УДК 519.87

Об одном подходе к оценке погрешности и значимости измерений в линейных задачах

© С. И. Спивак[1], О. Г. Кантор[2], Д. С. Юнусова[3]

Аннотация. Рассматриваются задачи восстановления линейных зависимостей по экспериментальным данным, для которых получение точного решения в силу невозможности удовлетворения ряда объективных условий неосуществимо. В работе предлагается метод, позволяющий не только получать приближенные решения, но и проводить оценку погрешности измерений и их значимости, что является существенным для совершенствования процедуры построения функциональных зависимостей на стадии планирования экспериментов в части уточнения экспериментальных данных или их исключения из рассмотрения. Приведены результаты апробации предлагаемого метода.

Ключевые слова: задачи восстановления линейных зависимостей, погрешность измерений, значимость измерений, двойственные оценки.

1. Введение

Задачи восстановления линейных зависимостей по экспериментальным данным сводятся к системам линейных алгебраических уравнений вида:

$$AX = B \tag{1.1}$$

где $A = (a_{ij})$ и $B = (b_i)$ – экспериментальные данные $(i = 1, \ldots, m, j = 1, \ldots, n)$, а $X = (x_1, x_2, \ldots, x_n)^T$ – искомые параметры модели.

Точное решение системы (1.1) может не удовлетворять ограничениям модели или некоторым другим требованиям. В этой связи актуальными являются определение приближенного решения таких систем и связанная с ней оценка величины погрешности измерений, под которой будем понимать расхождение значений расчетных и экспериментальных величин не в каждом отдельном наблюдении, а в целом по всей совокупности наблюдений. Способы формализации такого рода задач, очевидно, зависят от многих факторов: имеющейся информации, предъявляемых к модели требований, предпочтений и компетенции исследователя и пр. В настоящей работе проблема приближенного решения системы (1.1) с последующим вычислением оценок погрешности и значимости измерений рассмотрена с позиций линейных задач.

2. Описание подхода для определения погрешности измерений

Будем предполагать, что на параметры модели X наложены условия неотрицательности:

$$X \geq 0 \tag{2.1}$$

[1] Заведующий кафедрой математического моделирования, ФГБОУ ВПО Башкирский государственный университет, г. Уфа; semen.spivak@mail.ru.

[2] Старший научный сотрудник, ФГБУН Институт социально-экономических исследований Уфимского научного центра РАН, г. Уфа; o_kantor@mail.ru.

[3] Аспирант, ФГБОУ ВПО Башкирский государственный университет, г. Уфа; kazakova_d_s@mail.ru.

Также, будем считать, что на значения параметров накладываются ограничения, выражающие их принадлежность какому-либо множеству значений, другими словами, параметры модели удовлетворяют системе ограничений:

$$CX \geq D \tag{2.2}$$

где $C = (c_{lj})$ – это матрица, состоящая из коэффициентов при параметрах модели в системе ограничений $(l = 1, \ldots, k, j = 1, \ldots, n)$, а $D = (d_l)$ – концы промежутков значений, которым принадлежат параметры модели $(l = 1, \ldots, k)$.

Тогда задача определения приближенного решения системы (1.1) с учетом ограничений (2.1) и (2.2) может быть сведена к задаче линейного программирования:

$$\begin{gathered} \varepsilon \to min \\ |AX - B| \leq \varepsilon \\ CX \geq D \\ X \geq 0 \end{gathered} \tag{2.3}$$

Здесь ε – параметр, характеризующий величину погрешности измерений.

В результате решения задачи (2.3) должны быть определены параметры модели, удовлетворяющие требуемым ограничениям и величина погрешности измерений.

Следует заметить, что матрицы A и B формируются по результатам наблюдений, а потому не могут рассматриваться как абсолютно точные, так как результаты их измерений неминуемо сопряжены с некоторыми ошибками, которые приводят к отклонениям измеряемых значений величин A и B от их истинных значений.

В случае предполагаемых ошибок в матрице B, различных для каждого отдельного измерения $i = 1, \ldots, m$, задача определения параметров модели, обеспечивающих минимальное отклонение от экспериментальных величин (b_i) может быть сведена к задаче линейного программирования следующего вида:

$$\begin{gathered} \varepsilon \to min \\ AX = \Delta B \\ CX \geq D \\ X \geq 0 \\ |\delta_i - 1| \leq \varepsilon, i = 1, \ldots, m \end{gathered} \tag{2.4}$$

Здесь ε – это величина погрешности измерений, а Δ – диагональная матрица, на главной диагонали которой стоят элементы δ_i $(i = 1, \ldots, m)$ – неизвестные величины, отождествляемые с параметрами, характеризующими ошибки в измерениях элементов матрицы B. Величины ε и δ_i связаны очевидным соотношением: $\varepsilon = \max\limits_i |\delta_i - 1|$.

