТЕХНОЛОГИИ

DOI:10.25743/ICT.2024.29.4.006

О восстановлении функциональных зависимостей по ненакрывающим интервальным данным

С. П. Шарый[1,*], М. А. Звягин[2]

[1]Федеральный исследовательский центр информационных и вычислительных технологий, 630090, Новосибирск, Россия

[2]Новосибирский государственный университет, 630090, Новосибирск, Россия

*Контактный автор: Шарый Сергей Петрович, e-mail: shary@ict.nsc.ru

Поступила 07 июля 2022 г., доработана 25 марта 2024 г., принята в печать 01 апреля 2024 г.

Цель работы — представить простой и естественный подход к восстановлению функциональных зависимостей по данным с интервальной неопределенностью, которые не являются накрывающими. Решение задачи сводится к нахождению минимума выпуклой негладкой функции и может быть эффективно найдено с помощью методов негладкой оптимизации. Приведен численный пример, показывающий разительное отличие результата решения задачи восстановления линейной зависимости по накрывающей и ненакрывающей выборкам. Обсуждаются особенности практического применения новой методики.

Ключевые слова: интервал, интервальный анализ данных, задача восстановления зависимости, накрывающие измерения, ненакрывающие измерения, метод прямой интервальной аппроксимации.

Цитирование: Шарый С.П., Звягин М.А. О восстановлении функциональных зависимостей по ненакрывающим интервальным данным. Вычислительные технологии. 2024; 29(4):71–94. DOI:10.25743/ICT.2024.29.4.006.

Введение

Одна из основных целей математического моделирования — построение функциональных зависимостей между различными величинами, участвующими в описании интересующих нас процессов и явлений. Соответствующая постановка задачи называется *задачей восстановления зависимостей* (см., например, [1]), хотя часто встречаются и другие ее названия — задача выравнивания данных или сглаживания данных, задача подгонки данных, задача построения эмпирических формул, задача идентификации и др. В контексте теоретико-вероятностной статистики эта задача является предметом регрессионного анализа, где ее называют задачей построения регрессии. В последние десятилетия эта задача сделалась одной из центральных задач машинного обучения. В целом задача восстановления зависимостей, пожалуй, принадлежит к наиболее популярным и востребованным задачам практики.

Но экспериментальные данные, как правило, всегда неточны, и эту неточность можно описывать и обрабатывать по-разному. Классический теоретико-вероятностный подход опирается на предположение о том, что погрешности в данных являются “случайными величинами”, в том смысле, как их понимает теория вероятностей, с более или менее известными характеристиками их распределений. Такая постановка приводит к широко

известному методу наименьших квадратов и другим популярным методам теоретико-вероятностной статистики. В настоящей статье рассматривается другой подход, основанный на предположении, что погрешности ограниченны, причем нам более или менее известны границы их возможных значений. Это равносильно заданию интервалов возможных значений для результатов измерений, что вызывает необходимость привлечения методов интервального анализа для обработки таких данных.

Наша работа посвящена развитию методов интервального анализа данных для одного частного, но характерного случая, когда обрабатываемые интервалы не удовлетворяют свойству накрытия истинных значений измеряемых величин. С момента своего возникновения в 60-е годы прошлого века в интервальном анализе данных традиционно рассматривались лишь накрывающие интервальные измерения, но в последнее время было осознано, что ненакрывающие данные также должны быть предметом серьезного изучения и анализа. Мы исследуем задачу восстанавления простейшей линейной зависимости на основе интервальных данных, которые не обязательно содержат истинные значения измеренных величин.

Обозначения интервалов и других интервальных объектов, а также связанных с ними величин даются в согласии с неформальным международным стандартом [2]. В частности, интервалы и интервальные объекты выделяются в работе буквами жирного шрифта.

1. Накрывающие и ненакрывающие интервальные измерения и выборки

Напомним одно из базовых понятий метрологии — науки об измерениях [3–5]:

Определение 1. *Истинным значением измеряемой величины* называется значение, идеально отражающее эту величину в рамках принятой модели (теории) рассматриваемого объекта или явления.

Важно отметить, что измеряемая величина существует лишь в рамках принятой модели, т. е. имеет смысл только до тех пор, пока модель признается адекватной объекту или явлению. Принципиальным положением классической метрологии является утверждение о существовании истинного значения. Но получение этого истинного значения на практике часто невозможно, так как измерения могут искажаться неизбежными помехами, измерительные приборы могут быть несовершенны и давать не вполне точные результаты и т. п.

В последние десятилетия в западной науке наметилась явная тенденция к отходу от использования понятия “истинного значения” на том основании, что оно труднодоступно или вообще недоступно, не может быть грубо материально осязаемо и т. п. Как следствие, вместо понятия “погрешность измерения” современные западные методики (к примеру, [6]) предлагают говорить про “неопределенность измерения” саму по себе, вне связи с какими-то объективными значениями величин и т. д. Реакция профессионального сообщества метрологов России на эти новшества является сложной и неоднозначной. Можно даже говорить об определенном неприятии этих методологических установок. Мы далее придерживаемся точки зрения классической метрологии, которая выражена, например, в учебнике [3] и современных стандартах [4, 5].

В современной теории интервальных измерений различают измерения накрывающие и ненакрывающие [7, 8]:

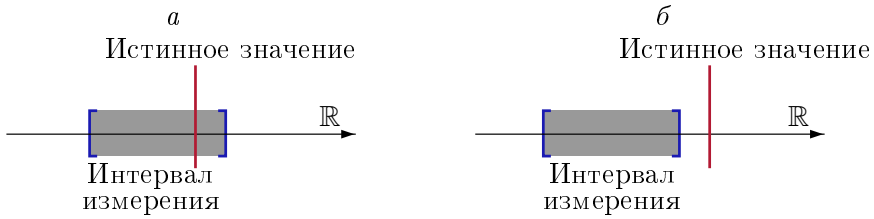

Рис. 1. Накрывающее (*а*) и ненакрывающее (*б*) измерения точечного истинного значения некоторой физической величины
Fig. 1. Enclosing (*а*) and non-enclosing (*б*) measurements of the point true value of a physical quantity at some point

Определение 2. *Накрывающее измерение* (*накрывающий замер*) — это интервальная оценка измеряемой величины, которая гарантированно содержит ее истинное значение. Измерение, относительно которого нельзя гарантировать, что оно содержит истинное значение измеряемой величины, называется *ненакрывающим* (рис. 1).

Таким образом, ненакрывающее измерение может содержать истинное значение, а может и не содержать. Какая из этих возможностей реализуется, нам точно неизвестно.

Измерения обычно проводят группами (сериями), а результатами таких серий являются *выборки* интервальных данных [8]. Свойства выборки решающим образом зависят от свойств составляющих ее измерений.

Определение 3. Выборка интервальных результатов измерений называется *накрывающей*, если доминирующая часть (подавляющее большинство) входящих в нее интервалов — накрывающие. В противном случае, когда бóльшая часть входящих в выборку интервалов измерений (в пределе — все) не являются накрывающими, выборка называется *ненакрывающей*.

В этом определении фигурируют не вполне строгие выражения — “большинство”, “доминирующая часть” и т. п., что вызвано существом задачи и желанием сделать определение практичным. Дело в том, что реальные измерения и наблюдения часто сопровождаются так называемыми промахами или грубыми ошибками — такими результатами, которые сильно искажены и никак не отражают свойства исследуемого объекта. Они, естественно, почти всегда не удовлетворяют свойству накрытия истинного значения, но не допускать их присутствие в выборке нереалистично. Такие промахи стараются выявить и отсеять в процессе предварительной обработки данных.

К промахам близки по смыслу выбросы в данных, которые являются аномальными измерениями, выбивающимися из общего характера выборки и которые требуют дополнительного исследования.

Накрывающее измерение ценно тем, что дает “двустороннюю вилку” для возможных значений интересующей нас величины, и эта оценка может служить отправной точкой, основой для применения к обработке данных мощных методов интервального анализа. Если же измерение — ненакрывающее, то содержание интервальных данных обедняется: вместо двусторонних “вилок” имеются просто “растекшиеся” точки. Арсенал методов, которые могут быть применены к таким данным, также обедняется и сужается.

Вынесенный на обложку книги [8] рис. 2 наглядно иллюстрирует интересующую нас задачу восстановления линейной зависимости по интервальным данным. Многомерные интервалы данных, по которым требуется построить искомую функцию, называют также *брусами неопределенности* измерений или *отрезками неопределенности* измерений

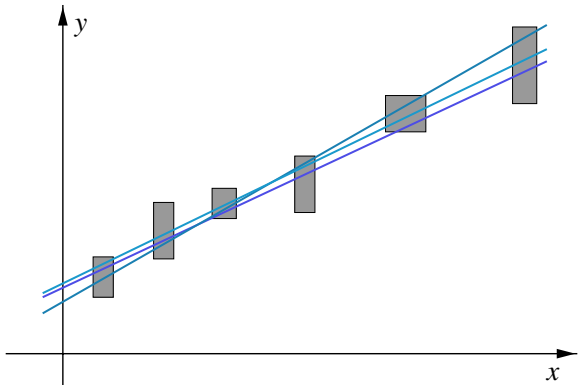

Рис. 2. Восстановление линейной зависимости по интервальным данным
Fig. 2. Constructing a linear functional dependency from interval data

в зависимости от их размерности [8]. Если условие прохождения графика восстанавливаемой зависимости через брусы неопределенности измерений не работает, то остается единственный критерий, согласно которому можно определять, насколько график "подходит" к данным, и это — расстояние от графика до брусов неопределенности, рассматриваемых как целостные объекты.

Напомним, что интервальный анализ данных зародился в начале 60-х гг. прошлого века и его началом можно считать пионерную работу Л.В. Канторовича [9]. В ней сформулированы основы нового подхода к обработке неточностей и неопределенностей, которые предлагалось описывать в виде двусторонних оценок, т. е. фактически интервалов. При этом молчаливо считалось, что интервальные результаты измерений являются накрывающими и всегда (или почти всегда) содержат истинные значения измеряемой величины. С годами постепенно пришло понимание того факта, что ненакрывающие измерения и ненакрывающие выборки тоже существуют, они составляют заметную долю в общем числе измерений с интервальными результатами. Более того, было осознано, что ненакрывающие интервальные измерения и выборки могут быть полезными и должны серьезно рассматриваться наряду с накрывающими [7, 8]. Это вызвано, в частности, тем фактом, что обеспечить свойство накрытия истинного значения — это не вполне тривиальная задача (и даже не однозначная) и ее решение требует отдельных усилий. С другой стороны, трактовка ненакрывающих интервальных данных как промахов или выбросов неоправданно упрощает ситуацию, в некотором смысле даже вульгаризует ее.

