



# Прикладная математика

СКИФ



Кафедра «Эксплуатация транспортных  
систем и логистика»

**Лекционный курс**

**Автор**

**Короткий А. А.**

Ростов-на-Дону,  
2018

## **Аннотация**

Лекционный курс предназначен предназначен для студентов очной, заочной формы обучения по направлениям 23.04.02 Наземные транспортно-технологические комплексы

## **Автор**

**Короткий Анатолий Аркадьевич –**

**д.т.н., профессор кафедры «ЭТСиЛ»**

## ОГЛАВЛЕНИЕ

2.1. СТРУКТУРА И СВОЙСТВА ДВОЙСТВЕННОЙ ЗАДАЧИ.....	4
3.1 ОПТИМИЗАЦИЯ БЕЗ ОГРАНИЧЕНИЙ. ГРАДИЕНТНЫЙ СПУСК.....	10
6.1 ПАРНЫЕ И МНОЖЕСТВЕННЫЕ КОРРЕЛЯЦИИ. НЕЛИНЕЙНЫЕ РЕГРЕССИИ. ....	16
4.1 РАСПЛЫВЧАТЫЕ ЦЕЛИ. ....	114
5.1 КОНКУРЕНТНЫЕ МОДЕЛИ . МОДЕЛИ ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ.....	119
7.1 ДЕТЕРМИНИРОВАННЫЕ ВРЕМЕННЫЕ РЯДЫ. ....	135

## 2.1. СТРУКТУРА И СВОЙСТВА ДВОЙСТВЕННОЙ ЗАДАЧИ

Задачу максимизации ЛП с экономической точки зрения можно рассматривать как задачу о распределении ограниченных ресурсов  $b_1, \dots, b_m$  между различными потребителями, например между определенными технологическими процессами, которые представляются столбцами  $A_1, \dots, A_n$  матрицы ограничений задачи.

Любое допустимое решение задачи ЛП  $x_1, \dots, x_n$  дает конкретное распределение, которое указывает ту долю каждого из ресурсов, которая должна быть использована при осуществлении соответствующего технологического процесса.

Рассмотрим **пример**. Предприятие производит три вида продукции  $x_1$ ,  $x_2$  и  $x_3$ , каждый из которых проходит обработку на токарном, фрезеровальном и сверлильном станках. Общий фонд машинного времени для каждого из станков ограничен. Пусть  $c_1$ ,  $c_2$  и  $c_3$  - прибыль от реализации единицы соответствующего вида продукции. Необходимо определить, какое количество продукции каждого вида следует производить каждую неделю, чтобы обеспечить максимальную прибыль.

Эта задача имеет такой вид:

$$\text{максимизировать } c_1x_1 + c_2x_2 + c_3x_3 \quad (2.1.1)$$

при ограничениях

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{13}x_3 &\leq b_1; \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{23}x_3 &\leq b_2; \\ a_{31}x_1 + a_{32}x_2 + \dots + a_{33}x_3 &\leq b_3. \end{aligned} \quad (2.1.2)$$

где  $a_{1j}$ ,  $a_{2j}$ ,  $a_{3j}$  - время, необходимое для обработки единицы  $j$ -го вида продукции на токарном, фрезеровальном и сверлильном станках соответственно ( $j=1, 2, 3$ ),  $b_1$ ,  $b_2$ ,  $b_3$  - недельный ресурс машинного времени для группы токарных, фрезеровальных и сверлильных станков соответственно.

Обозначим через  $y_1$ ,  $y_2$ ,  $y_3$  - цену единицы времени работы токарного, сверлильного и фрезеровального оборудования.

Тогда  $a_{11}y_1 + a_{21}y_2 + a_{31}y_3$  можно трактовать как затраты на изготовление единицы продукции первого вида;

Предположим, что цены ресурсов  $y_1$ ,  $y_2$ ,  $y_3$  выбраны так, что выполняются соотношения

Прикладная математика

$$\begin{aligned} a_{11}y_1 + a_{12}y_2 + \dots + a_{13}y_3 &\geq c_1; \\ a_{21}y_1 + a_{22}y_2 + \dots + a_{23}y_3 &\geq c_2; \\ a_{31}y_1 + a_{32}y_2 + \dots + a_{33}y_3 &\geq c_3. \end{aligned} \quad (2.1.3.)$$

Поскольку  $b_1, b_2, b_3$  - общий использованный ресурс машинного времени для каждого из станков, то  $b_1y_1 + b_2y_2 + b_3y_3$  - общие затраты на производство (общая стоимость использованных ресурсов).

Тогда можно сформулировать следующую задачу.

Требуется определить цены  $y_1, y_2, y_3$ , удовлетворяющие условиям (2.1.3), при которых минимизируются суммарные затраты на производство, а именно:

$$\text{минимизировать } g(y_1, y_2, y_3) = b_1y_1 + b_2y_2 + b_3y_3 \quad (2.1.4.)$$

при ограничениях (2.1.3) и

$$y_1 \geq 0, y_2 \geq 0, y_3 \geq 0.$$

Задачу (2.1.3), (2.1.4) называют двойственной относительно задачи (2.1.1), называемой прямой.

Запишем прямую и двойственную задачи в общем виде.

Прямая задача:

$$\text{максимизировать } \sum_{j=1}^n c_j x_j \quad (2.1.5.)$$

при ограничениях

$$\sum_{j=1}^n a_{ij} x_j, i = 1, 2, \dots, m, \quad (2.1.6.)$$

$$x_j \geq 0, j = 1, 2, \dots, n. \quad (2.1.7.)$$

Двойственная задача:

$$\text{минимизировать } \sum_{i=1}^m b_i y_i \quad (2.1.8.)$$

при ограничениях

$$\sum_{i=1}^m a_{ij} y_i \geq c_j, j = 1, 2, \dots, n, \quad (2.1.9.)$$

$$y_i \geq 0, i = 1, 2, \dots, m. \quad (2.1.10.)$$

В матричном виде пара двойственных задач записывается следующим образом:

Прямая задача:

Прикладная математика

$$\text{максимизировать } c^T x \quad (2.1.11.)$$

при ограничениях

$$Ax \leq b; \quad (2.1.12.)$$

$$x \geq 0 \quad (2.1.13.)$$

Двойственная задача:

$$\text{максимизировать } b^T y \quad (2.1.14.)$$

при условиях

$$A^T y \geq c; \quad (2.1.15.)$$

$$y \geq 0 \quad (2.1.16.)$$

Сопоставляя формы записи прямой и двойственной задач, можно установить между ними следующие взаимосвязи.

1. Если прямая задача является задачей максимизации, то двойственная будет задачей минимизации, и наоборот.
2. Коэффициенты целевой функции прямой задачи  $c_1, \dots, c_n$  становятся свободными членами ограничений двойственной задачи.
3. Свободные члены ограничений прямой задачи  $b_1, \dots, b_m$  становятся коэффициентами целевой функции двойственной задачи.
4. Матрица ограничений двойственной задачи получается путем транспонирования матрицы ограничений прямой задачи.
5. Знаки неравенств в ограничениях изменяются на противоположные.
6. Число ограничений прямой задачи равно числу переменных двойственной задачи, и наоборот.
7. Переменные  $y_1, \dots, y_m$  двойственной задачи иногда называют теневыми ценами.
8. Двойственную задачу выгоднее решать, чем прямую, если в прямой задаче при малом количестве переменных имеется большое количество ограничений ( $m > n$ ).
9. Связь между оптимальными решениями прямой и двойственной задач устанавливают, анализируя следующие теоремы теории двойственности.
10. **Теорема 2.1.1.** Если  $x_0$  и  $y_0$  допустимые решения прямой и двойственной задач, то есть  $Ax_0 \leq b$  и  $A^T y_0 \geq c$ , то

$$c^T x_0 \leq b^T y_0, \quad (2.1.17)$$

11. то есть значения целевой функции прямой задачи никогда не превышают значения целевой функции двойственной задачи.
12. **Доказательство.** Умножим выражение (2.1.12) на  $y_0^T$ , получим

Прикладная математика

$$y_0^T A x_0 \leq y_0^T b. \quad (2.1.18)$$

13. Аналогично умножим (2.1.15) на  $x_0^T$ :

$$x_0^T A^T y_0 \geq x_0^T c. \quad (2.1.19)$$

14. Но  $y_0^T A x_0 = (y_0^T A x_0)^T = x_0^T A^T y_0$ , а кроме того  $x_0^T c = c^T x_0$ .

15. Поэтому, сравнивая (2.1.19) и (2.1.18), получим

$$y_0^T b \geq y_0^T A x_0 \geq x_0^T c \text{ или } c^T x_0 \leq b^T y_0.$$

17. Теорема доказана.

18. **Теорема 2.1.2. (основная теорема двойственности).** Если  $x_0$  и  $y_0$  допустимые решения прямой и двойственной задач и кроме того, если  $c^T x_0 = b^T y_0$ , то  $x_0$  и  $y_0$  - оптимальные решения пары двойственных задач.

19. **Доказательство.** Согласно теореме 2.1.1 для всех допустимых решений  $x$  и  $y$  справедливо неравенство (2.1.17). В частности, для всех допустимых решений  $x$  справедливо  $c^T x \leq b^T y_0$ . Однако из условия теоремы  $c^T x_0 = b^T y_0$  следует  $c^T x_0 \leq c^T x_0$ . Следовательно,  $x_0$  - оптимальное решение.

20. Вторая часть теоремы доказывается аналогично.

21. В силу теоремы 2.1.1 для всех допустимых  $y$  справедливо  $c^T x_0 \leq b^T y$ . Но из условия  $c^T x_0 = b^T y_0$  следует, что  $b^T y \geq b^T y_0$  для всех  $y \geq 0$ . Таким образом,  $y_0$  - оптимальное решение.

22. **Теорема 2.1.3.** Если в оптимальном решении прямой задачи (2.1.5) - (2.1.7)  $i$ -тое ограничение выполняется как строгое неравенство, то оптимальное значение соответствующей двойственной переменной равно нулю, то есть

$$\text{если } \sum_{j=1}^n a_{ij} x_j \text{ опт} = A^I x \text{ опт} < b_i, \text{ то } y_i \text{ опт} = 0, \quad (2.1.20)$$

23. где  $A_i$  -  $i$ -я строка матрицы  $A$ .

24. Смысл теоремы 2.1.3 состоит в следующем. Если некоторый ресурс  $b_i$  имеется в избытке, и  $i$ -тое ограничение при оптимальном решении выполняется как строгое неравенство, то это ограничение становится несущественным, и оптимальная цена соответствующего ресурса равна нулю.

25. Теорему 2.1.3. дополняет теорема 2.1.4, устанавливающая взаимосвязь между оптимальным решением прямой задачи и ограничениями двойственной.

26. **Теорема 2.1.4.** Если в оптимальном решении двойственной задачи ограничение  $j$  выполняется как строгое неравенство, то оптимальное значение соответствующей переменной прямой задачи должно быть равно нулю, то есть

Прикладная математика

$$\text{если } A - j^T y_{\text{опт}} - c_j > 0, \text{ то } x_j_{\text{опт}} = 0. \quad (2.1.21)$$

27. Дадим экономическую интерпретацию теоремы 2.1.4.

28. Поскольку величины  $y_i$  ( $i=1,2,..,m$ ) представляют собой

цены соответствующих ресурсов, то  $A_j^T y = \sum_{i=1}^n a_{ij} y_i$  - это затраты на  $j$ -й технологический процесс, а величина  $c_j$  - прибыль от реализации единицы соответствующего продукта. Поэтому с экономической точки зрения теорема 2.1.4 означает следующее: если  $j$ -й технологический процесс оказывается строго невыгодным относительно оптимальных цен ресурсов  $y_{\text{опт}}$ , то в оптимальном решении прямой задачи интенсивность использования данного технологического процесса должна быть равна нулю, и соответствующий вид продукции не выпускается как нерентабельный.

29. Таким образом, теорема 2.1.4 выражает *принцип рентабельности* для оптимально организованного производства.

30. Из этой теоремы вытекает также, что если  $x_j_{\text{опт}} > 0$ , то

$$A_j^T y_{\text{опт}} - c_j = 0 \quad (2.1.22)$$

31. Предположим, что среди переменных  $x_1, x_2, .., x_n$  прямой задачи есть множество из  $m$  переменных, которые в оптимальном решении прямой задачи имеют ненулевые значения. Пусть, например, такими переменными оказались первые по порядку  $m$  переменных.

32. Тогда на основании уравнения (2.1.22) получаем  $m$  условий рентабельности:

$$\begin{aligned} A_1^T y_{\text{опт}} - c_1 &= 0; \\ A_2^T y_{\text{опт}} - c_2 &= 0; \\ &\dots\dots\dots \\ A_m^T y_{\text{опт}} - c_m &= 0; \end{aligned} \quad (2.1.23)$$

33. где  $A_1^T = (a_{11}, a_{21}, .., a_{m1});$   
 $\dots\dots\dots$

34.  $A_m^T = (a_{1m}, a_{2m}, .., a_{mm}).$

35. Доказательства теорем 2.1.3 и 2.1.4 проведем последовательно.

36. Пусть  $x_{\text{опт}}$  и  $y_{\text{опт}}$  - оптимальные решения прямой и двойственной задач. Поскольку эти решения допустимые, то

$$Ax_{\text{опт}} - b \leq 0; \quad (2.1.24)$$

$$A^T y_{\text{опт}} - c \geq 0. \quad (2.1.25)$$

Прикладная математика

37. Умножив неравенство (2.1.24) на  $y_{опт}^T$ , а неравенство (2.1.25) - на  $x_{опт}^T$ , получим

$$y_{опт}^T Ax_{опт} - y_{опт}^T b \leq 0; \quad (2.1.26)$$

$$x_{опт}^T A^T y_{опт} - x_{опт}^T c \geq 0. \quad (2.1.27)$$

38. Так как в силу теоремы 2.2  $y_{опт}^T b = x_{опт}^T c$  и  $y_{опт}^T Ax_{опт} = x_{опт}^T A^T y_{опт}$ , то выражения (2.1.26), (2.1.27) строго равны нулю.

39. Расписав левую часть неравенства (2.1.26), получим

$$y_{опт}^T (Ax_{опт} - b) = y_{1\text{ опт}} (A^{(1)}x_{опт} - b_1) + y_{2\text{ опт}} (A^{(2)}x_{опт} - b_2) + \dots + y_{m\text{ опт}} (A^{(m)}x_{опт} - b_m) = 0. \quad (2.1.28)$$

40. Поскольку  $y_{i\text{ опт}} \geq 0$  и  $A^{(i)}x_{опт} - b_i \leq 0$  для всех  $i = 1, 2, \dots, m$ , то левая часть уравнения (2.1.28) может быть равна 0 только в том случае, если каждое слагаемое в отдельности равно нулю.

41. Таким образом, для каждого  $i$ , при котором  $A^{(i)}x_{опт} - b_i < 0$ , имеем  $y_{i\text{ опт}} = 0$ , что и требовалось доказать в теореме 2.1.3.

42. Рассмотрим теперь левую часть неравенства (2.1.27), предварительно расписав ее

$$x_{опт}^T A^T y_{опт} - x_{опт}^T c = x_{опт}^T (A^T y_{опт} - c) = x_{1\text{ опт}} (A_1^T y_{опт} - c_1) + x_{2\text{ опт}} (A_2^T y_{опт} - c_2) + \dots + x_{n\text{ опт}} (A_n^T y_{опт} - c_n) = 0, \quad (2.1.29)$$

43. где  $A = [A_1, A_2, \dots, A_n]$ .

44. Так как все  $x_{j\text{ опт}} \geq 0$  и  $A_j^T y_{опт} - c_j \geq 0$  для всех  $j = 1, \dots, n$ , то уравнение (2.1.29) строго равно нулю, если для каждого  $j$ , при котором  $(A_j^T y_{опт} - c_j) > 0$ , соответствующая переменная  $x_{j\text{ опт}}$  равна нулю.

45. Приведем еще две важные теоремы теории двойственности.

46. **Теорема 2.1.5. (теорема существования).** *Прямая и двойственная задачи имеют оптимальные решения тогда и только тогда, когда обе они имеют допустимые решения.*

47. **Теорема 2.1.6. (теорема двойственности).** *Допустимый вектор  $x_0$  оптимальный тогда и только тогда, когда в двойственной задаче имеется такое допустимое решение  $y_0$ , что*

$$c^T x_0 = b^T y_0. \quad (2.1.30)$$

48. Между оптимальными решениями прямой и двойственной задач и элементами индексных строк симплекс-таблиц, соответствующих этим решениям, существует следующая взаимосвязь:

$$\begin{aligned} \Delta_{n+1}^{\text{нр}} &= y_{i \text{ onm}}, \\ -\Delta_{m+j}^{\text{дв}} &= x_{j \text{ onm}}, \\ i &= 1, 2, \dots, m; \quad j = 1, 2, \dots, n, \end{aligned} \quad (2.1.31)$$

49. где  $n$  - количество переменных прямой задачи;  $m$  - количество ее ограничений;

50.  $\Delta_{n+1}^{\text{нр}}$ ,  $\Delta_{m+j}^{\text{дв}}$  - соответствующие элементы индексной строки симплекс-таблицы прямой и двойственной задач соответственно.

51. При этом, если  $n+i$ , где  $1 \leq i \leq m$ , больше числа векторов-столбцов матрицы ограничений расширенной формы соответствующей задачи, то элементы  $\Delta_{n+i}$ ,  $\Delta_{m+j}$  находятся путем циклической перестановки, начиная с элемента  $\Delta_1$ .

### 3.1 ОПТИМИЗАЦИЯ БЕЗ ОГРАНИЧЕНИЙ. ГРАДИЕНТНЫЙ СПУСК.

Рассматривается задача поиска минимума функции  $f(x): \mathbb{R}^n \rightarrow \mathbb{R}$ , записываемая в виде:

$$(1) \quad f(x) \rightarrow \min_{x \in \mathbb{R}^n}$$

Пусть функция  $f(x)$  такова, что можно вычислить ее градиент. Тогда можно применить метод градиентного спуска, описанный в данной статье.

В статье приведены теоремы сходимости метода градиентного спуска, а также рассмотрена его варианты:

- Метод градиентного спуска с постоянным шагом
- Метод градиентного спуска с дроблением шага
- Метод наискорейшего спуска

Метод градиентного спуска

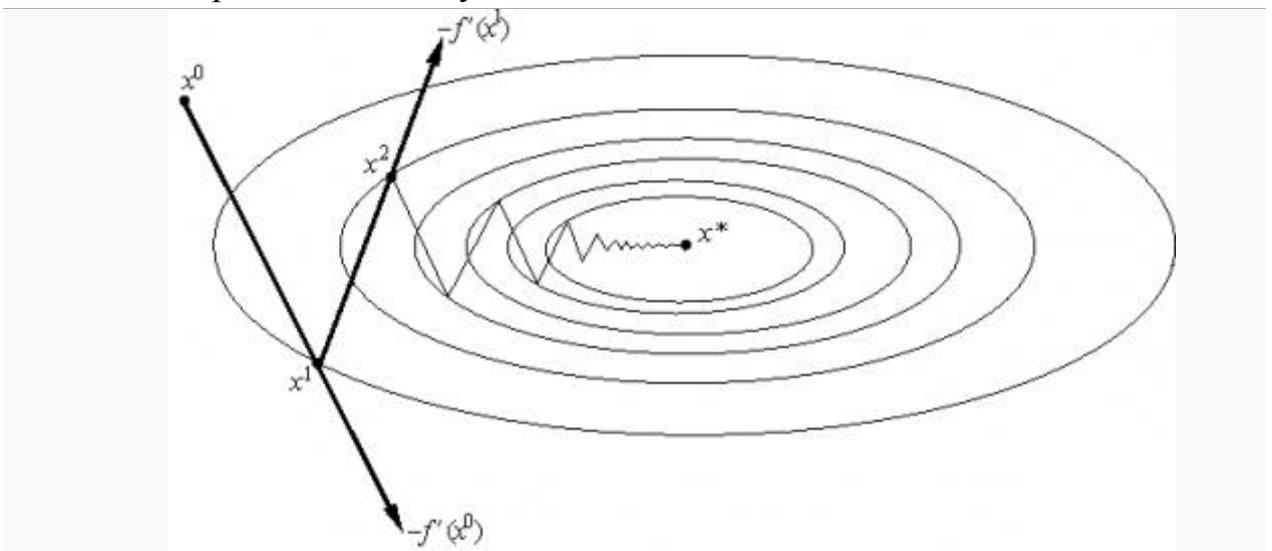




Рис.1 Геометрическая интерпретация метода градиентного спуска с постоянным шагом. На каждом шаге мы сдвигаемся по вектору антиградиента, "уменьшенному в  $\lambda$  раз".

### Идея метода

Основная идея метода заключается в том, чтобы осуществлять оптимизацию в направлении наискорейшего спуска, а это направление задаётся антиградиентом  $-\nabla f$ :

$$x^{[k+1]} = x^{[k]} - \lambda^{[k]} \nabla f(x^{[k]})$$

где  $\lambda^{[k]}$  выбирается

- постоянной, в этом случае метод может расходиться;
- дробным шагом, т.е. длина шага в процессе спуска делится на некое число;

- наискорейшим спуском:  $\lambda^{[k]} = \operatorname{arg\,min}_{\lambda} f(x^{[k]} - \lambda \nabla f(x^{[k]}))$

### Алгоритм

**Вход:** функция  $f: \mathbb{R}^n \rightarrow \mathbb{R}$

**Выход:** найденная точка оптимума  $x$

1. Повторять:
2.  $x^{[k+1]} = x^{[k]} - \lambda^{[k]} \nabla f(x^{[k]})$ , где  $\lambda^{[k]}$  выбирается одним из описанных выше способов
3. если выполнен критерий останова, то возвращаем текущее значение  $x^{[k+1]}$

### Критерий останова

Критерии остановки процесса приближенного нахождения минимума могут быть основаны на различных соображениях. Некоторые из них:

1.  $\|x^{[k+1]} - x^{[k]}\| \leq \epsilon$
2.  $\|f(x^{[k+1]}) - f(x^{[k]})\| \leq \epsilon$

Здесь  $x^{[k]} \in \mathbb{R}^n$  - значение, полученное после  $k$ -го шага оптимизации.  $\epsilon$  - наперед заданное положительное число.

### Сходимость градиентного спуска с постоянным шагом

**Теорема 1 о сходимости метода градиентного спуска с постоянным шагом.**

Пусть  $\lambda^{[k]} = \lambda = \text{const}$ , функция  $f$  дифференцируема, ограничена снизу. Пусть выполняется условие Липшица для градиента  $f'(x)$ :  $\|f'(x) - f'(y)\| \leq L\|x - y\|$ . Пусть  $0 < \lambda < 2/L$ .

Тогда  $\lim_{k \rightarrow \infty} f'(x^{[k]}) = 0$ ,  $f(x^{[k+1]}) < f(x^{[k]})$  при любом выборе начального приближения.