В случае предполагаемых ошибок в каждом элементе матрицы А задача (1.1) может быть формализована в следующем виде:

$$\begin{gathered} \varepsilon \to min \\ A'X = B \\ CX \geq D \\ X \geq 0 \\ |\gamma_{ij} - 1| \leq \varepsilon, i = 1, \ldots, m, j = 1, \ldots, n \end{gathered} \tag{2.5}$$

Здесь матрица $A' = (\gamma_{ij} a_{ij})$ $(i = 1, \ldots, m, j = 1, \ldots, n)$.

Задача (2.5) является нелинейной, поскольку в выражении $A'X = B$ элементы вектора X – неизвестные величины, а A' – матрица, которая зависит от $n \times m$ неизвестных величин. Данное обстоятельство создает существенные проблемы для численной реализации модели (2.5). Вместе с тем относительно элементов матрицы A может быть известна информация, которая позволит упростить процесс получения решения, например, при наличии индивидуальных систематических ошибок, имеющих место при измерении каждой из наблюдаемых величин, или индивидуальных ошибок каждого отдельного наблюдения. Первая из перечисленных выше ситуаций отражается в пропорциональных изменениях всех элементов каждого из столбцов $(A' = A\Gamma, \Gamma = (\gamma_{jj}), j = 1, \dots, n)$, что никак не отражается на погрешности измерений ε, а влияет только на параметры X, а вторая – строк $(A' = \Gamma A, \Gamma = (\gamma_{ii}), i = 1, \dots, m)$, что сводит задачу определения неизвестных параметров к задаче (2.4).

3. Методика определения значимости измерений

Важное практическое значение имеет оценка влияния погрешности экспериментальных данных модели на погрешность измерений ε [1], что позволяет осуществлять анализ информационной ценности измерений и, как следствие, выявлять те, которые следует рассматривать как наиболее недостоверные или значимые и пр. Результатами такого анализа могут быть, например, выводы, о необходимости, при наличии соответствующих возможностей, уточнения некоторых экспериментальных данных, или рекомендации об их исключении из рассмотрения при непосредственном построении функциональных зависимостей.

Из теории линейного программирования известно, что для этих целей может быть использована теория двойственности [2]. Применительно к исследуемой в настоящей работе проблеме решение соответствующих двойственных задач позволит оценить влияние элементов матрицы экспериментальных данных B и матрицы ограничений на параметры модели D в прямых задачах линейного программирования на величину минимального значения погрешности измерений ε. Двойственная задача для задачи линейного программирования (2.3) имеет вид:

$$\begin{gathered}
(B, y^1) - (B, y^2) + (D, y^3) \to max \\
A^T y^1 - A^T y^2 + C^T y^3 \le 0 \\
\sum_{i=1}^{m} y_i^1 + \sum_{i=1}^{m} y_i^2 \le 1 \\
y^1 \ge 0, y^2 \ge 0, y^3 \ge 0.
\end{gathered} \tag{3.1}$$

Здесь $y^1 = (y_i^1)$, $y^2 = (y_i^2)$ $(i = 1, \dots, m)$, $y^3 = (y_l^3)$ $(l = 1, \dots, k)$ – векторы оптимального решения двойственной задачи.

Заметим, что в задаче (2.3) для оценки степени влияния i-го соотношения из системы неравенств $|AX - B| \le \varepsilon$ на значение погрешности измерений ε необходимо рассмотреть соответствующие компоненты векторов y^1 и y^2, являющихся решением задачи (3.1), и выбрать из них максимальный.

Аналогичным образом может быть выписана двойственная задача для задачи линейного программирования (2.4).

4. Результаты апробации

Описанный выше подход был применен для определения степени влияния погрешностей наблюдаемых величин на погрешность измерения при решении задачи моделирования численности населения Российской Федерации методом системной динамики [3] - [6]. Общий вид исследованной модели системной динамики следующий:

$$\begin{aligned} \frac{dN}{dt} &= a_1 N^{\alpha_1} D^{\beta_1} I^{\gamma_1} - a_2 N^{\alpha_2} D^{\beta_2} I^{\gamma_2} \\ \frac{dD}{dt} &= a_3 N^{\alpha_3} D^{\beta_3} I^{\gamma_3} - a_4 N^{\alpha_4} D^{\beta_4} I^{\gamma_4} \\ \frac{dI}{dt} &= a_5 N^{\alpha_5} D^{\beta_5} I^{\gamma_5} - a_6 N^{\alpha_6} D^{\beta_6} I^{\gamma_6} \end{aligned} \tag{4.1}$$

где N – численность населения РФ; D – душевые доходы за год; I – индекс потребительских цен. Информационную базу исследования составили данные официальной статистической отчетности за период с 1998 по 2009 гг.

