

О ВЫЯВЛЕНИИ И ИСПОЛЬЗОВАНИИ СТРУКТУРНЫХ ОСОБЕННОСТЕЙ МАТРИЦ В ЗАДАЧАХ ПЛАНИРОВАНИЯ

Э. Б. ЕРШОВ

(Москва)

Принципиальные представления о направлении исследований, излагаемых в настоящей работе, разрабатывались в непосредственном контакте с В. А. Волконским [1].

В первой части предлагаемой статьи итерационный процесс, описанный в [1] для произвольных систем линейных уравнений и для задач выпуклого программирования с линейными ограничениями, исследуется применительно к системе уравнений статической модели межотраслевого баланса.

Во второй части статьи описан удобный в случае задач большой размерности алгоритм выделения в матрицах неприводимых и неразложимых блоков. Этот алгоритм может применяться для практического выяснения структурных особенностей рассматриваемых матриц.

РЕШЕНИЕ СИСТЕМ УРАВНЕНИЯ МЕЖОТРАСЛЕВОГО БАЛАНСА МЕТОДОМ ВЫДЕЛЕНИЯ СУЩЕСТВЕННЫХ ЭЛЕМЕНТОВ

1. Для нахождения решения системы линейных уравнений

$$(A + B)_{m \times m} x_{m \times 1} = b_{m \times 1} \quad (1)$$

с невырожденной матрицей $(A + B)$ и достаточно малыми элементами матрицы B можно применить итерационный процесс, определяемый формулами

$$Ax^n = b - Bx^{n-1}, \quad (2)$$

где x^n — n -е приближение.

В [1] показано, что последовательность $\{x^n\}$ сходится по выбранной норме векторов $\|v\|$ к решению системы (1), если

$$\|B \cdot A^{-1}\| < 1, \quad (3)$$

где $\|a\|$ — норма матриц, согласованная с нормой $\|v\|$. Неравенство (3) показывает, как следует понимать малость элементов матрицы B при выбранных нормах векторов и матриц и заданной матрице $(A + B)$.

Впрочем, малость элементов матрицы B можно не связывать с какой-либо конкретной нормой матриц, согласованной или подчиненной одной из норм векторов. Будем называть разбиение матрицы $(A + B)$ на сумму матриц A и B итеративно-сходящимся, а матрицу B — матрицей итеративно малых элементов в $(A + B)$, если

$$\rho(BA^{-1}) \equiv \rho(A^{-1}B) < 1, \quad (4)$$

($\rho(a)$ — спектральный радиус матрицы (a)). Выполнение неравенства (4) необходимо и достаточно для сходимости последовательности векторов невязок $(v^n) \equiv b - (A + B)x^n$ к нулевому вектору, причем последовательность $\|v^n\|$ при любой норме мажорируется прогрессией со знаменателем $\lambda = \rho(BA^{-1})$.

Систему (1) можно представить в виде

$$x = (-A^{-1}B)x + A^{-1}b. \quad (5)$$

Очевидно, что описанный выше метод дает ту же последовательность приближенных решений $\{x^n\}$, что и метод последовательных приближений для системы (5), а условие (4) представляет собой необходимое и достаточное условие сходимости метода последовательных приближений для системы (5). Отличие предлагаемого итерационного процесса от метода последовательных приближений, примененного непосредственно к системе (5), состоит только в том, что предварительное нахождение обратной матрицы (A^{-1}) , необходимое для преобразования системы (2) в систему (5), заменяется многократным решением системы (2) с матрицей коэффициентов A . Такой прием может оказаться эффективным при решении на ЭВМ задач с большими m и выделении таких матриц A существенных элементов, чтобы решение системы (5) было значительно проще непосредственного решения системы (1), но в то же время чтобы $\rho(A^{-1}B)$ было достаточно мало, обеспечивая получение нужной точности за приемлемое число итераций.

Заметим также, что матрица $(-A^{-1}B)$ является матрицей невязок для матрицы $(A + B)^{-1}$ и приближенно обратной к ней матрицы (A^{-1}) , так как

$$E - A^{-1}(A + B) = -A^{-1}B.$$

Неравенство (4) дает необходимое и достаточное условие сходимости к матрице $(A + B)^{-1}$ последовательности матриц, получаемых в соответствии с итерационным методом исправления элементов приближенно обратной матрицы (см. [2], стр. 182). Поэтому выделение в матрице $(A + B)$ матрицы B итеративно-малых элементов может быть использовано не только при решении систем линейных уравнений (без обращения матриц), но и непосредственно при нахождении обратных матриц.

Приведенные соображения поясняют смысл вводимого термина «итеративно-сходящееся разбиение матрицы $(A + B)$ на сумму матриц A и B ».

Интересно обратить внимание на тот факт, что такие общеизвестные итерационные методы решения систем линейных уравнений, как метод простой итерации (метод Якоби), одношаговый циклический процесс (метод Зейделя) и метод Некрасова, являясь модификациями метода последовательных приближений, получаются из процесса, задаваемого формулой (2), при некоторых частных видах матрицы A . Впрочем это очевидно из того, что приближения $\hat{x}^n$ предлагаемого процесса совпадают с приближениями x^n , получаемыми из так называемого стационарного итерационного процесса $\hat{x}^n = \hat{x}^{n-1} + H(b - (A + B)\hat{x}^{n-1})$ с фиксированной матрицей H при $H = (E + BA^{-1})(A + B)^{-1}$.

Выбор итеративно-сходящихся разбиений матриц может основываться только на изучении классов матриц для конкретных задач. Ниже приведенный итерационный процесс применяется к частному виду системы (1), а именно к системе уравнений межотраслевого или межпродуктового баланса.