В условиях теоремы градиентный метод обеспечивает сходимость  $\{x^{[k]}\}$  либо к точной нижней грани  $\inf_x f(x)$  (если функция  $f(x)$  не имеет минимума) либо к значению  $f(x^*)$ :  $\lim_{k \rightarrow \infty} x^{[k]} = x^*$ ;  $f'(x^*) = 0$ . Существуют примеры, когда в точке  $x^*$  реализуется седло, а не минимум. Тем не менее, на практике методы градиентного спуска обычно обходят седловые точки и находят локальные минимумы целевой функции.

**Определение.** Дифференцируемая функция  $f$  называется сильно выпуклой (с константой  $\Lambda > 0$ ), если для любых  $x$  и  $y$  из  $\mathbb{R}^n$  справедливо  $f(x+y) \geq f(x) + \langle f'(x), y \rangle + \Lambda \|y\|^2 / 2$ .

**Теорема 2 о сходимости метода градиентного спуска с постоянным шагом.**

Пусть функция  $f$  дифференцируема, сильно выпукла с константой  $\Lambda$ . Пусть выполняется условие Липшица для градиента  $f'(x)$ :  $\|f'(x) - f'(y)\| \leq L \|x - y\|$ . Пусть  $0 < \lambda < 2/L$ .

Тогда  $\lim_{k \rightarrow \infty} x^{[k]} = x^*$ ;  $\|x^{[k]} - x^*\| \leq q^k \|x^{[0]} - x^*\|$ ,  $q = \max\{|1 - \lambda\Lambda|, |1 - \lambda L|\}$

при любом выборе начального приближения.

**Выбор оптимального шага**



Рис.2 Ситуация, когда метод градиентного спуска сходится плохо.

Константа  $q$ , фигурирующая в теореме 2 и характеризующая скорость сходимости метода, зависит от шага  $\lambda$ . Нетрудно видеть, что величина  $q = q(\lambda) = \max\{|1 - \lambda\Lambda|, |1 - \lambda L|\}$  минимальна, если шаг  $\lambda$  выбирается из условия  $|1 - \lambda\Lambda| = |1 - \lambda L|$ , т. е. если  $\lambda = \lambda^* = 2/(\Lambda + L)$ .

При таком выборе шага оценка сходимости будет наилучшей и будет характеризоваться величиной:

$$q = q^* = \frac{L - \Lambda}{L + \Lambda}.$$

В качестве  $\Lambda$  и  $L$  могут выступать равномерные по  $x$  оценки сверху и снизу собственных значений оператора  $f''(x)$ . Если  $\lambda \ll \Lambda$ , то  $q^* \approx 1$  и метод сходится очень медленно. Геометрически случай  $\lambda \ll \Lambda$  соответствует функциям с сильно вытянутыми линиями уровня (см. рис. 2). Простейшим

примером такой функции может служить функция на  $\mathbb{R}^2$ , задаваемая формулой:

$$f(x_1, x_2) = \Lambda x_1^2 + L x_2^2, \quad \Lambda \ll L.$$

Поведение итераций градиентного метода для этой функции изображено на рис. 2 — они, быстро спустившись на "дно оврага", затем медленно "зигзагообразно" приближаются к точке минимума. Число  $\mu = L/\Lambda$  (характеризующее, грубо говоря, разброс собственных значений оператора  $f''(x)$ ) называют числом обусловленности функции  $f$ . Если  $\mu \gg 1$ , то функции называют плохо обусловленными или овражными. Для таких функций градиентный метод сходится медленно.

Но даже для хорошо обусловленных функций проблема выбора шага нетривиальна в силу отсутствия априорной информации о минимизируемой функции. Если шаг выбирается малым (чтобы гарантировать сходимость), то метод сходится медленно. Увеличение же шага (с целью ускорения сходимости) может привести к расходимости метода. Далее будут описаны два алгоритма автоматического выбора шага, позволяющие частично обойти указанные трудности.

### Градиентный метод с дроблением шага

В этом варианте градиентного метода величина шага  $\lambda^{[k]}$  на каждой итерации выбирается из условия выполнения неравенства

$$(2) \quad f(x^{[k+1]}) = f(x^{[k]} - \lambda^{[k]} f'(x^{[k]})) \leq f(x^{[k]}) - \epsilon \lambda^{[k]} \|f'(x^{[k]})\|^2,$$

где  $\epsilon \in (0,1)$  — некоторая заранее выбранная константа.

Процедуру нахождения такого  $\lambda^{[k]}$  обычно оформляют так. Выбирается число  $\delta \in (0,1)$  и некоторый начальный шаг  $\lambda^{[0]}$ . Теперь для каждого  $k$  полагают  $\lambda^{[k]} = \lambda^{[0]}$  и делают шаг градиентного метода. Если с таким  $\lambda^{[k]}$  условие (2) выполняется, то переходят к следующему  $k$ . Если же (2) не выполняется, то умножают  $\lambda^{[k]}$  на  $\delta$  ("дробят шаг") и повторяют эту процедуру до тех пор пока неравенство (2) не будет выполняться. В условиях теоремы 1 эта процедура для каждого  $k$  за конечное число шагов приводит к нужному  $\lambda^{[k]}$ .

Можно показать, что в условиях теоремы 2 градиентный метод с дроблением шага линейно сходится. Описанный алгоритм избавляет нас от проблемы выбора  $\lambda^{[k]}$  на каждом шаге, заменяя ее на проблему выбора параметров  $\epsilon$ ,  $\delta$  и  $\lambda^{[0]}$ , к которым градиентный метод менее чувствителен. При этом, разумеется, объем вычислений возрастает (в связи с необходимостью процедуры дробления шага), впрочем, не очень сильно, поскольку в большинстве задач основные вычислительные затраты ложатся на вычисление градиента.

**Метод наискорейшего спуска**

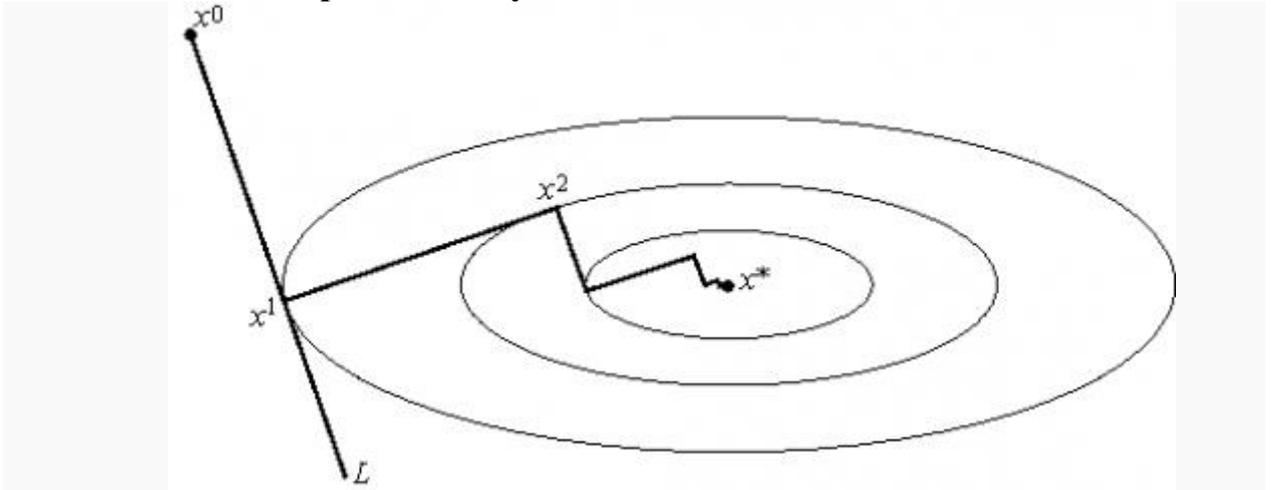


Рис.3 Геометрическая интерпретация метода наискорейшего спуска. На каждом шаге  $\lambda^{[k]}$  выбирается так, чтобы следующая итерация была точкой минимума функции  $f$  на луче  $L$ .

Этот вариант градиентного метода основывается на выборе шага из следующего соображения. Из точки  $x^{[k]}$  будем двигаться в направлении антиградиента до тех пор пока не достигнем минимума функции  $f$  на этом направлении, т. е. на луче  $L = \{x = x^{[k]} - \lambda f'(x^{[k]}); \lambda \leq 0\}$ :

$$\lambda^{[k]} = \arg \min_{\lambda \in [0, \infty)} f(x^{[k]} - \lambda f'(x^{[k]}))$$

Другими словами,  $\lambda^{[k]}$  выбирается так, чтобы следующая итерация была точкой минимума функции  $f$  на луче  $L$  (см. рис. 3). Такой вариант градиентного метода называется методом наискорейшего спуска. Заметим, кстати, что в этом методе направления соседних шагов ортогональны.

Метод наискорейшего спуска требует решения на каждом шаге задачи одномерной оптимизации. Практика показывает, что этот метод часто требует меньшего числа операций, чем градиентный метод с постоянным шагом.

В общей ситуации, тем не менее, теоретическая скорость сходимости метода наискорейшего спуска не выше скорости сходимости градиентного метода с постоянным (оптимальным) шагом.

**Числовые примеры**

**Метод градиентного спуска с постоянным шагом**

Для исследования сходимости метода градиентного спуска с постоянным шагом была выбрана функция:

$$f(x_1, x_2) = 10x_1^2 + x_2^2.$$

Начальное приближение - точка (10,10). Использован критерий останова:  $\|f(x^{[k+1]}) - f(x^{[k]})\| \leq 10^{-5}$

Результаты эксперимента отражены в таблице:

Значение шага $\lambda$	Достигнутая точность	Количество итераций
0.1	метод расходится	
0.01	2e-4	320
0.001	2e-3	2648
0.0001	1e-2	20734

Из полученных результатов можно сделать вывод, что при слишком большом шаге метод расходится, при слишком малом сходится медленно и точность хуже. Надо выбирать шаг наибольшим из тех, при которых метод сходится.

### Градиентный метод с дроблением шага

Для исследования сходимости метода градиентного спуска с дроблением шага (2) была выбрана функция:

$$f(x_1, x_2) = 10 * x_1^2 + x_2^2.$$

Начальное приближение - точка (10,10). Использован критерий останова:  $\|f(x^{[k+1]}) - f(x^{[k]})\| \leq 10^{-5}$

Результаты эксперимента отражены в таблице:

Значение параметра $\epsilon$	Значение параметра $\delta$	Значение параметра $\lambda^{[0]}$	Достигну тая точность	Количес тво итераций
0.95	0.95	1	5e-4	629
0.1	0.95	1	1e-5	41
0.1	0.1	1	2e-4	320
0.1	0.95	0.01	2e-4	320

Из полученных результатов можно сделать вывод об оптимальном выборе параметров:  $\epsilon = 0.1$ ,  $\delta = 0.95$ ,  $\lambda^{[0]} = 1$ , хотя метод не сильно чувствителен к выбору параметров.

### Метод наискорейшего спуска

Для исследования сходимости метода наискорейшего спуска была выбрана функция:

$$f(x_1, x_2) = 10 * x_1^2 + x_2^2.$$

Начальное приближение - точка (10,10). Использован критерий останова:  $\|f(x^{[k+1]}) - f(x^{[k]})\| \leq 10^{-5}$

Для решения одномерных задач оптимизации использован метод золотого сечения.

Метод получил точность  $6e-8$  за 9 итераций.

Отсюда можно сделать вывод, что метод наискорейшего спуска сходится быстрее, чем градиентный метод с дроблением шага и метод градиентного спуска с постоянным шагом.

Недостатком методом наискорейшего спуска является необходимость решать одномерную задачу оптимизации.

Рекомендации программисту

При программировании методов градиентного спуска следует аккуратно относиться к выбору параметров, а именно

- Метод градиентного спуска с постоянным шагом: шаг  $\lambda$  следует выбирать меньше 0.01, иначе метод расходится (метод может расходиться и при таком шаге в зависимости от исследуемой функции).

- Градиентный метод с дроблением шага не очень чувствителен к выбору параметров. Один из вариантов выбора параметров:

$$\epsilon = 0.1, \delta = 0.95, \lambda^{[0]} = 1$$

- Метод наискорейшего спуска: в качестве метода одномерной оптимизации можно использовать метод золотого сечения (когда он применим).

## 6.1 ПАРНЫЕ И МНОЖЕСТВЕННЫЕ КОРРЕЛЯЦИИ. НЕЛИНЕЙНЫЕ РЕГРЕССИИ.

### 1. Парная регрессия и корреляция

Парная регрессия представляет собой регрессию между двумя переменными –  $y$  и  $x$ , т. е. модель вида:  $y = \hat{f}(x)$ , где  $y$  – зависимая переменная (результативный признак);  $x$  – независимая, или объясняющая, переменная (признак-фактор). Знак « $\hat{\phantom{f}}$ » означает, что между переменными  $x$  и  $y$  нет строгой функциональной зависимости, поэтому практически в каждом отдельном случае величина  $y$  складывается из двух слагаемых:  $y = \hat{y}_x + \varepsilon$ , где  $y$  – фактическое значение результативного признака;  $\hat{y}_x$  – теоретическое значение результативного признака, найденное исходя из уравнения регрессии;  $\varepsilon$  – случайная величина, характеризующая отклонения реального значения результативного признака от теоретического, найденного по уравнению регрессии.

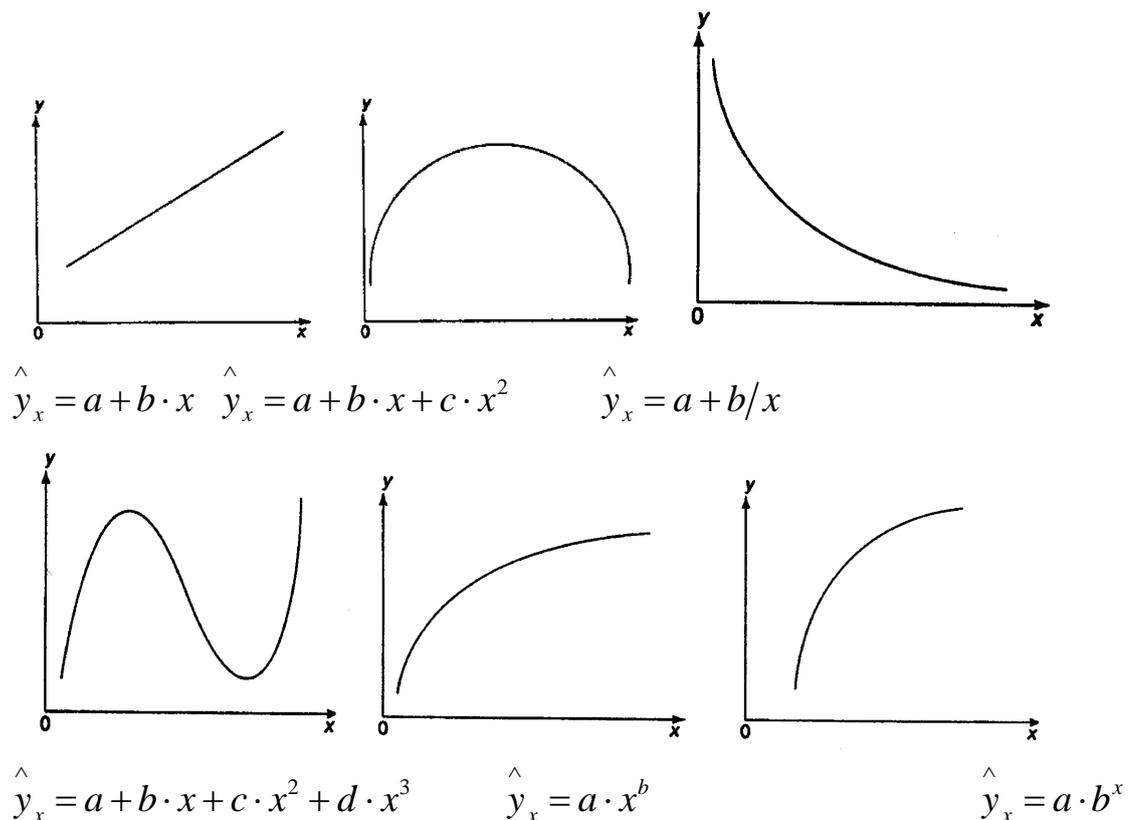
Случайная величина  $\varepsilon$  называется также возмущением. Она включает влияние не учтенных в модели факторов, случайных ошибок и особенностей измерения. Ее присутствие в модели порождено тремя источниками:

спецификацией модели, выборочным характером исходных данных, особенностями измерения переменных.

В парной регрессии выбор вида математической функции  $\hat{y}_x = f(x)$  может быть осуществлен тремя методами:

- 1) графическим;
- 2) аналитическим, т.е. исходя из теории изучаемой взаимосвязи;
- 3) экспериментальным.

При изучении зависимости между двумя признаками графический метод подбора вида уравнения регрессии достаточно нагляден. Он основан на поле корреляции. Основные типы кривых, используемые при количественной оценке связей, представлены на рис. 1.1:



**Рис. 1.1.** Основные типы кривых, используемые при количественной оценке связей между двумя переменными.

Значительный интерес представляет аналитический метод выбора типа уравнения регрессии. Он основан на изучении материальной природы связи исследуемых признаков.

При обработке информации на компьютере выбор вида уравнения регрессии обычно осуществляется экспериментальным методом, т. е. путем сравнения величины остаточной дисперсии  $\sigma_{\text{нб}}^2$ , рассчитанной при разных моделях.

Если уравнение регрессии проходит через все точки корреляционного поля, что возможно только при функциональной связи, когда все точки лежат на линии регрессии  $\hat{y}_x = f(x)$ , то фактические значения результативного

признака совпадают с теоретическими  $y = \hat{y}_x$ , т.е. они полностью обусловлены влиянием фактора  $x$ . В этом случае остаточная дисперсия  $\sigma_{\text{нб}}^2 = 0$ .

В практических исследованиях, как правило, имеет место некоторое рассеяние точек относительно линии регрессии. Оно обусловлено влиянием прочих, не учитываемых в уравнении регрессии, факторов. Иными словами, имеют место отклонения фактических данных от теоретических  $\left(y - \hat{y}_x\right)$ .

Величина этих отклонений и лежит в основе расчета остаточной дисперсии:

$$\sigma_{\text{нб}}^2 = \frac{1}{n} \sum \left(y - \hat{y}_x\right)^2.$$

Чем меньше величина остаточной дисперсии, тем меньше влияние не учитываемых в уравнении регрессии факторов и тем лучше уравнение регрессии подходит к исходным данным.

Считается, что число наблюдений должно в 7-8 раз превышать число рассчитываемых параметров при переменной  $x$ . Это означает, что искать линейную регрессию, имея менее 7 наблюдений, вообще не имеет смысла. Если вид функции усложняется, то требуется увеличение объема наблюдений, ибо каждый параметр при  $x$  должен рассчитываться хотя бы по 7 наблюдениям. Значит, если мы выбираем параболу второй степени  $\hat{y}_x = a + b \cdot x + c \cdot x^2$ , то требуется объем информации уже не менее 14 наблюдений.

### 1.1. Линейная модель парной регрессии и корреляции

Рассмотрим простейшую модель парной регрессии – линейную регрессию. Линейная регрессия находит широкое применение в эконометрике ввиду четкой экономической интерпретации ее параметров.

Линейная регрессия сводится к нахождению уравнения вида

$$\hat{y}_x = a + b \cdot x \text{ или } y = a + b \cdot x + \varepsilon. \quad (1.1)$$

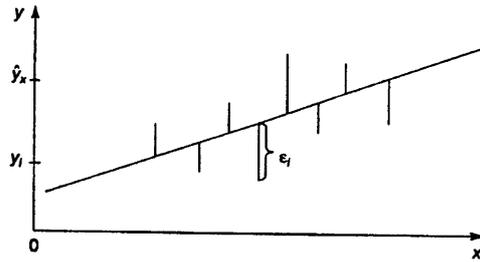
Уравнение вида  $\hat{y}_x = a + b \cdot x$  позволяет по заданным значениям фактора  $x$  находить теоретические значения результативного признака, подставляя в него фактические значения фактора  $x$ .

Построение линейной регрессии сводится к оценке ее параметров –  $a$  и  $b$ . Классический подход к оцениванию параметров линейной регрессии основан на методе наименьших квадратов (МНК). МНК позволяет получить такие оценки параметров  $a$  и  $b$ , при которых сумма квадратов отклонений фактических значений результативного признака  $y$  от теоретических  $\hat{y}_x$  минимальна:

$$\sum_{i=1}^n \left(y_i - \hat{y}_{x_i}\right)^2 = \sum_{i=1}^n \varepsilon_i^2 \rightarrow \min. \quad (1.2)$$

Прикладная математика

Т.е. из всего множества линий линия регрессии на графике выбирается так, чтобы сумма квадратов расстояний по вертикали между точками и этой линией была бы минимальной (рис. 1.2):



**Рис. 1.2.** Линия регрессии с минимальной дисперсией остатков.

Как известно из курса математического анализа, чтобы найти минимум функции (1.2), надо вычислить частные производные по каждому из параметров  $a$  и  $b$  и приравнять их к нулю. Обозначим  $\sum_i \varepsilon_i^2$  через  $S(a, b)$ ,

тогда:  $S(a, b) = \sum (y - a - b \cdot x)^2$ .

$$\begin{cases} \frac{\partial S}{\partial a} = -2 \sum (y - a - b \cdot x) = 0; \\ \frac{\partial S}{\partial b} = -2 \sum x(y - a - b \cdot x) = 0. \end{cases} \quad (1.3)$$

После несложных преобразований, получим следующую систему линейных уравнений для оценки параметров  $a$  и  $b$ :

$$\begin{cases} a \cdot n + b \cdot \sum x = \sum y; \\ a \cdot \sum x + b \cdot \sum x^2 = \sum x \cdot y. \end{cases} \quad (1.4)$$

Решая систему уравнений (1.4), найдем искомые оценки параметров  $a$  и  $b$ . Можно воспользоваться следующими готовыми формулами, которые следуют непосредственно из решения системы (1.4):

$$a = \bar{y} - b \cdot \bar{x}, \quad b = \frac{\text{cov}(x, y)}{\sigma_x^2} = \frac{\text{cov}(x, y)}{\text{var}(x)}, \quad (1.5)$$

где  $\text{cov}(x, y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \overline{\tilde{o}\tilde{o}} - \bar{o} \cdot \bar{o}$  – ковариация признаков  $x$  и

$y$ ,  $\sigma_x^2 = \overline{x^2} - \bar{x}^2$  – дисперсия признака  $x$  и

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1 + x_2 + \dots + x_n}{n}, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = \frac{y_1 + y_2 + \dots + y_n}{n},$$

$$\overline{y \cdot x} = \frac{1}{n} \sum y \cdot x, \quad \overline{x^2} = \frac{1}{n} \sum x^2$$

Прикладная математика

$$\text{var}(x) = \frac{1}{n} \sum_{i=1}^i (x_i - \bar{x})^2 = \bar{\delta}^2 - \bar{\delta}^2 = \sigma_{\delta}^2, \quad \text{var}(y) = \frac{1}{n} \sum_{i=1}^i (y_i - \bar{y})^2 = \bar{\delta}^2 - \bar{\delta}^2 = \sigma_{\delta}^2,$$

$$\sigma_y = \sqrt{\frac{\sum_{k=1}^n (y_k - \bar{y})^2}{n}} = \sqrt{\text{var}(y)}, \quad \sigma_{\delta} = \sqrt{\frac{\sum_{k=1}^n (x_k - \bar{x})^2}{n}} = \sqrt{\text{var}(x)}.$$

Ковариация – числовая характеристика совместного распределения двух случайных величин, равная математическому ожиданию произведения отклонений этих случайных величин от их математических ожиданий. Дисперсия – характеристика случайной величины, определяемая как математическое ожидание квадрата отклонения случайной величины от ее математического ожидания. Математическое ожидание – сумма произведений значений случайной величины на соответствующие вероятности.