По результатам специально организованного численного эксперимента была получена модель следующего вида:

$$\begin{gathered} \frac{dN}{dt} = 8,139 \cdot 10^{-22} \cdot \frac{N^{0,05} \cdot D^2}{I^2} - 64,01 \cdot \frac{N^{0,33} \cdot D^{0,3}}{I^{0,3}} \\ \frac{dD}{dt} = 560 \cdot D^{0,35} - 9900 \cdot I \\ \frac{dI}{dt} = 0,131 \cdot I^{-0,4} - 0,0072 \cdot \frac{N^{0,092} \cdot D^{0,092}}{I^{0,092}} \end{gathered} \tag{4.2}$$

Модель (4.2) с достаточно высокой точностью описывает экспериментальные данные, о чем свидетельствуют низкие значения средних ошибок аппроксимации для каждого из уравнений модели: $\overline{A}_N = 0,13\%$, $\overline{A}_D = 3,33\%$, $\overline{A}_I = 5,86\%$.

Вместе с тем, как показывает практика использования статистической информации, исходные данные для модели (4.1) нельзя рассматривать как абсолютно точные. Это обусловлено рядом причин объективного характера: запаздывание сбора и обработки данных, неточности (а иногда и умышленные искажения) предоставляемой информации, аккумулируемой органами статистики, и пр. В этой связи, несмотря на хорошую точность построенной модели, целесообразной является оценка значимости имеющихся экспериментальных данных. И одним из возможных способов реализации этого является описанный в настоящей работе подход.

Ниже приводится описание его применения для первого уравнения модели (4.1). Для этого потребовалось осуществить линеаризацию данного уравнения, для чего было использовано разложение в ряд Тейлора с центром в точке $\{a_i^0, \alpha_i = 0, \beta_i = 0, \gamma_i = 0\}$, $i = \overline{1,2}$:

$$\begin{aligned} \Delta N \approx a_1 + a_1^0 \cdot \ln N \cdot \alpha_1 + a_1^0 \cdot \ln D \cdot \beta_1 + a_1^0 \cdot \ln I \cdot \gamma_1 - \\ - a_2 - a_2^0 \cdot \ln N \cdot \alpha_2 + a_2^0 \cdot \ln D \cdot \beta_2 + a_2^0 \cdot \ln I \cdot \gamma_2 \end{aligned} \tag{4.3}$$

Были введены следующие обозначения: $x_1 = a_1$, $x_2 = a_2$, $x_3 = \alpha_1$, $x_4 = \alpha_2$, $x_5 = \beta_1$, $x_6 = \beta_2$, $x_7 = \gamma_1$, $x_8 = \gamma_2$. На основании исходной информации для определения параметров $\{a_i, \alpha_i, \beta_i, \gamma_i\}$, $i = \overline{1,2}$ были сформированы матрицы A и B:

$$A = \begin{pmatrix} 1 & 1,881 & 0,692 & 0,000 & -1 & 1128,683 & 415,074 & 0,0000 \\ 1 & 1,881 & 0,741 & -0,031 & -1 & 1128,576 & 444,835 & -18,669 \\ 1 & 1,881 & 0,773 & -0,018 & -1 & 1128,312 & 463,945 & -11,039 \\ 1 & 1,880 & 0,803 & -0,017 & -1 & 1128,072 & 481,609 & -10,235 \\ 1 & 1,880 & 0,828 & -0,014 & -1 & 1127,803 & 496,846 & -8,438 \\ 1 & 1,879 & 0,855 & -0,011 & -1 & 1127,520 & 513,042 & -6,800 \\ 1 & 1,879 & 0,877 & -0,011 & -1 & 1127,189 & 525,940 & -6,639 \\ 1 & 1,878 & 0,900 & -0,010 & -1 & 1126,900 & 540,065 & -6,208 \\ 1 & 1,878 & 0,923 & -0,009 & -1 & 1126,598 & 553,785 & -5,171 \\ 1 & 1,877 & 0,944 & -0,011 & -1 & 1126,374 & 566,500 & -6,746 \\ 1 & 1,877 & 0,961 & -0,012 & -1 & 1126,284 & 576,740 & -7,492 \\ 1 & 1,877 & 0,974 & -0,008 & -1 & 1126,340 & 584,490 & -5,060 \end{pmatrix} B = \begin{pmatrix} -262647,1 \\ -649238,1 \\ -586457,1 \\ -654217,1 \\ -685624,1 \\ -795385,1 \\ -693926,1 \\ -720608,1 \\ -532523,1 \\ -212070,1 \\ -104799,1 \\ 10589,9 \end{pmatrix}$$