2. Теория: случай точных независимых переменных

Нам необходимо определить линейную функцию вида

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_m x_m \tag{1}$$

по наборам измеренных значений независимых переменных x_1, ..., x_m и зависимой переменной y. Предполагаем, что имеется всего n измерений, в результате которых получены наборы значений

$$(x_{i1}, x_{i2}, \ldots, x_{im}, \boldsymbol{y}_i), \qquad i = 1, \ldots, n, \tag{2}$$

где x_{ij} — значение j-й независимой переменной x_j в i-м измерении; $\boldsymbol{y}_i$ — интервал оценки значения функции в i-м измерении, $j = 1, 2, \ldots, m$. Иными словами, рассматриваем ситуацию, когда x_{ij} — известные точно вещественные числа, а y_i заданы неточно и для них известны интервальные оценки значений $\boldsymbol{y}_i = [\underline{\boldsymbol{y}}_i, \overline{\boldsymbol{y}}_i]$ (рис. 3). При этом интервалы $\boldsymbol{y}_i$, вообще говоря, не обязательно являются накрывающими для истинных значений измеряемых величин. В этих условиях необходимо найти вещественные параметры β_0, β_1, ..., β_m для выражения (1), чтобы оно "наилучшим образом" приближало (аппроксимировало и т. п.) измеренные данные (2).

Ниже нам понадобится много работать с вектором значений независимых переменных в i-м измерении, и будем обозначать

$$x_{i:} = \left(x_{i1}, x_{i2}, \ldots, x_{im}\right)$$

в соответствии с нотацией, идущей от MATLAB'а и других систем компьютерной математики.

Как уже отмечалось, если не требуется, чтобы график восстанавливаемой функции каким-либо образом обязательно проходил через брусы непределенности измерений (что соответствовало бы накрывающим измерениям), то остается лишь один критерий соответствия конструируемой функции исходным данным. Это — расстояние от графика функции до данных, понимаемое как некоторая мера отклонения брусов неопределенности (или отрезков неопределенности) от соответствующих им точек графика. Фактически расстояния от отдельных брусов неопределенности измерений до графика восстанавливаемой функции являются аналогами так называемых "остатков" в регрессионном анализе [10].

В рассматриваемой ситуации это расстояние до графика, во-первых, нужно определить для отдельного интервала неопределенности $\boldsymbol{y}_i$ и, во-вторых, нужно определить его для всей выборки (2) по набору расстояний до каждого $\boldsymbol{y}_i$, $i = 1, 2, \ldots, n$. Второй пункт этой программы может быть реализован, к примеру, способом, которым задается расстояние (метрика) на прямом декартовом произведении метрических пространств. Расстояние от графика до выборки в целом можно также построить с помощью какой-либо нормы вектора отдельных расстояний, подходящей по смыслу задачи. Займемся поэтому расстоянием от графика восстанавливаемой линейной функции до интервала неопределенности измерения.

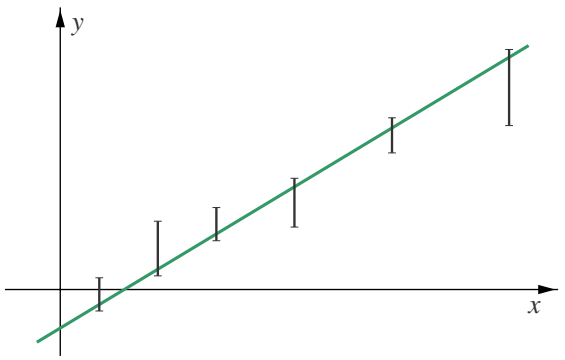

Рис. 3. Восстановление линейной зависимости по интервальным данным в случае точных значений независимых переменных

Fig. 3. Constructing a linear functional dependency from interval data for the case of exact values of independent variables

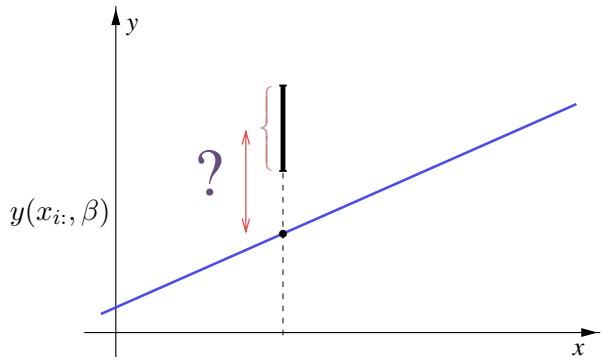

Рис. 4. Расстояние от графика восстанавливаемой функции до интервального результата измерения (стрелка)
Fig. 4. Distance from the graph of the constructed function to an interval measurement result (arrow)

Из выборки (2) возьмем значения аргументов i-го измерения, т. е. $(x_{i1}, x_{i2}, \ldots, x_{im}) = x_{i:}$. Подставим их в уравнение восстанавливаемой линейной функции (1). Будет получено ее значение, которое обозначим

$$y(x_{i:},\beta) = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_m x_{im},$$

где $(\beta_0, \beta_1, \ldots, \beta_m) = \beta$. Расстояние dist по вертикали от точки $y(x_{i:},\beta)$ до отрезка неопределенности $(x_{i1}, x_{i2}, \ldots, x_{im}, \boldsymbol{y}_i)$ (рис. 4) естественно определять так же, как это делается в интервальном анализе [11], т. е. как расстояние от числа $y(x_{i:},\beta)$, которое является вырожденным интервалом $[\,y(x_{i:},\beta), y(x_{i:},\beta)]$, до интервала $\boldsymbol{y}_i$:

$$\operatorname{dist}\big(y(x_{i:},\beta), \boldsymbol{y}_i\big) = \max\big\{\big|y(x_{i:},\beta) - \underline{\boldsymbol{y}}_i\big|, \big|y(x_{i:},\beta) - \overline{\boldsymbol{y}}_i\big|\big\}. \tag{3}$$

Переходя ко всей выборке, получаем вектор

$$\Big(\operatorname{dist}\big(y(x_{1:},\beta), \boldsymbol{y}_1\big), \operatorname{dist}\big(y(x_{2:},\beta), \boldsymbol{y}_2\big), \ldots, \operatorname{dist}\big(y(x_{n:},\beta), \boldsymbol{y}_n\big)\Big)^{\top},$$

образованный расстояниями от отрезков неопределенности измерений до графика. Взяв какую-нибудь норму $\|\cdot\|$ этого вектора, получим меру отклонения всей выборки от графика функции:

$$\Phi(\beta) := \Big\|\Big(\operatorname{dist}\big(y(x_{1:},\beta), \boldsymbol{y}_1\big), \operatorname{dist}\big(y(x_{2:},\beta), \boldsymbol{y}_2\big), \ldots, \operatorname{dist}\big(y(x_{n:},\beta), \boldsymbol{y}_n\big)\Big)^{\top}\Big\|.$$

Назовем $\Phi(\beta)$ *функцией отклонения* данных от графика восстанавливаемой зависимости.

В зависимости от конкретного выбора векторной нормы получаются различные версии этой функции Φ, которые будем обозначать необходимыми модифицирующими индексами. В частности, для популярных норм, принимая во внимание неотрицательность расстояния dist, имеем

$$\Phi_1(\beta) = \sum_{i=1}^{n} \operatorname{dist}\big(y(x_{i:},\beta), \boldsymbol{y}_i\big) \qquad \text{— для 1-нормы,}$$

$$\Phi_2(\beta) = \sqrt{\sum_{i=1}^{n} \big(\operatorname{dist}\big(y(x_{i:},\beta), \boldsymbol{y}_i\big)\big)^2} \qquad \text{— для 2-нормы (евклидовой нормы),}$$

$$\Phi_\infty(\beta) = \max_{1\le i\le n} \operatorname{dist}\big(y(x_{i:},\beta), \boldsymbol{y}_i\big) \qquad \text{— для чебышёвской нормы (максимум-нормы),}$$

где $\beta = (\beta_0, \beta_1, \ldots, \beta_m)$. Значения аргументов β_0, β_1, ..., β_m, доставляющие минимум функции Φ, являются искомой оценкой параметров восстанавливаемой функции (1). Назовем описанный выше метод оценивания параметров *прямой интервальной аппроксимацией* данных, обозначая его для краткости аббревиатурой ПИА.

Введенным выше конструкциям можно придать другую форму, более удобную при исследовании функции отклонения и ее конкретных реализаций Φ_1, Φ_2, Φ_∞ и др.

Напомним, что для интервалов $\boldsymbol{a}$ и $\boldsymbol{b}$ операция *внутреннего вычитания* (называемая также алгебраическим вычитанием [11]), обратная к интервальному сложению, определяется как

$$\boldsymbol{a} \ominus \boldsymbol{b} = \big[\underline{\boldsymbol{a}} - \underline{\boldsymbol{b}}, \overline{\boldsymbol{a}} - \overline{\boldsymbol{b}}\big].$$

Результатом этой операции может быть как обычный классический интервал, так и "неправильный интервал" из арифметики Каухера [11]. Модуль интервала, правильного или неправильного, понимают в интервальном анализе как максимум модулей его концов. Тогда необходимое расстояние между интервалами (3), как известно, можно представить в эквивалентном виде:

$$\operatorname{dist}\big(y(x_{1:}, \beta), \boldsymbol{y}_i\big) = |y(x_{1:}, \beta) \ominus \boldsymbol{y}_i|.$$

Естественным покомпонентным образом операция "$\ominus$" распространяется на интервальные векторы. Если теперь воспользоваться понятием нормы интервальных векторов [11], обобщающим модуль интервалов, то функцию отклонения Φ можно переписать в виде

$$\Phi(\beta) = \Phi\big(\beta_0, \beta_1, \ldots, \beta_m\big) = \|y(x, \beta) \ominus \boldsymbol{y}\|, \tag{4}$$

где

$$y(x, \beta) = \big(y(x_{1:}, \beta), y(x_{2:}, \beta), \ldots, y(x_{n:}, \beta)\big)^\top, \qquad \boldsymbol{y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n)^\top.$$

Для нахождения параметров линейной функции (1), наилучшим образом подходящей к данным (2), необходимо минимизировать в выбранной норме значение $\Phi(\beta) = \|y(x, \beta) \ominus \boldsymbol{y}\|$ как функцию от $\beta = (\beta_0, \beta_1, \ldots, \beta_m)$. Аргумент найденного минимума даст искомую оценку параметров.

Конкретное выражение для Φ зависит от того, какую конкретно норму интервальных векторов берем в выражении (4). В качестве нормы, которая агрегирует отдельные расстояния от графика до интервалов неопределенности, можно взять, например, одну из следующих:

$$\|\boldsymbol{a}\|_1 = \sum_{i=1}^{n} |\boldsymbol{a}_i|, \qquad \|\boldsymbol{a}\|_p = \left(\sum_{i=1}^{n} |\boldsymbol{a}_i|^p\right)^{1/p}, \qquad \|\boldsymbol{a}\|_\infty = \max_{1 \le i \le n} |\boldsymbol{a}_i|.$$

Чтобы дать явные развернутые выражения для функции $\Phi(\beta_0, \beta_1, \ldots, \beta_m)$, соответствующие этим нормам, необходим вспомогательный результат.