Параметр  $b$  называется *коэффициентом регрессии*. Его величина показывает среднее изменение результата с изменением фактора на одну единицу.

Формально  $a$  – значение  $y$  при  $x=0$ . Если признак-фактор  $x$  не может иметь нулевого значения, то вышеуказанная трактовка свободного члена  $a$  не имеет смысла, т.е. параметр  $a$  может не иметь экономического содержания.

Уравнение регрессии всегда дополняется показателем тесноты связи. При использовании линейной регрессии в качестве такого показателя выступает линейный *коэффициент корреляции*  $r_{xy}$ , который можно рассчитать по следующим формулам:

$$r_{xy} = b \cdot \frac{\sigma_x}{\sigma_y} = \frac{\text{cov}(x, y)}{\sigma_x \cdot \sigma_y} = \frac{\bar{\delta\delta} - \bar{\delta} \cdot \bar{\delta}}{\sqrt{(\bar{\delta}^2 - \bar{\delta}^2)(\bar{\delta}^2 - \bar{\delta}^2)}} = \frac{\text{cov}(x, y)}{\sqrt{\text{var}(x) \text{var}(y)}}. \quad (1.6)$$

Линейный коэффициент корреляции находится в пределах:  $-1 \leq r_{xy} \leq 1$ . Чем ближе абсолютное значение  $r_{xy}$  к единице, тем сильнее линейная связь между факторами (при  $r_{xy} = \pm 1$  имеем строгую функциональную зависимость).

Теснота линейной связи между переменными может быть оценена на основании шкалы Чеддока:

Теснота связи	Значение коэффициента корреляции при наличии:	
	Прямой связи	Обратной связи
Слабая	0,1–0,3	(–0,3)–(–0,1)
Умеренная	0,3–0,5	(–0,5)–(–0,3)
Заметная	0,5–0,7	(–0,7)–(–0,5)
Высокая	0,7–0,9	(–0,9)–(–0,7)
Весьма высокая	0,9–1	(–1)–(–0,9)

Положительное значение коэффициента корреляции говорит о положительной связи между  $x$  и  $y$ , когда с ростом одной из переменных другая тоже растет. Отрицательное значение коэффициента корреляции означает, с ростом одной из переменных другая убывает, с убыванием одной из переменных другая растет.

Для оценки качества подбора линейной функции рассчитывается квадрат линейного коэффициента корреляции  $r_{xy}^2$ , называемый *коэффициентом детерминации*. Коэффициент детерминации характеризует долю дисперсии результативного признака  $y$ , объясняемую регрессией, в общей дисперсии результативного признака:

$$r_{xy}^2 = 1 - \frac{\sigma_{\text{инд}}^2}{\sigma_y^2}, \quad (1.7)$$

$$\text{где } \sigma_{\text{инд}}^2 = \frac{1}{n} \sum (y - \hat{y}_x)^2, \quad \sigma_y^2 = \frac{1}{n} \sum (y - \bar{y})^2 = \overline{y^2} - \bar{y}^2.$$

Соответственно величина  $1 - r_{xy}^2$  характеризует долю дисперсии  $y$ , вызванную влиянием остальных, не учтенных в модели, факторов.

После того как найдено уравнение линейной регрессии, проводится оценка значимости как уравнения в целом, так и отдельных его параметров.

Проверить значимость уравнения регрессии – значит установить, соответствует ли математическая модель, выражающая зависимость между переменными, экспериментальным данным и достаточно ли включенных в уравнение объясняющих переменных (одной или нескольких) для описания зависимой переменной.

Чтобы иметь общее суждение о качестве модели из относительных отклонений по каждому наблюдению, определяют среднюю ошибку аппроксимации:

$$\bar{A} = \frac{1}{n} \sum \left| \frac{y - \hat{y}_x}{y} \right| \cdot 100\%. \quad (1.8)$$

Средняя ошибка аппроксимации не должна превышать 8–10%.

Оценка значимости уравнения регрессии в целом производится на основе  $F$ -критерия Фишера, которому предшествует дисперсионный анализ.

Прикладная математика

Согласно основной идее дисперсионного анализа, общая сумма квадратов отклонений переменной  $y$  от среднего значения  $\bar{y}$  раскладывается на две части – «объясненную» и «необъясненную»:

$$\sum (y - \bar{y})^2 = \sum (\hat{y}_x - \bar{y})^2 + \sum (y - \hat{y}_x)^2,$$

где  $\sum (y - \bar{y})^2$  – общая сумма квадратов отклонений;  $\sum (\hat{y}_x - \bar{y})^2$  – сумма квадратов отклонений, объясненная регрессией (или факторная сумма квадратов отклонений);  $\sum (y - \hat{y}_x)^2$  – остаточная сумма квадратов отклонений, характеризующая влияние неучтенных в модели факторов.

Схема дисперсионного анализа имеет вид, представленный в таблице 1.1 ( $n$  – число наблюдений,  $m$  – число параметров при переменной  $x$ ).

Таблица 1.1

Компоненты дисперсии	Сумма квадратов	Число степеней свободы	Дисперсия на одну степень свободы
Общая	$\sum (y - \bar{y})^2$	$n - 1$	$S_{\text{об}}^2 = \frac{\sum (y - \bar{y})^2}{n - 1}$
Факторная	$\sum (\hat{y}_x - \bar{y})^2$	$m$	$S_{\text{фак}}^2 = \frac{\sum (\hat{y}_x - \bar{y})^2}{m}$
Остаточная	$\sum (y - \hat{y}_x)^2$	$n - m - 1$	$S_{\text{ост}}^2 = \frac{\sum (y - \hat{y}_x)^2}{n - m - 1}$

Определение дисперсии на одну степень свободы приводит дисперсии к сравнимому виду. Сопоставляя факторную и остаточную дисперсии в расчете на одну степень свободы, получим величину  $F$ -критерия Фишера:

$$F = \frac{S_{\text{фак}}^2}{S_{\text{ост}}^2}. \quad (1.9)$$

Фактическое значение  $F$ -критерия Фишера (1.9) сравнивается с табличным значением  $F_{\text{таб}}(\alpha; k_1; k_2)$  при уровне значимости  $\alpha$  и степенях свободы  $k_1 = m$  и  $k_2 = n - m - 1$ . При этом, если фактическое значение

$F$ -критерия больше табличного, то признается статистическая значимость уравнения в целом.

Для парной линейной регрессии  $m = 1$ , поэтому

$$F = \frac{S_{\hat{y}}^2}{S_{\text{ост}}^2} = \frac{\sum (y_x - \bar{y})^2}{\sum (y - \hat{y}_x)^2} \cdot (n - 2). \quad (1.10)$$

Величина  $F$ -критерия связана с коэффициентом детерминации  $r_{xy}^2$ , и ее можно рассчитать по следующей формуле:

$$F = \frac{r_{xy}^2}{1 - r_{xy}^2} \cdot (n - 2). \quad (1.11)$$

В парной линейной регрессии оценивается значимость не только уравнения в целом, но и отдельных его параметров. С этой целью по каждому из параметров определяется его стандартная ошибка:  $m_b$  и  $m_a$ .

Стандартная ошибка коэффициента регрессии определяется по формуле:

$$m_b = \sqrt{\frac{S_{\text{ост}}^2}{\sum (x - \bar{x})^2}} = \frac{S_{\text{ост}}}{\sigma_x \cdot \sqrt{n}}, \quad (1.12)$$

где  $S_{\text{ост}}^2 = \frac{\sum (y - \hat{y}_x)^2}{n - 2}$  – остаточная дисперсия на одну степень свободы.

Величина стандартной ошибки совместно с  $t$ -распределением Стьюдента при  $n - 2$  степенях свободы применяется для проверки существенности коэффициента регрессии и для расчета его доверительного интервала.

Для оценки существенности коэффициента регрессии его величина сравнивается с его стандартной ошибкой, т.е. определяется фактическое значение  $t$ -критерия Стьюдента:  $t_b = \frac{b}{m_b}$  которое затем сравнивается с

табличным значением при определенном уровне значимости  $\alpha$  и числе степеней свободы  $(n - 2)$ . Доверительный интервал для коэффициента

регрессии определяется как  $b \pm t_{\alpha} \cdot m_b$ . Границы доверительного интервала

для коэффициента регрессии не должны содержать противоречивых результатов, например,  $-1,5 \leq b \leq 0,8$ . Такого рода запись указывает, что истинное значение коэффициента регрессии одновременно содержит положительные и отрицательные величины и даже ноль, чего не может быть, и коэффициент признается незначимым.

Стандартная ошибка параметра  $a$  определяется по формуле:

$$m_a = \sqrt{S_{\text{ин}}^2 \cdot \frac{\sum x^2}{n \cdot \sum (x - \bar{x})^2}} = S_{\text{ин}} \cdot \frac{\sqrt{\sum x^2}}{\sigma_x \cdot n}. \quad (1.13)$$

Процедура оценивания существенности данного параметра не отличается от рассмотренной выше для коэффициента регрессии.

Вычисляется  $t$ -критерий:  $t_a = \frac{a}{m_a}$ , его величина сравнивается с табличным значением при  $n - 2$  степенях свободы.

Значимость линейного коэффициента корреляции проверяется на основе величины ошибки коэффициента корреляции  $m_r$ :

$$m_r = \sqrt{\frac{1 - r^2}{n - 2}}. \quad (1.14)$$

Фактическое значение  $t$ -критерия Стьюдента определяется как  $t_r = \frac{r}{m_r}$ .

Существует связь между  $t$ -критерием Стьюдента и  $F$ -критерием Фишера:

$$t_b = t_r = \sqrt{F}. \quad (1.15)$$

В прогнозных расчетах по уравнению регрессии определяется предсказываемое  $\hat{y}_p$  значение как точечный прогноз  $\hat{y}_x$  при  $x_p = x_k$ , т.е. путем подстановки в уравнение регрессии  $\hat{y}_x = a + b \cdot x$  соответствующего значения  $x$ . Однако точечный прогноз явно не реален. Поэтому он дополняется расчетом стандартной ошибки  $\hat{y}_p$ , т.е.  $m_{\hat{y}_p}$ , и соответственно интервальной оценкой прогнозного значения  $\hat{y}_p$ :

Прикладная математика

$$\hat{y}_p - \Delta_{y_p} \leq \hat{y}_p \leq \hat{y}_p + \Delta_{y_p},$$

где  $\Delta_{y_p} = m_{y_p} \cdot t_{\text{доверия}}$ , а  $m_{y_p}$  – средняя ошибка прогнозируемого индивидуального значения:

$$m_{y_p} = S_{\text{инд}} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{n \cdot \sigma_x^2}}. \quad (1.16)$$

**1.2. Нелинейные модели парной регрессии и корреляции**

Различают два класса нелинейных регрессий:

1. Регрессии, нелинейные относительно включенных в анализ объясняющих переменных, но линейные по оцениваемым параметрам, например

- полиномы различных степеней –  $\hat{y}_x = a + b \cdot x + c \cdot x^2$ ,  
 $\hat{y}_x = a + b \cdot x + c \cdot x^2 + d \cdot x^3$ ;

- равнобочная гиперболола –  $\hat{y}_x = a + b/x$ ;
- полулогарифмическая функция –  $\hat{y}_x = a + b \cdot \ln x$ .

2. Регрессии, нелинейные по оцениваемым параметрам, например

- степенная –  $\hat{y}_x = a \cdot x^b$ ;
- показательная –  $\hat{y}_x = a \cdot b^x$ ;
- экспоненциальная –  $\hat{y}_x = e^{a+b \cdot x}$ .

Регрессии нелинейные по включенным переменным приводятся к линейному виду простой заменой переменных (*линеаризация*), а дальнейшая оценка параметров производится с помощью метода наименьших квадратов. Рассмотрим некоторые функции.

Парабола второй степени  $\hat{y}_x = a + b \cdot x + c \cdot x^2$  приводится к линейному виду с помощью замены:  $x = x_1$ ,  $x^2 = x_2$ . В результате приходим к двухфакторному уравнению  $\hat{y}_x = a + b \cdot x_1 + c \cdot x_2$ , оценка параметров которого при помощи МНК, приводит к системе следующих нормальных уравнений:

Прикладная математика

$$\begin{cases} a \cdot n + b \cdot \sum x_1 + c \cdot \sum x_2 = \sum y; \\ a \cdot \sum x_1 + b \cdot \sum x_1^2 + c \cdot \sum x_1 \cdot x_2 = \sum x_1 \cdot y; \\ a \cdot \sum x_2 + b \cdot \sum x_1 \cdot x_2 + c \cdot \sum x_2^2 = \sum x_2 \cdot y. \end{cases}$$

А после обратной замены переменных получим

$$\begin{cases} a \cdot n + b \cdot \sum x + c \cdot \sum x^2 = \sum y; \\ a \cdot \sum x + b \cdot \sum x^2 + c \cdot \sum x^3 = \sum x \cdot y; \\ a \cdot \sum x^2 + b \cdot \sum x^3 + c \cdot \sum x^4 = \sum x^2 \cdot y. \end{cases} \quad (1.17)$$

Парабола второй степени обычно применяется в случаях, когда для определенного интервала значений фактора меняется характер связи рассматриваемых признаков: прямая связь меняется на обратную или обратная на прямую.

Равносторонняя гипербола  $\hat{y}_x = a + b/x$  приводится к линейному уравнению простой заменой:  $z = 1/x$ . Система линейных уравнений при применении МНК будет выглядеть следующим образом:

$$\begin{cases} a \cdot n + b \cdot \sum \frac{1}{x} = \sum y; \\ a \cdot \sum \frac{1}{x} + b \cdot \sum \frac{1}{x^2} = \sum \frac{1}{x} \cdot y. \end{cases} \quad (1.18)$$

Аналогичным образом приводятся к линейному виду зависимости  $\hat{y}_x = a + b \cdot \ln x$ ,  $\hat{y}_x = a + b \cdot \sqrt{x}$  и другие.

Несколько иначе обстоит дело с регрессиями нелинейными по оцениваемым параметрам, которые делятся на два типа: *нелинейные модели внутренне линейные* (приводятся к линейному виду с помощью соответствующих преобразований, например, логарифмированием) и *нелинейные модели внутренне нелинейные* (к линейному виду не приводятся).

К внутренне линейным моделям относятся, например, степенная функция –  $\hat{y}_x = a \cdot x^b$ , показательная –  $\hat{y}_x = a \cdot b^x$ , экспоненциальная –  $\hat{y}_x = e^{a+b \cdot x}$ , логистическая –  $\hat{y}_x = \frac{a}{1 + b \cdot e^{-c \cdot x}}$ , обратная –  $\hat{y}_x = \frac{1}{a + b \cdot x}$ .

Прикладная математика

К внутренне нелинейным моделям можно, например, отнести следующие модели:  $\hat{y}_x = a + b \cdot x^c$ ,  $\hat{y}_x = a \cdot \left(1 - \frac{1}{1 - x^b}\right)$ .

Среди нелинейных моделей наиболее часто используется степенная функция  $y = a \cdot x^b \cdot \varepsilon$ , которая приводится к линейному виду логарифмированием:

$$Y = A + b \cdot X + E,$$

где  $Y = \ln y$ ,  $X = \ln x$ ,  $A = \ln a$ ,  $E = \ln \varepsilon$ . Т.е. МНК мы применяем для преобразованных данных:

$$\begin{cases} A \cdot n + b \cdot \sum X = \sum Y, \\ A \cdot \sum X + b \cdot \sum X^2 = \sum X \cdot Y, \end{cases}$$

а затем потенцированием находим искомое уравнение.

Широкое использование степенной функции связано с тем, что параметр  $b$  в ней имеет четкое экономическое истолкование – он является коэффициентом эластичности.

*Коэффициент эластичности* показывает, на сколько процентов измениться в среднем результат, если фактор изменится на 1%. Формула для расчета коэффициента эластичности имеет вид:

$$\dot{Y} = f'(x) \cdot \frac{x}{y}. \tag{1.19}$$

Так как для остальных функций коэффициент эластичности не является постоянной величиной, а зависит от соответствующего значения фактора  $x$ , то обычно рассчитывается средний коэффициент эластичности:

$$\bar{Y} = f'(\bar{x}) \cdot \frac{\bar{x}}{\bar{y}}. \tag{1.20}$$

Приведем формулы для расчета средних коэффициентов эластичности для наиболее часто используемых типов уравнений регрессии:

Таблица 1.5

Вид функции, $y$	Первая производная, $y'$	Средний коэффициент эластичности, $\bar{Y}$	Линеаризация
------------------	--------------------------	---	--------------

Прикладная математика

1	2	3	4
$y = a + b \cdot x + \varepsilon$	$b$	$\frac{b \cdot \bar{x}}{a + b \cdot \bar{x}}$	-
$y = a + b \cdot x + c \cdot x^2$	$b + 2c \cdot x$	$\frac{(b + 2c \cdot \bar{x}) \cdot \bar{x}}{a + b \cdot \bar{x} + c \cdot \bar{x}^2}$	$X_1 = x, X_2 = x^2$
$y = a + \frac{b}{x} + \varepsilon$	$-\frac{b}{x^2}$	$-\frac{b}{a \cdot \bar{x} + b}$	$X = 1/x, Y = y$
$y = a \cdot x^b \cdot \varepsilon$	$a \cdot b \cdot x^{b-1}$	$b$	$X = \ln x, Y = \ln y$
$y = a \cdot b^x \cdot \varepsilon$	$a \cdot \ln b \cdot b^x$	$\bar{x} \cdot \ln b$	$X = x, Y = \ln y$
$y = a + b \cdot \ln x + \varepsilon$	$\frac{b}{x}$	$\frac{b}{a + b \cdot \ln \bar{x}}$	$X = \ln x, Y = y$
$y = e^{a+bx} \cdot \varepsilon$	$b e^{a+bx}$	$\bar{x} \cdot b$	$X = x, Y = \ln y$
$y = \frac{a}{1 + b \cdot e^{-c \cdot x + \varepsilon}}$	$\frac{a \cdot b \cdot c \cdot e^{-c \cdot x}}{(1 + b \cdot e^{-c \cdot x})^2}$	$\frac{b \cdot c \cdot \bar{x}}{b + e^{c \cdot \bar{x}}}$	
$y = \frac{1}{a + b \cdot x + \varepsilon}$	$-\frac{b}{(a + b \cdot x)^2}$	$-\frac{b \cdot \bar{x}}{a + b \cdot \bar{x}}$	$X = x, Y = 1/y$

Уравнение нелинейной регрессии, так же, как и в случае линейной зависимости, дополняется показателем тесноты связи. В данном случае это индекс корреляции:

$$\rho_{xy} = \sqrt{1 - \frac{\sigma_{\text{нò}}^2}{\sigma_y^2}}. \quad (1.21)$$

Величина данного показателя находится в пределах:  $0 \leq \rho_{xy} \leq 1$ . Чем ближе значение индекса корреляции к единице, тем теснее связь рассматриваемых признаков, тем более надежно уравнение регрессии.

Квадрат индекса корреляции носит название индекса детерминации и характеризует долю дисперсии результативного признака  $y$ , объясняемую регрессией, в общей дисперсии результативного признака:

$$\rho_{xy}^2 = 1 - \frac{\sigma_{\text{нò}}^2}{\sigma_y^2} = \frac{\sigma_{\text{áúýñí}}^2}{\sigma_y^2}, \quad (1.22)$$

## Прикладная математика

т.е. имеет тот же смысл, что и в линейной регрессии;

$$\sigma_{\hat{y}_x}^2 = \frac{1}{n} \sum \left( \hat{y}_x - \bar{y} \right)^2.$$

Индекс детерминации  $\rho_{xy}^2$  можно сравнивать с коэффициентом детерминации  $r_{xy}^2$  для обоснования возможности применения линейной функции. Чем больше кривизна линии регрессии, тем величина  $r_{xy}^2$  меньше  $\rho_{xy}^2$ . А близость этих показателей указывает на то, что нет необходимости усложнять форму уравнения регрессии и можно использовать линейную функцию.

Индекс детерминации используется для проверки существенности в целом уравнения регрессии по  $F$ -критерию Фишера:

$$F = \frac{\rho_{xy}^2}{1 - \rho_{xy}^2} \cdot \frac{n - m - 1}{m}, \quad (1.23)$$

где  $\rho_{xy}^2$  – индекс детерминации,  $n$  – число наблюдений,  $m$  – число параметров при переменной  $x$ . Фактическое значение  $F$ -критерия (1.23) сравнивается с табличным при уровне значимости  $\alpha$  и числе степеней свободы  $k_2 = n - m - 1$  (для остаточной суммы квадратов) и  $k_1 = m$  (для факторной суммы квадратов).

О качестве нелинейного уравнения регрессии можно также судить и по средней ошибке аппроксимации, которая, так же как и в линейном случае, вычисляется по формуле (1.8).

## 2. Множественная регрессия и корреляция

Парная регрессия может дать хороший результат при моделировании, если влиянием других факторов, воздействующих на объект исследования, можно пренебречь. Если же этим влиянием пренебречь нельзя, то в этом случае следует попытаться выявить влияние других факторов, введя их в модель, т.е. построить уравнение множественной регрессии

$$y = \hat{f}(x_1, x_2, \dots, x_m),$$

где  $y$  – зависимая переменная (результативный признак),  $x_i$  – независимые, или объясняющие, переменные (признаки-факторы).

### 2.1. Спецификация модели. Отбор факторов при построении уравнения множественной регрессии

Построение уравнения множественной регрессии начинается с решения вопроса о спецификации модели. Он включает в себя два круга вопросов: отбор факторов и выбор вида уравнения регрессии.

Включение в уравнение множественной регрессии того или иного набора факторов связано прежде всего с представлением исследователя о природе взаимосвязи моделируемого показателя с другими экономическими явлениями. Факторы, включаемые во множественную регрессию, должны отвечать следующим требованиям.

