УДК 512.64

Эллипсоидная аппроксимация множества решений интервальных систем линейных алгебраических уравнений

Е.К. Ергалиев[1], М.Н. Мадияров[1], Н.М. Оскорбин[2], Л.Л. Смолякова[2]

[1]Восточно-Казахстанский университет им. С. Аманжолова (Усть-Каменогорск, Казахстан)
[2]Алтайский государственный университет (Барнаул, Россия)

Ellipsoidal Approximation of the Set of Solutions of Interval Systems of Linear Algebraic Equations

E.R. Ergaliev[1], M.N. Madiyarov[1], N.M. Oskorbin[2], L.L. Smolyakova[2]

[1]Sarsen Amanzholov East Kazakhstan University (Ust-Kamenogorsk, Kazakhstan)
[2]Altai State University (Barnaul, Russia)

Представлены результаты аппроксимации множества решений интервальных систем линейных алгебраических уравнений (ИСЛАУ), которые используются в задачах моделирования линейных по параметрам детерминированных процессов. Предполагается, что моделируемый процесс описывается выходной переменной и совокупностью входных переменных, ошибки измерения которых по предположению задаются известными интервалами, симметричными относительно нулевого значения. Традиционно множества решений ИСЛАУ в прикладных задачах аппроксимируются внешне моделью бруса, минимального гиперпрямоугольника, стороны которого параллельны осям выбранной системы координат. В данной работе предлагается использовать эллипсоидную аппроксимацию этих множеств, которая является более эффективной. К основным результатам работы относятся обоснование предположений относительно свойств моделируемого процесса, выбор математического метода построения аппроксимирующего эллипсоида, предложенный способ формирования граничных точек и численный метод решения задачи. Выполнено компьютерное моделирование в среде Excel задачи оценки параметров линейного процесса, которое использовано для сравнительного исследования аппроксимаций решений ИСЛАУ брусом и эллипсом.

Ключевые слова: моделирование процессов, интервальные системы линейных алгебраических уравнений, объединенное множество решений, аппроксимация гиперпрямоугольником, эллипсоидная аппроксимация.

The article presents the results of the approximation of the set of solutions of interval systems of linear algebraic equations. These systems are used in the problems of modeling linear deterministic processes. It is assumed that the modeled process is described by an output variable and a set of input variables, the measurement errors of which are assumed to be set by known intervals symmetric with respect to the zero value. Traditionally, the sets of solutions of interval systems of linear algebraic equations in applied problems are approximated by a hyper-rectangular whose sides are parallel to the axes of the selected coordinate system. In this paper, we propose to use an ellipsoidal approximation of these sets, which is more efficient. The main results of the work include the substantiation of assumptions about the properties of the modeled process, the choice of a mathematical method for constructing an approximating ellipsoid, the proposed method for forming boundary points, and a numerical method for solving the problem. A computer simulation of the problem of estimating the parameters of a linear process is performed in Excel, which is used for a comparative study of approximations of solutions of interval systems of linear algebraic equations by a hyper-rectangular and an ellipse.

Key words: process modeling, interval systems of linear algebraic equations, united solution set, hyper-rectangular approximation, ellipse approximation.

DOI: 10.14258/izvasu(2021)4-15

Введение

В статье представлены результаты исследования проблемы аппроксимации множества решений интервальных систем линейных алгебраических уравнений (ИСЛАУ), которые используются в задачах моделирования линейных по параметрам детерминированных процессов [1–5]. Предполагается, что моделируемый процесс описывается выходной переменной и совокупностью входных переменных, ошибки измерения которых по предположению задаются известными интервалами, симметричными относительно нулевого значения.

При моделировании указанных линейных процессов используется теоретический подход, при котором исходные предположения относительно структуры модели и границы интервалов ошибок измерения всех переменных являются достоверными и не требуют проверки их выполнимости методами разведочного анализа. Указанная постановка задачи моделирования процессов позволяет сосредоточить внимание на исследовании задачи эллипсоидной аппроксимации информационных множеств, которое проводится на примере объединенного множества решений ИСЛАУ [2]. Моделирование процессов в условиях возможного нарушения исходных предположений метода анализа данных рассматривалось, например, в работах [6–8].