Предложение 1. Если $\boldsymbol{a} \in \mathbb{IR}$ и $b \in \mathbb{R}$, то $|\boldsymbol{a} \ominus b| = \operatorname{rad} \boldsymbol{a} + |\operatorname{mid} \boldsymbol{a} - b|$.

Доказательство. Пусть $b = t + \operatorname{mid} \boldsymbol{a} = t + \dfrac{1}{2}(\overline{\boldsymbol{a}} + \underline{\boldsymbol{a}})$ для некоторого вещественного числа t. Тогда

$$|\boldsymbol{a} \ominus b| = \max\{|\underline{\boldsymbol{a}} - b|, |\overline{\boldsymbol{a}} - b|\} = \max\left\{\left|\frac{\underline{\boldsymbol{a}} - \overline{\boldsymbol{a}}}{2} - t\right|, \left|\frac{\overline{\boldsymbol{a}} - \underline{\boldsymbol{a}}}{2} - t\right|\right\} = \max\{|-\operatorname{rad} \boldsymbol{a} - t|, |\operatorname{rad} \boldsymbol{a} - t|\}.$$

Обозначим последнее выражение через $g(t)$. Оно является максимумом расстояний от t до двух симметричных относительно нуля точек вещественной оси: $-\operatorname{rad}\boldsymbol{a}$ и $\operatorname{rad}\boldsymbol{a}$. Поэтому

$$g(t)=\begin{cases}|-\operatorname{rad}\boldsymbol{a}-t|, & \text{если } t\geq 0\\ |\operatorname{rad}\boldsymbol{a}-t|, & \text{если } t\leq 0\end{cases}=\begin{cases}t+\operatorname{rad}\boldsymbol{a}, & \text{если } t\geq 0\\ -t+\operatorname{rad}\boldsymbol{a}, & \text{если } t\leq 0\end{cases}=|t|+\operatorname{rad}\boldsymbol{a}.$$

Следовательно, получаем $|\boldsymbol{a}\ominus b|=\operatorname{rad}\boldsymbol{a}+|\operatorname{mid}\boldsymbol{a}-b|$. ■

Теперь можно записать функцию отклонения $\Phi\big(\beta_0,\ldots,\beta_m\big)$ в конкретных нормах (нижний индекс указывает, какой норме соответствует Φ):

$$\Phi_1\big(\beta_0,\ldots,\beta_m\big)=\sum_{i=1}^{n}\operatorname{rad}\boldsymbol{y}_i+\sum_{i=1}^{n}\left|\operatorname{mid}\boldsymbol{y}_i-\left(\beta_0+\sum_{j=1}^{m}\beta_j x_{ij}\right)\right|, \tag{5}$$

$$\Phi_p\big(\beta_0,\ldots,\beta_m\big)=\left(\sum_{i=1}^{n}\left(\operatorname{rad}\boldsymbol{y}_i+\left|\operatorname{mid}\boldsymbol{y}_i-\left(\beta_0+\sum_{j=1}^{m}\beta_j x_{ij}\right)\right|\right)^p\right)^{1/p}, \tag{6}$$

$$\Phi_\infty\big(\beta_0,\ldots,\beta_m\big)=\max_{1\leq i\leq n}\left\{\operatorname{rad}\boldsymbol{y}_i+\left|\operatorname{mid}\boldsymbol{y}_i-\left(\beta_0+\sum_{j=1}^{m}\beta_j x_{ij}\right)\right|\right\}. \tag{7}$$

Пример 1. Пусть $m=1$, т. е. восстанавливаем линейную зависимость вида $y=\beta_0+\beta_1 x$ от одной переменной x.

Решим задачу отыскания минимума функции $\Phi_\infty(\beta_0,\beta_1)$ для $n=2$, т. е. для двух измерений. В этом случае заданы интервалы неопределенности $(x_{11},\boldsymbol{y}_1)$ и $(x_{21},\boldsymbol{y}_2)$, так что имеем

$$\Phi_\infty(\beta_0,\beta_1)=\max\big\{\operatorname{rad}\boldsymbol{y}_1+|\operatorname{mid}\boldsymbol{y}_1-(\beta_0+\beta_1 x_1)|,\operatorname{rad}\boldsymbol{y}_2+|\operatorname{mid}\boldsymbol{y}_2-(\beta_0+\beta_1 x_2)|\big\}.$$

Ясно, что минимум выражения достигается на решении системы линейных уравнений

$$\begin{cases}\beta_0+\beta_1 x_1=\operatorname{mid}\boldsymbol{y}_1,\\ \beta_0+\beta_1 x_2=\operatorname{mid}\boldsymbol{y}_2,\end{cases}$$

которое одновременно обнуляет выражения под модулями. Это решение имеет вид

$$\beta_0=\frac{x_2\operatorname{mid}\boldsymbol{y}_1-x_1\operatorname{mid}\boldsymbol{y}_2}{x_2-x_1},\qquad \beta_1=\frac{\operatorname{mid}\boldsymbol{y}_2-\operatorname{mid}\boldsymbol{y}_1}{x_2-x_1}.$$

При $\operatorname{rad}\boldsymbol{y}_1=\operatorname{rad}\boldsymbol{y}_2=0$ оно превращается в решение задачи о проведении прямой через две точки, что вполне естественно. ■

Заметим, что для неинтервальных данных, когда $\operatorname{rad}\boldsymbol{y}=0$ и $\operatorname{mid}\boldsymbol{y}=y$, введенные выше функционалы Φ_1, Φ_p и Φ_∞ получают следующий вид:

$$\Phi_1(\beta_0,\ldots,\beta_m)=\sum_{i=1}^{n}\left|y_i-\left(\beta_0+\sum_{j=1}^{m}\beta_j x_{ij}\right)\right|,$$

$$\Phi_p(\beta_0,\ldots,\beta_m)=\left(\sum_{i=1}^{n}\left|y_i-\left(\beta_0+\sum_{j=1}^{m}\beta_j x_{ij}\right)\right|^p\right)^{1/p},$$

$$\Phi_\infty(\beta_0,\ldots,\beta_m)=\max_{1\leq i\leq n}\left|y_i-\left(\beta_0+\sum_{j=1}^{m}\beta_j x_{ij}\right)\right|.$$

Поиск минимума этих выражений соответствует задачам о наилучшем приближении точек гиперплоскостью для 1-нормы, p-нормы и чебышёвской нормы (∞-нормы) соответственно.

Отметим, что описанный выше метод ПИА удовлетворяет принципу соответствия, сформулированному Н. Бором и переосмысленному на случай интервального анализа данных в книге [8]. При стягивании ширины интервалов неопределенности к нулю получим в пределе метод восстановления зависимости по точечным данным, минимизирующий какую-то норму вектора остатков наблюдений. В частности, если в качестве нормы $\|\cdot\|$ взята евклидова норма (2-норма), то в пределе получим классический метод наименьших квадратов.

3. Свойства функции отклонения

Исследуем свойства функции отклонения $\Phi(\beta_0, \beta_1, \ldots, \beta_m)$ как в общем случае, так и для отдельных ее реализаций (5)–(7). Будем существенно использовать результаты выпуклого анализа [12, 13].

Определение 4. Множество $S \subseteq \mathbb{R}^n$ называется *выпуклым*, если одновременно с любыми двумя своими точками содержит отрезок прямой, который их соединяет. Иными словами, множество $S \subseteq \mathbb{R}^n$ называется *выпуклым*, если для любых x, $y \in S$ и любого $\lambda \in [0,1]$ точка $\lambda x + (1-\lambda)y$ также лежит в S.

Определение 5. Пусть S — выпуклое множество в $\mathbb{R}^n$. Функция $f : S \to \mathbb{R}$ называется *выпуклой*, если для любых $x, y \in S$ и $\lambda \in [0,1]$ выполняется неравенство

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda) f(y).$$

Из математического анализа известны признаки выпуклости функции, основанные на исследовании знака второй производной и т. п. К сожалению, мы не можем воспользоваться ими, так как рассматриваемые функции отклонения не являются дифференцируемыми всюду на своей области определения.

Предложение 2. Функция $\Phi(\beta_0, \beta_1, \ldots, \beta_m)$, определяемая посредством (4), является выпуклой для любой векторной нормы $\|\cdot\|$.

Доказательство. Пусть $X_j \in \mathbb{R}^n$, $j = 1, 2, \ldots, m$, — векторы с компонентами x_{ij}, так что $X_j = (x_{ij})_{i=1}^n$, и пусть $0 \le \lambda \le 1$. Тогда, опираясь на свойства норм векторов, можем утверждать справедливость следующей цепочки соотношений:

$$\begin{aligned}
&\Phi\big(\lambda\alpha_0 + (1-\lambda)\beta_0, \lambda\alpha_1 + (1-\lambda)\beta_1, \ldots, \lambda\alpha_m + (1-\lambda)\beta_m\big) = \\
&\quad = \left\| \lambda\alpha_0 + (1-\lambda)\beta_0 + \sum_{j=1}^{m} \big(\lambda\alpha_j + (1-\lambda)\beta_j\big) X_j \ominus \boldsymbol{y} \right\| = \\
&\quad = \left\| \lambda\alpha_0 + (1-\lambda)\beta_0 + \sum_{j=1}^{m} \big(\lambda\alpha_j + (1-\lambda)\beta_j\big) X_j \ominus \big(\lambda\boldsymbol{y} + (1-\lambda)\boldsymbol{y}\big) \right\| = \\
&\quad = \left\| \lambda\left(\alpha_0 + \sum_{j=1}^{m} \alpha_j X_j\right) + (1-\lambda)\left(\beta_0 + \sum_{j=1}^{m} \beta_j X_j\right) \ominus \big(\lambda\boldsymbol{y} + (1-\lambda)\boldsymbol{y}\big) \right\| = \\
&\quad = \left\| \lambda\left(\alpha_0 + \sum_{j=1}^{m} \alpha_j X_j\right) \ominus \lambda\boldsymbol{y} + (1-\lambda)\left(\beta_0 + \sum_{j=1}^{m} \beta_j X_j\right) \ominus (1-\lambda)\boldsymbol{y} \right\| \le
\end{aligned}$$

$$\leq \lambda \left\| \left(\alpha_0 + \sum_{j=1}^{m} \alpha_j X_j \right) \ominus \boldsymbol{y} \right\| + (1-\lambda) \left\| \left(\beta_0 + \sum_{j=1}^{m} \beta_j X_j \right) \ominus \boldsymbol{y} \right\| =$$
$$= \lambda \Phi(\alpha_0, \alpha_1, \ldots, \alpha_m) + (1-\lambda)\Phi(\beta_0, \beta_1, \ldots, \beta_m).$$

Здесь переход от второй строки к третьей возможен потому, что из-за неотрицательности коэффициентов λ и $(1-\lambda)$ вместо соотношения субдистрибутивности выполнена дистрибутивность $\lambda \boldsymbol{y} + (1-\lambda)\boldsymbol{y} = \boldsymbol{y}$.