1. Они должны быть количественно измеримы. Если необходимо включить в модель качественный фактор, не имеющий количественного измерения, то ему нужно придать количественную определенность.

2. Факторы не должны быть интеркоррелированы и тем более находиться в точной функциональной связи.

Отбор факторов производится на основе качественного теоретико-экономического анализа. Однако теоретический анализ часто не позволяет однозначно ответить на вопрос о количественной взаимосвязи рассматриваемых признаков и целесообразности включения фактора в модель. Поэтому отбор факторов обычно осуществляется в две стадии: на первой подбираются факторы исходя из сущности проблемы; на второй – на основе матрицы показателей корреляции определяют статистики для параметров регрессии.

Коэффициенты интеркорреляции (т.е. корреляции между объясняющими переменными) позволяют исключать из модели дублирующие факторы. Считается, что две переменные явно *коллинеарны*,

т.е. находятся между собой в линейной зависимости, если  $r_{x_i, x_j} \geq 0,7$ . Если факторы явно коллинеарны, то они дублируют друг друга и один из них рекомендуется исключить из регрессии. Предпочтение при этом отдается не фактору, более тесно связанному с результатом, а тому фактору, который при достаточно тесной связи с результатом имеет наименьшую тесноту связи с другими факторами. В этом требовании проявляется специфика множественной регрессии как метода исследования комплексного воздействия факторов в условиях их независимости друг от друга.

По величине парных коэффициентов корреляции обнаруживается лишь явная коллинеарность факторов. Наибольшие трудности в использовании аппарата множественной регрессии возникают при наличии мультиколлинеарности факторов, когда более чем два фактора связаны между собой линейной зависимостью, т.е. имеет место совокупное воздействие факторов друг на друга.

Для оценки мультиколлинеарности факторов может использоваться определитель матрицы парных коэффициентов корреляции между факторами.

Чем ближе к нулю определитель матрицы межфакторной корреляции, тем сильнее мультиколлинеарность факторов и ненадежнее результаты множественной регрессии. И, наоборот, чем ближе к единице определитель матрицы межфакторной корреляции, тем меньше мультиколлинеарность факторов.

Существует ряд подходов преодоления сильной межфакторной корреляции. Самый простой путь устранения мультиколлинеарности состоит в исключении из модели одного или нескольких факторов. Другой подход связан с преобразованием факторов, при котором уменьшается корреляция между ними.

При отборе факторов также рекомендуется пользоваться следующим правилом: число включаемых факторов обычно в 6–7 раз меньше объема совокупности, по которой строится регрессия. Если это соотношение нарушено, то число степеней свободы остаточной дисперсии очень мало. Это приводит к тому, что параметры уравнения регрессии оказываются статистически незначимыми, а  $F$ -критерий меньше табличного значения.

## 2.2. Метод наименьших квадратов (МНК).

### Свойства оценок на основе МНК

Возможны разные виды уравнений множественной регрессии: линейные и нелинейные.

Ввиду четкой интерпретации параметров наиболее широко используется линейная функция. В линейной множественной регрессии

$\hat{y}_x = a + b_1x_1 + b_2x_2 + \dots + b_mx_m$  параметры при  $x$  называются *коэффициентами «чистой» регрессии*. Они характеризуют среднее изменение результата с изменением соответствующего фактора на единицу при неизменном значении других факторов, закрепленных на среднем уровне.

Рассмотрим линейную модель множественной регрессии

$$y = a + b_1x_1 + b_2x_2 + \dots + b_mx_m + \varepsilon. \quad (2.1)$$

Классический подход к оцениванию параметров линейной модели множественной регрессии основан на методе наименьших квадратов (МНК). МНК позволяет получить такие оценки параметров, при которых сумма квадратов отклонений фактических значений результативного признака  $y$  от расчетных  $\hat{y}$  минимальна:

$$\sum_i \left( y_i - \hat{y}_{x_i} \right)^2 \rightarrow \min. \quad (2.2)$$

Как известно из курса математического анализа, для того чтобы найти экстремум функции нескольких переменных, надо вычислить частные производные первого порядка по каждому из параметров и приравнять их к нулю.

Имеем функцию  $m + 1$  аргумента:

$$S(a, b_1, b_2, \dots, b_m) = \sum (y - a - b_1x_1 - b_2x_2 - \dots - b_mx_m)^2.$$

Находим частные производные первого порядка:

Прикладная математика

$$\begin{cases} \frac{\partial S}{\partial a} = -2 \sum (y - a - b_1 x_1 - b_2 x_2 - \dots - b_m x_m) = 0; \\ \frac{\partial S}{\partial b_1} = -2 \sum x_1 (y - a - b_1 x_1 - b_2 x_2 - \dots - b_m x_m) = 0; \\ \dots \\ \frac{\partial S}{\partial b_m} = -2 \sum x_m (y - a - b_1 x_1 - b_2 x_2 - \dots - b_m x_m) = 0. \end{cases}$$

После элементарных преобразований приходим к системе линейных нормальных уравнений для нахождения параметров линейного уравнения множественной регрессии (2.1):

$$\begin{cases} na + b_1 \sum x_1 + b_2 \sum x_2 + \dots + b_m \sum x_m = \sum y, \\ a \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_1 x_2 + \dots + b_m \sum x_1 x_m = \sum y x_1, \\ \dots \\ a \sum x_m + b_1 \sum x_1 x_m + b_2 \sum x_2 x_m + \dots + b_m \sum x_m^2 = \sum y x_m. \end{cases} \quad (2.3)$$

Для двухфакторной модели данная система будет иметь вид:

$$\begin{cases} na + b_1 \sum x_1 + b_2 \sum x_2 = \sum y, \\ a \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_1 x_2 = \sum y x_1, \\ a \sum x_2 + b_1 \sum x_1 x_2 + b_2 \sum x_2^2 = \sum y x_2. \end{cases}$$

Метод наименьших квадратов применим и к уравнению множественной регрессии в *стандартизированном масштабе*:

$$t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + \dots + \beta_m t_{x_m} + \varepsilon, \quad (2.4)$$

где  $t_y, t_{x_1}, \dots, t_{x_m}$  – стандартизированные переменные:  $t_y = \frac{y - \bar{y}}{\sigma_y}$ ,  $t_{x_i} = \frac{x_i - \bar{x}_i}{\sigma_{x_i}}$ , для которых среднее значение равно нулю:  $\bar{t}_y = \bar{t}_{x_i} = 0$ , а среднее квадратическое отклонение равно единице:  $\sigma_{t_y} = \sigma_{t_{x_i}} = 1$ ;  $\beta_i$  – стандартизированные коэффициенты регрессии.

Стандартизированные коэффициенты регрессии показывают, на сколько единиц изменится в среднем результат, если соответствующий фактор  $x_i$  изменится на одну единицу при неизменном среднем уровне других



Прикладная математика

$$\begin{cases} y_{x_1 \cdot x_2, x_3, \dots, x_m} = \hat{f}(x_1), \\ y_{x_2 \cdot x_1, x_3, \dots, x_m} = \hat{f}(x_2), \\ \dots \dots \dots \\ y_{x_m \cdot x_1, x_2, \dots, x_{m-1}} = \hat{f}(x_m), \end{cases} \quad (2.8)$$

т.е. уравнения регрессии, которые связывают результативный признак с соответствующим фактором  $x_i$  при закреплении остальных факторов на среднем уровне. В развернутом виде систему (2.8) можно переписать в виде:

$$\begin{cases} y_{x_1 \cdot x_2, x_3, \dots, x_m} = a + b_1 x_1 + b_2 \bar{x}_2 + b_3 \bar{x}_3 + \dots + b_m \bar{x}_m + \varepsilon, \\ y_{x_2 \cdot x_1, x_3, \dots, x_m} = a + b_1 \bar{x}_1 + b_2 x_2 + b_3 \bar{x}_3 + \dots + b_m \bar{x}_m + \varepsilon, \\ \dots \dots \dots \\ y_{x_m \cdot x_1, x_2, \dots, x_{m-1}} = a + b_1 \bar{x}_1 + b_2 \bar{x}_2 + b_3 \bar{x}_3 + \dots + b_m x_m + \varepsilon. \end{cases}$$

При подстановке в эти уравнения средних значений соответствующих факторов они принимают вид парных уравнений линейной регрессии, т.е. имеем

$$\begin{cases} y_{x_1 \cdot x_2, x_3, \dots, x_m} = A_1 + b_1 x_1, \\ y_{x_2 \cdot x_1, x_3, \dots, x_m} = A_2 + b_2 x_2, \\ \dots \dots \dots \\ y_{x_m \cdot x_1, x_2, \dots, x_{m-1}} = A_m + b_m x_m, \end{cases} \quad (2.9)$$

где

$$\begin{cases} A_1 = a + b_2 \bar{x}_2 + b_3 \bar{x}_3 + \dots + b_m \bar{x}_m, \\ A_2 = a + b_1 \bar{x}_1 + b_3 \bar{x}_3 + \dots + b_m \bar{x}_m, \\ \dots \dots \dots \\ A_m = a + b_1 \bar{x}_1 + b_2 \bar{x}_2 + b_3 \bar{x}_3 + \dots + b_{m-1} x_{m-1}. \end{cases}$$

В отличие от парной регрессии частные уравнения регрессии характеризуют изолированное влияние фактора на результат, ибо другие факторы закреплены на неизменном уровне. Эффекты влияния других факторов присоединены в них к свободному члену уравнения множественной регрессии. Это позволяет на основе частных уравнений регрессии определять *частные коэффициенты эластичности*:

Прикладная математика

$$\hat{Y}_{y_{x_i}} = b_i \cdot \frac{x_i}{\hat{y}_{x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_m}}, \quad (2.10)$$

где  $b_i$  – коэффициент регрессии для фактора  $x_i$  в уравнении множественной регрессии,  $\hat{y}_{x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_m}$  – частное уравнение регрессии.

Наряду с частными коэффициентами эластичности могут быть найдены *средние по совокупности показатели эластичности*:

$$\bar{Y}_i = b_i \cdot \frac{\bar{x}_i}{\bar{y}_{x_i}}, \quad (2.11)$$

которые показывают на сколько процентов в среднем изменится результат, при изменении соответствующего фактора на 1%. Средние показатели эластичности можно сравнивать друг с другом и соответственно ранжировать факторы по силе их воздействия на результат.

### 2.3. Проверка существенности факторов и показатели качества регрессии

Практическая значимость уравнения множественной регрессии оценивается с помощью показателя множественной корреляции и его квадрата – показателя детерминации.

Показатель множественной корреляции характеризует тесноту связи рассматриваемого набора факторов с исследуемым признаком или, иначе, оценивает тесноту совместного влияния факторов на результат.

Независимо от формы связи показатель множественной корреляции может быть найден как индекс множественной корреляции:

$$R_{y_{x_1 x_2 \dots x_m}} = \sqrt{1 - \frac{\sigma_{\hat{y}}^2}{\sigma_y^2}}. \quad (2.12)$$

Границы изменения индекса множественной корреляции от 0 до 1. Чем ближе его значение к 1, тем теснее связь результативного признака со всем набором исследуемых факторов. Величина индекса множественной корреляции должна быть больше или равна максимальному парному индексу корреляции:

Прикладная математика

$$R_{yx_1x_2\dots x_m} \geq r_{yx_i(\max)} \quad (i = \overline{1, m}).$$

Расчет индекса множественной корреляции предполагает определение уравнения множественной регрессии и на его основе остаточной дисперсии:

$$\sigma_{\hat{y}}^2 = \frac{1}{n} \sum \left( y - \hat{y}_{x_1x_2\dots x_m} \right)^2. \quad (2.13)$$

Можно пользоваться следующей формулой индекса множественной детерминации:

$$R_{yx_1x_2\dots x_m}^2 = 1 - \frac{\sum \left( y - \hat{y}_{x_1x_2\dots x_m} \right)^2}{\sum (y - \bar{y})^2}. \quad (2.14)$$

При линейной зависимости признаков формула индекса множественной корреляции может быть представлена следующим выражением:

$$R_{yx_1x_2\dots x_m} = \sqrt{\sum \beta_i \cdot r_{yx_i}}, \quad (2.15)$$

где  $\beta_i$  – стандартизованные коэффициенты регрессии;  $r_{yx_i}$  – парные коэффициенты корреляции результата с каждым фактором.

Формула индекса множественной корреляции для линейной регрессии получила название *линейного коэффициента множественной корреляции*, или, что то же самое, *совокупного коэффициента корреляции*.

Возможно также при линейной зависимости определение совокупного коэффициента корреляции через матрицу парных коэффициентов корреляции:

$$R_{yx_1x_2\dots x_p} = \sqrt{1 - \frac{\Delta r}{\Delta r_{11}}}, \quad (2.16)$$

где

$$\Delta r = \begin{vmatrix} 1 & r_{yx_1} & r_{yx_2} & \dots & r_{yx_p} \\ r_{yx_1} & 1 & r_{x_1x_2} & \dots & r_{x_1x_p} \\ r_{yx_2} & r_{x_2x_1} & 1 & \dots & r_{x_2x_p} \\ \dots & \dots & \dots & \dots & \dots \\ r_{yx_p} & r_{x_px_1} & r_{x_px_2} & \dots & 1 \end{vmatrix}$$

Прикладная математика

– определитель матрицы парных коэффициентов корреляции;

$$\Delta r_{11} = \begin{vmatrix} 1 & r_{x_1x_2} & \dots & r_{x_1x_p} \\ r_{x_2x_1} & 1 & \dots & r_{x_2x_p} \\ \dots & \dots & \dots & \dots \\ r_{x_px_1} & r_{x_px_2} & \dots & 1 \end{vmatrix}$$

– определитель матрицы межфакторной корреляции.

В рассмотренных показателях множественной корреляции (индекс и коэффициент) используется остаточная дисперсия, которая имеет систематическую ошибку в сторону преуменьшения, тем более значительную, чем больше параметров определяется в уравнении регрессии при заданном объеме наблюдений  $n$ . Если число параметров при  $x_i$  равно  $m$  и приближается к объему наблюдений, то остаточная дисперсия будет близка к нулю и коэффициент (индекс) корреляции приблизится к единице даже при слабой связи факторов с результатом. Для того чтобы не допустить возможного преувеличения тесноты связи, используется скорректированный индекс (коэффициент) множественной корреляции.

Скорректированный индекс множественной корреляции содержит поправку на число степеней свободы, а именно остаточная сумма квадратов  $\sum \left( y - \hat{y}_{x_1x_2\dots x_m} \right)^2$  делится на число степеней свободы остаточной вариации  $(n - m - 1)$ , а общая сумма квадратов отклонений  $\sum (y - \bar{y})^2$  на число степеней свободы в целом по совокупности  $(n - 1)$ .

Формула скорректированного индекса множественной детерминации имеет вид:

$$\hat{R}^2 = 1 - \frac{\sum (y - \hat{y})^2 / (n - m - 1)}{\sum (y - \bar{y})^2 / (n - 1)}, \quad (2.17)$$

где  $m$  – число параметров при переменных  $x$ ;  $n$  – число наблюдений.

Поскольку  $\frac{\sum (y - \hat{y}_{x_1 x_2 \dots x_m})^2}{\sum (y - \bar{y})^2} = 1 - R^2$ , то величину скорректированного

индекса детерминации можно представить в виде:

$$\hat{R}^2 = 1 - (1 - R^2) \cdot \frac{n-1}{n-m-1}. \quad (2.17a)$$

Чем больше величина  $m$ , тем сильнее различия  $\hat{R}^2$  и  $R^2$ .

Как было показано выше, ранжирование факторов, участвующих во множественной линейной регрессии, может быть проведено через стандартизованные коэффициенты регрессии ( $\beta$ -коэффициенты). Эта же цель может быть достигнута с помощью частных коэффициентов корреляции (для линейных связей). Кроме того, частные показатели корреляции широко используются при решении проблемы отбора факторов: целесообразность включения того или иного фактора в модель можно доказать величиной показателя частной корреляции.

*Частные коэффициенты корреляции* характеризуют тесноту связи между результатом и соответствующим фактором при элиминировании (устранении влияния) других факторов, включенных в уравнение регрессии.

Показатели частной корреляции представляют собой отношение сокращения остаточной дисперсии за счет дополнительного включения в анализ нового фактора к остаточной дисперсии, имевшей место до введения его в модель.

В общем виде при наличии  $m$  факторов для уравнения

$$y = a + b_1 x_1 + b_2 x_2 \dots + b_m x_m + \varepsilon$$

коэффициент частной корреляции, измеряющий влияние на  $y$  фактора  $x_i$ , при неизменном уровне других факторов, можно определить по формуле:

$$r_{y x_i \cdot x_1 x_2 \dots x_{i-1} x_{i+1} \dots x_m} = \sqrt{1 - \frac{1 - R^2_{y x_1 x_2 \dots x_i \dots x_m}}{1 - R^2_{y x_1 x_2 \dots x_{i-1} x_{i+1} \dots x_m}}}, \quad (2.18)$$

Прикладная математика

где  $R^2_{yx_1x_2\dots x_i\dots x_m}$  – множественный коэффициент детерминации всех  $m$  факторов с результатом;  $R^2_{yx_1x_2\dots x_{i-1}x_{i+1}\dots x_m}$  – тот же показатель детерминации, но без введения в модель фактора  $x_i$ .

При двух факторах формула (2.18) примет вид:

$$r_{yx_1x_2} = \sqrt{1 - \frac{1 - R^2_{yx_1x_2}}{1 - r^2_{yx_2}}}; \quad r_{yx_2x_1} = \sqrt{1 - \frac{1 - R^2_{yx_1x_2}}{1 - r^2_{yx_1}}}. \quad (2.18a)$$

Порядок частного коэффициента корреляции определяется количеством факторов, влияние которых исключается. Например,  $r_{yx_1x_2}$  – коэффициент частной корреляции первого порядка. Соответственно коэффициенты парной корреляции называются коэффициентами нулевого порядка. Коэффициенты частной корреляции более высоких порядков можно определить через коэффициенты частной корреляции более низких порядков по рекуррентной формуле:

$$r_{yx_i x_1 x_2 \dots x_{i-1} x_{i+1} \dots x_m} = \frac{r_{yx_i x_1 x_2 \dots x_{i-1} x_{i+1} \dots x_{m-1}} - r_{yx_m x_1 x_2 \dots x_{m-1}} \cdot r_{x_i x_m x_1 x_2 \dots x_{i-1} x_{i+1} \dots x_{m-1}}}{\sqrt{(1 - r^2_{yx_m x_1 x_2 \dots x_{m-1}}) \cdot (1 - r^2_{x_i x_m x_1 x_2 \dots x_{i-1} x_{i+1} \dots x_{m-1}})}}. \quad (2.19)$$

При двух факторах данная формула примет вид:

$$r_{yx_1x_2} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{\sqrt{(1 - r^2_{yx_2}) \cdot (1 - r^2_{x_1x_2})}}; \quad r_{yx_2x_1} = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1x_2}}{\sqrt{(1 - r^2_{yx_1}) \cdot (1 - r^2_{x_1x_2})}}. \quad (2.19a)$$

Для уравнения регрессии с тремя факторами частные коэффициенты корреляции второго порядка определяются на основе частных коэффициентов корреляции первого порядка. Так, по уравнению  $y = a + b_1x_1 + b_2x_2 + b_3x_3 + \varepsilon$  возможно исчисление трех частных коэффициентов корреляции второго порядка:

$$r_{yx_1x_2x_3}, \quad r_{yx_2x_1x_3}, \quad r_{yx_3x_1x_2},$$

каждый из которых определяется по рекуррентной формуле. Например, при  $i = 1$  имеем формулу для расчета  $r_{yx_1x_2x_3}$ :

$$r_{yx_1x_2x_3} = \frac{r_{yx_1x_2} - r_{yx_3x_2} \cdot r_{x_1x_3x_2}}{\sqrt{(1 - r^2_{yx_3x_2}) \cdot (1 - r^2_{x_1x_3x_2})}}. \quad (2.20)$$

Прикладная математика

Рассчитанные по рекуррентной формуле частные коэффициенты корреляции изменяются в пределах от  $-1$  до  $+1$ , а по формулам через множественные коэффициенты детерминации – от  $0$  до  $1$ . Сравнение их друг с другом позволяет ранжировать факторы по тесноте их связи с результатом. Частные коэффициенты корреляции дают меру тесноты связи каждого фактора с результатом в чистом виде. Если из стандартизованного уравнения регрессии  $t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + \beta_3 t_{x_3} + \varepsilon$  следует, что  $\beta_1 > \beta_2 > \beta_3$ , т.е. по силе влияния на результат порядок факторов таков:  $x_1, x_2, x_3$ , то этот же порядок факторов определяется и по соотношению частных коэффициентов корреляции,  $r_{yx_1 \cdot x_2 \cdot x_3} > r_{yx_2 \cdot x_1 \cdot x_3} > r_{yx_3 \cdot x_1 \cdot x_2}$ .

Из приведенных выше формул частных коэффициентов корреляции видна связь этих показателей с совокупным коэффициентом корреляции. Зная частные коэффициенты корреляции (последовательно первого, второго и более высокого порядка), можно определить совокупный коэффициент корреляции по формуле:

$$R_{yx_1 x_2 \dots x_m} = \sqrt{1 - (1 - r_{yx_1}^2) \cdot (1 - r_{yx_2 \cdot x_1}^2) \cdot (1 - r_{yx_3 \cdot x_1 x_2}^2) \cdot \dots \cdot (1 - r_{yx_m \cdot x_1 x_2 \dots x_{m-1}}^2)}. \quad (2.21)$$

В частности, для двухфакторного уравнения формула (2.21) принимает вид:

$$R_{yx_1 x_2} = \sqrt{1 - (1 - r_{yx_1}^2) \cdot (1 - r_{yx_2 \cdot x_1}^2)}. \quad (2.21)$$

При полной зависимости результативного признака от исследуемых факторов коэффициент совокупного их влияния равен единице. Из единицы вычитается доля остаточной вариации результативного признака  $(1 - r^2)$ , обусловленная последовательно включенными в анализ факторами. В результате подкоренное выражение характеризует совокупное действие всех исследуемых факторов.

Значимость уравнения множественной регрессии в целом, так же как и в парной регрессии, оценивается с помощью  $F$ -критерия Фишера:

$$F = \frac{S_{\hat{y}}}{S_{\hat{\varepsilon}}} = \frac{R^2}{1 - R^2} \cdot \frac{n - m - 1}{m}, \quad (2.22)$$

Прикладная математика

где  $S_{\text{оаёо}}$  – факторная сумма квадратов на одну степень свободы;  $S_{\text{іно}}$  – остаточная сумма квадратов на одну степень свободы;  $R^2$  – коэффициент (индекс) множественной детерминации;  $m$  – число параметров при переменных  $x$  (в линейной регрессии совпадает с числом включенных в модель факторов);  $n$  – число наблюдений.