Традиционно множества решений ИСЛАУ в прикладных задачах аппроксимируются внешне моделью бруса, минимального гиперпрямоугольника, стороны которого параллельны осям выбранной системы координат. В данной статье предлагается использовать эллипсоидную аппроксимацию информационных множеств, которая существенно лучше характеризует область неопределенности значений параметров моделируемого процесса, и этот эффект возрастает с ростом размерности факторного пространства.

В общем случае исследование множеств решений ИСЛАУ встречает значительные сложности, в том числе и при построении его внешней или внутренней аппроксимации брусом. Так, в работе [2, с. 323] утверждается, что «вычисление для объединенного множества решений внешних покоординатных оценок с любой заданной абсолютной или относительной точностью есть NP-трудная задача». Однако в частном случае положительных компонент решения ИСЛАУ объединенное множество решений задается системой линейных неравенств [5, с. 112] и его исследование можно проводить методами линейного программирования (ЛП). Отметим, что использование этого подхода к анализу данных не ограничивает его общности в прикладном интервальном анализе данных. Это утверждение поясним так: во-первых, при отрицательном направлении влияния некоторых факторов искомой зависимости можно воспользоваться заменой переменных и, во-вторых, если интервальная оценка параметра одного или нескольких факторов содержит нулевое значение, то следует проверить гипотезу об их несущественном влиянии на выходную переменную процесса. Таким образом, эллипсоидная аппроксимация в данной работе рассматривается для ограниченного многогранного множества в пространстве R^n_+.

Задача погружения многогранника в эллипсоид минимального объема рассматривалась, например, в работах [9–10]. В них рассмотрены задачи построения экстремальных эллипсоидов в двух случаях: когда аппроксимируемое множество задано совокупностью точек, в том числе граничных, и когда множество задано системой линейных неравенств.

Нами проводится исследование эллипсоидной аппроксимации информационных множеств в задачах прикладного интервального анализа экспериментальных данных. Рассматриваются вопросы обоснования предположений относительно свойств моделируемого процесса, выбора математического метода построения аппроксимирующего эллипсоида, выбора способа формирования граничных точек и численного метода решения задачи. Выполнено компьютерное моделирование в среде Excel задачи оценки параметров линейного процесса, которое использовано для сравнительного исследования аппроксимаций решений ИСЛАУ брусом и эллипсоидом методами вычислительных экспериментов.

Информационное множество в задачах интервального анализа данных

Моделируемый процесс представляется структурно в виде «черного ящика», входные и выходная переменные которого измеряются с известными интервальными ошибками, а искомая математическая модель по предположению имеет следующий вид:

$$b = \mathrm{x}_1\mathrm{a}_1 + \cdots + \mathrm{x}_n\mathrm{a}_n. \qquad (1)$$

Следует пояснить используемые обозначения формулы (1) и ниже, которые отличаются от обозначений, применяемых при анализе данных, в пользу обозначений в линейном программировании. Входные переменные моделируемого процесса обозначаются как вектор $a = (a_1, \ldots , a_n)$, а b — выходная переменная. Тогда вектор x обозначает неизвестные коэффициенты линейной модели (1).

Прикладной интервальный анализ этих процессов проводится с использованием множеств решений ИСЛАУ, коэффициенты и правая часть которой записана по результатам интервальных наблюдений.

В матричной форме ИСЛАУ записывается интервальной $(N \times n)$ матрицей $\boldsymbol{A}$ коэффициентов и интервальным $(N \times 1)$ вектором правой части $\boldsymbol{B}$ в следующем виде: $\boldsymbol{A}x = \boldsymbol{B}$, где элементы матриц заданы интервалами: $A^H \le \boldsymbol{A} \le A^V$; $B^H \le \boldsymbol{B} \le B^V$. Размерности матриц определяются числом наблюдений N и числом n оцениваемых коэффициентов моделируемого процесса. Матрицы A^H, A^V и векторы B^H, B^V при анализе данных записываются с использованием результатов наблюдений и оценок интервальных ошибок [5].