Заметим, что в практических задачах, по-видимому, имеет смысл ограничиваться только такими итеративно-сходящимися разбиениями матриц $(A + B)$, для которых выполняются требования

$$\text{sign } A_{ij} = \text{sign } B_{ij} \quad (i, j = 1, 2, \dots, m)$$

или даже требования

$$A_{ij} \cdot B_{ij} = 0 \quad (i, j = 1, 2, \dots, m).$$

2. Частным, но достаточно важным для экономических приложений случаем системы (1) является система уравнений межотраслевого или межпродуктового баланса

$$(E - a)x = y \quad (6)$$

с продуктивной матрицей коэффициентов прямых затрат (a) (см. [3], стр. 367—376), т. е. с неотрицательной матрицей, удовлетворяющей условию

$$\rho(a) < 1. \quad (7)$$

При применении к системе (6) итерационного процесса, описанного в п. 1, естественно рассматривать лишь разбиения матрицы $(E - a)$ на сумму двух слагаемых вида

$$(E - a) = (E - f) + (-b),$$

где (f) и (b) — матрицы с неотрицательными элементами.

Каждое такое разбиение будет итеративно-сходящимся, так как имеет место неравенство

$$\rho(bF) = \rho(Fb) \leq \rho(a), \quad (8)$$

где $F = (E - f)^{-1}$. Матрица F всегда существует и имеет (так же как и матрица $(E - a)^{-1}$) неотрицательные элементы, так как из неравенств

$$0 \leq f_{ij} \leq a_{ij}, \quad i, j = 1, 2, \dots, m \quad (9)$$

следует [4, стр. 296] неравенство

$$\rho(f) \leq \rho(a).$$

Докажем неравенство (8). Для этого предварительно докажем более слабое утверждение:

$$\rho(Fb) = \rho(bF) < 1. \quad (10)$$

Заметим, что равенство спектральных радиусов для матриц (bF) и (Fb) следует из подобия этих матриц.

Рассмотрим систему уравнений, определяющую спектральный радиус $\rho(Fb) = \rho$ матрицы (Fb) с неотрицательными элементами (и соответствующий ему собственный вектор z)

$$\rho z = (Fb)z,$$

которую после умножения слева на $(E - f)$ представим в виде

$$\rho z = (\rho f + b)z. \quad (11)$$

Допустим, что существует матрица $\hat{f}$, удовлетворяющая неравенствам (9), для которой $\rho(\hat{F}\hat{b}) \geq 1$, где $\hat{F} = (E - \hat{f})^{-1}$, $\hat{b} = a - \hat{f}$. Определим однопараметрические семейства матриц, $\{f(t)\}$ и $\{b(t)\}$ при $0 \leq t \leq 1$ формулами $f(t) = t\hat{f}$ и $b(t) = a - t\hat{f}$.

Рассмотрим спектральный радиус $\rho((E - f(t))^{-1}b(t)) = g(t)$ как функцию параметра t . Поскольку $g(0) = \rho(a) < 1$ и $g(1) = \rho(\hat{F}\hat{b}) \geq 1$, то из непрерывности функции $g(t)$ по t найдется такое t^0 ($0 < t^0 \leq 1$), для которого $g(t^0) = 1$.

Но тогда при $b = b(t^0)$ и $F = (E - f(t^0))^{-1} = F(t^0)$ система (11) для собственного вектора $z(t^0)$ матрицы $(F(t^0)b(t^0))$ имеет вид:

$$z(t^0) = (f(t^0) + b(t^0))z(t^0),$$

или

$$z(t^0) = az(t^0),$$

что невозможно в силу (7). Это противоречие доказывает неравенство (10).

Перейдем к доказательству неравенства (8). Из (9) следует представление матрицы F с помощью сходящегося матричного ряда по степеням матрицы f :

$$F = E + f + f^2 + \dots,$$

что, в свою очередь, обеспечивает выполнение следующего матричного неравенства (поэлементно):

$$F \geq E.$$

Но тогда из легко проверяемого тождества

$$(E - a)^{-1} = (E - Fb)^{-1}F$$

следует (здесь и далее неравенство $A \geq B$ для векторов или матриц A и B является краткой формой записи поэлементных неравенств):

$$(E - a)^{-1} \geq (E - Fb)^{-1}, \quad \rho((E - a)^{-1}) \geq \rho((E - Fb)^{-1})$$

и

$$\frac{1}{1 - \rho(a)} \geq \frac{1}{1 - \rho(Fb)}. \quad (12)$$

В силу (7) и (10) из (12) получаем неравенство (8).

Напомним, что $\rho(Fb) = \lambda$ является знаменателем геометрической прогрессии, мажорирующей последовательность норм невязок v^n для приближений x^n , получаемых как решение системы

$$(E - f)x^n = bx^{n-1} + y. \quad (13)$$

При $f = 0$ и $a = b$ формула (13) определяет обычный процесс последовательных приближений, и в этом случае $\lambda = \rho(a)$. Поэтому из неравенства (8) очевидно, что сходимость последовательности $\{x^n\}$ к решению системы (6) в методе последовательных приближений в общем случае медленнее, чем в случае процесса, определяемого формулой (13), с ненулевой матрицей (f) . Это утверждение можно уточнить, если, не ограничиваясь сравнением знаменателей прогрессий, мажорирующих последовательности норм невязок, сопоставить непосредственно приближения x^n и $\hat{x}^n$, где $\hat{x}^n$ — вектор, получаемый на n -й итерации метода последовательных приближений (конечно, при $x^0 = \hat{x}^0$).