Оценивается значимость не только уравнения в целом, но и фактора, дополнительно включенного в регрессионную модель. Необходимость такой оценки связана с тем, что не каждый фактор, вошедший в модель, может существенно увеличивать долю объясненной вариации результативного признака. Кроме того, при наличии в модели нескольких факторов они могут вводиться в модель в разной последовательности. Ввиду корреляции между факторами значимость одного и того же фактора может быть разной в зависимости от последовательности его введения в модель. Мерой для оценки включения фактора в модель служит *частный  $F$ -критерий*, т.е.  $F_{x_i}$ .

Частный  $F$ -критерий построен на сравнении прироста факторной дисперсии, обусловленного влиянием дополнительно включенного фактора, с остаточной дисперсией на одну степень свободы по регрессионной модели в целом. В общем виде для фактора  $x_i$  частный  $F$ -критерий определится как

$$F_{x_i} = \frac{R^2_{yx_1 \dots x_i \dots x_m} - R^2_{yx_1 \dots x_{i-1} x_{i+1} \dots x_m}}{1 - R^2_{yx_1 \dots x_i \dots x_m}} \cdot \frac{n - m - 1}{1}, \quad (2.23)$$

где  $R^2_{yx_1 \dots x_i \dots x_m}$  – коэффициент множественной детерминации для модели с полным набором факторов,  $R^2_{yx_1 \dots x_{i-1} x_{i+1} \dots x_m}$  – тот же показатель, но без включения в модель фактора  $x_i$ ,  $n$  – число наблюдений,  $m$  – число параметров в модели (без свободного члена).

Фактическое значение частного  $F$ -критерия сравнивается с табличным при уровне значимости  $\alpha$  и числе степеней свободы: 1 и  $n - m - 1$ . Если фактическое значение  $F_{x_i}$  превышает  $F_{\text{оаёё}}(\alpha, k_1, k_2)$ , то дополнительное включение фактора  $x_i$  в модель статистически оправданно и коэффициент чистой регрессии  $b_i$  при факторе  $x_i$  статистически значим. Если же фактическое значение  $F_{x_i}$  меньше табличного, то дополнительное включение в модель фактора  $x_i$  не увеличивает существенно долю объясненной вариации признака  $y$ , следовательно, нецелесообразно его включение в модель; коэффициент регрессии при данном факторе в этом случае статистически незначим.

Прикладная математика

Для двухфакторного уравнения частные  $F$ -критерии имеют вид:

$$F_{x_1} = \frac{R_{yx_1x_2}^2 - r_{yx_2}^2}{1 - R_{yx_1x_2}^2} \cdot (n - 3), \quad F_{x_2} = \frac{R_{yx_1x_2}^2 - r_{yx_1}^2}{1 - R_{yx_1x_2}^2} \cdot (n - 3). \quad (2.23a)$$

С помощью частного  $F$ -критерия можно проверить значимость всех коэффициентов регрессии в предположении, что каждый соответствующий фактор  $x_i$  вводился в уравнение множественной регрессии последним.

Частный  $F$ -критерий оценивает значимость коэффициентов чистой регрессии. Зная величину  $F_{x_i}$ , можно определить и  $t$ -критерий для коэффициента регрессии при  $i$ -м факторе,  $t_{b_i}$ , а именно:

$$t_{b_i} = \sqrt{F_{x_i}}. \quad (2.24)$$

Оценка значимости коэффициентов чистой регрессии по  $t$ -критерию Стьюдента может быть проведена и без расчета частных  $F$ -критериев. В этом случае, как и в парной регрессии, для каждого фактора используется формула:

$$t_{b_i} = \frac{b_i}{m_{b_i}}, \quad (2.25)$$

где  $b_i$  – коэффициент чистой регрессии при факторе  $x_i$ ,  $m_{b_i}$  – средняя квадратическая (стандартная) ошибка коэффициента регрессии  $b_i$ .

Для уравнения множественной регрессии  $\hat{y} = a + b_1x_1 + b_2x_2 + \dots + b_mx_m$  средняя квадратическая ошибка коэффициента регрессии может быть определена по следующей формуле:

$$m_{b_i} = \frac{\sigma_y \sqrt{1 - R_{yx_1 \dots x_m}^2}}{\sigma_{x_i} \sqrt{1 - R_{x_i x_1 \dots x_m}^2}} \cdot \frac{1}{\sqrt{n - m - 1}}, \quad (2.26)$$

где  $\sigma_y$  – среднее квадратическое отклонение для признака  $y$ ,  $\sigma_{x_i}$  – среднее квадратическое отклонение для признака  $x_i$ ,  $R_{yx_1 \dots x_m}^2$  – коэффициент детерминации для уравнения множественной регрессии,  $R_{x_i x_1 \dots x_m}^2$  – коэффициент детерминации для зависимости фактора  $x_i$  со всеми другими

Прикладная математика

факторами уравнения множественной регрессии;  $n - m - 1$  – число степеней свободы для остаточной суммы квадратов отклонений.



Наибольшее распространение в эконометрических исследованиях получила *система взаимосвязанных уравнений*. В ней одни и те же зависимые переменные в одних уравнениях входят в левую часть, а в других уравнениях – в правую часть системы:

$$\begin{cases} y_1 = b_{12}y_2 + b_{13}y_3 + \dots + b_{1m}y_m & + a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n + \varepsilon_1, \\ y_2 = b_{21}y_1 + b_{23}y_3 + \dots + b_{2m}y_m & + a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n + \varepsilon_2, \\ y_3 = b_{31}y_1 + b_{32}y_2 + \dots + b_{3m}y_m & + a_{31}x_1 + a_{32}x_2 + \dots + a_{3n}x_n + \varepsilon_3, \\ \dots & \dots \\ y_m = b_{m1}y_1 + b_{m2}y_2 + \dots + b_{m,m-1}y_{m-1} & + a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n + \varepsilon_n. \end{cases} \quad (3.3)$$

Система взаимосвязанных уравнений получила название *системы совместных, одновременных уравнений*. В эконометрике эта система уравнений называется также *структурной формой модели*. В отличие от предыдущих систем каждое уравнение системы одновременных уравнений не может рассматриваться самостоятельно, и для нахождения его параметров традиционный МНК неприменим. С этой целью используются специальные приемы оценивания.

### 3.1. Структурная и приведенная формы модели

Система совместных, одновременных уравнений (или структурная форма модели) обычно содержит эндогенные и экзогенные переменные.

*Эндогенные переменные* – это зависимые переменные, число которых равно числу уравнений в системе и которые обозначаются через  $y$ , взаимосвязанные переменные, которые определяются внутри модели (системы).

*Экзогенные переменные* – это предопределенные переменные, влияющие на эндогенные переменные, но не зависящие от них, независимые переменные, которые определяются вне системы. Обозначаются через  $x$ .

В качестве экзогенных переменных могут рассматриваться значения эндогенных переменных за предшествующий период времени (*лаговые переменные*).

*Предопределенными переменными* называются экзогенные и лаговые эндогенные переменные системы.



Прикладная математика

С позиции идентифицируемости структурные модели можно подразделить на три вида:

- 1) идентифицируемые;
- 2) неидентифицируемые;
- 3) сверхидентифицируемые.

Модель *идентифицируема*, если все структурные ее коэффициенты определяются однозначно, единственным образом по коэффициентам приведенной формы модели, т. е. если число параметров структурной модели равно числу параметров приведенной формы модели. В этом случае структурные коэффициенты модели оцениваются через параметры приведенной формы модели и модель идентифицируема.

Модель *неидентифицируема*, если число приведенных коэффициентов меньше числа структурных коэффициентов, и в результате структурные коэффициенты не могут быть оценены через коэффициенты приведенной формы модели.

Модель *сверхидентифицируема*, если число приведенных коэффициентов больше числа структурных коэффициентов. В этом случае на основе коэффициентов приведенной формы можно получить два или более значений одного структурного коэффициента. В этой модели число структурных коэффициентов меньше числа коэффициентов приведенной формы. Сверхидентифицируемая модель в отличие от неидентифицируемой модели практически решается, но требует для этого специальных методов исчисления параметров.

Структурная модель всегда представляет собой систему совместных уравнений, каждое из которых требуется проверять на идентификацию. Модель считается идентифицируемой, если каждое уравнение системы идентифицируемо. Если хотя бы одно из уравнений системы неидентифицируемо, то и вся модель считается неидентифицируемой. Сверхидентифицируемая модель содержит хотя бы одно сверхидентифицируемое уравнение.

Выполнение условия идентифицируемости модели проверяется для каждого уравнения системы. Чтобы уравнение было идентифицируемо, необходимо, чтобы число предопределенных переменных, отсутствующих в данном уравнении, но присутствующих в системе, было равно числу эндогенных переменных в данном уравнении без одного.

Прикладная математика

Если обозначить число эндогенных переменных в  $i$ -м уравнении системы через  $H$ , а число экзогенных (предопределенных) переменных, которые содержатся в системе, но не входят в данное уравнение, — через  $D$ , то условие идентифицируемости модели может быть записано в виде следующего счетного правила:

Таблица 3.1

$D + 1 = H$	уравнение идентифицируемо
$D + 1 < H$	уравнение неидентифицируемо
$D + 1 > H$	уравнение сверхидентифицируемо

Для оценки параметров структурной модели система должна быть идентифицируема или сверхидентифицируема.

Если необходимое условие выполнено, то далее проверяется достаточное условие идентификации.

Достаточное условие идентификации. Уравнение идентифицируемо, если определитель матрицы, составленной из коэффициентов при переменных, отсутствующих в исследуемом уравнении, не равен нулю, и ранг этой матрицы не менее числа эндогенных переменных системы без единицы.

### 3.3. Методы оценки параметров структурной формы модели

Коэффициенты структурной модели могут быть оценены разными способами в зависимости от вида системы одновременных уравнений. Наибольшее распространение в литературе получили следующие методы оценивания коэффициентов структурной модели:

- 1) косвенный метод наименьших квадратов;
- 2) двухшаговый метод наименьших квадратов;
- 3) трехшаговый метод наименьших квадратов;
- 4) метод максимального правдоподобия с полной информацией;
- 5) метод максимального правдоподобия при ограниченной информации.

Рассмотрим вкратце сущность каждого из этих методов.

Прикладная математика

*Косвенный метод наименьших квадратов* (КМНК) применяется в случае точно идентифицируемой структурной модели. Процедура применения КМНК предполагает выполнение следующих этапов работы.

1. Структурная модель преобразовывается в приведенную форму модели.
2. Для каждого уравнения приведенной формы модели обычным МНК оцениваются приведенные коэффициенты  $\delta_{ij}$ .
3. Коэффициенты приведенной формы модели трансформируются в параметры структурной модели.

Если система свержидентифицируема, то КМНК не используется, ибо он не дает однозначных оценок для параметров структурной модели. В этом случае могут использоваться разные методы оценивания, среди которых наиболее распространенным и простым является двухшаговый метод наименьших квадратов (ДМНК).

Основная идея ДМНК – на основе приведенной формы модели получить для свержидентифицируемого уравнения теоретические значения эндогенных переменных, содержащихся в правой части уравнения.

Далее, подставив их вместо фактических значений, можно применить обычный МНК к структурной форме свержидентифицируемого уравнения. Метод получил название ДМНК, ибо дважды используется МНК: на первом шаге при определении приведенной формы модели и нахождении на ее основе оценок теоретических значений эндогенной переменной  $\tilde{y}_i = \delta_{i1}x_1 + \delta_{i2}x_2 + \dots + \delta_{in}x_n$  и на втором шаге применительно к структурному свержидентифицируемому уравнению при определении структурных коэффициентов модели по данным теоретических (расчетных) значений эндогенных переменных.

Сверхидентифицируемая структурная модель может быть двух типов:

- 1) все уравнения системы свержидентифицируемы;
- 2) система содержит наряду со свержидентифицируемыми точно идентифицируемые уравнения.

Если все уравнения системы свержидентифицируемые, то для оценки структурных коэффициентов каждого уравнения используется ДМНК. Если в системе есть точно идентифицируемые уравнения, то структурные коэффициенты по ним находятся из системы приведенных уравнений.

#### 4. Временные ряды

При построении эконометрической модели используются два типа данных:

- 1) данные, характеризующие совокупность различных объектов в определенный момент времени;
- 2) данные, характеризующие один объект за ряд последовательных моментов времени.

Модели, построенные по данным первого типа, называются *пространственными моделями*. Модели, построенные на основе второго типа данных, называются *моделями временных рядов*.

*Временной ряд (ряд динамики)* – это совокупность значений какого-либо показателя за несколько последовательных моментов или периодов времени. Каждый уровень временного ряда формируется под воздействием большого числа факторов, которые условно можно подразделить на три группы:

- 1) факторы, формирующие тенденцию ряда;
- 2) факторы, формирующие циклические колебания ряда;
- 3) случайные факторы.

Каждый уровень временного ряда формируется под воздействием тенденции, сезонных колебаний и случайной компоненты.

1) *трендовой (T)*, описывающей общее изменение со временем результативного признака;

2) *сезонной (S)*, отражающей повторяемость данных через небольшой промежуток времени;

3) *случайной (E)*, отражающей влияние случайных факторов.

В большинстве случаев фактический уровень временного ряда можно представить как сумму или произведение трендовой, циклической и случайной компонент. Модель, в которой временной ряд представлен как сумма перечисленных компонент, называется *аддитивной моделью* временного ряда. Модель, в которой временной ряд представлен как произведение перечисленных компонент, называется *мультипликативной моделью* временного ряда.

*Основная задача эконометрического исследования* отдельного временного ряда – выявление и придание количественного выражения каждой из перечисленных выше компонент с тем, чтобы использовать

полученную информацию для прогнозирования будущих значений ряда или при построении моделей взаимосвязи двух или более временных рядов.

#### 4.1. Автокорреляция уровней временного ряда

При наличии во временном ряде тенденции и циклических колебаний значения каждого последующего уровня ряда зависят от предыдущих. Корреляционную зависимость между последовательными уровнями временного ряда называют *автокорреляцией уровней ряда*.

Количественно ее можно измерить с помощью линейного коэффициента корреляции между уровнями исходного временного ряда и уровнями этого ряда, сдвинутыми на несколько шагов во времени.

Формула для расчета коэффициента автокорреляции имеет вид:

$$r_1 = \frac{\sum_{t=2}^n (y_t - \bar{y}_1)(y_{t-1} - \bar{y}_2)}{\sqrt{\sum_{t=2}^n (y_t - \bar{y}_1)^2 \sum_{t=2}^n (y_{t-1} - \bar{y}_2)^2}}, \quad (4.1)$$

где

$$\bar{y}_1 = \frac{1}{n-1} \sum_{t=2}^n y_t, \quad \bar{y}_2 = \frac{1}{n-1} \sum_{t=2}^n y_{t-1}.$$

Эту величину называют *коэффициентом автокорреляции* уровней ряда первого порядка, так как он измеряет зависимость между соседними уровнями ряда  $t$  и  $y_{t-1}$ .

Аналогично можно определить коэффициенты автокорреляции второго и более высоких порядков. Так, коэффициент автокорреляции второго порядка характеризует тесноту связи между уровнями  $y_t$  и  $y_{t-2}$  и определяется по формуле:

$$r_2 = \frac{\sum_{t=3}^n (y_t - \bar{y}_3)(y_{t-2} - \bar{y}_4)}{\sqrt{\sum_{t=3}^n (y_t - \bar{y}_3)^2 \sum_{t=3}^n (y_{t-2} - \bar{y}_4)^2}}, \quad (4.2)$$

где

Прикладная математика

$$\bar{y}_3 = \frac{1}{n-2} \sum_{t=3}^n y_t, \quad \bar{y}_4 = \frac{1}{n-2} \sum_{t=3}^n y_{t-2}.$$

Если наиболее высоким оказался коэффициент автокорреляции первого порядка, исследуемый ряд содержит только тенденцию. Если наиболее высоким оказался коэффициент автокорреляции порядка  $\tau$ , то ряд содержит циклические колебания с периодичностью в  $\tau$  моментов времени. Если ни один из коэффициентов автокорреляции не является значимым, можно сделать одно из двух предположений относительно структуры этого ряда: либо ряд не содержит тенденции и циклических колебаний, либо ряд содержит сильную нелинейную тенденцию, для выявления которой нужно провести дополнительный анализ.

Свойства коэффициента автокорреляции.

1. Он строится по аналогии с линейным коэффициентом корреляции и таким образом характеризует тесноту только линейной связи текущего и предыдущего уровней ряда. Поэтому по коэффициенту автокорреляции можно судить о наличии линейной (или близкой к линейной) тенденции.

2. По знаку коэффициента автокорреляции нельзя делать вывод о возрастающей или убывающей тенденции в уровнях ряда. Большинство временных рядов экономических данных содержат положительную автокорреляцию уровней, однако при этом могут иметь убывающую тенденцию.

Число периодов, по которым рассчитывается коэффициент автокорреляции, называют *лагом*.

Последовательность коэффициентов автокорреляции уровней первого, второго и т.д. порядков называют *автокорреляционной функцией* временного ряда. График зависимости ее значений от величины лага (порядка коэффициента автокорреляции) называется *коррелограммой*.

#### 4. 2. Моделирование тенденции временного ряда

Распространенным способом моделирования тенденции временного ряда является построение аналитической функции, характеризующей зависимость уровней ряда от времени, или тренда. Этот способ называют *аналитическим выравниванием временного ряда*.

Для построения трендов чаще всего применяются следующие функции: линейный тренд:  $y_t = a + b \cdot t$ ; гипербола:  $y_t = a + \frac{b}{t}$ ; экспоненциальный

Прикладная математика

тренд:  $y_t = e^{a+bt}$  (или  $y_t = a \cdot b^t$ ); степенная функция:  $y_t = a \cdot t^b$ ; полиномы различных степеней:  $y_t = a + b_1 \cdot t + b_2 \cdot t^2 + \dots + b_m \cdot t^m$ .

Параметры каждого из перечисленных выше трендов можно определить обычным МНК, используя в качестве независимой переменной время  $t = 1, 2, \dots, n$ , а в качестве зависимой переменной – фактические уровни временного ряда  $y_t$ . Для нелинейных трендов предварительно проводят стандартную процедуру их линеаризации.

Наиболее простую экономическую интерпретацию имеет линейная функция  $y_t = a + bt$ :  $a$  – начальный уровень временного ряда в момент времени  $t = 0$ ;  $b$  – средний за период абсолютный прирост уровней ряда.

Параметры  $a$  и  $b$  находятся по формулам:

$$a = \frac{\sum_{i=1}^n y_i - b \sum_{i=1}^n t_i}{n}; \quad b = \frac{n \sum_{i=1}^n t_i y_i - \sum_{i=1}^n t_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n t_i^2 - (\sum_{i=1}^n t_i)^2}.$$

Существует несколько способов определения типа тенденции. К числу наиболее распространенных способов относятся качественный анализ изучаемого процесса, построение и визуальный анализ графика зависимости уровней ряда от времени. В этих же целях можно использовать и коэффициенты автокорреляции уровней ряда. Тип тенденции можно определить путем сравнения коэффициентов автокорреляции первого порядка, рассчитанных по исходным и преобразованным уровням ряда. Если временной ряд имеет линейную тенденцию, то его соседние уровни  $y_t$  и  $y_{t-1}$  тесно коррелируют. В этом случае коэффициент автокорреляции первого порядка уровней исходного ряда должен быть высоким. Если временной ряд содержит нелинейную тенденцию, например, в форме экспоненты, то коэффициент автокорреляции первого порядка по логарифмам уровней исходного ряда будет выше, чем соответствующий коэффициент, рассчитанный по уровням ряда. Чем сильнее выражена нелинейная тенденция в изучаемом временном ряде, тем в большей степени будут различаться значения указанных коэффициентов.

Выбор наилучшего уравнения в случае, когда ряд содержит нелинейную тенденцию, можно осуществить путем перебора основных форм тренда,

расчета по каждому уравнению скорректированного коэффициента детерминации и средней ошибки аппроксимации. Этот метод легко реализуется при компьютерной обработке данных.

### 4.3. Моделирование сезонных колебаний

Простейший подход к моделированию сезонных колебаний – это расчет значений сезонной компоненты методом скользящей средней и построение аддитивной или мультипликативной модели временного ряда.

Общий вид аддитивной модели следующий:

$$Y = T + S + E. \quad (4.3)$$

Эта модель предполагает, что каждый уровень временного ряда может быть представлен как сумма трендовой ( $T$ ), сезонной ( $S$ ) и случайной ( $E$ ) компонент.

Общий вид мультипликативной модели выглядит так:

$$Y = T \cdot S \cdot E. \quad (4.4)$$

Эта модель предполагает, что каждый уровень временного ряда может быть представлен как произведение трендовой ( $T$ ), сезонной ( $S$ ) и случайной ( $E$ ) компонент.

Выбор одной из двух моделей осуществляется на основе анализа структуры сезонных колебаний. Если амплитуда колебаний приблизительно постоянна, строят аддитивную модель временного ряда, в которой значения сезонной компоненты предполагаются постоянными для различных циклов. Если амплитуда сезонных колебаний возрастает или уменьшается, строят мультипликативную модель временного ряда, которая ставит уровни ряда в зависимость от значений сезонной компоненты.

Построение аддитивной и мультипликативной моделей сводится к расчету значений  $T$ ,  $S$  и  $E$  для каждого уровня ряда.

Процесс построения модели включает в себя следующие шаги.

- 1) Выравнивание исходного ряда методом скользящей средней.
- 2) Расчет значений сезонной компоненты  $S$ .
- 3) Устранение сезонной компоненты из исходных уровней ряда и получение выровненных данных ( $T + E$ ) в аддитивной или ( $T \cdot E$ ) в мультипликативной модели.
- 4) Аналитическое выравнивание уровней ( $T + E$ ) или ( $T \cdot E$ ) и расчет значений  $T$  с использованием полученного уравнения тренда.

Прикладная математика

5) Расчет полученных по модели значений  $(T + E)$  или  $(T \cdot E)$ .

6) Расчет абсолютных и/или относительных ошибок. Если полученные значения ошибок не содержат автокорреляции, ими можно заменить исходные уровни ряда и в дальнейшем использовать временной ряд ошибок  $E$  для анализа взаимосвязи исходного ряда и других временных рядов.

В моделях с сезонной компонентой обычно предполагается, что сезонные воздействия за период взаимопогашаются. В аддитивной модели это выражается в том, что сумма значений сезонной компоненты по всем кварталам должна быть равна нулю. В мультипликативной модели это выражается в том, что сумма значений сезонной компоненты по всем кварталам должна быть равна числу периодов в цикле.

Скорректированные значения сезонной компоненты в аддитивной

модели равны  $S_i = \bar{S}_i - k$ , где  $k = \frac{\sum_{i=1}^n S_i}{n}$ , в мультипликативной модели  $S_i$

получаются при умножении ее средней оценки  $\bar{S}_i$  на корректирующий

коэффициент  $k$ , где  $k = \frac{n}{\sum_{i=1}^n S_i}$ .