В целом сравнивая начало и конец цепочки, можем видеть, что полученное неравенство на значения функции Φ, справедливое для любых аргументов, как раз и означает выпуклость этой функции. ■

Нам понадобится

Определение 6. Для функции $f : \mathbb{R}^n \to \mathbb{R}$ ее *надграфиком* называется множество $\operatorname{epi} f = \{(x, y) \in \mathbb{R}^{n+1} \mid x \in \mathbb{R}^n, y \in \mathbb{R}, y \geq f(x)\}$, т. е. множество точек в $\mathbb{R}^{n+1}$, лежащих на графике функции f и выше его.

Напомним, что *полупространством* в линейном пространстве $\mathbb{R}^n$ называют одну из двух частей, на которые оно разделяется гиперплоскостью, т. е. плоскостью коразмерности 1. Если разделяющая гиперплоскость принадлежит полупространству, то оно называется *замкнутым*. Гиперплоскость в $\mathbb{R}^n$ задается, как известно, линейным алгебраическим уравнением вида

$$a_1 x_1 + a_2 x_2 + \ldots + a_n x_n = b$$

с какими-то коэффициентами a_1, a_2, …, a_n и свободным членом b. Поэтому замкнутое полупространство — это множество точек из $\mathbb{R}^n$, удовлетворяющих нестрогому линейному алгебраическому неравенству

$$a_1 x_1 + a_2 x_2 + \ldots + a_n x_n \leq b \tag{8}$$

или же аналогичному нестрогому неравенству противоположного смысла.

Определение 7. *Выпуклым полиэдральным множеством* в $\mathbb{R}^n$ называется пересечение конечного набора замкнутых полупространств или, что равносильно, множество решений конечной системы нестрогих линейных алгебраических неравенств вида (8).

Определение 8. Функция $f : \mathbb{R}^n \to \mathbb{R}$ называется *выпуклой полиэдральной функцией*, если ее надграфик — выпуклое полиэдральное множество в $\mathbb{R}^{n+1}$.

Понятие выпуклой полиэдральной функции является дальнейшей конкретизацией понятия выпуклой функции. Фактически это функции, графики которых составлены из кусков гиперплоскостей (рис. 5). Покажем, что введенные выше функции Φ_1 и Φ_∞ не просто выпуклые, но дополнительно удовлетворяют этому усиленному условию.

Предложение 3. Функции $\Phi_1(\beta_0, \ldots, \beta_m)$ и $\Phi_\infty(\beta_0, \ldots, \beta_m)$ являются выпуклыми и полиэдральными.

Доказательство. Оно основано на том, что сумма выпуклых функций и поточечный максимум выпуклых функций также являются выпуклыми (см., к примеру, [12, 13]). Это же самое справедливо, как нетрудно понять, для свойства полиэдральности. По этой причине достаточно доказать выпуклость и полиэдральность для отдельных "строительных блоков", из которых сконструированы выражения (5) и (7) для $\Phi_1(\beta_0, \ldots, \beta_m)$ и $\Phi_\infty(\beta_0, \ldots, \beta_m)$, т. е. для выражений

$$\operatorname{rad} \boldsymbol{y}_i + \left| \operatorname{mid} \boldsymbol{y}_i - \left(\beta_0 + \sum_{j=1}^{m} \beta_j x_{ij} \right) \right|.$$

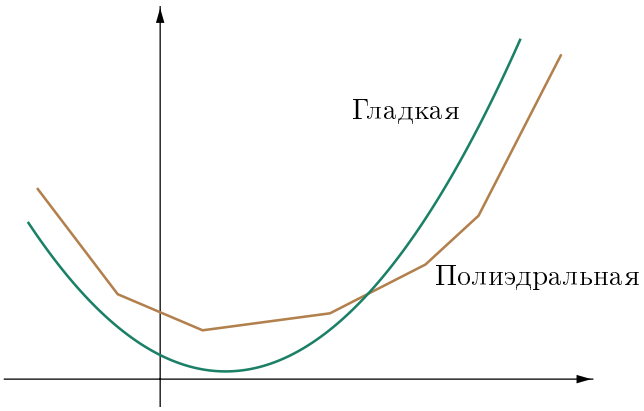

Рис. 5. Выпуклые функции — гладкая и полиэдральная
Fig. 5. Convex functions — smooth and polyhedral

Если $(\beta'_0,\ldots,\beta'_m)$ и $(\beta''_0,\ldots,\beta''_m)$ — два каких-то набора параметров, $0 \le \lambda \le 1$, то в силу неравенства треугольника

$$\begin{aligned}
&\left|\operatorname{mid} \boldsymbol{y}_i - \left(\big(\lambda\beta'_0 + (1-\lambda)\beta''_0\big) + \sum_{j=1}^{m}\big(\lambda\beta'_j + (1-\lambda)\beta''_j\big)x_{ij}\right)\right| \le \\
&\le \left|\lambda\operatorname{mid} \boldsymbol{y}_i - \left(\lambda\beta'_0 + \sum_{j=1}^{m}\lambda\beta'_j x_{ij}\right)\right| + \left|(1-\lambda)\operatorname{mid} \boldsymbol{y}_i - \left((1-\lambda)\beta''_0 + \sum_{j=1}^{m}(1-\lambda)\beta''_j x_{ij}\right)\right| = \\
&= \lambda\left|\operatorname{mid} \boldsymbol{y}_i - \left(\beta'_0 + \sum_{j=1}^{m}\beta'_j x_{ij}\right)\right| + (1-\lambda)\left|\operatorname{mid} \boldsymbol{y}_i - \left(\beta''_0 + \sum_{j=1}^{m}\beta''_j x_{ij}\right)\right|,
\end{aligned}$$

откуда следует выпуклость. Полиэдральность вытекает из представления для модуля $|a| = \max\{a, -a\}$. ■

4. Теория: случай неточных независимых переменных

Разумеется, неточность и неопределенность могут присутствовать не только в значениях функции y, но и в значениях независимых переменных $x_1, x_2, \ldots, x_m$. Для них в этом случае по результатам i-го измерения задаются соответствующие интервалы неопределенности $\boldsymbol{x}_{i1}$, $\boldsymbol{x}_{i2}$, ..., $\boldsymbol{x}_{im}$, образующие в совокупности интервальный вектор $\boldsymbol{x}_{i:}$, так что

$$\boldsymbol{x}_{i:} = (\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \ldots, \boldsymbol{x}_{im}).$$

При точном задании аргумента $x_{i:}$ мы получали конкретную точку $(x_{i:}, y(x_{i:}, \beta))$ на графике функции и считали расстояние от нее до интервала данных $\boldsymbol{y}_i$. Теперь же значения аргументов образуют целый брус $\boldsymbol{x}_{i:}$, и для этой новой ситуации нужно определить способ расчета "остатков" согласно терминологии регрессионного анализа, т. е. расстояний от графика восстанавливаемой функциональной зависимости до бруса неопределенности данных. Возможны различные подходы к этому определению.

Один из простых и естественных способов — посчитать расстояние для каждого фиксированного аргумента $x \in \boldsymbol{x}$ и выбрать максимум полученных расстояний. Это согласуется с общим "чебышёвским" (минимаксным) смыслом наших конструкций, так

как расстояние от точки до интервала — это тоже максимум расстояний от точки до представителей интервала. Пусть, как и ранее,

$$y(x_{i:},\beta) = y(x_{i1},x_{i2},\ldots,x_{im}) = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_m x_{im}.$$

Тогда i-м остатком, т. е. расстоянием от бруса неопределенности i-го измерения до графика восстанавливаемой линейной функции положим значение

$$\max_{x_{i:}\in \boldsymbol{x}_{i:}} |y(x_{i:},\beta) \ominus \boldsymbol{y}| = \max_{x_{i:}\in \boldsymbol{x}_{i:}} \left\{ |\underline{\boldsymbol{y}} - y(x_{i:},\beta)|, |\overline{\boldsymbol{y}} - y(x_{i:},\beta)| \right\}. \tag{9}$$

Минимизация определенного таким способом расстояния соответствует поиску минимума максимального отклонения точек бруса от аппроксимирующей плоскости “в вертикальном направлении” (рис. 6). Ясно, что возможны также другие способы определения расстояния от брусов неопределенности данных до графика восстанавливаемой зависимости (“остатков”), причем в каких-то прикладных задачах они могут оказаться даже предпочтительнее выбранного нами.

Используя предложение 1, получим

$$\begin{aligned}
\max_{x_{i:}\in \boldsymbol{x}_{i:}} |y(x_{i:},\beta) \ominus \boldsymbol{y}_i| &= \max_{x_{i:}\in \boldsymbol{x}_{i:}} \left\{ \operatorname{rad} \boldsymbol{y}_i + \left| \operatorname{mid} \boldsymbol{y}_i - \left(\beta_0 + \sum_{j=1}^{m} \beta_j x_{ij} \right) \right| \right\} = \\
&= \max_{x_{i:}\in \boldsymbol{x}_{i:}} \left\{ \operatorname{rad} \boldsymbol{y}_i + \max \left\{ \operatorname{mid} \boldsymbol{y}_i - \left(\beta_0 + \sum_{j=1}^{m} \beta_j x_{ij} \right), -\operatorname{mid} \boldsymbol{y}_i + \left(\beta_0 + \sum_{j=1}^{m} \beta_j x_{ij} \right) \right\} \right\} = \\
&= \max \left\{ \max_{x_{i:}\in \boldsymbol{x}_{i:}} \left\{ \overline{\boldsymbol{y}}_i - \left(\beta_0 + \sum_{j=1}^{m} \beta_j x_{ij} \right) \right\}, \max_{x_{i:}\in \boldsymbol{x}_{i:}} \left\{ -\underline{\boldsymbol{y}}_i + \left(\beta_0 + \sum_{j=1}^{m} \beta_j x_{ij} \right) \right\} \right\}.
\end{aligned} \tag{10}$$

Стоящие внутри внешних фигурных скобок два максимума по $x_{i:} \in \boldsymbol{x}_{i:}$ — это задачи нахождения максимумов линейных функций по брусу $\boldsymbol{x}_{i:}$, т. е. задачи линейного программирования. Они могут быть относительно просто решены стандарными средствами, например с помощью готовых программ для решения задач линейного программирования.