В литературе [2, 4] задачи интервального анализа данных решаются с использованием трех базовых множеств решений ИСЛАУ: объединенное, допусковое и управляемое. Для рассматриваемых линейных процессов эти информационные множества можно задавать системами линейных неравенств и для оценки работоспособности моделей использовать их эллипсоидные аппроксимации. В данной работе ограничимся исследованием объединенного множества решений, которое в предикатной форме записывается следующим образом [2]:

$$\Xi_{uni} = \{x \in R^n : (\exists \boldsymbol{A} \in \boldsymbol{A})(\exists \boldsymbol{B} \in \boldsymbol{B})(Ax = B)\}. \qquad (2)$$

Это множество в пространстве R^n_+ задается системой линейных неравенств, которую запишем в следующем виде:

$$\Xi_{uni} = \{x \in R^n_+ : A^V x \ge B^H ;\ A^H x \le B^V\}. \qquad (3)$$

Рассмотрим прикладные задачи анализа данных, которые возникают при моделировании процессов.

1. Задача прогноза выходной переменной моделируемого процесса в заданной точке факторного пространства

Как следует из практики математического моделирования процессов при условии достаточно точных установкок входных переменных, в этом случае требуется оценить интервалы изменения выходной переменной с учетом полученного по базе данных и базе знаний информационного множества.

Запишем следующие задачи ЛП для интервальной оценки значения выходной переменной b в заданной точке $a_p \in R^n$ факторного пространства:

$$b^H(a_p) = \min\nolimits_{x \in \Xi_{uni}} a_p x;\ b^V(a_p) = \max\nolimits_{x \in \Xi_{uni}} a_p x. \qquad (4)$$

2. Интервальные оценки коэффициентов линейной зависимости моделируемого процесса

В прикладном интервальном анализе множество истинных значений вектора x задается информационным множеством, но для его визуализации на практике традиционно используется аппроксимация брусом. Данное представление можно рассматривать в качестве независимых интервальных оценок компонент вектора x. Для их вычислений достаточно решить $2n$ задач линейного программирования.

Например, x_1 принадлежит интервалу $[x_1^H, x_1^V]$, граничные точки которого имеют следующие выражения:

$$x_1^H = \min\nolimits_{x \in \Xi_{uni}} x_1;\ x_1^V = \max\nolimits_{x \in \Xi_{uni}} x_1. \qquad (5)$$

Заметим, что если полученный интервал содержит нулевое значение оценки исследуемого параметра, то можно ставить вопрос о существенности влияния соответствующего фактора процесса на выходную переменную.

3. Задача проверки принадлежности заданной точки информационному множеству

Точка x^d принадлежит множеству Ξ_{uni} тогда и только тогда, когда $\delta(x^d) = 0$, где $\delta(x^d)$ — решение следующей задачи квадратичного программирования:

$$\delta(x^d) = \min\nolimits_{x \in \Xi_{uni}} \left\| x^d - x \right\|. \qquad (6)$$

Задача 3 может решаться на практике для исследования свойств объединенного множества решений и для проверки выполнимости исходных предположений прикладного интервального анализа данных, в том числе для оценки значимости выбранных входных переменных.

Аппроксимация эллипсоидом объединенного множества решений ИСЛАУ

Рассмотрим математические задачи построения минимального эллипсоида по заданным граничным точкам множества Ξ_{uni}.

Искомый эллипсоид Ξ_ε будем задавать следующим образом:

$$\Xi_\varepsilon = \{x \in R^n : M(x - c), (x - c) \le 1\}. \qquad (7)$$

В формуле (7) записано произведение квадратной матрицы на вектор и скалярное произведение векторов, а сама формула задает внутреннюю часть эллипсоида вместе с границей; M — симметричная положительно-определенная матрица, c — центр эллипсоида в n-мерном пространстве.

Объем эллипсоида вычисляется по формуле:

$$vol(\Xi_\varepsilon) = \omega_n (detM)^{-1/2}. \qquad (8)$$

В формуле (8) ω_n — объем единичного шара в n-мерном пространстве; $detM$ — определитель матрицы M.