Прогнозное значение  $F_t$  уровня временного ряда в аддитивной модели есть сумма трендовой и сезонной компонент, в мультипликативной модели есть произведение трендовой и сезонной компонент.

#### 4.4. Автокорреляция в остатках. Критерий Дарбина-Уотсона

Автокорреляция в остатках может быть вызвана несколькими причинами, имеющими различную природу.

1. Она может быть связана с исходными данными и вызвана наличием ошибок измерения в значениях результативного признака.

2. В ряде случаев автокорреляция может быть следствием неправильной спецификации модели. Модель может не включать фактор, который оказывает существенное воздействие на результат и влияние которого отражается в остатках, вследствие чего последние могут оказаться автокоррелированными. Очень часто этим фактором является фактор времени  $t$ .

Один из более распространенных методов определения автокорреляции в остатках – это расчет критерия Дарбина-Уотсона:

Прикладная математика

$$d = \frac{\sum_{t=2}^n (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^n \varepsilon_t^2}. \quad (4.5)$$

Т.е. величина  $d$  есть отношение суммы квадратов разностей последовательных значений остатков к остаточной сумме квадратов по модели регрессии.

Можно показать, что при больших значениях  $n$  существует следующее соотношение между критерием Дарбина-Уотсона  $d$  и коэффициентом автокорреляции остатков первого порядка  $r_1$ :  $d = 2 \cdot (1 - r_1)$ .

Таким образом, если в остатках существует полная положительная автокорреляция и  $r_1 = 1$ , то  $d = 0$ . Если в остатках полная отрицательная автокорреляция, то  $r_1 = -1$  и, следовательно,  $d = 4$ . Если автокорреляция остатков отсутствует, то  $r_1 = 0$  и  $d = 2$ . Т.е.  $0 \leq d \leq 4$ .

Алгоритм выявления автокорреляции остатков на основе критерия Дарбина-Уотсона следующий. Выдвигается гипотеза  $H_0$  об отсутствии автокорреляции остатков. Альтернативные гипотезы  $H_1$  и  $H_1^*$  состоят, соответственно, в наличии положительной или отрицательной автокорреляции в остатках. Далее по специальным таблицам определяются критические значения критерия Дарбина-Уотсона  $d_L$  и  $d_U$  для заданного числа наблюдений  $n$ , числа независимых переменных модели  $m$  и уровня значимости  $\alpha$ . По этим значениям числовой промежуток  $[0; 4]$  разбивают на пять отрезков. Принятие или отклонение каждой из гипотез с вероятностью  $1 - \alpha$  осуществляется следующим образом:

$0 < d < d_L$  – есть положительная автокорреляция остатков,  $H_0$  отклоняется, с вероятностью  $P = 1 - \alpha$  принимается  $H_1$ ;

$d_L < d < d_U$  – зона неопределенности;

$d_U < d < 4 - d_U$  – нет оснований отклонять  $H_0$ , т.е. автокорреляция остатков отсутствует;

$4 - d_U < d < 4 - d_L$  – зона неопределенности;

$4 - d_L < d < 4$  – есть отрицательная автокорреляция остатков,  $H_0$  отклоняется, с вероятностью  $P = 1 - \alpha$  принимается  $H_1^*$ .

Если фактическое значение критерия Дарбина-Уотсона попадает в зону неопределенности, то на практике предполагают существование автокорреляции остатков и отклоняют гипотезу  $H_0$ .

Формула	Пояснение
$\sigma_{\text{інб}}^2 = \frac{1}{n} \sum (y - \hat{y}_x)^2$	Остаточная дисперсия
$a = \bar{y} - b \cdot \bar{x}$	Параметр регрессии $a$
$b = \frac{\text{cov}(x, y)}{\sigma_x^2} = \frac{\text{cov}(x, y)}{\text{var}(x)}$	Коэффициент регрессии
$\text{cov}(x, y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \overline{\delta\delta} - \bar{\delta} \cdot \bar{\delta}$	Ковариация
$\text{var}(x) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \overline{\delta^2} - \bar{\delta}^2 = \sigma_{\delta}^2$	Вариация $x$
$\text{var}(y) = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 = \overline{\delta^2} - \bar{\delta}^2 = \sigma_{\delta}^2$	Вариация $y$
$\sigma_y = \sqrt{\frac{\sum_{k=1}^n (y_k - \bar{y})^2}{n}} = \sqrt{\text{var}(y)}$	Среднее квадратическое отклонение $x$
$\sigma_{\delta} = \sqrt{\frac{\sum_{k=1}^n (x_k - \bar{x})^2}{n}} = \sqrt{\text{var}(x)}$	Среднее квадратическое отклонение $y$
$r_{xy} = b \cdot \frac{\sigma_x}{\sigma_y} = \frac{\text{cov}(x, y)}{\sigma_x \cdot \sigma_y} = \frac{\overline{\delta\delta} - \bar{\delta} \cdot \bar{\delta}}{\sqrt{(\overline{\delta^2} - \bar{\delta}^2)(\overline{\delta^2} - \bar{\delta}^2)}} = \frac{\text{cov}(x, y)}{\sqrt{\text{var}(x) \text{var}(y)}}$	Коэффициент корреляции
$r_{xy}^2 = 1 - \frac{\sigma_{\text{інб}}^2}{\sigma_y^2}$	Коэффициент детерминации
$\bar{A} = \frac{1}{n} \sum \left  \frac{y - \hat{y}_x}{y} \right  \cdot 100\%$	Средняя ошибка аппроксимации
$\sum (y - \bar{y})^2 = \sum (\hat{y}_x - \bar{y})^2 + \sum (y - \hat{y}_x)^2$	Общая сумма квадратов отклонений равна сумме факторной и остаточной сумм квадратов отклонений
$\sum (y - \bar{y})^2$	Общая сумма квадратов отклонений

Прикладная математика

$\sum \left( \hat{y}_x - \bar{y} \right)^2$	Факторная сумма квадратов отклонений
$\sum \left( y - \hat{y}_x \right)^2$	Остаточная сумма квадратов отклонений
$S_{\text{факт}}^2 = \frac{\sum (y - \bar{y})^2}{n - 1}$	Общая дисперсия на одну степень свободы
$S_{\text{факт}}^2 = \frac{\sum \left( \hat{y}_x - \bar{y} \right)^2}{m}$	Факторная дисперсия на одну степень свободы
$S_{\text{ост}}^2 = \frac{\sum \left( y - \hat{y}_x \right)^2}{n - m - 1}$	Остаточная дисперсия на одну степень свободы
$F = \frac{S_{\text{факт}}^2}{S_{\text{ост}}^2} = \frac{\sum \left( \hat{y}_x - \bar{y} \right)^2}{\sum \left( y - \hat{y}_x \right)^2} \cdot (n - 2) = \frac{r_{xy}^2}{1 - r_{xy}^2} \cdot (n - 2)$	Расчетное значение критерия Фишера
$F_{\text{табл}}(\alpha; k_1; k_2), k_1 = m \text{ и } k_2 = n - m - 1.$	Табличное значение критерия Фишера
$m_b = \sqrt{\frac{S_{\text{ост}}^2}{\sum (x - \bar{x})^2}} = \frac{S_{\text{ост}}}{\sigma_x \cdot \sqrt{n}}$	Стандартная ошибка коэффициента регрессии
$S_{\text{ост}}^2 = \frac{\sum \left( y - \hat{y}_x \right)^2}{n - 2}$	Остаточная дисперсия на одну степень свободы
$t_b = \frac{b}{m_b}$	t-статистика коэффициента регрессии
$b \pm t_{\text{табл}} \cdot m_b$	Доверительный интервал коэффициента регрессии
$m_a = \sqrt{S_{\text{ост}}^2 \cdot \frac{\sum x^2}{n \cdot \sum (x - \bar{x})^2}} = S_{\text{ост}} \cdot \frac{\sqrt{\sum x^2}}{\sigma_x \cdot n}$	Стандартная ошибка параметра регрессии

Прикладная математика

$t_a = \frac{a}{m_a}$	$t$ -статистика параметра регрессии
$m_r = \sqrt{\frac{1-r^2}{n-2}}$	Стандартная ошибка коэффициента корреляции
$t_r = \frac{r}{m_r}$	$t$ -статистика коэффициента корреляции
$t_b = t_r = \sqrt{F}$	Связь между критерием Стьюдента и критерием Фишера
$\hat{y}_p - \Delta_{y_p} \leq y_p \leq \hat{y}_p + \Delta_{y_p}$	Доверительный интервал прогноза
$\Delta_{y_p} = m_{y_p} \cdot t_{\text{доверия}}$	Предельная ошибка прогноза
$m_{y_p} = S_{i \text{ нб}} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{n \cdot \sigma_x^2}}$	Стандартная ошибка прогноза
$\bar{Y} = f'(\bar{x}) \cdot \frac{\bar{x}}{\bar{y}}$	Коэффициент эластичности
$\rho_{xy} = \sqrt{1 - \frac{\sigma_{i \text{ нб}}^2}{\sigma_y^2}}$	Индекс корреляции
$\rho_{xy}^2 = 1 - \frac{\sigma_{i \text{ нб}}^2}{\sigma_y^2} = \frac{\sigma_{i \text{ аúýñí}}^2}{\sigma_y^2}$	Индекс детерминации
$F = \frac{\rho_{xy}^2}{1 - \rho_{xy}^2} \cdot \frac{n - m - 1}{m}$	Расчетное значение критерия Фишера для нелинейной регрессии
$t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + \dots + \beta_m t_{x_m} + \varepsilon,$	Стандартизованный вид множественной регрессии
$b_i = \beta_i \frac{\sigma_y}{\sigma_{x_i}}$	Связь между коэффициентами «чистой» регрессии и стандартизованными

Прикладная математика

$\hat{Y}_{y_{x_1}} = b_i \cdot \frac{x_i}{y_{x_1 \cdot x_2 \dots x_{i-1} \cdot x_{i+1} \dots x_m}}$	<p>Частный коэффициент эластичности</p>
$\bar{Y}_i = b_i \cdot \frac{\bar{x}_i}{\bar{y}_{x_i}}$	<p>Средний показатель эластичности</p>
$R_{yx_1 x_2 \dots x_m} = \sqrt{1 - \frac{\sigma_{\hat{y}}^2}{\sigma_y^2}} = \sqrt{1 - \frac{\Delta r}{\Delta r_{11}}} = \sqrt{\sum \beta_i \cdot r_{yx_i}}$	<p>Множественный коэффициент корреляции</p>
$R^2_{yx_1 x_2 \dots x_m} = 1 - \frac{\sum (y - \hat{y}_{x_1 x_2 \dots x_m})^2}{\sum (y - \bar{y})^2}$	<p>Множественный коэффициент детерминации</p>
$\Delta r = \begin{vmatrix} 1 & r_{yx_1} & r_{yx_2} & \dots & r_{yx_p} \\ r_{yx_1} & 1 & r_{x_1 x_2} & \dots & r_{x_1 x_p} \\ r_{yx_2} & r_{x_2 x_1} & 1 & \dots & r_{x_2 x_p} \\ \dots & \dots & \dots & \dots & \dots \\ r_{yx_p} & r_{x_p x_1} & r_{x_p x_2} & \dots & 1 \end{vmatrix}$	<p>Определитель матрицы парных коэффициентов</p>
$\Delta r_{11} = \begin{vmatrix} 1 & r_{x_1 x_2} & \dots & r_{x_1 x_p} \\ r_{x_2 x_1} & 1 & \dots & r_{x_2 x_p} \\ \dots & \dots & \dots & \dots \\ r_{x_p x_1} & r_{x_p x_2} & \dots & 1 \end{vmatrix}$	<p>Определитель матрицы межфакторной корреляции</p>
$\hat{R}^2 = 1 - \frac{\sum (y - \hat{y})^2 / (n - m - 1)}{\sum (y - \bar{y})^2 / (n - 1)} = 1 - (1 - R^2) \cdot \frac{n - 1}{n - m - 1}$	<p>Скорректированный индекс множественной детерминации</p>
$r_{yx_1 \cdot x_2 \dots x_{i-1} \cdot x_{i+1} \dots x_m} = \sqrt{1 - \frac{1 - R^2_{yx_1 x_2 \dots x_i \dots x_m}}{1 - R^2_{yx_1 x_2 \dots x_{i-1} \cdot x_{i+1} \dots x_m}}}$	<p>Частный коэффициент корреляции</p>
$r_{yx_1 \cdot x_2} = \sqrt{1 - \frac{1 - R^2_{yx_1 x_2}}{1 - r_{yx_2}^2}}, \quad r_{yx_1 \cdot x_2} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1 x_2}}{\sqrt{(1 - r_{yx_2}^2) \cdot (1 - r_{x_1 x_2}^2)}}$	<p>Частный коэффициент корреляции</p>
$r_{yx_1 \cdot x_2 \dots x_{i-1} \cdot x_{i+1} \dots x_m} = \frac{r_{yx_1 \cdot x_1 x_2 \dots x_{i-1} \cdot x_{i+1} \dots x_{m-1}} - r_{yx_1 \cdot x_1 x_2 \dots x_{m-1}} \cdot r_{x_1 x_m \cdot x_1 x_2 \dots x_{i-1} \cdot x_{i+1} \dots x_m}}{\sqrt{(1 - r_{yx_1 \cdot x_1 x_2 \dots x_{m-1}}^2) \cdot (1 - r_{x_1 x_m \cdot x_1 x_2 \dots x_{i-1} \cdot x_{i+1} \dots x_{m-1}}^2)}}$	<p>Частный коэффициент корреляции</p>

Прикладная математика

$r_{y_{x_1 \cdot x_2 \cdot x_3}} = \frac{r_{y_{x_1 \cdot x_2}} - r_{y_{x_3 \cdot x_2}} \cdot r_{x_1 \cdot x_3 \cdot x_2}}{\sqrt{(1 - r_{y_{x_3 \cdot x_2}}^2)(1 - r_{x_1 \cdot x_3 \cdot x_2}^2)}}$	<p>Частный коэффициент корреляции</p>
$R_{y_{x_1 \cdot x_2 \dots x_m}} = \sqrt{1 - (1 - r_{y_{x_1}}^2) \cdot (1 - r_{y_{x_2 \cdot x_1}}^2) \cdot (1 - r_{y_{x_3 \cdot x_1 \cdot x_2}}^2) \cdot \dots \cdot (1 - r_{y_{x_m \cdot x_1 \cdot x_2 \dots x_{m-1}}}^2)}$	<p>Множественный коэффициент корреляции</p>
$R_{y_{x_1 \cdot x_2}} = \sqrt{1 - (1 - r_{y_{x_1}}^2) \cdot (1 - r_{y_{x_2 \cdot x_1}}^2)}$	<p>Множественный коэффициент корреляции</p>
$F = \frac{S_{\text{объ}}}{S_{\text{ост}}} = \frac{R^2}{1 - R^2} \cdot \frac{n - m - 1}{m}$	<p>Расчетное значение критерия Фишера для множественной регрессии</p>
$F_{x_i} = \frac{R_{y_{x_1 \dots x_i \dots x_m}}^2 - R_{y_{x_1 \dots x_{i-1} x_{i+1} \dots x_m}}^2}{1 - R_{y_{x_1 \dots x_i \dots x_m}}^2} \cdot \frac{n - m - 1}{1}$	<p>Частный <math>F</math>-критерий</p>
$F_{x_1} = \frac{R_{y_{x_1 \cdot x_2}}^2 - r_{y_{x_2}}^2}{1 - R_{y_{x_1 \cdot x_2}}^2} \cdot (n - 3), \quad F_{x_2} = \frac{R_{y_{x_1 \cdot x_2}}^2 - r_{y_{x_1}}^2}{1 - R_{y_{x_1 \cdot x_2}}^2} \cdot (n - 3)$	<p>Частный <math>F</math>-критерий</p>
$t_{b_i} = \frac{b_i}{m_{b_i}}$	<p><math>t</math>-статистика коэффициента множественной регрессии</p>
$m_{b_i} = \frac{\sigma_y \sqrt{1 - R_{y_{x_1 \dots x_m}}^2}}{\sigma_{x_i} \sqrt{1 - R_{x_i \cdot x_1 \dots x_m}^2}} \cdot \frac{1}{\sqrt{n - m - 1}}$	<p>Стандартная ошибка коэффициента множественной регрессии</p>

1.1 ГЕОМЕТРИЧЕСКАЯ ИНТЕРПРЕТАЦИЯ РЕШЕНИЯ

**1. Геометрическая интерпретация задач линейного программирования**

Для понимания всего дальнейшего полезно знать и представлять себе геометрическую интерпретацию задач линейного программирования, которую можно дать для случаев  $n = 2$  и  $n = 3$ .

Прикладная математика

Наиболее наглядна эта интерпретация для случая  $n = 2$ , т.е. для случая двух переменных  $x_1$  и  $x_2$ . Пусть нам задана задача линейного программирования в стандартной форме

$$\begin{aligned}
 & c_1x_1 + c_2x_2 \Rightarrow \max \\
 & a_{11}x_1 + a_{12}x_2 \leq b_1, \\
 & a_{21}x_1 + a_{22}x_2 \leq b_2, \\
 & \dots \quad \dots\dots \\
 & a_{m1}x_1 + a_{m2}x_2 \leq b_m,
 \end{aligned}
 \tag{1.19}$$

$$x_1 \geq 0; \quad x_2 \geq 0.$$

Возьмём на плоскости декартову систему координат и каждой паре чисел  $(x_1, x_2)$  поставим в соответствие точку на этой плоскости.

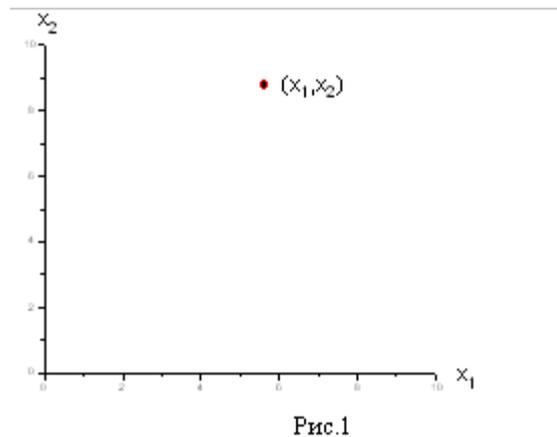


Рис.1

Обратим прежде всего внимание на ограничения  $x_1 \geq 0$  и  $x_2 \geq 0$ . Они из всей плоскости вырезают лишь её первую четверть (см. рис. 1). Рассмотрим теперь, какие области соответствуют неравенствам вида  $a_1x_1 + a_2x_2 \leq b$ . Сначала рассмотрим область, соответствующую равенству  $a_1x_1 + a_2x_2 = b$ . Как Вы, конечно, знаете, это прямая линия. Строить её проще всего по двум точкам.

Пусть  $b \neq 0$ . Если взять  $x_1 = 0$ , то получится  $x_2 = b/a_2$ . Если взять  $x_2 = 0$ , то получится  $x_1 = b/a_1$ . Таким образом, на прямой лежат две точки  $(0, b/a_2)$  и  $(b/a_1, 0)$ . Дальше через эти две точки можно по линейке провести прямую линию (смотри рисунок 2).

Прикладная математика

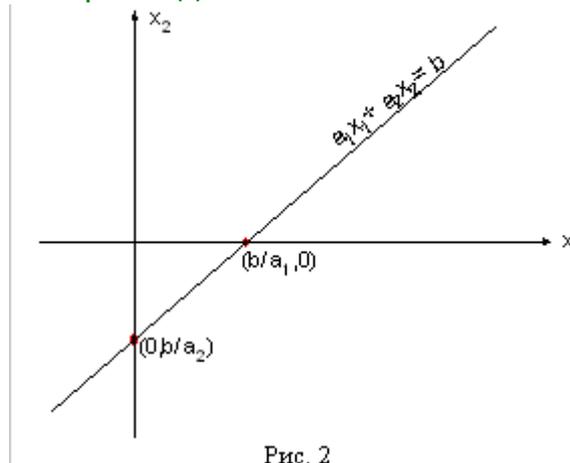


Рис. 2

Если же  $b=0$ , то на прямой лежит точка  $(0,0)$ . Чтобы найти другую точку, можно взять любое отличное от нуля значение  $x_1$  и вычислить

соответствующее ему значение  $x_2$ .

Эта построенная прямая разбивает всю плоскость на две полуплоскости. В одной её части  $a_1x_1 + a_2x_2 < b$ , а в другой наоборот  $a_1x_1 + a_2x_2 > b$ . Узнать, в какой полуплоскости какой знак имеет место проще всего посмотрев, какому неравенству удовлетворяет какая-то точка плоскости, например, начало координат, т.е. точка  $(0,0)$ .

Пример

Определить полуплоскость, определяемую неравенством  $4x_1 - 6x_2 \leq 3$ .

Решение

Сначала строим прямую  $4x_1 - 6x_2 = 3$ . Полагая  $x_1 = 0$  получим  $-6x_2 = 3$  или  $x_2 = -1/2$ . Полагая  $x_2 = 0$  получим  $4x_1 = 3$  или  $x_1 = 3/4$ . Таким образом, наша прямая проходит через точки  $(0, -1/2)$  и  $(3/4, 0)$  (см. рис. 3)

Теперь посмотрим, в какой полуплоскости лежит точка  $(0,0)$ , т.е. начало координат. Имеем  $4 \cdot 0 - 6 \cdot 0 < 3$ , т.е. начало координат принадлежит полуплоскости, где  $4x_1 - 6x_2 \leq 3$ . Тем самым определилась и нужная нам полуплоскость (см. рис. 3).