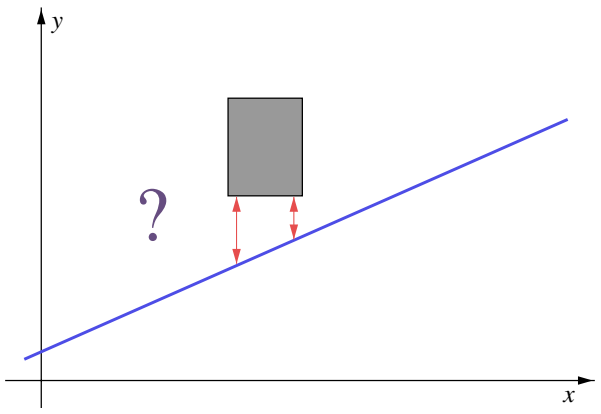

Рис. 6. Расстояние от графика восстанавливаемой функции до интервального результата измерения (стрелки)

Fig. 6. Distance from the graph of the constructed function to an interval measurement result (arrows)

С другой стороны, из свойств задачи линейного программирования следует, что фигурирующие в выражении (9) максимумы достигаются в угловых точках бруса $\boldsymbol{x}_{i:}$. В случае малых размерностей m можно найти эти максимумы с помощью полного перебора вершин интервального бруса $\boldsymbol{x}_{i:}$. В частности, при $m = 1$ имеем

$$\max_{x_{i1} \in \boldsymbol{x}_{i1}} |y(x_{i1}, \beta) \ominus \boldsymbol{y}_i| = \operatorname{rad} \boldsymbol{y}_i + \max \left\{ \left| \operatorname{mid} \boldsymbol{y}_i - \left(\beta_0 + \beta_1 \underline{\boldsymbol{x}}_{i1} \right) \right|, \left| \operatorname{mid} \boldsymbol{y}_i - \left(\beta_0 + \beta_1 \overline{\boldsymbol{x}}_{i1} \right) \right| \right\}. \quad (11)$$

При небольших m также можно выписать аналогичные общие формулы, выражающие расстояние от прямой до бруса через максимум из $2^m n$ чисел, где m — размерность пространства входных данных, а n — длина выборки.

В целом метод прямой интервальной аппроксимации (ПИА) для рассмотренного случая заключается в минимизации нормы вектора отклонений ("остатков"), т. е. некоторой функции $\Phi(\beta)$, агрегирующей отдельные расстояния (9) от брусов данных до прямой, и аргумент найденного минимума дает оценку параметров восстанавливаемой линейной функции (1). Ниже, для случая интервального задания независимых переменных рассмотрим один и, по-видимому, наиболее естественный вариант ее конструкции, когда берется чебышёвская норма (максимум-норма) вектора расстояний:

$$\Phi_\infty(\beta) = \max_{1 \le i \le n} \max_{x_{i:} \in \boldsymbol{x}_{i:}} |y(x_{i:}, \beta) \ominus \boldsymbol{y}_i|.$$

Развернутое представление Φ_∞ усложняется в сравнении с (7), но принципиально практически не изменяется: теперь оно является максимумом по всем угловым точкам бруса $\boldsymbol{x}_{i:}$ от выражений вида (7). Например, если воспользуемся (11), то можно выписать представление Φ_∞ для $m = 1$, т. е. для случая восстановления линейных функций одной переменной:

$$\Phi_\infty(\beta_0, \beta_1) = \max_{1 \le i \le n} \left\{ \operatorname{rad} \boldsymbol{y}_i + \left| \operatorname{mid} \boldsymbol{y}_i - \left(\beta_0 + \beta_1 \underline{\boldsymbol{x}}_{i1} \right) \right|, \operatorname{rad} \boldsymbol{y}_i + \left| \operatorname{mid} \boldsymbol{y}_i - \left(\beta_0 + \beta_1 \overline{\boldsymbol{x}}_{i1} \right) \right| \right\}.$$

Из этих представлений функции Φ_∞ для интервальных $\boldsymbol{x}_{i:}$ можно сделать вывод о том, что она выпуклая и полиэдральная, так как сконструирована из выражений того же вида, что и для точечных $x_{i:}$ с помощью операции взятия максимума по конечному множеству.

Далее для определения параметров β_0, β_1, ..., β_m необходимо найти безусловный минимум функции Φ_∞ по аргументам β_0, β_1, ..., β_m. Если применять для этого оптимизационные методы первого порядка, использующие кроме значений функции еще ее субградиенты (напомним, что Φ_∞ — негладкая), то для нахождения субградиента (см. [12]) необходимо определять аргумент, на котором достигается

$$\max_{x_{i:} \in \boldsymbol{x}_{i:}} \left| \operatorname{mid} \boldsymbol{y}_i - \left(\beta_0 + \sum_{j=1}^{m} \beta_j x_{ij} \right) \right|.$$

Этот аргумент может быть своим для каждого отдельного вектора $(\beta_0, \beta_1, \ldots, \beta_m)$. Напомним, что для выпуклых функций субградиенты совпадают с градиентами в тех точках, где эти градиенты существуют. Для выпуклой полиэдральной функции, которая составлена из конечного числа "кусков линейных функций", градиенты заведомо

существуют на некотором конечном множестве областей D_i, $D_i \subset \mathbb{R}^{m+1}$, $i = 1, 2, \ldots, I$ (где I — какое-то натуральное число), таких что

$$\bigcup_{i \in I} D_i = \mathbb{R}^{m+1}.$$

Представленная в этом разделе версия метода ПИА также удовлетворяет принципу соответствия для методов обработки интервальных данных, который сформулирован в книге [8]. При стягивании брусов неопределенности в точки получим в пределе метод восстановления зависимости по точечным данным, минимизирующий чебышёвскую норму вектора остатков наблюдений, т. е. популярное чебышёвское сглаживание данных.

5. Реализация

Для реализации представленного выше метода прямой интервальной аппроксимации требуется численное решение задачи безусловной выпуклой оптимизации с негладкой целевой функцией. Это хорошо развитое направление вычислительной оптимизации, в котором предложено немало эффективных методов и значительная часть из них опирается на использование градиентов или субградиентов целевой функции.

Для вычисления минимума функции Φ_∞ — чебышёвской нормы вектора отклонений от выборки интервальных данных — авторами реализована программа `sapprindat` для систем компьютерной математики Octave и MATLAB, которая в настоящее время свободно доступна на веб-сайте "Интервальный анализ и его приложения" [14]. В качестве "движка" в этой программе использован код `ralgb5`, созданный П.И. Стецюком (Институт кибернетики НАН Украины, Киев) и реализующий так называемый r-алгоритм — метод субградиентного спуска с растяжениями пространства [15, 16]. Отметим, что тот же самый движок использован в популярной программе `tolsolvty` [17], предназначенной для нахождения максимума распознающего функционала допускового множества решений для интервальных линейных систем. Она реализует сильную версию метода максимума совместности для оценивания параметров линейной функциональной зависимости по интервальным данным [18, 19].

В таблице приведен текст процедуры-функции на языке системы компьютерной математики Octave, которая вычисляет значение целевой функции Φ_∞, т. е. расстояние от интервальной выборки до графика восстанавливаемой функции, а также ее субградиент. Эта реализация существенно опирается на функцию `glpk`, стандартную свободно распространяемую функцию, предназначенную для решения общей задачи линейного программирования, которая входит в систему Octave и другие библиотеки, распространяемые по лицензии GPL. С ее помощью вычисляются значения внутренних максимумов в выражении (10), как это описано в предыдущем разделе работы. Предполагается, что входные интервальные данные — это интервальная $n \times m$-матрица $\boldsymbol{X}$ и интервальный n-вектор $\boldsymbol{y}$, которые задаются парами матриц и парами векторов нижних и верхних концов. Более точно, в головной программе `sapprindat` создаются матрицы `infX` и `supX` и векторы `infy` и `supy` тех же размеров, что $\boldsymbol{X}$ и $\boldsymbol{b}$ соответственно, такие что

$$\texttt{infX} = \underline{\boldsymbol{A}}, \qquad \texttt{supX} = \overline{\boldsymbol{A}}, \qquad \texttt{infy} = \underline{\boldsymbol{y}}, \qquad \texttt{supy} = \overline{\boldsymbol{y}}.$$

Листинг процедуры вычисления расстояния от графика линейной функции
до интервальной выборки и его субградиента
Listing of the procedure for calculating the distance from the graph of a linear function
to an interval sample and its subgradient

```
bc = 0.5 * (infb + supb)
br = 0.5 * (supb - infb)

function [f,g] = calcfg(x)
%
%   функция, которая вычисляет значение f  минимизируемой
%   чебышёвской нормы вектора отклонений и ее субградиент g
%
    A_opt = zeros(length(bc), length(x));
%
%   для каждого  интервального наблюдения i  с  помощью  стандартной  функции  glpk
%   из  системы  Octave  решаем  задачу  линейного  программирования  по  отысканию
%   матрицы A_opt, максимизирующей выражение |bc(i) - (A(i,:)x)|,  где A ограничена
%   условиями infA <= A <= supA
%
    for i = 1:length(bc)
        %
        %   подготовка условий для задачи линейного программирования
        %
        matrix_of_conditions =[ eye(length(x)); eye(length(x)) ];
        vector_of_conditions = [ infA(i , :), supA(i, :) ]';

        ctype = "";
        vartype = "";
        for j = 1:length(x)
            vartype = strcat(vartype , "C");
            ctype = strcat("L", ctype, "U");
        end
        %
        %   находим  максимум и минимум  произведения  A(i,:)*x,
        %   затем выбираем  максимизирующее |bc(i) - (A(i,:)*x)|
        %
        sense = -1;
        [a_max, f_max, status] = glpk (x, matrix_of_conditions,
                                  vector_of_conditions, [], [], ctype, vartype, sense);
        sense = 1;
        [a_min, f_min, status] = glpk (x, matrix_of_conditions,
                                  vector_of_conditions, [], [], ctype, vartype, sense);

        if (bc(i) - f_min >= f_max - bc(i))
            A_opt(i, :) = a_min;
        else
            A_opt(i, :) = a_max;
        end
    end

%   вычисление функции calcfg
    [f, index]   = max(br + abs(bc - A_opt * x));

%   вычисление субградиента для calcfg
    g = A_opt(index,:)' * sign(A_opt(index,:)*x - bc(index));

endfunction
```

6. Вычислительные эксперименты

Пример 2. Рассмотрим восстановление линейной зависимости вида

$$y = \beta_1 x_1 + \beta_0 \tag{12}$$

по данным

x	1	2	3
y	$[1, 2.5]$	$[2, 3]$	$[1.5, 2]$

. (13)

Они изображены на рис. 7 в виде вертикальных отрезков неопределенности данных [8]. Применим сначала для решения задачи традиционные интервальные методы, которые опираются на допущение, что интервалы значений функции — накрывающие.