Выразим приближения x^n и $\hat{x}^n$ непосредственно через матрицы a , b , F и вектор x^0 . Легко видеть, что

$$x^n = (E + Fb + \dots + (Fb)^{n-1})Fy + (Fb)^nx^0 \quad (14)$$

и

$$\hat{x}^n = (E + a + \dots + a^{n-1})y + a^nx^0. \quad (15)$$

Как показывают простые примеры, для векторов $(Fb)^nx^0$ и a^nx^0 в силу произвольности вектора x^0 и матрицы f , удовлетворяющей неравенствам (9), невозможно при всех n ($n = 1, 2, 3, \dots$) доказать ни одно из двух неравенств $(Fb)^nx^0 \geq a^nx^0$ и $(Fb)^nx^0 \leq a^nx^0$.

Поэтому в формулах (14) — (15) имеет смысл сравнивать только члены, содержащие вектор y . Это сравнение основывается на следующем неравенстве:

$$\begin{aligned} (E + a + \dots + a^n) &\leq (E + Fb + \dots + (Fb)^n)F = \\ &= F(E + bF + \dots + (bF)^n). \end{aligned} \quad (16)$$

Равенство двух вариантов правых частей в (16) проверяется непосредственно. Доказательство самого неравенства (16) основывается на легко проверяемом соотношении

$$(E - f)(E + a + \dots + a^n) = E + b(E + a + \dots + a^{n-1}) - fa^n$$

или

$$(E + a + \dots + a^n) = F + Fb(E + a + \dots + a^{n-1}) - Ffa^n,$$

из которого с помощью индукции по n получаем

$$(E + a + \dots + a^n) = (E + Fb + \dots + (Fb)^n)F - F[(bF)^n f + (bF)^{n-1} fa + \dots + (bF) fa^{n-1} + fa^n].$$

Отсюда в силу неотрицательности матриц b , f и F следует (16).

Таким образом, при обычном для межотраслевого или межпродуктового баланса предположении о неотрицательности вектора y и при нулевом начальном приближении $x^0 = 0$ из (14) — (16) получаем

$$0 \leq \hat{x}_i^n \leq x_i^n \leq \tilde{x}_i \quad (i = 1, 2, \dots, m; n = 1, 2, \dots),$$

где $\tilde{x}$ — решение системы (6). Очевидно, что при $x^0 \neq 0$ подобное сравнение приближений x^n и $\hat{x}^n$ уже невозможно.

Следующее утверждение естественным образом обобщает неравенства (8) и (16): для двух разложений матрицы $(E - a)$, задаваемых в виде

$$(E - a) = (E - f_1) + (-b_1) \quad \text{и} \quad (E - a) = (E - f_2) + (-b_2), \quad \text{при} \quad 0 \leq$$

$$\leq f_2 \leq f_1 \leq a \quad \text{имеем} \quad \rho(F_1 b_1) \leq \rho(F_2 b_2),$$

$$[E + F_1 b_1 + \dots + (F_1 b_1)^n] F_1 \geq [E + F_2 b_2 + \dots + (F_2 b_2)^n] F_2$$

$$(n = 1, 2, \dots),$$

где

$$F_1 = (E - f_1)^{-1}, \quad F_2 = (E - f_2)^{-1}.$$

Доказательства этих неравенств здесь не приводятся (см. [5]), так как они мало отличаются от доказательств неравенств (8) и (16).

Все приведенные выше утверждения очевидным образом переформулируются для системы линейных уравнений $Q(E - a) = q$, используемой в моделях межотраслевого баланса для расчета вектора полных затрат Q труда, заработной платы или вектора среднеотраслевых индексов цен (для стоимостного баланса) по соответствующему вектору q прямых затрат или удельных весов чистой продукции в продукции отраслей.

3. Выигрыш в скорости сходимости процесса (13) по сравнению с методом последовательных приближений связан с тем, что предлагаемый метод основан на многократном решении систем уравнений с матрицей $(E - f)$ коэффициентов при неизвестных. В настоящее время трудно сколько-нибудь обоснованно сравнить «трудоемкость» решения системы (6) методом последовательных приближений и методом, описываемым формулами (13), при векторе y правых частей из некоторого фиксированного множества, которое в случае системы уравнений межотраслевого баланса может предполагаться заданным, например, неравенствами $y_i \geq 0$ ($i = 1, 2, \dots, m$). При таком сравнении, помимо оценки числа итераций, достаточного для достижения требуемой точности, было бы необходимо оценить «трудоемкость» каждой итерации, в том числе «трудоемкость» однократного решения системы (13) с некоторой матрицей $(E - f)$. Теоретически и экспериментально такая характеристика даже для системы

(13) еще не построена, и вопрос о сравнительной эффективности указанных методов и о выборе итеративно сходящегося разбиения матрицы $(E - a)$, по возможности минимизирующего общую «трудоемкость» решения системы (6), может решаться только после рассмотрения достаточно большого числа конкретных задач.

Несмотря на указанные выше трудности, все же можно предложить еще далеко не совершенные, упрощенные оценки трудоемкости решения системы уравнений межотраслевых балансов и обращения соответствующих им матриц. Примем,

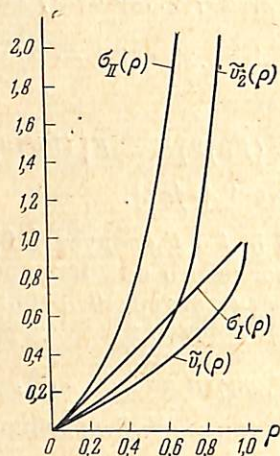


Рис. 1

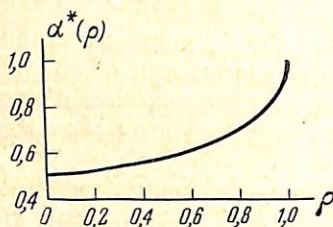


Рис. 2

например, спектральный радиус $\rho(a)$ или $\rho((E - a)^{-1} - E)$ в качестве показателя трудоемкости σ метода последовательных приближений для системы (6). Тогда, поскольку метод решения системы (6) с помощью итеративно-сходящегося разбиения матрицы $(E - a)$ в силу (14) в определенном смысле эквивалентен представлению матрицы $(E - a)^{-1}$ в виде:

$$(E - a)^{-1} = (E - Fb)^{-1}(E - f)^{-1},$$

условимся считать показателем трудоемкости для процесса (13) одну из следующих трех величин $v_k(f, b)$:

$$v_1(f, b) = \max[\rho(f), \rho(Fb)] \quad \text{при } \sigma_I(a) = \rho(a)$$

и

$$v_2(f, b) = \max[\rho(F - E); \rho((E - Fb)^{-1} - E)]$$

или

$$v_3(f, b) = \rho(F - E) + \rho((E - Fb)^{-1} - E) \quad \left. \begin{array}{l} \text{при } \sigma_{II}(a) = \rho((E - a)^{-1} - E). \end{array} \right\}$$

Для каждого из показателей $v_k(f, b)$ естественно считать наилучшим такое итеративно-сходящееся разбиение матрицы $(E - a)$, для которого достигается

$$\min_{0 \leq f \leq a} v_k(f, b) \equiv v_k(a).$$

Нахождение таких наилучших разбиений нами не рассматривается. Ограничимся указанием некоторых оценок сверху для $v_k(a)$ ($k = 1, 2, 3$) и тех разбиений, для которых они достигаются. Для этого рассмотрим разбиение матрицы $(E - a)$ на сумму двух матриц следующего частного вида:

$$E - f = E - aa, \quad -b = -(1 - a)a, \quad (17)$$

где число a удовлетворяет неравенствам $0 \leq a \leq 1$.

Тогда, как показано в [5], для функций $\tilde{v}_k(a)$, определенных формулами

$$\tilde{v}_k(a) = \min_{0 \leq \alpha \leq 1} v_k(\alpha a, (1 - \alpha)a) \equiv v_k(\alpha_k a, (1 - \alpha_k)a),$$

имеем

$$\tilde{v}_1(a) = 1 - \sqrt{1 - \rho(a)}, \quad \tilde{v}_2(a) = \frac{\tilde{v}_1(a)}{1 - \tilde{v}_1(a)}, \quad \tilde{v}_3(a) = [2\tilde{v}_2(a),$$

и

$$\alpha_1 = \alpha_2 = \alpha_3 = \frac{1 - \sqrt{1 - \rho(a)}}{\rho(a)} \equiv \alpha^*.$$

Функции $\tilde{v}_k(a)$ дают оценки сверху для $v_k(a)$, так как из определений очевидно, что $v_k(a) \leq \tilde{v}_k(a)$ ($k = 1, 2, 3$).

Из сравнения показателей $\tilde{v}_k(a)$, которые оказываются функциями только от $\rho(a)$ (их графики изображены на рис. 1), с принятыми показателями трудоемкости метода последовательных приближений $\sigma_I(a)$ и $\sigma_{II}(a)$ видно, что применение даже простейших разбиений матрицы $(E - a)$, какими являются разбиения (17), позволяет значительно снизить трудоемкость решения системы (6). Зависимость параметра α^* , определяющего оптимальное разбиение среди разбиений (17), от $\rho(a)$ изображена на рис. 2.

Заметим, что наибольший выигрыш в трудоемкости решения системы (6) методом (13) при условии (17) по сравнению с методом последовательных приближений, определяемый как максимум по $\rho(a)$ ($0 \leq \rho(a) \leq 1$) функции $[\sigma_I(a) - \tilde{v}_1(a)]$, достигается при $\rho(a) = 0,75$. Это соображение может иметь практическое значение, так как для реальных межотраслевых балансов в стоимостном выражении спектральный радиус матрицы прямых затрат $\rho(a)$ заключен, по-видимому, между 0,5 и 0,8.

Не следует переоценивать значение таких абстрактных характеристик трудоемкости решения систем линейных уравнений или обращения матриц, какими в случае предлагаемого метода являются величины $v_k(f, b)$. В них не учитывается, например, возможность использования особенностей строения матрицы $(E - f)$ с целью экономии памяти ЭВМ и упрощения алгоритма решения. Кроме того, число итераций, достаточное при произвольном начальном приближении для достижения заданной точности с помощью метода последовательных приближений или приведенности его модификации, хотя и может считаться функцией меры точности и спектральных радиусов $\rho(a)$, $\rho(f)$ и $\rho(Fb)$, но, конечно, более сложного вида, чем рассмотренные показатели $\sigma(a)$ и $v_k(a)$. Между тем указанное число итераций, точнее его оценка, также могло бы рассматриваться в качестве простейшего показателя трудоемкости решения системы (6) этими методами. Поэтому задача построения числовой характеристики трудоемкости нахождения решения системы уравнений межотраслевого баланса тем или иным из рассматриваемых методов не может считаться решенной и, как уже указывалось, нуждается в дальнейшем исследовании.

АЛГОРИТМ ПРИВЕДЕНИЯ МАТРИЦ К БЛОЧНО-ТРЕУГОЛЬНОМУ ВИДУ И ЕГО ИСПОЛЬЗОВАНИЕ ПРИ РЕШЕНИИ ЗАДАЧ ЛИНЕЙНОЙ АЛГЕБРЫ

4. При решении систем линейных уравнений большой размерности, а также других задач линейной алгебры существенные упрощения могут быть достигнуты с помощью предварительного анализа расположения в рассматриваемых матрицах нулевых элементов и последующего использования выясненных таким образом особенностей строения этих матриц.