Прикладная математика

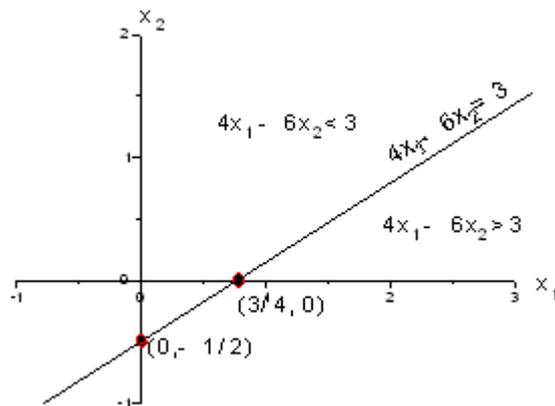


Рис. 3

Вернёмся теперь к задаче линейного программирования. Там имеют место  $m$  неравенств

$$\begin{aligned}
 a_{11}x_1 + a_{12}x_2 &\leq b_1, \\
 a_{21}x_1 + a_{22}x_2 &\leq b_2, \\
 &\dots\dots\dots \\
 a_{m1}x_1 + a_{m2}x_2 &\leq b_m.
 \end{aligned}
 \tag{1.20}$$

Каждое из них задает на плоскости некоторую полуплоскость. Нас интересуют те точки, которые удовлетворяют всем этим  $m$  неравенствам, т.е. точки, которые принадлежат **всем этим полуплоскостям одновременно**. Следовательно, область, определяемая неравенствами вида (1.20), геометрически изображается общей частью (пересечением) всех полуплоскостей, определяемых отдельными ограничениями (к ним,

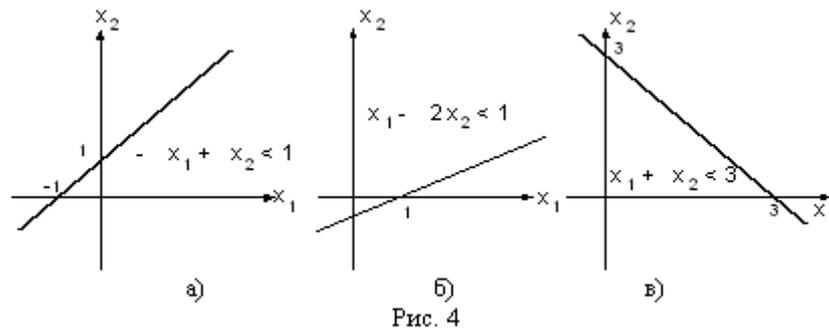
естественно, надо добавить ограничения  $x_1 \geq 0$  и  $x_2 \geq 0$ ).

Как уже говорилось выше, эта область называется **допустимой областью** задачи линейного программирования.

Пример

Найти допустимую область задачи линейного программирования, определяемую ограничениями

$$\begin{cases}
 -x_1 + x_2 \leq 1, \\
 x_1 - 2x_2 \leq 1, \\
 x_1 + x_2 \leq 3, \\
 x_1 \geq 0, x_2 \geq 0.
 \end{cases}
 \tag{1.21}$$



Решение

1. Рассмотрим прямую  $-x_1 + x_2 = 1$ . При  $x_1 = 0$   $x_2 = 1$ , а при  $x_2 = 0$   $x_1 = -1$ . Таким образом, эта прямая проходит через точки  $(0,1)$  и  $(-1,0)$ . Беря  $x_1 = x_2 = 0$  получим, что  $-0+0 < 1$  и поэтому интересующая нас полуплоскость лежит ниже прямой, изображенной на рис. 4а.

2. Рассмотрим прямую  $x_1 - 2x_2 = 1$ . При  $x_1 = 0$   $x_2 = -1/2$ , а при  $x_2 = 0$   $x_1 = 1$ . Таким образом, эта прямая проходит через точки  $(0, -1/2)$  и  $(1,0)$ . так как  $0 - 2 \cdot 0 < 1$  (4.б).

3. Наконец, рассмотрим  $x_2 = 0$   $x_1 = 1$  прямую  $x_1 + x_2 = 3$ . Она проходит через точки  $(0,3)$  и  $(3,0)$  и так как  $0+0 < 3$ , то интересующая нас полуплоскость лежит ниже прямой, изображенной на рис. 4.в).

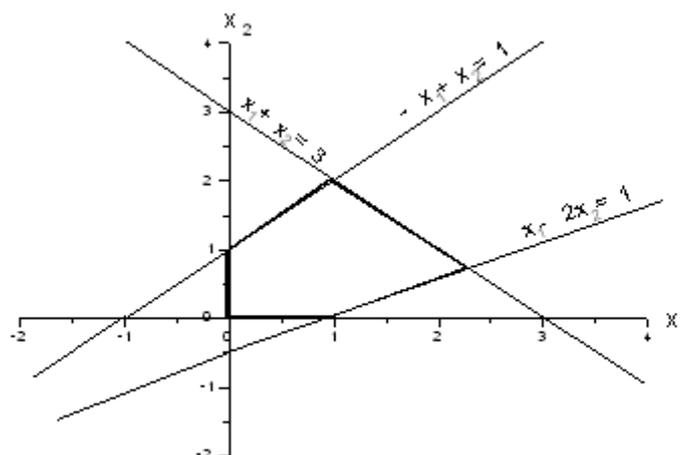


Рис. 5

Сводя все вместе и добавляя условия  $x_1 \geq 0$ ,  $x_2 \geq 0$  получим рисунок 5, где выделена область, в которой выполняются одновременно все ограничения (1.21). Обратите внимание на то, что получившаяся область имеет вид **выпуклого многоугольника**.

Прикладная математика

Вернемся теперь к общему случаю, когда одновременно выполняются неравенства

$$\begin{aligned}
 a_{11}x_1 + a_{12}x_2 &\leq b_1, \\
 a_{21}x_1 + a_{22}x_2 &\leq b_2, \\
 &\dots\dots\dots \\
 a_{m1}x_1 + a_{m2}x_2 &\leq b_m, \\
 x_1 &\geq 0, \quad x_2 \geq 0.
 \end{aligned}
 \tag{1.22}$$

Не приводя строгих доказательств, укажем те случаи, которые тут могут получиться.

1. **Основной случай** - получающаяся область имеет вид ограниченного выпуклого многоугольника ( см. рис. 6).
2. **Неосновной случай**  $\square$  получается неограниченный выпуклый многоугольник, имеющий вид, подобный изображенному на рис. 7. Подобная ситуация, например, получится, если в рассмотренном выше примере убрать ограничение  $x_1 + x_2 \leq 3$ . Оставшаяся часть будет неограниченным выпуклым многоугольником.

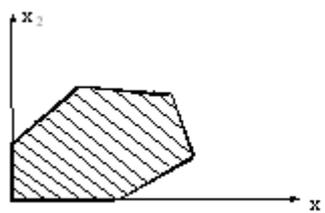


Рис. 6

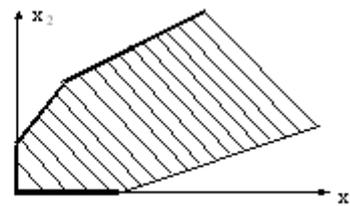
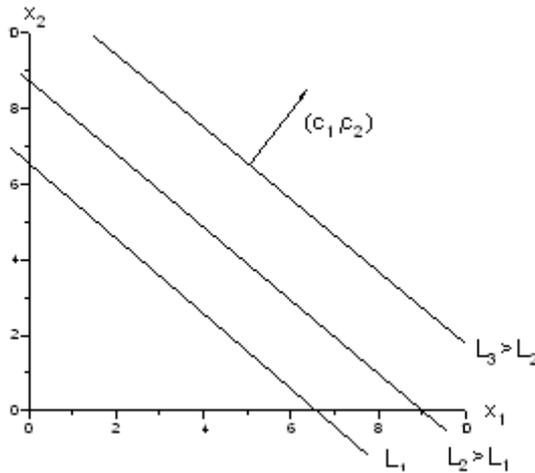


Рис. 7

3. Наконец, возможен случай, когда неравенства (1.22) **противоречат друг другу**, и допустимая область вообще **пуста**.

Вернёмся теперь к исходной задаче линейного программирования. В ней, кроме системы неравенств, есть еще **целевая функция**  $c_1x_1 + c_2x_2 \Rightarrow \max$ .

Прикладная математика



Рассмотрим прямую  $c_1x_1 + c_2x_2 = L$ . Будем увеличивать  $L$ . Что будет происходить с нашей прямой?

Легко догадаться, что прямая будет двигаться параллельно самой себе в том направлении, которое дается вектором  $(c_1, c_2)$ , так как это  $\square$  вектор нормали к нашей прямой и одновременно вектор градиента функции  $f(x_1, x_2) = c_1x_1 + c_2x_2$ .

А теперь сведем всё вместе. Итак, надо решить задачу

$$\begin{aligned}
 c_1x_1 + c_2x_2 &\Rightarrow \max \\
 a_{11}x_1 + a_{12}x_2 &\leq b_1, \\
 a_{21}x_1 + a_{22}x_2 &\leq b_2, \\
 &\dots\dots\dots \\
 a_{m1}x_1 + a_{m2}x_2 &\leq b_m, \\
 x_1 \geq 0, \quad x_2 &\geq 0.
 \end{aligned}$$

Ограничения задачи вырезают на плоскости некоторый многоугольник. Пусть при некотором  $L$  прямая  $c_1x_1 + c_2x_2 = L$  пересекает допустимую область. Это пересечение дает какие-то значения переменных  $(x_1, x_2)$ , которые являются планами.

Увеличивая  $L$  мы начнем двигать нашу прямую и её пересечение с допустимой областью будет изменяться (см. рис. 9). В конце концов эта прямая выйдет на границу допустимой области  $\square$  как правило, это будет одна из **вершин многоугольника**. Дальнейшее увеличение  $L$  приведёт к тому, что пересечение

Прикладная математика

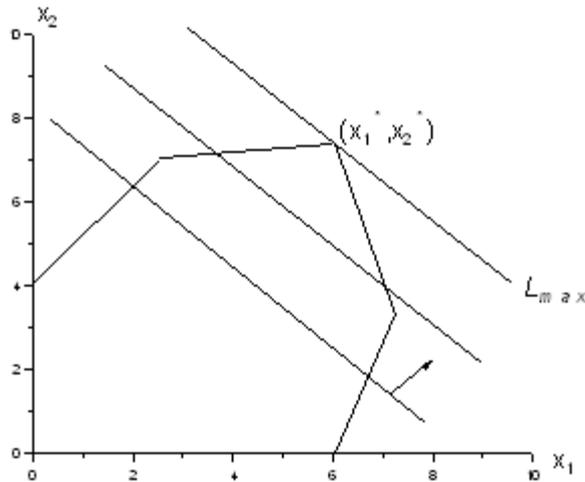


Рис. 9

прямой  $c_1x_1 + c_2x_2 = L$  с допустимой областью будет пустым. Поэтому то положение прямой  $c_1x_1 + c_2x_2 = L$ , при котором она вышла на граничную точку допустимой области, и даст решение задачи, а соответствующее значение  $L$  и будет оптимальным значением целевой функции.

**Пример**

Решить задачу

$$\begin{aligned}
 &x_1 + 2x_2 \Rightarrow \max \\
 &-x_1 + x_2 \leq 1, \\
 &x_1 - 2x_2 \leq 1, \\
 &x_1 + x_2 \leq 3, \\
 &x_1 \geq 0, \quad x_2 \geq 0.
 \end{aligned}
 \tag{1.23}$$

**Решение**

Допустимую область мы уже строили  $\square$  она изображена на рис. 5.

Повторим еще раз этот рисунок, оставив только допустимую область и

нарисовав дополнительно прямые  $c_1x_1 + c_2x_2 = L$  (см. рис. 10).

Прикладная математика

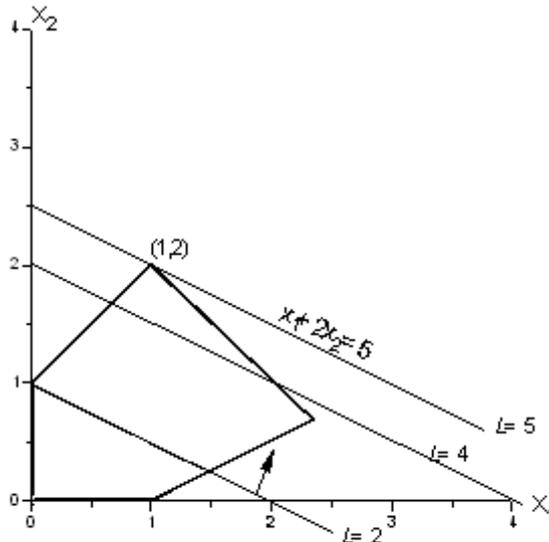


Рис.10

Пусть, например,  $L=2$ . Тогда прямая  $x_1 + 2x_2 = 2$  проходит через точки  $(2,0)$  и  $(0,1)$  и изображена на рис. 10. Будем теперь увеличивать  $L$ . Тогда прямая начнёт двигаться параллельно самой себе в направлении, указанном стрелкой. Легко догадаться, что максимальное значение  $L$  получится тогда, когда прямая пройдет через вершину многоугольника, указанную на рисунке, и дальнейшее увеличение  $L$  приведет к тому, что прямая выйдет за пределы многоугольника и её пересечение с допустимой областью будет пустым.

Выделенная вершина лежит на пересечении прямых

$$\begin{aligned} -x_1 + x_2 &= 1, \\ x_1 + x_2 &= 3 \end{aligned}$$

и поэтому имеет координаты  $x_1 = 1, x_2 = 2$ . Это и есть решение нашей задачи, т.е.  $x_1 = 1, x_2 = 2$  есть **оптимальный план** задачи (1.23). При этом значение целевой функции  $L = 1 + 2 \cdot 2 = 5$ , что и дает её максимальное значение.

Обратите внимание на то, что оптимальный план, как правило, соответствует какой-то вершине многоугольника, изображающего допустимую область. И лишь в том случае, когда прямая  $c_1x_1 + c_2x_2 = L$  случится так, что решение не будет единственным. Но и в этом случае вершины, соответствующие границам этой стороны, дают оптимальные планы нашей задачи линейного программирования. Таким образом, вершины допустимой области играют в решении задач линейного программирования особую роль.

Прикладная математика

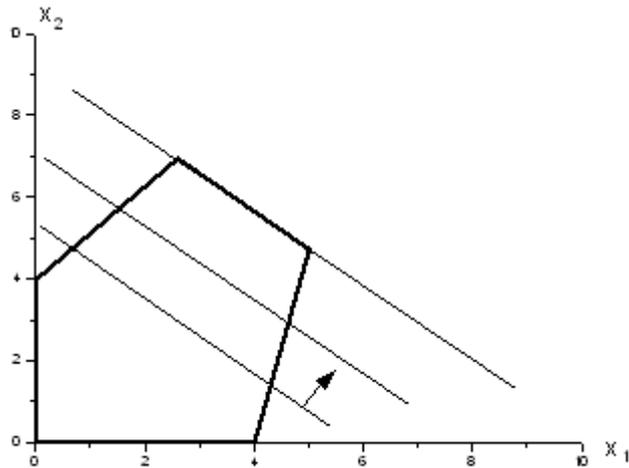


Рис. 11

Ну, а если допустимая область неограничена, то и значение целевой функции может быть неограниченным.

Подводя итог этим примерам, можно сформулировать следующие положения:

1. допустимая область  $\square$  это выпуклый многоугольник;
2. оптимум достигается в вершине допустимой области (если допустимая область ограничена и не пуста);
3. ограниченность целевой функции в допустимой области является необходимым и достаточным условием разрешимости задачи.

Дальнейшее будет посвящено более строгому обоснованию этих утверждений и формулировке алгоритма решения.

---

Задачи

**Решить графически следующие задачи линейного программирования.**

Прикладная математика

$$\begin{aligned}
 &4x_1 + 2x_2 \Rightarrow \max \\
 &2x_1 + 3x_2 \leq 18 \\
 &-x_1 + 3x_2 \leq 9 \\
 &2x_1 - x_2 \leq 10 \\
 1. &x_1 \geq 0 \quad x_2 \geq 0
 \end{aligned}$$

$$\begin{aligned}
 &2x_1 + 4x_2 \Rightarrow \max \\
 &3x_1 + 2x_2 \leq 11 \\
 &-2x_1 + x_2 \leq 2 \\
 &x_1 - 3x_2 \leq 0 \\
 2. &x_1 \geq 0 \quad x_2 \geq 0
 \end{aligned}$$

$$\begin{aligned}
 &x_1 - 2x_2 \Rightarrow \min \\
 &x_1 - x_2 \leq 1 \\
 &x_1 + x_2 \geq 2 \\
 &x_1 - 2x_2 \leq 0 \\
 3. &x_1 \geq 0 \quad x_2 \geq 0
 \end{aligned}$$

$$\begin{aligned}
 &2x_1 + 4x_2 \Rightarrow \max \\
 &-x_1 + x_2 \leq 3 \\
 &x_1 + 2x_2 \leq 12 \\
 &3x_1 - x_2 \leq 15 \\
 4. &x_1 \geq 0 \quad x_2 \geq 0
 \end{aligned}$$

$$\begin{aligned}
 &x_1 + 3x_2 \Rightarrow \max \\
 &x_1 - x_2 \leq 1 \\
 &2x_1 + x_2 \leq 2 \\
 &x_1 - x_2 \geq 0 \\
 5. &x_1 \geq 0 \quad x_2 \geq 0
 \end{aligned}$$

$$\begin{aligned}
 &x_1 + x_2 \Rightarrow \max \\
 &x_1 + 2x_2 \leq 10 \\
 &x_1 + 2x_2 \geq 2 \\
 &2x_1 + x_2 \leq 10 \\
 6. &x_1 \geq 0 \quad x_2 \geq 0
 \end{aligned}$$

Прикладная математика

$$\begin{aligned}x_1 + x_2 &\Rightarrow \max(\min) \\2x_1 + 4x_2 &\leq 16 \\-4x_1 + 2x_2 &\leq 8 \\x_1 + 3x_2 &\geq 9 \\7. \quad x_1 \geq 0 \quad x_2 \geq 0\end{aligned}$$

$$\begin{aligned}2x_1 + 3x_2 &\Rightarrow \max \\2x_1 + x_2 &\leq 10 \\-2x_1 + 3x_2 &\leq 6 \\2x_1 + 4x_2 &\geq 8 \\8. \quad x_1 \geq 0 \quad x_2 \geq 0\end{aligned}$$

$$\begin{aligned}x_1 + x_2 &\Rightarrow \max \\x_1 + 2x_2 &\leq 14 \\-5x_1 + 3x_2 &\leq 15 \\4x_1 + 6x_2 &\geq 24 \\9. \quad x_1 \geq 0 \quad x_2 \geq 0\end{aligned}$$

$$\begin{aligned}x_1 + 2x_2 &\Rightarrow \max \\4x_1 - 2x_2 &\leq 12 \\-x_1 + 3x_2 &\leq 6 \\2x_1 + 4x_2 &\geq 16 \\10. \quad x_1 \geq 0 \quad x_2 \geq 0\end{aligned}$$

## 1.2 БАЗИС ОПОРНОГО ПЛАНА

**Базисное решение** (опорный план) [basic solution] — термин линейного программирования, одно из допустимых решений, находящихся в вершинах области допустимых решений, либо (если линия уровня параллельна одному из отрезков границы области) Б. р. — весь этот отрезок (см. рис. Л. 2 к ст. “Линейное программирование”). Оно является решением системы линейных ограничений, которое нельзя представить в виде линейной комбинации никаких других решений.

При решении задачи линейного программирования можно поступить следующим образом: найти любое из таких “вершинных” решений — не обязательно оптимальное — и принять его за исходный пункт расчетов. Такое решение и будет базисным. Если оно окажется оптимальным, расчет на этом закончен, если нет — последовательно проверяют, не будут ли оптимальными соседние вершинные точки: ту из них, в которой план эффективнее, принимают снова за исходную точку; и так, последовательно проверяя на оптимальность аналогичные вершины, приходят к искомому оптимуму. На этом принципе строятся т. н. симплексный метод решения задач линейного программирования, а также ряд других способов, объединенных общим названием “**методы последовательного улучшения допустимого решения (МПУ)**”: метод обратной матрицы, или модифицированный симплекс-метод, **метод потенциалов** для транспортной задачи и др. Они отличаются друг от друга вычислительными особенностями перехода от одного базисного решения к другому, улучшенному

Решение любой задачи линейного программирования можно найти симплексным методом. Прежде чем применять указанный метод, следует записать исходную задачу в форме основной задачи линейного программирования, если она не имеет такой формы записи.

Симплексный метод решения задачи линейного программирования основан на переходе от одного опорного плана к другому, при котором значение целевой функции возрастает (при условии, что данная задача имеет оптимальный план и каждый ее опорный план является невырожденным). Указанный переход возможен, если известен какой-нибудь исходный опорный план. Рассмотрим задачу, для которой этот план можно непосредственно записать.

Пусть требуется найти максимальное значение функции



Прикладная математика

$P_1, P_2, \dots, P_m$ , а также вектор  $P_0$  могут быть представлены в виде линейной комбинации векторов данного базиса. Пусть

$$P_j = \sum_{i=1}^m x_{ij} P_i \quad (j = \overline{0, m}).$$

Положим  $z_j = \sum_{i=1}^m c_i x_{ij} \quad (j = \overline{1, n}), \Delta_j = z_j - c_j \quad (j = \overline{1, n})$ . Так как векторы  $P_1, P_2, \dots, P_m$  –

единичные, то  $x_{ij} = a_{ij}$  и  $z_j = \sum_{i=1}^m c_i a_{ij}$ , а

$$\Delta_j = \sum_{i=1}^m c_i a_{ij} - c_j.$$

**Теорема 5**

(признак оптимальности опорного плана). *Опорный план  $X^* = (x_1^*, x_2^*, \dots, x_m^*; 0; 0; \dots; 0)$  задачи (22) – (24) является оптимальным, если  $\Delta_j \geq 0$  для любого  $j \in \overline{1, n}$*

**Теорема 6.**

*Если  $\Delta_k < 0$  для некоторого  $j=k$  и среди чисел  $a_{ik} \quad (i = \overline{1, m})$  нет положительных ( $a_{ik} \leq 0$ ), то целевая функция (22) задачи (22) – (24) не ограничена на множестве ее планов.*

**Теорема 7.**

*Если опорный план  $X$  задачи (22) – (24) невырожден и  $\Delta_k < 0$ , но среди чисел  $a_{ik}$  есть положительные (не все  $a_{ik} \leq 0$ ), то существует опорный план  $X'$  такой, что  $F(X') > F(X)$ .*

Сформулированные теоремы позволяют проверить, является ли найденный опорный план оптимальным, и выявить целесообразность перехода к новому опорному плану.

Исследование опорного плана на оптимальность, а также дальнейший вычислительный процесс удобнее вести, если условия задачи и первоначальные данные, полученные после определения исходного опорного плана, записать так, как показано в табл. 3.

Прикладная математика

В столбце  $C_6$  этой таблицы записывают коэффициенты при неизвестных целевой функции, имеющие те же индексы, что и векторы данного базиса.

В столбце  $P_0$  записывают положительные компоненты исходного опорного плана, в нем же в результате вычислений получают положительные компоненты оптимального плана. Столбцы векторов  $P_j$  представляют собой коэффициенты разложения этих векторов по векторам данного базиса.

В табл. 3 первые  $m$  строк определяются исходными данными задачи, а показатели  $(m+1)$ -й строки вычисляют. В этой строке в столбце вектора  $P_0$  записывают значение целевой функции, которое она принимает при данном опорном плане, а в столбце вектора  $P_j$  – значение  $\Delta_j = z_j - c_j$ .