Нетрудно видеть, что через отрезки неопределенности данных на рис. 7 можно провести прямую линию с неположительным угловым коэффициентом. Иными словами, множество параметров (β_0, β_1) линейных функций (12), совместных с данными (13), непусто. Оно называется *информационным множеством* задачи восстановления зависимости [8], и его можно нарисовать, например, с помощью пакета `IntLinIncR2` [20]. Результат визуализации представлен на рис. 8.

С помощью метода максимума совместности [18, 19, 21] можно получить точечную оценку параметров "наиболее подходящей" линейной функции, которая наилучшим образом совместна с данными (13). Она находится как аргумент максимума специального "распознающего функционала", дающего количественную меру совместности, и на практике для этой цели можно применить, к примеру, известную программу `tolsolvty` для какой-нибудь системы компьютерной математики — Octave, MATLAB и т. п. [17]. Результатом расчетов является точка максимума $(2.625, -0.25)$, так что искомая функция задается выражением

$$y = -0.25x + 2.625 \tag{14}$$

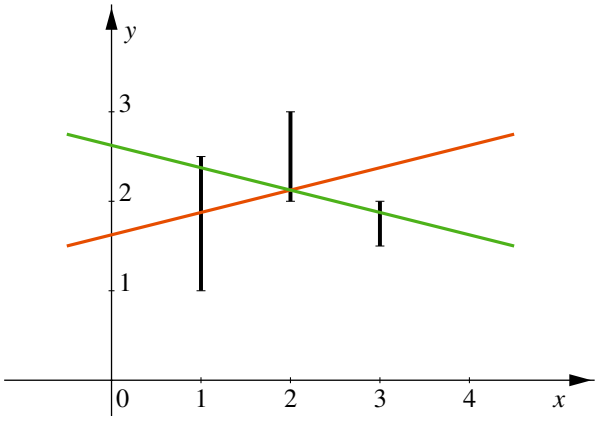

Рис. 7. Восстановление линейной зависимости по ненакрывающей интервальной выборке (13)
Fig. 7. Constructing a linear functional dependency from non-enclosing interval sample (13)

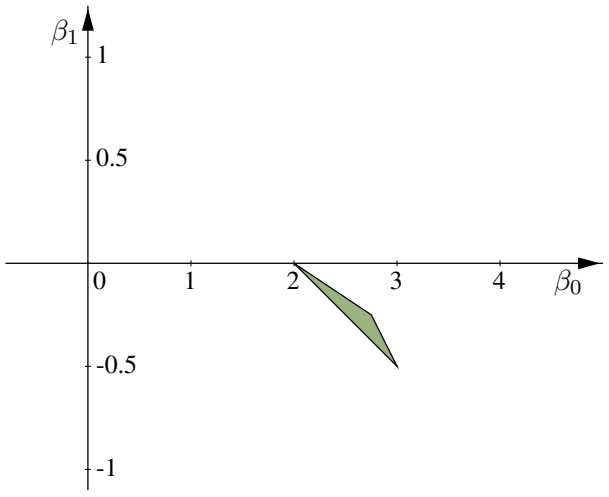

Рис. 8. Информационное множество задачи восстановления зависимости в примере 2 и точка параметров, полученная методом ПИА
Fig. 8. Information set of the line fitting problem in example 2 and the parameter point obtained by the DIA method

(зеленая линия на рис. 7), ее график проходит через все интервалы неопределенности. Но метод ПИА с чебышёвской нормой выдает в качестве наилучшей, с его точки зрения, линейную функцию

$$y = 0.25x + 1.625, \tag{15}$$

которая резко отличается от (14) (красная линия на рис. 7). Задающие ее параметры $(\beta_0, \beta_1) = (1.625, 0.25)$, как можно видеть из рис. 8, не лежат в информационном множестве. ■

Различие в результатах методов максимума совместности и прямой интервальной аппроксимации кажется разительным и непонятным. Особенно шокирует факт грубого игнорирования построенной линейной функцией (15) коридора совместных зависимостей для задачи с данными (13).

В действительности все это вполне объясняется тем принципиальным фактом, что при прямой интервальной аппроксимации данных совершенно игнорируется привычный смысл интервалов, который присущ традиционному интервальному анализу данных и подразумевает, что обрабатываемые интервалы являются вместилищами для истинных значений величин. Теперь это просто какие-то брусы, "болванки" (или что-то аналогичное) без какого-либо дополнительного смысла, между которыми наилучшим образом нужно провести график восстанавливаемой зависимости, и никаких других данных для решения задачи у нас нет.

Более того, метод прямой интервальной аппроксимации данных с чебышёвской метрикой в приведенном выше примере честно построил прямую линию наилучшего приближения, которая отстоит от каждого из отрезков неопределенности на расстояние 0.875, и это наименьшая возможная величина в данном случае (выполнены условия чебышёвского альтернанса [22]). Но цена отказа от свойства накрытия выборки оказывается очень чувствительной, что мы увидим ниже.

7. Сравнение с символьным анализом данных

Идея применения интервалов для представления данных измерений и наблюдений имеет давнюю историю, и часто она реализуется весьма различными способами. В частности, в символьном анализе данных — научном направлении, которое активно развивается на Западе с конца 90-х гг. прошлого века [23], интервалы рассматриваются не как двусторонние оценки диапазонов возможных значений измеряемых величин, а просто как какие-то размазанные значения, среди которых не обязательно присутствуют истинные. Иными словами, в символьном анализе данных интервальные результаты измерений и наблюдений являются ненакрывающими именно в том смысле, который мы обсуждали выше в разд. 1. Таким образом, эти методы символьного анализа данных имеют то же целевое назначение, что и представленный выше метод ПИА. Естественно сравнить результаты их применения. Ниже мы слегка коснемся этого обширного вопроса, опираясь на обзор [24], представляющий наиболее значимые методы символьного анализа данных.

Основная идея методов символьного анализа данных в применении к интервалам состоит в том, чтобы использовать методы традиционного регрессионного анализа для восстановления зависимостей по характерным точкам этих интервалов данных или же по каким-то их характеристикам (центрам, радиусам, нижним и верхним границам). Далее на основе полученных результатов теми или иными способами строится предсказательная модель.

“Метод центров”, предложенный в символьном анализе данных в 2000 г., заключался просто в применении метода наименьших квадратов к центрам интервалов данных. В дальнейшем в рамках символьного анализа данных предложено немало других методик, обзор которых можно увидеть, к примеру, в работе [24]. На рис. 9 наглядно показаны результаты, которые дают для данных (13) так называемые метод минимумов и максимумов (min-max method) и ограниченный метод центров и диапазонов (constrained center and range method).

Важной конструкцией, возникающей в связи с задачей восстановления зависимостей, является так называемый *прогнозный коридор*, который описывает возможную неопределенность предсказания значений функциональной зависимости [8]. В разных подходах к обработке данных этот коридор строится по-разному. В частности, в методах восстановления зависимостей по накрывающим интервальным данным прогнозный коридор часто берется в виде *коридора совместных зависимостей* [8], если информационное множество задачи непусто. Иными словами, прогнозный коридор получается объединением графиков всех функциональных зависимостей из заданного параметрического семейства, совместных (согласующихся и т. п.) с интервальными данными.

В методах символьного анализа данных ситуация с прогнозным коридором неоднозначная. Единого подхода к его построению, который был бы присущ всем этим методам, не существует, и в каждом отдельном методе он строится по-своему. Например, в методе минимумов и максимумов отдельно решаются две задачи восстановления зависимостей — по нижним и верхним границам исходных данных соответственно. В результате их решения находятся наборы коэффициентов $\underline{\beta}_0, \ldots, \underline{\beta}_m$ и $\overline{\beta}_0, \ldots, \overline{\beta}_m$, по которым строятся зависимости вида

$$y = \underline{\beta}_0 + \underline{\beta}_1 x_1 + \ldots + \underline{\beta}_m x_m,$$
$$y = \overline{\beta}_0 + \overline{\beta}_1 x_1 + \ldots + \overline{\beta}_m x_m.$$

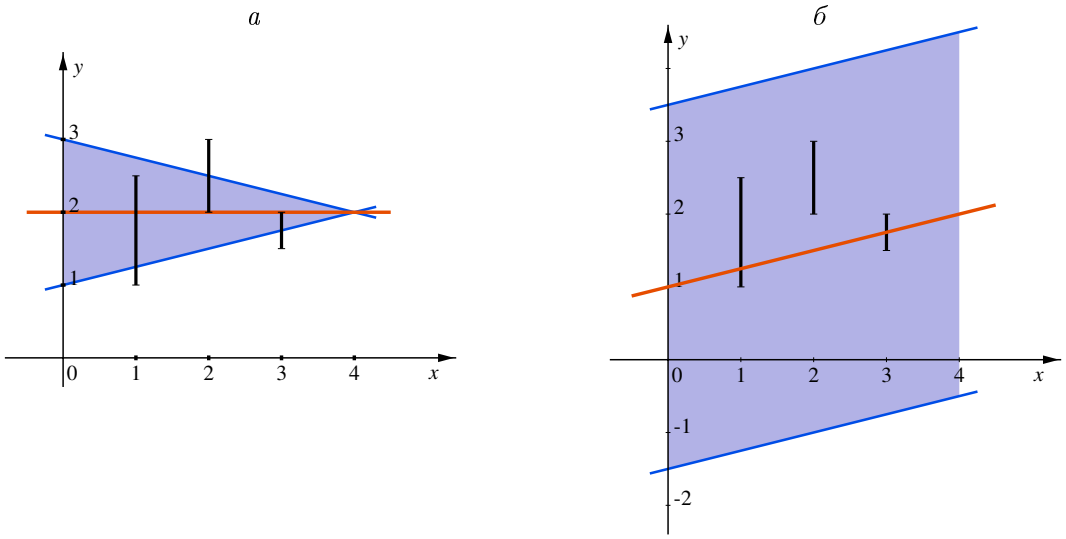

Рис. 9. Восстановление линейной зависимости различными методами символьного анализа данных по выборке (13) и соответствующие прогнозные коридоры: *а* — метод минимумов и максимумов; *б* — ограниченный метод центров и диапазонов

Fig. 9. Constructing a linear functional dependency by various methods of Symbolic Data Analysis from the sample (13), and the corresponding forecast corridors: *а* — min-max method; *б* — constrained center and range method