При таком анализе основными понятиями являются понятия неразложимости и неприводимости матриц. Неразложимой матрицей называется

такая квадратная матрица (a) , которая с помощью одновременной перестановки строк и столбцов не может быть преобразована в матрицу вида

$$\begin{pmatrix} (b)_{k \times k} & (0)_{k \times l} \\ (d)_{l \times k} & (e)_{l \times l} \end{pmatrix},$$

где $(b)_{k \times k}$, $(e)_{l \times l}$ — квадратные матрицы ($k, l \geq 1$), $(0)_{k \times l}$ — матрица с нулевыми элементами. В определении неприводимой матрицы добавляется требование

$$(d)_{l \times k} = (0)_{l \times k}.$$

Одновременной перестановкой строк и столбцов произвольная матрица (a) преобразуется к блочно-треугольному виду (a^∇) с неразложимыми блоками по главной диагонали, нулевыми матрицами выше блочной диагонали и выделенными неприводимыми суперблоками. Заметим, что неразложимая матрица по определению оказывается приведенной к блочно-треугольному виду.

Выделение неприводимых суперблоков приводит к распадению системы линейных уравнений с матрицей коэффициентов (a) на независимые подсистемы меньшего порядка. Блочно-треугольная структура неприводимой матрицы (a)

$$(a) = \begin{pmatrix} (a^I, I) & & & & \\ (a^{II}, I) & (a^{II}, II) & & & \\ (a^{III}, I) & (a^{III}, II) & (a^{III}, III) & & \\ \dots & \dots & \dots & \dots & \\ (a^0, I) & (a^0, II) & (a^0, III) & \dots & (a^0, 0) \end{pmatrix} \begin{matrix} \\ \\ \\ \\ 0 \end{matrix}$$

позволяет находить решение системы уравнений с невырожденной матрицей коэффициентов (a) , последовательно решая подсистемы вида

$$(a^{J, J})x^J = \tilde{b}^J,$$

где $\tilde{b}^J = b^J - \sum_{K < J} \{a^{J, K}\}x^K$ и x^J , b^J — векторы, состоящие из координат векторов x и b , соответствующих диагональному блоку $(a^{J, J})$.

Для матриц с выделенными неприводимыми и неразложимыми блоками очевидным образом упрощается нахождение обратных матриц, решение полной и частной проблем собственных значений. Поэтому предварительное приведение матриц к блочно-треугольному виду при решении задач линейной алгебры большой размерности вполне оправдано. Но определения неразложимости и неприводимости матриц еще не дают эффективного алгоритма приведения к блочно-треугольному виду матриц достаточно высоких порядков. Этим объясняется необходимость разработки алгоритмов приведения матриц к блочно-треугольному виду, пригодных для реализации на ЭВМ в расчете на системы с сотнями и тысячами неизвестных.

Примеры подобных алгоритмов рассматривались в [6—8]. Они по существу решают задачу анализа связности вершин некоторого графа или задачу классификации состояний однородных цепей Маркова.

Связь этих задач с задачей приведения матриц к блочно-треугольному виду, а также некоторые общепринятые для графов и цепей Маркова термины (см. [9], стр. 7—16) мы будем постоянно использовать.

Приведем некоторые необходимые определения.

Матрице n -го порядка $(a)_{n \times n}$ поставим в соответствие граф $\Gamma(a)$ с n вершинами V_i ($i = 1, 2, \dots, n$), у которого по определению существует дуга (V_i, V_j) с началом V_i и концом V_j тогда и только тогда, когда $a_{ij} \neq 0$,

и такая дуга единственна. Для графа $\Gamma(a)$ матрица с элементами

$$\Phi_0(a)_{ij} = \begin{cases} 0 & \text{при } a_{ij} = 0, \\ 1 & \text{при } a_{ij} \neq 0 \end{cases}$$

является матрицей смежности.

Вершины V_i и V_j назовем эквивалентными, если существует как путь (по дугам) из V_j в V_i , так и путь из V_i в V_j . Множество вершин графа $\Gamma(a)$ разбивается на такие непересекающиеся классы эквивалентных между собой вершин, что вершины из разных классов уже не являются эквивалентными. Эти классы вершин будем называть классами эквивалентности и обозначать (после присвоения им номеров) через K_I ($I = 1, 2, \dots, m$). Очевидно, что выделение для матрицы (a) наибольших неразложимых блоков сводится к нахождению классов эквивалентности вершин графа $\Gamma(a)$.

Классы K_I и K_J будем называть несвязанными, если в $\Gamma(a)$ не существует ни одной цепи (по ребрам, а не дугам), включающей одновременно хотя бы по одной вершине из этих классов. Классы K_I и K_J не связаны тогда и только тогда, когда в $\Gamma(a + a^T)$ не существует путей из $V_i \in K_I$ в $V_j \in K_J$.

Множества вершин графа $\Gamma(a)$, образующих классы эквивалентности для $\Gamma(a + a^T)$, будем называть классами связности для графа $\Gamma(a)$ и обозначать их через H^α . Каждый класс связности состоит из всех вершин нескольких (возможно, одного) классов эквивалентности. Распадению множества вершин графа $\Gamma(a)$ на непересекающиеся классы связности соответствует, как легко видеть, выделение в матрице (a) наибольших неприводимых подматриц, опирающихся на главную диагональ, а в $\Gamma(a)$ — выделению компонент связности.

Для классов K_I и K_J из одного класса связности можно ввести отношение предшествования или отношения равноправия. По определению классы K_I предшествуют классу K_J ($K_I \leq K_J$), если существует путь из $V_i \in K_I$ в $V_j \in K_J$. Классы K_I и K_J из одного класса связности равноправны ($K_I \rightarrow K_J$), если между ними нет отношения предшествования.

Пользуясь приведенными определениями, дадим точное определение преобразования матрицы к блочно-треугольному виду.