Задачу поиска эллипсоида минимального объема, содержащего все L заданных точек, запишем в следующем виде:

$$\begin{aligned} &detM \to max; \\ &M(x^j - c), (x^j - c) \le 1,\ j = 1, \ldots, L; \\ &M \text{ положительно определена}. \end{aligned} \qquad (9)$$

Математическая модель минимального эллипсоида предполагает в качестве искомых переменных элементы матрицы M и вектор c — центр эллипсоида. С учетом симметрии матрицы M общее число переменных в задаче (9) составляет $0{,}5n(n + 3)$. Число переменных с ростом размерности пространства резко увеличивается, и уже при n больше 10 решение задачи (9) средствами Excel становится невозможным. В данной работе это ограничение не является сдерживающим фактором проведения исследований. Мы ограничимся пространствами размерности до 5, на которых проведем исследование поставленных задач в среде Excel.

Компьютерное моделирование эллипсоидной аппроксимации объединенного множества решений ИСЛАУ

Компьютерная модель анализа данных в нашем случае разработана в одной книге Excel и включает на отдельных листах генерацию линейного процесса с заданными свойствами; задание интервальных ошибок наблюдений; формирование таблицы данных; задание условий объединенного множества решений и вычисление *L* координат его граничных точек; вычисление параметров требуемого эллипсоида. Генерация равномерно распределенных чисел на интервале [0, 1] в MS Excel осуществляется с помощью функции СЛЧИС().

Исследование касается выбора граничных точек для характеристики информационного множества. Среди многовариантных расчетов предпочтение отдано следующему: *2n* точек получено при решении задач (5); *2n(n-1)* получено при решении этих задач с фиксированной одной переменной на среднем уровне. Например, при 5-мерном пространстве общее число точек для представления объединенного множества решений получается равным 50. Задачи (5) при линейных ограничениях (3) решались симплекс-методом линейного программирования.

Решение задачи (9) проводилось с использованием инструмента «Поиск решения», который находит параметры эллипсоида встроенным методом математического программирования. Пример таблиц Excel при нахождении внешней эллипсоидной аппроксимации объединенного множества решений представлен на рисунке, где выбрано 18 точек (*L=18*) и вычислены оценки объемов бруса и эллипсоида. В данном случае относительное уменьшение объема бруса при эллипсоидной аппроксимации получилось равным 43,2 %.

Возможность и целесообразность использования эллипсоидной аппроксимации множеств решений ИСЛАУ подтверждена многовариантными компьютерными расчетами. Даже в случаях малых ошибок наблюдения входных переменных и в сравнимых по вариации влияниях факторов на выходную переменную сокращение объема бруса является существенным. В отдельных случаях модельных процессов указанное сокращение увеличивается до 90 %.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Эллипсоидная аппроксимация объединенного множества решений ИСЛАУ												
2		M=			Объем единичного шара=			4,189		Определитель=		47,422	
3	6,926	4,952	5,893		C1	c2	c3	M11	M12	M13	M22	M23	M33
4	4,952	6,590	6,723		0,964	0,926	1,028	6,926	4,952	5,893	6,590	6,723	9,326
5	5,893	6,723	9,326										
6					Объем Бруса=		1,07	Объем Эллипса=			0,61	W=	43,2%
7	Таблица крайних точек ОМР				Векторы (X-C)			Векторы M*(X-C)				Ограничения	
8	№ п/п	X1	X2	X3	X1-C1	X2-C2	X3-C3	1	2	3		<=	
9	1	0,40	1,07	1,23	-0,56	0,14	0,20	-2,03	-0,52	-0,52		0,9717	1
10	2	1,34	0,86	0,59	0,37	-0,06	-0,44	-0,33	-1,53	-2,33		1,0000	1
11	3	0,88	1,53	0,54	-0,08	0,60	-0,49	-0,48	0,26	-1,02		0,6954	1
12	4	1,13	0,23	1,47	0,17	-0,70	0,44	0,31	-0,80	0,41		0,7913	1
13	5	1,32	1,14	0,47	0,36	0,22	-0,56	0,23	-0,58	-1,68		0,8988	1
14	6	0,88	0,32	1,69	-0,09	-0,61	0,66	0,27	-0,01	1,54		1,0000	1
15	7	0,87	0,33	1,68	-0,09	-0,60	0,66	0,24	-0,02	1,52		0,9831	1
16	8	0,87	1,53	0,54	-0,09	0,60	-0,48	-0,53	0,24	-1,03		0,6916	1
17	9	1,29	0,38	1,08	0,32	-0,55	0,05	-0,16	-1,66	-1,29		0,7871	1
18	10	0,42	1,22	1,08	-0,54	0,30	0,05	-2,00	-0,40	-0,74		0,9348	1
19	11	0,42	0,88	1,40	-0,55	-0,05	0,37	-1,83	-0,52	-0,07		1,0000	1
20	12	1,33	0,88	0,58	0,37	-0,05	-0,45	-0,29	-1,47	-2,29		0,9822	1
21	13	1,31	0,44	0,94	0,35	-0,49	-0,09	-0,49	-2,05	-2,02		1,0000	1
22	14	0,92	0,44	1,63	-0,05	-0,49	0,60	0,83	0,62	2,08		0,9140	1
23	15	0,43	1,30	1,00	-0,53	0,37	-0,02	-1,99	-0,34	-0,85		0,9562	1
24	16	1,14	1,30	0,50	0,18	0,37	-0,53	-0,06	-0,24	-1,41		0,6503	1
25	17	1,20	1,30	0,83	0,23	0,37	-0,20	2,32	2,31	2,07		1,0000	1
26	18	0,50	0,76	1,48	-0,46	-0,17	0,45	-1,37	-0,35	0,38		0,8620	1