Тогда в качестве интервала неопределенности предсказания (а для точечных данных получаем сечение прогнозного коридора [8]) для входного $\boldsymbol{x} = [\underline{\boldsymbol{x}}, \overline{\boldsymbol{x}}]$, т. е. выбирается интервал неопределенности $\boldsymbol{y} = [\underline{\boldsymbol{y}}, \overline{\boldsymbol{y}}]$, в котором

$$\underline{\boldsymbol{y}} = \min\bigl\{\underline{\beta}_0 + \underline{\beta}_1\underline{\boldsymbol{x}}_1 + \ldots + \underline{\beta}_m\underline{\boldsymbol{x}}_m, \underline{\beta}_0 + \underline{\beta}_1\overline{\boldsymbol{x}}_1 + \ldots + \underline{\beta}_m\overline{\boldsymbol{x}}_m\bigr\},$$
$$\overline{\boldsymbol{y}} = \max\bigl\{\overline{\beta}_0 + \overline{\beta}_1\underline{\boldsymbol{x}}_1 + \ldots + \overline{\beta}_m\underline{\boldsymbol{x}}_m, \overline{\beta}_0 + \overline{\beta}_1\overline{\boldsymbol{x}}_1 + \ldots + \overline{\beta}_m\overline{\boldsymbol{x}}_m\bigr\}.$$

В ограниченном методе центров и диапазонов ищутся два набора коэффициентов — $\beta_0^c, \ldots, \beta_m^c$ и $\beta_0^r, \ldots, \beta_m^r$, решающих задачи восстановления зависимостей вида

$$\operatorname{mid}\boldsymbol{y} = \beta_0^c + \beta_1^c\operatorname{mid}\boldsymbol{x}_1 + \ldots + \beta_m^c\operatorname{mid}\boldsymbol{x}_m,$$
$$\operatorname{rad}\boldsymbol{y} = \beta_0^r + \beta_1^r\operatorname{rad}\boldsymbol{x}_1 + \ldots + \beta_m^r\operatorname{rad}\boldsymbol{x}_m$$

при дополнительном условии $\beta_j^r \geq 0$, $j = 1, 2, \ldots, m$. Тогда в качестве интервала неопределенности предсказания берется такой интервал $\boldsymbol{y} = [\underline{\boldsymbol{y}}, \overline{\boldsymbol{y}}]$, что

$$\underline{\boldsymbol{y}} = (\beta_0^c + \beta_1^c\operatorname{mid}\boldsymbol{x}_1 + \ldots + \beta_m^c\operatorname{mid}\boldsymbol{x}_m) - (\beta_0^r + \beta_1^r\operatorname{rad}\boldsymbol{x}_1 + \ldots + \beta_m^r\operatorname{rad}\boldsymbol{x}_m),$$
$$\overline{\boldsymbol{y}} = (\beta_0^c + \beta_1^c\operatorname{mid}\boldsymbol{x}_1 + \ldots + \beta_m^c\operatorname{mid}\boldsymbol{x}_m) + (\beta_0^r + \beta_1^r\operatorname{rad}\boldsymbol{x}_1 + \ldots + \beta_m^r\operatorname{rad}\boldsymbol{x}_m).$$

В случае точечных входных данных получаем

$$\underline{\boldsymbol{y}} = \beta_0^c + \beta_1^c x_1 + \ldots + \beta_m^c x_m - \beta_0^r,$$
$$\underline{\boldsymbol{y}} = \beta_0^c + \beta_1^c x_1 + \ldots + \beta_m^c x_m + \beta_0^r.$$

На рис. 9 видно, что прогнозный коридор в обоих случаях ограничен сверху и снизу прямыми линиями, что делает его применение весьма ограниченным. В частности, для метода минимумов и максимумов этот коридор вырождается в точку при значениях аргумента больше 4, а дальше становится бессмысленным, поскольку его нижняя и верхняя границы меняются местами. Для ограниченного метода центров и диапазонов прогнозный коридор вообще непомерно широк и по этой причине малоинформативен.

Как уже отмечалось, метод ПИА удовлетворяет “принципу соответствия”: он дает корректные и осмысленные результаты оценивания при неограниченном уменьшении ширины интервальных данных, т. е. стягивании интервалов данных в точки. Как следствие, для метода прямой интервальной аппроксимации можно построить прогнозный коридор в виде объединения графиков линейных функций, решающих всевозможные точечные задачи восстановления зависимостей для точечных данных из обрабатываемых интервалов [8]. Для построения такого множества пока не создано эффективных численных алгоритмов, и поэтому мы воспользуемся перебором на достаточно мелкой сетке в интервалах данных. Иными словми, чтобы построить приближенный вид прогнозного коридора, разобьем интервалы равномерной сеткой на достаточно представительные конечные множества точек и построим для них линейные функции, решающие задачи восстановления зависимостей. Результаты такого построения приведены на рис. 10.

Заметим, что построенный таким образом прогнозный коридор лишен недостатков прогнозных коридоров символьного анализа данных, которые представлены на рис. 9. Он не вырождается в точку, и на интервале заданных значений аргументов он значительно у́же, чем прогнозный коридор ограниченного метода центров и диапазонов. К тому же прогнозный коридор прямой интервальной аппроксимации расширяется по мере

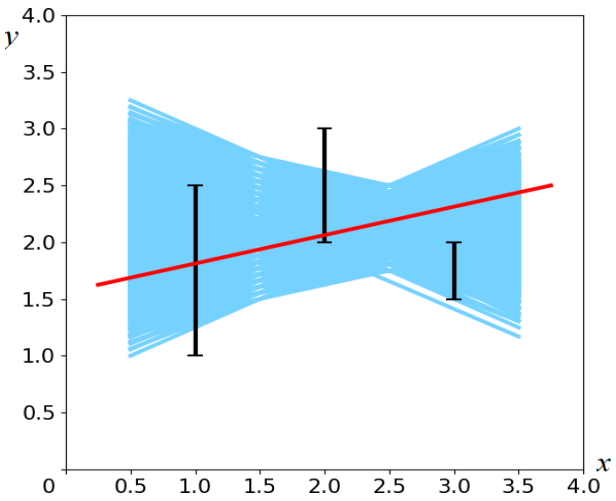

Рис. 10. Прогнозный коридор при решении задачи из примера 2 методом прямой интервальной аппроксимации

Fig. 10. Forecast corridor for the solution of the problem from example 2 by direct interval approximation method

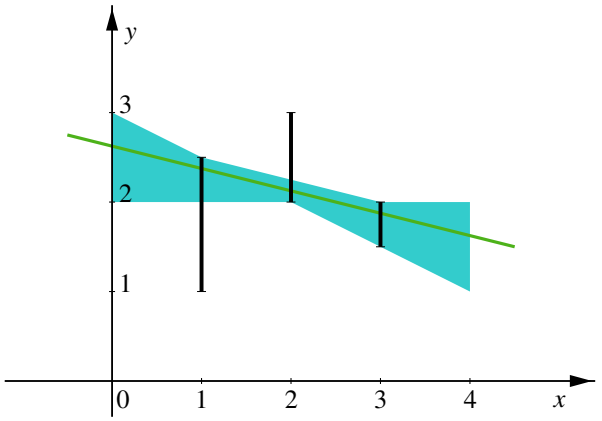

Рис. 11. Коридор совместных зависимостей для примера 2 при накрывающих интервалах данных

Fig. 11. Corridor of compatible functional dependencies for example 2 with enclosing data intervals

удаления от заданных значений аргументов, что вполне естественно: по мере удаления от тех данных, по которым построена функция, увеличивается погрешность прогнозов.

Наконец, рассмотрим на рис. 11 коридор совместных зависимостей для задачи из примера 2 и для области определения аргумента от 0 до 4. Он образован всевозможными линейными функциями, графики которых проходят через отрезки неопределенности данных из рис. 7 (которые считаются накрывающими). Этот коридор существенно у́же прогнозного коридора на рис. 10, что зримо демонстрирует ценность свойства накрытия интервальными данными истинных значений измеряемых величин.

По поводу прогнозного коридора на рис. 10 необходимо отметить, что он все-таки дает довольно большую неопределенность предсказания, так как угловой коэффициент прямых из этого коридора переходит через нуль и не имеет определенного знака. Применяя терминологию книги [8], можно сказать, что вариабельность оценок параметров восстанавливаемой зависимости здесь велика. В некоторых практических задачах такие оценки следует признать неинформативными. Но вот в коридоре совместных зависимостей (см. рис. 11), построенном по накрывающим данным, все прямые имеют один и тот же характер монотонности. Мы опять убеждаемся в полезности свойства накрытия истинных значений интервалами данных, которое позволяет строить более точные оценки и прогнозы.

В заключение работы рассмотрим еще один характерный пример.

Пример 3. Рассмотрим восстановление линейной зависимости вида

$$y = \beta_1 x + \beta_0$$

по данным

x	1	2	3
y	$[1, 2.5]$	$[2, 3]$	$[3, 3.5]$

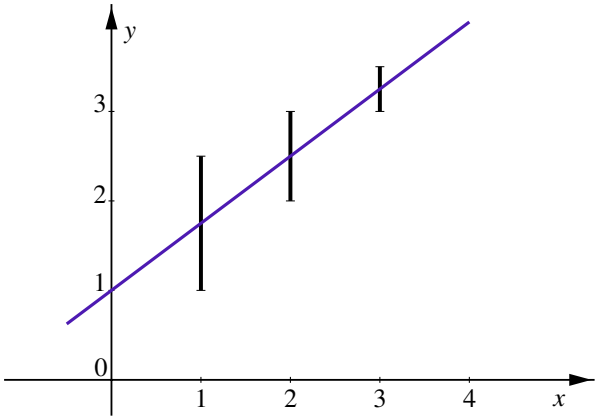

Рис. 12. Решение задачи восстановления зависимостей для примера 3 с помощью метода ПИА и метода максимума совместности
Fig. 12. Solving the line fitting problem from example 3 using the DIA method and the maximum compatibility method

Они изображены на рис. 12 в виде отрезков неопределенности. Применение как метода максимума совместности, так и метода ПИА дает одинаковые результаты — линейную функцию

$$y = 0.75x + 1,$$

график которой построен на том же рис. 12 синей линией.

Совершенно аналогичная картина — близость результатов, полученных методом ПИА и методами восстановления зависимостей для накрывающих интервальных данных, наблюдается при обработке любых интервальных выборок, в которых отслеживается общая четкая тенденция в данных. ■

Список литературы

[1] **Вапник В.Н.** Восстановление зависимостей по эмпирическим данным. М.: Наука; 1979: 448.

[2] **Kearfott B., Nakao M., Neumaier A., Rump S., Shary S.P., van Hentenryck P.** Standardized notation in interval analysis. Computational Technologies. 2010; 15(1):7–13.

[3] **Бурдун М.Д., Марков Б.Н.** Основы метрологии. М.: Издательство стандартов; 1985: 256.

[4] РМГ 83-2007. Государственная система обеспечения единства измерений. Шкалы измерений. Термины и определения. Рекомендации по межгосударственной стандартизации В№ 83-2007: 19. М.: Стандартинформ; 2008: 19.