Матрицу $(a^\nabla)_{n \times n}$ назовем нижним блочно-треугольным видом матрицы $(a)_{n \times n}$, если

а) $a^\nabla = P^T a P$, где P — матрица перестановки элементов $1, \dots, n$, т. е. квадратная матрица n -го порядка, в каждой строке и каждом столбце которой стоит только один ненулевой элемент, равный $+1$ (матрицы (a) и (a^∇) подобны, так как $P^T = P^{-1}$);

б) матрица (a^∇) разбивается на блоки согласно разбиению множества $(1, \dots, n)$ на m непересекающихся подмножеств:

$$\Phi_1 = (1, 2, \dots, n_1), \quad \Phi_2 = (n_1 + 1; \dots; n_1 + n_2), \dots, \\ \dots, \quad \Phi_m = (n_1 + \dots + n_{m-1}; \dots; n)$$

таких, что для графа $\Gamma(a^\nabla)$ множества вершин V_i при $i \in \Phi_I$ ($I = 1, 2, \dots, m$) являются классами эквивалентности K_I ($I = 1, 2, \dots, m$);

в) разбиение множества вершин графа $\Gamma(a^\nabla)$ на классы связности H^α ($\alpha = 1, 2, \dots, p$) задается в виде

$$K_I \subset H^\alpha \quad \text{при} \quad I = \left(\sum_{r=0}^{\alpha-1} m_r \right) + 1, \quad \left(\sum_{r=0}^{\alpha-1} m_r \right) + 2, \dots, \left(\sum_{r=0}^{\alpha} m_r \right) \\ \left(m_0 = 0, m_r > 0 \text{ при } r = 1, 2, \dots, p; \sum_{r=0}^p m_r = m \right);$$

г) для $K_I \subset H^\alpha$, $K_J \subset H^\alpha$ при $I < J$ имеет место одно из двух отношений: $K_I \leq K_J$ или $K_I \rightarrow \leftarrow K_J$.

Данное определение не дает возможности построить по матрице (a) единственную матрицу (a^∇) , так как оставляет произвольным порядок классов связности и порядок вершин графа $\Gamma(a)$ из одного класса эквивалентности (номера соответствующих строк и столбцов матрицы (a) после преобразования в матрицу (a^∇)). В известной степени остается произвольной и нумерация классов эквивалентности, входящих в один класс связности. Конкретные алгоритмы приводят матрицу (a) к одному из ее нижних блочно-треугольных видов.

5. Нахождение матрицы (a^∇) по известной матрице (a) естественно заменить нахождением матрицы $\varphi_0(a^\nabla)$ по матрице $\varphi_0(a)$, так как каждый элемент матриц $\varphi_0(a^\nabla)$ и $\varphi_0(a)$ равен или 0, или 1 и для него в памяти ЭВМ достаточно отвести всего лишь один двоичный разряд. Любой алгоритм приведения матриц к блочно-треугольному виду должен решать четыре задачи, а именно:

- а) выделять классы эквивалентности;
- б) определять классы связности;
- в) определять порядок нумерации классов эквивалентности в классах связности, соответствующий блочно-треугольному виду матрицы;
- г) преобразовывать матрицу (a) в матрицу (a^∇) .

Выделение классов эквивалентности наиболее естественно вести по матрице $\Psi(a)$ связности вершин графа $\Gamma(a)$, для которой по определению $\Psi_{ij}(a) = 1$, если в $\Gamma(a)$ существует путь из V_i в V_j , и $\Psi_{ij}(a) = 0$, если такого пути не существует.

Очевидно, что для вершин V_i и V_j из одного класса эквивалентности и только для них

$$\Psi_{ij}(a)\Psi_{ji}(a) = 1. \quad (18)$$

Поэтому первый этап предлагаемого алгоритма состоит в построении матрицы связности $\Psi(a)$ по известной матрице смежности $\varphi_0(a)$.

Один из наиболее очевидных, но не самый эффективный способ нахождения матрицы $\Psi(a)$, заключается в использовании легко проверяемой формулы:

$$\Psi(a) = \varphi_0\left(\sum_{k=1}^r \varphi_0^k(a)\right), \quad (19)$$

где r определяется из

$$\varphi_0\left(\sum_{k=1}^{r-1} \varphi_0^k(a)\right) \neq \varphi_0\left(\sum_{k=1}^r \varphi_0^k(a)\right) = \varphi_0\left(\sum_{k=1}^{r+1} \varphi_0^k(a)\right).$$

В практических расчетах формула (19) преобразуется в формулу последовательного нахождения матриц $\varphi_0\left(\sum_{k=1}^l \varphi_0^k(a)\right)$. Такой процесс может быть задан, например, в виде

$$g_1 = \varphi_0(a), \quad g_2 = \varphi_0((E + g_1)\varphi_0(a)), \dots, \quad g_{l+1} = \varphi_0((E + g_l)\varphi_0(a)), \quad (20)$$

или

$$h_0 = \varphi_0(a), \quad h_1 = \varphi_0((E + h_0)h_0), \dots, \quad h_{l+1} = \varphi_0((E + h_l)h_l), \quad (21)$$

где, как легко видеть,

$$g_l = \varphi_0\left(\sum_{k=1}^l \varphi_0^k(a)\right) \quad \text{и} \quad h_l = \varphi_0\left(\sum_{k=1}^{2^l} \varphi_0^k(a)\right).$$

Но даже если при перемножении матриц $(E + g_i)$ и $\varphi_0(a)$ или $(E + h_i)$ и h_i сразу находить элементы матриц g_{i+1} или h_{i+1} (логическим умножением i -й строки первого сомножителя на j -й столбец второго сомножителя до первого совпадения в них единиц на k -м месте), то и в этом случае приведенные способы нахождения $\Psi(a)$ для матриц (a) высоких порядков недостаточно удобны.