С учетом всех предъявляемых требований к параметрам модели, задача определения приближенного значения параметров модели была формализована в следующем виде:

$$\begin{gathered} \varepsilon \to min \\ |AX - B| \le \varepsilon \\ 0 \le x_1 \le 3 \\ 1,948 \le x_2 \le 2,050 \\ -14151439,74 \le x_1 - x_5 \le 23,4 \\ 0 \le x_6 \le 1 \\ 0 \le x_p \le 5, p = 3,4,7,8 \\ x_5 \ge 0 \end{gathered} \tag{4.4}$$

Решение задачи (4.4): $X = (0; 1,948; 5; 5; 395119,963; 0; 5; 5)^T$, погрешность измерений $\varepsilon = 402804,23$, решение двойственной к ней (см. модель (3.1)):

$$\begin{gathered} y^1 = (0;0;0;0;0;0,5;0;0;0;0;0;0)^T, \\ y^2 = (0;0;0;0;0;0;0;0;0;0;0;0,5)^T, \\ y^3 = (0;0;0,001;0,0595;0,0015;0;3,5724;0,87;0;0)^T. \end{gathered}$$

Полученные результаты позволили сделать вывод о том, что погрешность измерений в большей степени определяется верхним ограничением на переменную x_7 $(y_7^3 = 3,5724)$, в меньшей 6-м и 12-м элементами матрицы B $(y_6^1 = 0,5$ и $y_{12}^2 = 0,5)$, верхними ограничениями на переменные x_4, x_5, x_8 $(y_4^3 = 0,0595, y_5^3 = 0,0015, y_8^3 = 0,87)$ и нижними ограничением на переменную x_3 $(y_3^3 = 0,001)$. Даная информация может быть использована при планировании дальнейших экспериментов с исследуемой моделью. При этом должен соблюдаться следующий принцип: в каждом последующем эксперименте информативность новых измерений должна быть не менее значима, что должно отражаться при формировании соответствующих условий. Так, например, сужение диапазонов значений переменных x_4, x_5, x_7, x_8 целесообразно проводить за счет увеличения нижних границ, а переменной x_3 – верхней.

Список литературы

1. Спивак С. И., Тимошенко В. И., Слинько М. Г., "Методы построения кинетических моделей стационарных реакций", *Химическая промышленность*, 1979, № 3, 33–36.

2. Канторович Л. В., Горстко А. Б., *Оптимальные решения в экономике*, Наука, М., 1972, 231 с.

3. Спивак С. И., Кантор О. Г., "Оценка качества спецификации моделей системной динамики", *Журнал СВМО*, **14**:2 (2012), 34–39.

4. Спивак С. И., Кантор О. Г., "Оценка параметров моделей системной динамики", *Журнал СВМО*, **13**:3 (2011), 107–113.

5. Спивак С. И., Кантор О. Г., Салахов И. Р., "Алгоритм получения прогнозируемых параметров социально-экономических систем", *Системы управления и информационные технологии*, 2013, № 4(54), 43–45.

6. Спивак С. И., Кантор О. Г., Салахов И. Р., "Вычислительная реализация оценки управляющих параметров модели системной динамики", *Вестник Башкирского университета*, **17**:4 (2012), 1658–1660.

A approach to error estimation and significance of measurements in linear problems.

© S. I. Spivak[4], O. G. Kantor[5], D. S. Yunusova[6]

Abstract. We consider the problem of recovering the linear dependency on the experimental data, for which it is impossible to find exact solutions because of inability to meet a number of objective conditions. This paper proposes a method to not only obtain approximate solutions , but also to assess the accuracy of measurements and their significance, which is essential for improving the procedure for construction of functional dependencies in the planning stage of the experiments in verifying experimental data or their exclusion from consideration. The paper also provides results of testng of the proposed method.

Key Words: the problem of recovering the linear dependency, the accuracy of measurements, the significance of measurements, dual estimates.

[4] Head of the Chair of Mathematical Modelling, Bashkir State University, Ufa; semen.spivak@mail.ru.

[5] Senior Scientist, Institute of Social and Economic Research of RAS Ufa Scientific Center, Ufa; o_kantor@mail.ru.

[6] Postgraduate, Bashkir State University, Ufa; kazakova_d_s@mail.ru.