[5] РМГ 29-2013. Метрология. Основные термины и определения. Рекомендации по межгосударственной стандартизации 29-2013. М.: Стандартинформ; 2014: 56.

[6] Evaluation of measurement data — guide to the expression of uncertainty in measurement. JCGM, 2008. Available at: `https://www.bipm.org/documents/20126/2071204/JCGM_100_2008_E.pdf`.

[7] **Shary S.P.** Enclosing vs. non-enclosing measurements in interval data processing. A Presentation at International Online Seminar on Interval Methods in Control Engineering. January 14, 2022. DOI:10.13140/RG.2.2.34844.62087.

[8] **Баженов А.Н., Жилин С.И., Кумков С.И., Шарый С.П.** Обработка и анализ интервальных данных. Ижевск; М.: Издательство "ИКИ"; 2024: 355.

[9] **Канторович Л.В.** О некоторых новых подходах к вычислительным методам и обработке наблюдений. Сибирский математический журнал. 1962; 3(5):701–709.

[10] **Дрейпер Н., Смит Г.** Прикладной регрессионный анализ. М.: "Диалектика"; 2017: 912.

[11] **Шарый С.П.** Конечномерный интервальный анализ. Новосибирск: XYZ; 2024: 671. Адрес доступа: `http://www.nsc.ru/interval/Library/InteBooks/SharyBook.pdf`.

[12] **Пшеничный Б.Н.** Выпуклый анализ и экстремальные задачи. М.: Наука; 1980: 320.

[13] **Рокафеллар Р.** Выпуклый анализ. М.: Мир; 1973: 469.

[14] sapprindat. Адрес доступа: `http://www.nsc.ru/interval/Programing/OctCodes/sapprindat.m`.

[15] **Шор Н.З., Журбенко Н.Г.** Метод минимизации, использующий операцию растяжения пространства в направлении разности двух последовательных градиентов. Кибернетика. 1971; (3):51–59.

[16] **Стецюк П.И.** Субградиентные методы ralgb5 и ralgb4 для минимизации овражных выпуклых функций. Вычислительные технологии. 2017; 22(2):127–149.

[17] tolsolvty. Адрес доступа: `http://www.nsc.ru/interval/Programing/OctCodes/tolsolvty.m`.

[18] **Шарый С.П.** Сильная согласованность в задаче восстановления зависимостей при интервальной неопределенности данных. Вычислительные технологии. 2017; 22(2):150–172.

[19] **Shary S.P.** Weak and strong compatibility in data fitting problems under interval uncertainty. Advances in Data Science and Adaptive Analysis. 2020; 12(1):2050002.

[20] **Шарая И.А.** IntLinIncR2. Адрес доступа: `http://www.nsc.ru/interval/sharaya/irash.html#prog`.

[21] **Шарый С.П.** Метод максимума согласования для восстановления зависимостей по данным с интервальной неопределенностью. Известия Академии наук. Теория и системы управления. 2017; (6):3–19.

[22] **Бахвалов Н.С., Жидков Н.П., Кобельков Г.М.** Численные методы. М.: Изд-во "Лаборатория знаний"; 2020: 636.

[23] **Billard L., Diday E.** Symbolic data analysis. Conceptual statistics and data mining. Chichester: John Wiley & Sons; 2007: 328.

[24] **Kabir S., Wagner Ch., Ellerby Z.** Towards handling uncertainty-at-source in AI — a review and next steps for interval regression. IEEE Transactions on Artificial Intelligence. 2024; 5(1):3–22. DOI:10.1109/TAI.2023.3234930.

Вычислительные технологии, 2024, том 29, № 4, с. 71–94. © ФИЦ ИВТ, 2024 ISSN 1560-7534
Computational Technologies, 2024, vol. 29, no. 4, pp. 71–94. © FRC ICT, 2024 eISSN 2313-691X

COMPUTATIONAL TECHNOLOGIES

DOI:10.25743/ICT.2024.29.4.006

On constructing functional dependencies from non-enclosing interval data

S. P. Shary[1,*], M. A. Zvyagin[2]

[1]Federal Research Center for Information and Computational Technologies, 630090, Novosibirsk, Russia
[2]Novosibirsk State University, 630090, Novosibirsk, Russia
*Corresponding author: Sergey P. Shary, e-mail: shary@ict.nsc.ru
Received July 07, 2022, revised March 25, 2024, accepted April 01, 2024.

Abstract

The purpose of the paper is to present a simple and natural approach to reconstructing linear functional dependencies from non-enclosing data with interval uncertainty. It denotes interval data that is not guaranteed to contain the true values of the measured quantities, and therefore must be processed significantly differently than interval data that is certain to contain true values (enclosing). From the very definition of non-enclosing interval data it follows that they should be considered, rather, as integral objects without any internal structure, since it does not make sense for their point elements to require satisfaction of two-sided interval constraints, etc.

For this reason, the construction of functional dependencies from non-enclosing interval data should be performed on the basis of approaches that find the best approximation of the intervals under consideration without resorting to their internal content. This can be done, for example, using the approximation theory. In the present study, solving the line fitting problem is reduced to finding the minimum deviation of the graph of the constructed function from the interval data boxes.

The properties of the deviation functional for the most popular vector norms, which can be used to determine the distance between points, are investigated. It is shown that, under some conditions on the norm, the deviation functional is a convex polyhedral function. Its minimum can be efficiently found using existing non-smooth optimization methods. In particular, the paper presents a free program implemented by the authors for computing this minimum.

In conclusion, the work provides numerical examples demonstrating the behavior of the new technique in various situations, as well as its comparison with methods for solving the problem of line fitting from enclosing interval data. Finally, correlations with methods of Symbolic Data Analysis are discussed in detail.

Keywords: interval, interval data analysis, data fitting problem, enclosing measurements, non-enclosing measurements, method of direct interval approximation.

Citation: Shary S.P., Zvyagin M.A. On constructing functional dependencies from non-enclosing interval data. Computational Technologies. 2024; 29(4):71–94. DOI:10.25743/ICT.2024.29.4.006. (In Russ.)

References

1. **Vapnik V.N.** Estimation of dependencies based on empirical data. N.Y.: Springer; 1982: 400.
2. **Kearfott B., Nakao M., Neumaier A., Rump S., Shary S.P., van Hentenryck P.** Standardized notation in interval analysis. Computational Technologies. 2010; 15(1):7–13.
3. **Burdun M.D., Markov B.N.** Osnovy metrologii [Fundamentals of metrology]. Moscow: Izdatel'stvo Standartov; 1985: 256. (In Russ.)
4. RMG 83-2007. Gosudarstvennaya sistema obespecheniya edinstva izmereniy. Shkaly izmereniy. Terminy i opredeleniya. Rekomendatsii po mezhgosudarstvennoy standartizatsii [State system for ensuring the uniformity of measurements. Measurement scales. Terms and Definitions. Recommendations on interstate standardization]. No. 83-2007. Moscow: Standardinform; 2008: 19. (In Russ.)

5. RMG 29-2013. Metrologiya. Osnovnye terminy i opredeleniya. Rekomendatsii po mezhgosudarstvennoy standartizatsii [Metrology. Basic terms and definitions. Recommendations on interstate standardization]. No. 29-2013. Moscow: Standardinform; 2014: 56. (In Russ.)
6. Evaluation of measurement data — guide to the expression of uncertainty in measurement. JCGM, 2008. Available at: `https://www.bipm.org/documents/20126/2071204/JCGM_100_2008_E.pdf`.
7. **Shary S.P.** Enclosing vs. non-enclosing measurements in interval data processing. A Presentation at International Online Seminar on Interval Methods in Control Engineering. January 14, 2022. DOI:10.13140/RG.2.2.34844.62087.
8. **Bazhenov A.N., Zhilin S.I., Kumkov S.I., Shary S.P.** Obrabotka i analiz interval'nykh dannykh [Processing and analysis of interval data]. Izhevsk; Moscow: Izdatelstvo "IKI"; 2024: 355. (In Russ.)
9. **Kantorovich L.V.** Some new approaches to computational methods and the handling of observations. Sibirskii Matematicheskii Zhurnal. 1962; 3(5):701–709. (In Russ.)
10. **Draper N.R., Smith H.** Applied regression analysis. N.Y.: John Wiley & Sons; 1998: 736. DOI:10.1002/9781118625590.
11. **Shary S.P.** Konechnomernyy interval'nyy analiz [Finite-dimensional interval analysis]. Novosibirsk: XYZ; 2024: 671. Available at: `http://www.nsc.ru/interval/Library/InteBooks/SharyBook.pdf`. (In Russ.)
12. **Pshenichnyy B.N.** Vypuklyy analiz i ekstremal'nye zadachi [Convex analysis and extremal problems]. Moscow: Nauka; 1980: 320. (In Russ.)
13. **Rockafellar R.T.** Convex analysis. Princeton, NJ: Princeton University Press; 1970: 472.
14. sapprindat. Available at: `http://www.nsc.ru/interval/Programing/OctCodes/sapprindat.m`.
15. **Shor N.Z., Zhurbenko N.G.** Minimization method that uses space dilation in the direction of the difference of two successive gradients. Cybernetics. 1971; (3):51–59. (In Russ.)
16. **Stetsyuk P.I.** Subgradient methods ralgb5 and ralgb4 for minimization of ravine-like convex functions. Computational Technologies. 2017; 22(2):127–149. (In Russ.)
17. tolsolvty. Available at: `http://www.nsc.ru/interval/Programing/OctCodes/tolsolvty.m`.
18. **Shary S.P.** Strong compatibility in data fitting problem under interval data uncertainty. Computational Technologies. 2017; 22(2):150–172. (In Russ.)
19. **Shary S.P.** Weak and strong compatibility in data fitting problems under interval uncertainty. Advances in Data Science and Adaptive Analysis. 2020; 12(1):2050002.
20. **Sharaya I.A.** IntLinIncR2. Available at: `http://www.nsc.ru/interval/sharaya/irash.html#prog`.
21. **Shary S.P.** Maximum compatibility method for data fitting under interval uncertainty. Journal of Computer and Systems Sciences International. 2017; 56(6):897–913.
22. **Bakhvalov N.S., Zhidkov N.P., Kobelkov G.M.** Chislennye metody [Numerical methods]. Moscow: Laboratoriya Znaniy; 2020: 636. (In Russ.)
23. **Billard L., Diday E.** Symbolic data analysis. Conceptual Statistics and Data Mining. Chichester: John Wiley & Sons; 2007: 328.
24. **Kabir S., Wagner Ch., Ellerby Z.** Towards handling uncertainty-at-source in AI — a review and next steps for interval regression. IEEE Transactions on Artificial Intelligence. 2024; 5(1):3–22. DOI:10.1109/TAI.2023.3234930.