Во-первых, они связаны с большим числом логических операций, а во-вторых, с большим размером рабочих полей в памяти ЭВМ. Так, если не пользоваться внешними запоминающими устройствами и разбиением обрабатываемых матриц на блоки и зоны, то для реализации формул (20) и (21) потребуется рабочий массив ячеек памяти по крайней мере для двух матриц n -го порядка. Существуют приемы, позволяющие, усложнив алгоритм, сократить необходимый объем рабочего массива. Но более эффективным является следующий подход к нахождению матрицы $\Psi(a)$.

Существует связь между методами вычисления матрицы $\Psi(a)$ по $\varphi_0(a)$ и методами обращения матриц. Легко показать, пользуясь определением матрицы $\Psi(a)$, что

$$\Psi(a) = \varphi_0((E - \varepsilon\varphi_0(a))^{-1} - E),$$

где $0 < \varepsilon < 1/n$. В матрице $[(E - \varepsilon\varphi_0(a))^{-1} - E]$ для нас интересны не величины ее элементов, а только факт их отличия от нуля. Поэтому для нахождения матрицы $\Psi(a)$ имеет смысл так видоизменить какой-нибудь из известных итерационных алгоритмов нахождения обратных матриц, чтобы на каждом его этапе вычислялась матрица $\varphi_0(B_k - E)$, где B_k — k -е приближение к матрице $(E - \varepsilon\varphi_0(a))^{-1}$. При этом сами матрицы B_k могут не вычисляться, а значение параметра ε не фиксироваться. Процесс (20) является в этом смысле модификацией метода последовательных приближений для нахождения матрицы $(E - b)^{-1}$ при $\rho(b) < 1$, задаваемого в виде

$$B_1 = b, \quad B_2 = (E + B_1)b, \dots,$$

$$B_{k+1} = (E + B_k)b, \dots \{B_k \rightarrow (E - b)^{-1} \text{ при } k \rightarrow \infty\}.$$

Процесс, определенный формулами (21), является модификацией уже упоминавшегося в п. 1 метода нахождения обратных матриц уточнением элементов приближенно обратной матрицы.

Для нахождения матрицы $\Psi(a)$ можно модифицировать в указанном выше смысле не только итерационные, но и точные методы обращения матриц. В [8] показано, что на основе алгоритма нахождения матрицы $(E - b)^{-1}$ методом Жордана — Гаусса, т. е. методом исключения с единственным делением без обратного хода, может быть сконструирован следующий процесс для получения матрицы $\Psi(a)$:

$$(N^{(0)}) = \varphi_0(a), \quad (N^{(t+1)}) = (N_{ij}^{(t+1)}), \quad \left. \begin{array}{l} \text{где} \\ N_{ij}^{(t+1)} = \max(N_{ij}^{(t+1)}, N_{i, t+1}^{(t)} N_{t+1, j}^{(t)}) \\ (i, j = 1, 2, \dots, n; \quad t = 0, 1, \dots, n-1) \end{array} \right\} \quad (22)$$

$$(N^{(n)}) = \Psi(a). \quad (23)$$

Формула (23) доказывается непосредственно в терминах теории графов без использования аналогии с обращением матриц. Приведем это доказательство с некоторыми сокращениями.

Через Π_{ij} обозначим множество путей в графе $\Gamma(a)$ из вершины V_i в вершину V_j . Докажем, что элемент $N_{ij}^{(n)}$ матрицы $(N^{(n)})$ обязательно

равен 1, если множество Π_{ij} не пусто, и равен 0 в противном случае. Легко видеть, что в непустом множестве Π_{ij} существует путь, задаваемый последовательностью вершин $\{V_i, V_{i_2}, \dots, V_{i_s}, V_j\}$ ($s \geq 0$), у которого все вершины $V_{i_2}, \dots, V_{i_s}$ различны и отличны от вершин V_i и V_j (V_j может совпадать с V_i) и для которого минимален (по всем путям из Π_{ij}) наибольший номер входящей в него вершины, не являющейся началом и концом пути. Этот номер, рассматриваемый как функция от i и j , обозначим через $z(i, j)$ и будем называть весом пары вершин V_i и V_j при заданной нумерации вершин графа $\Gamma(a)$. По определению примем $z(i, j) = 0$ для вершин V_i и V_j с $(\varphi_0(a))_{ij} = 1$ и $z(i, j) = n + 1$ в случае пустого множества Π_{ij} .

Тогда формула (23) получается как следствие более общего предложения: для упорядоченной пары вершин (V_i, V_j) с $z(i, j) < n + 1$ имеем

$$N_{ij}^{(0)} = N_{ij}^{(1)} = \dots = N_{ij}^{(z(i, j)-1)} = 0, \quad (24)$$

$$N_{ij}^{(z(i, j))} = N_{ij}^{(z(i, j)+1)} = \dots = N_{ij}^{(n)} = 1.$$

Сформулированное утверждение доказывается индукцией по $z(i, j)$. Доказательство, которое мы здесь не приводим, основывается на том, что $N_{ij}^{(t+1)} = 1$ (при $N_{ij}^{(t)} = 0$) тогда и только тогда, когда $N_{i, t+1}^{(t)} = N_{t+1, j}^{(t)} = 1$.

Простота реализации на ЭВМ алгоритма нахождения матрицы связности вершин $\Psi(a)$ с помощью формул (22) очевидна.

По известной матрице $\Psi(a)$ легко выделяются классы эквивалентности вершин с помощью (18). Классы связности находятся как классы эквивалентности для графа $\Gamma(a + a^T)$. При этом удобно использовать матрицу $\hat{\Psi}(a)$, которая получается из $\Psi(a)$ вычеркиванием строк и столбцов, соответствующих всем вершинам каждого класса эквивалентности, кроме одной. Тогда по матрице $\varphi_0(\hat{\Psi}(a) + \hat{\Psi}^T(a))$ с помощью приведенного выше алгоритма находится матрица $\hat{\Psi}(a + a^T)$, для которой $\hat{\Psi}_{IJ}(a + a^T) = 1$ тогда и только тогда, когда K_I и K_J входят в один класс связности.