Пример эллипсоидной аппроксимации информационного множества при n=3

В процессе проведенного исследования затронуты прикладные проблемы интервального анализа, связанные с необходимостью визуализации информационных множеств и их сечений. Актуальными являются проблемы получения зависимых интервальных оценок параметров (зависимых интервалов оценок истинных значений коэффициентов моделируемого процесса). Полученные результаты позволяют уточнить методические подходы применения теоретических результатов ИСЛАУ в прикладных задачах анализа данных и математического моделирования реальных процессов.

Заключение

В данной работе проведено изучение возможностей эллипсоидной аппроксимации объединенного множества решений интервальной системы линейных алгебраических уравнений (ИСЛАУ) в прикладном интервальном анализе. При исследовании рассмотрены экспериментальные данные с интервальными ошибками наблюдений входных и выходной переменной линейного процесса.

Для модельных процессов без внутренних шумов в условиях правильных наблюдений методами вычислительного эксперимента проведено сравнение брусовой и эллипсоидной аппроксимации объединенного множества решений ИСЛАУ.

Полученные результаты позволяют уточнить методические подходы применения теоретических исследований ИСЛАУ в задачах анализа данных и использовать на практике методы и программную реализацию эллипсоидной аппроксимации множества неопределенности при математическом моделировании реальных процессов.

Библиографический список

1. Канторович Л.В. О некоторых новых подходах к вычислительным методам и обработке наблюдений // Сибирский математический журнал. 1962. Т. 3. № 5.

2. Шарый С.П. Конечномерный интервальный анализ. Новосибирск, 2017.

3. Шелудько А.С. Гарантированное оценивание параметров дискретных моделей хаотических процессов // Вестник ЮУрГУ. Серия : Вычислительная математика и информатика. 2018. Т. 7. № 1.

4. Мадияров М. Н., Оскорбин Н.М., Суханов С.И. Примеры интервального анализа данных в задачах моделирования процессов // Известия Алт. гос. ун-та. 2019. № 1 (105).

5. Жолен Л. Прикладной интервальный анализ. М. ; Ижевск, 2005.

6. Zhilin S.I. Simple method for outlier detection in źtting experimental data under interval error // Chemometrics and Intellectual Laboratory Systems. 2007. Vol. 88. № 1.

7. Singh K., Upadhyaya S. Outlier Detection: Applications And Techniques // International Journal of Computer Science Issues. 2012. Vol. 9. Issue 1. № 3.

8. Rana P., Pahuja D., Gautam R. A Critical Review on Outlier Detection Techniques // International Journal of Science and Research. 2014. Vol. 3. Issue 12.

9. Бедринцев А.А., Чепыжов В.В. Двухкритериальная задача построения оптимальных эллипсоидов для представления данных // Информационные технологии и системы – 2014 : сб. трудов. Н. Новгород. 2014.

10. Бедринцев А.А., Чепыжов В.В., Чернова С.С. Экстремальные эллипсоиды как аппроксиматоры пространства дизайна в задачах предсказательного метамоделирования // Искусственный интеллект и принятие решений. 2015. № 2.