Порядок нумерации классов эквивалентности в классах связности, соответствующий определению нижнего блочно-треугольного вида матриц, находится следующим образом. Через $\hat{\Psi}^\alpha(a)$ обозначим матрицу, стоящую на пересечении строк и столбцов матрицы $\hat{\Psi}(a)$, соответствующих классам эквивалентности $K_I \subset H^\alpha$. Матрицу $\hat{\Psi}^\alpha(a)$ можно называть матрицей связности классов эквивалентности из H^α .

Из определения классов эквивалентности очевидно, что в графе $\Gamma(\hat{\Psi}^\alpha(a))$ не существует контуров (см. [9], стр. 16—18), включающих не менее двух различных вершин (в $\Gamma(\hat{\Psi}^\alpha(a))$, роль вершин играют классы эквивалентности $K_I \subset H^\alpha$ графа $\Gamma(a)$). Поэтому в графе $\Gamma(\hat{\Psi}^\alpha(a))$ всегда найдется хотя бы одна вершина — тупик V_i , для которой нет ни одной дуги, имеющей V_i своим началом, кроме, быть может, петли $\{V_i, V_i\}$.

Это утверждение, доказательство которого опускаем ввиду его тривиальности, позволяет использовать следующий алгоритм выбора нумерации классов эквивалентности:

а) находим первую по порядку строку матрицы $\Psi^\alpha(a)$, состоящую из нулей вне главной диагонали (такая строка существует, так как в $\Gamma(\hat{\Psi}^\alpha(a))$ существует вершина — тупик); классу эквивалентности, соответствующему этой строке, присваиваем очередной номер I ;

б) вычеркивая из $\hat{\Psi}^{\alpha}(a)$ строку и столбец, соответствующие выделенному ранее классу эквивалентности, получаем новую матрицу, также являющуюся матрицей смежности некоторого графа без контуров, не являющихся петлями; к полученной матрице применяем процедуру пункта а) и т. д. до перенумерации всех классов эквивалентности $K_I \subset H^{\alpha}$.

Очевидно, что в результате перестановки строк и столбцов матрицы (а) согласно полученному порядку классов эквивалентности (с точностью до перенумерации строк и столбцов, соответствующих вершинам из одного класса эквивалентности, и до порядка классов H^{α}) матрица (а) приводится к нижнему блочно-треугольному виду.

Описанный алгоритм может быть использован и для анализа расположения нулевых элементов в прямоугольных матрицах. Для этого исходная матрица дополняется до квадратной введением нулевых строк или столбцов.

Программа, реализующая изложенный алгоритм на ЭВМ «Урал-2» и «Урал-4», составлена сотрудником НИЭИ при Госплане СССР Н. А. Толмачевой. Эта программа [10], используя только оперативную память и вводимая целиком (возможна экономия места под программу путем ее расчленения на части, вводимые на один и тот же рабочий массив памяти), позволяет находить матрицу $\Phi_0(a^{\nabla})$ для произвольной квадратной матрицы (а) до 244-го порядка включительно.

При анализе матриц больших порядков, возникающих в практических задачах, алгоритм приведения матриц к блочно-треугольному виду должен использоваться в комплексе с другими приемами. Так, в рассматриваемых матрицах целесообразно каким-либо из приближенных приемов выделить малые или не важные в каком-либо смысле элементы и лишь для матрицы с оставшимися элементами выделять неприводимые и неразложимые блоки.

Блочный вид получаемой матрицы может использоваться для упрощения решения задач различными способами, в том числе и изложенными в пп. 1—2. Непосредственное отношение выделение блоков имеет к использованию приемов итеративного агрегирования, один из которых рассматривался в работах [11] и [1].

ЛИТЕРАТУРА

1. В. А. Волконский. Использование специальной структуры матриц для решения задач планирования. Экономика и матем. методы, 1966, т. II, вып. 2.
2. Д. К. Фаддеев, В. Н. Фаддеева. Вычислительные методы линейной алгебры. М.—Л., Физматгиз, 1960.
3. Д. Гейл. Теория линейных экономических моделей. М., Изд-во иностр. лит., 1963.
4. С. Карлин. Математические методы в теории игр, программировании и экономике. М., «Мир», 1964.
5. Э. Б. Ершов. Итеративные методы в моделях межотраслевого баланса. НИЭИ при Госплане СССР, 1966 (ротапринт).
6. Т. А. Сарымсаков. Основы теории процессов Маркова. М., Гостехиздат, 1954.
7. K. Wenke. Praktische Anwendung linearer Wirtschaftsmodelle. Unternehmensforschung, 1964, Bd. 8, N 1. (РЖ Экономика промышленности, 1964, № 10, 10Д2).
8. Практическое применение линейных хозяйственных моделей. НИЭИ при Госплане СССР, 1965 (ротапринт).
9. Э. Б. Ершов. Приведение матриц к блочно-треугольному виду и анализ связности вершин графов. НИЭИ при Госплане СССР, 1965 (ротапринт).
10. К. Берж. Теория графов и ее применения. М., Изд-во иностр. лит., 1962.
11. Н. А. Толмачева. Программа приведения матриц к блочно-треугольному виду. НИЭИ при Госплане СССР, 1965 (ротапринт).
12. Л. М. Дудкин, Э. Б. Ершов. Межотраслевой баланс и материальные балансы отдельных продуктов. Плановое хозяйство, 1965, № 5.

Поступила в редакцию
25 VIII 1965