Значение  $Z_j$  находится как скалярное произведение вектора  $P_j (j = \overline{1, m})$  на вектор  $C_6 = (c_1; c_2; \dots; c_m)$ :

$$z_j = \sum_{i=1}^m c_i a_{ij} \quad (j = \overline{1, n}).$$

Значение  $P_0$  равно скалярному произведению вектора  $P_0$  на вектор  $C_6$ :

$$P_0 = \sum_{i=1}^m c_i b_i.$$

После заполнения таблицы 3 исходный опорный план проверяют на оптимальность. Для этого просматривают элементы  $(m+1)$ -й строки таблицы. В результате может иметь место один из следующих трех случаев:

- 1)  $\Delta_j \geq 0$  для  $j=m+1, m+2, \dots, n$  (при  $j = \overline{1, m}$   $z_j = c_j$ ). Поэтому в данном случае числа  $\Delta_j \geq 0$  для всех  $j$  от 1 до  $n$ ;
- 2)  $\Delta_j < 0$  для некоторого  $j$ , и все соответствующие этому индексу величины  $a_{ij} \leq 0 (i = \overline{1, m})$ ;
- 3)  $\Delta_j < 0$  для некоторых индексов  $j$ , и для каждого такого  $j$ , по крайней мере, одно из чисел  $a_{ij}$  положительно.

В первом случае на основании признака оптимальности исходный опорный план является оптимальным. Во втором случае целевая функция не ограничена сверху на множестве планов, а в третьем случае можно перейти от исходного плана к новому опорному плану, при котором значение целевой

Прикладная математика

функции увеличится. Этот переход от одного опорного плана к другому осуществляется исключением из исходного базиса какого-нибудь из векторов и введением в него нового вектора. В качестве вектора, вводимого в базис, можно взять любой из векторов  $F_j$  имеющий индекс  $j$ , для которого  $\Delta_j < 0$ . Пусть, например,  $\Delta_k < 0$  и решено ввести в базис вектор  $F_k$ .

Для определения вектора, подлежащего исключению из базиса, находят  $\min \frac{b_i}{a_{ik}}$  для всех  $a_{ik} > 0$ . Пусть этот минимум достигается при  $i=r$ . Тогда из базиса исключают вектор  $F_r$ , а число  $a_{rk}$  называют *разрешающим элементом*.

Столбец и строку, на пересечении которых находится разрешающий элемент, называют *направляющими*.

После выделения направляющей строки и направляющего столбца находят новый опорный план и коэффициенты разложения векторов  $F_j$  через векторы нового базиса, соответствующего новому опорному плану. Это легко реализовать, если воспользоваться методом Жордана–Гаусса. При этом можно показать, что положительные компоненты нового опорного плана вычисляются по формулам

$$b'_i = \begin{cases} b_i - \left(\frac{b_r}{a_{rk}}\right) a_{ik} & \text{при } i \neq r, \\ \frac{b_r}{a_{rk}} & \text{при } i = r, \end{cases} \quad (25)$$

а коэффициенты разложения векторов  $F_j$  через векторы нового базиса, соответствующего новому опорному плану, – по формулам

$$a'_{ij} = \begin{cases} a_{ij} - \left(\frac{a_{rj}}{a_{rk}}\right) a_{ik} & \text{при } i \neq r, \\ \frac{a_{rj}}{a_{rk}} & \text{при } i = r. \end{cases} \quad (26)$$

После вычисления  $b'_i$  и  $a'_{ij}$  согласно формулам (25) и (26) их значения заносят в табл. 4. Элементы  $(m+1)$ -й строки этой таблицы могут быть вычислены либо по формулам

$$F'_0 = F_0 - \left(\frac{b_r}{a_{rk}}\right) \Delta_k, \quad (27)$$

$$\Delta'_j = \Delta_j - \left(\frac{a_{rj}}{a_{rk}}\right) \Delta_k, \quad (28)$$

либо на основании их определения.

Таблица 3

i	Базис	C <sub>б</sub>	P <sub>0</sub>	c <sub>1</sub>	c <sub>2</sub>	...	c <sub>r</sub>	...	c <sub>m</sub>	c <sub>m+1</sub>	...	c <sub>k</sub>	...	c <sub>n</sub>
				P <sub>1</sub>	P <sub>2</sub>	...	P <sub>r</sub>	...	P <sub>m</sub>	P <sub>m+1</sub>	...	P <sub>k</sub>	...	P <sub>n</sub>
1	P <sub>1</sub>	c <sub>1</sub>	b <sub>1</sub>	1	0	...	0	...	0	a <sub>1m+1</sub>	...	a <sub>1k</sub>	...	a <sub>1n</sub>
2	P <sub>2</sub>	c <sub>2</sub>	b <sub>2</sub>	0	1	...	0	...	0	a <sub>2m+1</sub>	...	a <sub>2k</sub>	...	a <sub>2n</sub>
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
r	P <sub>r</sub>	c <sub>r</sub>	b <sub>r</sub>	0	0	...	1	...	0	a <sub>rm+1</sub>	...	a <sub>rk</sub>	...	a <sub>rn</sub>
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
m	P <sub>m</sub>	c <sub>m</sub>	b <sub>m</sub>	0	0	...	0	...	1	a <sub>mm+1</sub>	...	a <sub>mk</sub>	...	a <sub>mn</sub>
m+1			F <sub>m</sub>	0	0	...	0	...	0	Δ <sub>m+1</sub>	...	Δ <sub>k</sub>	...	Δ <sub>n</sub>

Таблица 4

i	Базис	C <sub>б</sub>	P <sub>0</sub>	c <sub>1</sub>	c <sub>2</sub>	...	c <sub>r</sub>	...	c <sub>m</sub>	c <sub>m+1</sub>	...	c <sub>k</sub>	...	c <sub>n</sub>
				P <sub>1</sub>	P <sub>2</sub>	...	P <sub>r</sub>	...	P <sub>m</sub>	P <sub>m+1</sub>	...	P <sub>k</sub>	...	P <sub>n</sub>
1	P <sub>1</sub>	c <sub>1</sub>	b' <sub>1</sub>	1	0	...	a' <sub>1r</sub>	...	0	a' <sub>1m+1</sub>	...	0	...	a' <sub>1n</sub>
2	P <sub>2</sub>	c <sub>2</sub>	b' <sub>2</sub>	0	1	...	a' <sub>2r</sub>	...	0	a' <sub>2m+1</sub>	...	0	...	a' <sub>2n</sub>
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
r	P <sub>r</sub>	c <sub>r</sub>	b' <sub>r</sub>	0	0	...	a' <sub>rr</sub>	...	0	a' <sub>rm+1</sub>	...	1	...	a' <sub>rn</sub>
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
m	P <sub>m</sub>	c <sub>m</sub>	b' <sub>m</sub>	0	0	...	a' <sub>mr</sub>	...	1	a' <sub>mm+1</sub>	...	0	...	a' <sub>mn</sub>
m+1			F <sub>m</sub>	0	0	...	z' <sub>r</sub> - c <sub>r</sub>	...	0	z' <sub>m+1</sub> - c <sub>m+1</sub>	...	0	...	z' <sub>n</sub> - c <sub>n</sub>

Наличие двух способов нахождения элементов  $(m+1)$ -й строки позволяет осуществлять контроль правильности проводимых вычислений.

Из формулы (27) следует, что при переходе от одного опорного плана к другому наиболее целесообразно ввести в базис вектор  $P_j$ , имеющий индекс  $j$ , при котором максимальным по абсолютной величине является число  $\left(\frac{b_r}{a_{rj}}\right) \Delta_j$  ( $\Delta_j < 0, a_{rj} > 0$ ). Однако с целью упрощения вычислительного процесса в дальнейшем будем вектор, вводимый в базис, определять, исходя из максимальной абсолютной величины отрицательных чисел  $\Delta_j$ . Если же таких чисел несколько, то в базис будем вводить вектор, имеющий такой же индекс, как и максимальное из чисел  $c_j$ , определяемых данными числами  $\Delta_j$  ( $\Delta_j < 0$ ).

Прикладная математика

Итак, переход от одного опорного плана к другому сводится к переходу от одной симплекс-таблицы к другой. Элементы новой симплекс-таблицы можно вычислить как с помощью рекуррентных формул (25)-(28), так и по правилам, непосредственно вытекающим из них. Эти правила состоят в следующем.

В столбцах векторов, входящих в базис, на пересечении строк и столбцов одноименных векторов проставляются единицы, а все остальные элементы данных столбцов полагают равными нулю.

Элементы векторов  $P_0$  и  $P_j$  в строке новой симплекс-таблицы, в которой записан вектор, вводимый в базис, получают из элементов этой же строки исходной таблицы делением их на величину разрешающего элемента. В столбце  $C_k$  в строке вводимого вектора проставляют величину  $c_k$ , где  $k$  – индекс вводимого вектора.

Остальные элементы столбцов вектора  $P_0$  и  $P_j$  новой симплекс-таблицы вычисляют по правилу треугольника. Для вычисления какого-нибудь из этих элементов находят три числа:

- 1) число, стоящее в исходной симплекс-таблице на месте искомого элемента новой симплекс-таблицы;
- 2) число, стоящее в исходной симплекс-таблице на пересечении строки, в которой находится искомый элемент новой симплекс-таблицы, и столбца, соответствующего вектору, вводимому в базис;
- 3) число, стоящее в новой симплекс-таблице на пересечении столбца, в котором стоит искомый элемент, и строки вновь вводимого в базис вектора (как отмечено выше, эта строка получается из строки исходной симплекс-таблицы делением ее элементов на разрешающий элемент).

Эти три числа образуют своеобразный треугольник, две вершины которого соответствуют числам, находящимся в исходной симплекс-таблице, а третья – числу, находящемуся в новой симплекс-таблице. Для определения искомого элемента новой симплекс-таблицы из первого числа вычитают произведение второго и третьего.

После заполнения новой симплекс-таблицы просматривают элементы  $(m+1)$ -й строки. Если все  $z_j' - c_j \geq 0$ , то новый опорный план является оптимальным. Если же среди указанных чисел имеются отрицательные, то, используя описанную выше последовательность действий, находят новый опорный план. Этот процесс продолжают до тех пор, пока либо не получают оптимальный план задачи, либо не устанавливают ее неразрешимость.

Прикладная математика

При нахождении решения задачи линейного программирования мы предполагали, что эта задача имеет опорные планы и каждый такой план является невырожденным. Если же задача имеет вырожденные опорные планы, то на одной из итераций одна или несколько переменных опорного плана могут оказаться равными нулю. Таким образом, при переходе от одного опорного плана к другому значение функции может остаться прежним. Более того, возможен случай, когда функция сохраняет свое значение в течение нескольких итераций, а также возможен возврат к первоначальному базису. В последнем случае обычно говорят, что произошло *защикливание*. Однако при решении практических задач этот случай встречается очень редко, поэтому мы на нем останавливаться не будем.

Итак, нахождение оптимального плана симплексным методом включает следующие этапы:

1. Находят опорный план.

2. Составляют симплекс-таблицу.

3. Выясняют, имеется ли хотя бы одно отрицательное число  $\Delta_j$ . Если нет, то найденный опорный план оптимален. Если же среди чисел  $\Delta_j$  имеются отрицательные, то либо устанавливают неразрешимость задачи, либо переходят к новому опорному плану.

4. Находят направляющие столбец и строку. Направляющий столбец определяется наибольшим по абсолютной величине отрицательным числом  $\Delta_j$ , а направляющая строка – минимальным из отношений компонент столбца вектора  $F_0$  к положительным компонентам направляющего столбца.

5. По формулам (25) – (28) определяют положительные компоненты нового опорного плана, коэффициенты разложения векторов  $P_j$  по векторам нового базиса и числа  $F_0'$ ,  $\Delta_j'$ . Все эти числа записываются в новой симплекс-таблице.

6. Проверяют найденный опорный план на оптимальность. Если план не оптимален и необходимо перейти к новому опорному плану, то возвращаются к этапу 4, а в случае получения оптимального плана или установления неразрешимости процесс решения задачи заканчивают.

**Пример 9.**

Прикладная математика

Для изготовления различных изделий  $A$ ,  $B$  и  $C$  предприятие использует три различных вида сырья. Нормы расхода сырья на производство одного изделия каждого вида, цена одного изделия  $A$ ,  $B$  и  $C$ , а также общее количество сырья каждого вида, которое может быть использовано предприятием, приведены в табл. 5.

Таблица 5

Вид сырья	Нормы затрат сырья (кг) на одно изделие			Общее количество сырья (кг)
	$A$	$B$	$C$	
I	18	15	12	360
II	6	4	8	192
III	5	3	3	180
Цена одного изделия (руб.)	9	10	16	

Изделия  $A$ ,  $B$  и  $C$  могут производиться в любых соотношениях (сбыт обеспечен), но производство ограничено выделенным предприятию сырьем каждого вида.

Составить план производства изделий, при котором общая стоимость всей произведенной предприятием продукции является максимальной.

**Решение.** Составим математическую модель задачи. Искомый выпуск изделий  $A$  обозначим через  $x_1$ , изделий  $B$  – через  $x_2$ , изделий  $C$  – через  $x_3$ . Поскольку имеются ограничения на выделенный предприятию фонд сырья каждого вида, переменные  $x_1, x_2, x_3$  должны удовлетворять следующей системе неравенств:

$$\begin{cases} 18x_1 + 15x_2 + 12x_3 \leq 360, \\ 6x_1 + 4x_2 + 8x_3 \leq 192, \\ 5x_1 + 3x_2 + 3x_3 \leq 180. \end{cases} \quad (29)$$

Прикладная математика

Общая стоимость произведенной предприятием продукции при условии выпуска  $x_1$  изделий  $A$ ,  $x_2$  изделий  $B$  и  $x_3$  изделий  $C$  составляет

$$F = 9x_1 + 10x_2 + 16x_3. \quad (30)$$

По своему экономическому содержанию переменные  $x_1, x_2, x_3$  могут принимать только лишь неотрицательные значения:

$$x_1, x_2, x_3 \geq 0. \quad (31)$$

Таким образом, приходим к следующей математической задаче: среди всех неотрицательных решений системы неравенств (29) требуется найти такое, при котором функция (30) принимает максимальное значение.

Запишем эту задачу в форме основной задачи линейного программирования. Для этого перейдем от ограничений-неравенств к ограничениям-равенствам. Введем три дополнительные переменные, в результате чего ограничения запишутся в виде системы уравнений

$$\begin{cases} 18x_1 + 15x_2 + 12x_3 + x_4 = 360, \\ 6x_1 + 4x_2 + 8x_3 + x_5 = 192, \\ 5x_1 + 3x_2 + 3x_3 + x_6 = 180. \end{cases}$$

Эти дополнительные переменные по экономическому смыслу означают не используемое при данном плане производства количество сырья того или иного вида. Например,  $x_4$  – это неиспользуемое количество сырья I вида.

Преобразованную систему уравнений запишем в векторной форме:

$$x_1P_1 + x_2P_2 + x_3P_3 + x_4P_4 + x_5P_5 + x_6P_6 = P_0,$$

где

$$P_1 = \begin{pmatrix} 18 \\ 6 \\ 5 \end{pmatrix}; P_2 = \begin{pmatrix} 15 \\ 4 \\ 3 \end{pmatrix}; P_3 = \begin{pmatrix} 12 \\ 8 \\ 3 \end{pmatrix}; P_4 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}; P_5 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}; P_6 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}; P_0 = \begin{pmatrix} 360 \\ 192 \\ 180 \end{pmatrix}.$$

Поскольку среди векторов  $P_1, P_2, P_3, P_4, P_5, P_6$  имеются три единичных вектора, для данной задачи можно непосредственно записать опорный план. Таковым является план  $X=(0; 0; 0; 360; 192; 180)$ , определяемый системой трехмерных единичных векторов  $P_4, P_5, P_6$  которые образуют базис трехмерного векторного пространства.

Прикладная математика

Составляем симплексную таблицу для I итерации (табл. 6), подсчитываем значения  $F_0, z_j - c_j$  и проверяем исходный опорный план на оптимальность:

$$F_0 = (C, P_0) = 0; z_1 = (C, P_1) = 0; z_2 = (C, P_2) = 0; z_3 = (C, P_3) = 0;$$

$$z_1 - c_1 = 0 - 9 = -9; z_2 - c_2 = 0 - 10 = -10; z_3 - c_3 = -16.$$

Для векторов базиса  $z_j - c_j = 0$ .

Таблица 6

i	Бази с	C b	P <sub>0</sub>	9	1	1	0	0	0
				0	0	6	0	0	0
				P	P	P	p	P	P
				1	2	3	4	5	6
1	P <sub>4</sub>	0	36	1	1	1	1	0	0
2	p <sub>5</sub>	0	0	8	5	2	0	1	0
3	p <sub>6</sub>	0	19	6	4	8	0	0	1
4			2	5	3	3	0	0	0
			18	-	-	-			
			0	9	10	16			
			0						

Как видно из таблицы 6, значения всех основных переменных  $x_1, x_2, x_3$  равны нулю, а дополнительные переменные принимают свои значения в соответствии с ограничениями задачи. Эти значения переменных отвечают такому “плану”, при котором ничего не производится, сырье не используется и значение целевой функции равно нулю (т. е. стоимость произведенной продукции отсутствует). Этот план, конечно, не является оптимальным.

Это видно и из 4-й строки табл. 6, так как в ней имеется три отрицательных числа:  $z_1 - c_1 = -9, z_2 - c_2 = -10$  и  $z_3 - c_3 = -16$ . Отрицательные числа не только свидетельствуют о возможности увеличения общей стоимости производимой продукции, но и показывают, на сколько увеличится эта сумма при введении в план единицы того или другого вида продукции.

Прикладная математика

Так, число – 9 означает, что при включении в план производства одного изделия  $A$  обеспечивается увеличение выпуска продукции на 9 руб. Если включить в план производства по одному изделию  $B$  и  $C$ , то общая стоимость изготавливаемой продукции возрастет соответственно на 10 и 16 руб. Поэтому с экономической точки зрения наиболее целесообразным является включение в план производства изделий  $C$ . Это же необходимо сделать и на основании формального признака симплексного метода, поскольку максимальное по абсолютной величине отрицательное число  $\Delta_j$  стоит в 4-й строке столбца вектора  $P_3$ . Следовательно, в базис введем вектор  $P_3$ . определяем вектор, подлежащий исключению из базиса. Для этого находим

$$\Theta_0 = \min \left( \frac{b_i}{a_{i3}} \right) \text{ для } a_{i3} > 0, \text{ т. е. } \Theta_0 = \min \left( \frac{360}{12}; \frac{192}{8}; \frac{180}{3} \right) = \frac{192}{8}.$$

Найдя число  $\frac{192}{8} = 24$ , мы тем самым с экономической точки зрения определили, какое количество изделий  $C$  предприятие может изготавливать с учетом норм расхода и имеющихся объемов сырья каждого вида. Так как сырья данного вида соответственно имеется 360, 192 и 180 кг, а на одно изделие  $C$  требуется затратить сырья каждого вида соответственно 12, 8 и 3 кг, то максимальное число изделий  $C$ , которое может быть изготовлено

предприятием, равно  $\min \left( \frac{360}{12}; \frac{192}{8}; \frac{180}{3} \right) = \frac{192}{8} = 24$ , т. е. ограничивающим фактором для производства изделий  $C$  является имеющийся объем сырья II вида. С учетом его наличия предприятие может изготовить 24 изделия  $C$ . При этом сырье II вида будет полностью использовано.

Следовательно, вектор  $P_5$  подлежит исключению из базиса. Столбец вектора  $P_3$  к 2-я строка являются направляющими. Составляем таблицу для II итерации (табл. 7).

Таблица 7

$i$	Бази с	$C$ $b$	$P_0$	9	1 0	1 6	0	0	0
				$P_1$	$P_2$	$P_3$	$P_4$	$P_5$	$P_6$
1	$P_4$	0	72	9	9	0	1	-	0
2	$P_3$	1	24	3/4	1/2	1	0	3/2	0
3	$P_6$	6	10	11/4		0	0	1/8	1
4		0	8	4		0	0		0

Прикладная математика

			38	3	3/			-	
			4		2			3/8	
					-2			2	

Сначала заполняем строку вектора, вновь введенного в базис, т. е. строку, номер которой совпадает с номером направляющей строки. Здесь направляющей является 2-я строка. Элементы этой строки табл. 7 получаются из соответствующих элементов таблицы 6 делением их на разрешающий элемент (т. е. на 8). При этом в столбце  $C_6$  записываем коэффициент  $E_0 E$ , стоящий в столбце вводимого в базис вектора  $E$ . Затем заполняем элементы столбцов для векторов, входящих в новый базис. В этих столбцах на пересечении строк и столбцов одноименных векторов проставляем единицы, а все остальные элементы полагаем равными нулю.

Для определения остальных элементов табл. 7 применяем правило треугольника. Эти элементы могут быть вычислены и непосредственно по рекуррентным формулам.

Вычислим элементы табл. 7, стоящие в столбце вектора  $P_0$ . Первый из них находится в 1-й строке этого столбца. Для его вычисления находим три числа:

- 1) число, стоящее в табл. 6 на пересечении столбца вектора  $P_0$  и 1-й строки (360);
- 2) число, стоящее в табл. 6 на пересечении столбца вектора  $P_3$  и 1-й строки (12);
- 3) число, стоящее в табл. 7 на пересечении столбца вектора  $P_0$  и 2-й строки (24).

Вычитая из первого числа произведение двух других, находим искомый элемент:  $360 - 12 \times 24 = 72$ ; записываем его в 1-й строке столбца вектора  $P_0$  табл. 7.

Второй элемент столбца вектора  $P_0$  табл. 7 был уже вычислен ранее. Для вычисления третьего элемента столбца вектора  $P_0$  также находим три числа. Первое из них (180) находится на пересечении 3-й строки и столбца вектора  $P_0$  табл. 6, второе (3) – на пересечении 3-й строки и столбца вектора  $P_3$  табл. 6, третье (24) – на пересечении 2-й строки и столбца вектора  $P_0$  табл. 8. Итак, указанный элемент есть  $180 - 24 \times 3 = 108$ . Число 108 записываем в 3-й строке столбца вектора  $P_0$  табл. 7.

Прикладная математика

Значение  $F_0$  в 4-й строке столбца этого же вектора можно найти двумя способами:

1) по формуле  $F_0 = (C, P_0)$ , т.е.  $F_0 = 0 \cdot 72 + 16 \cdot 24 + 0 \cdot 108 = 384$ ;

2) по правилу треугольника; в данном случае треугольник образован числами 0, -16, 24. Этот способ приводит к тому же результату:  $0 - (-16) \times 24 = 384$ .

При определении по правилу треугольника элементов столбца вектора  $P_0$  третье число, стоящее в нижней вершине треугольника, все время оставалось неизменным и менялись лишь первые два числа. Учтем это при нахождении элементов столбца вектора  $P_1$  табл. 7. Для вычисления указанных элементов первые два числа берем из столбцов векторов  $P_1$  и  $P_3$  табл. 6, а третье число – из табл. 7. Это число стоит на пересечении 2-й строки и столбца вектора  $P_1$  последней таблицы. В результате получаем значения искомых элементов:  $18 - 12 \times (3/4) = 9$ ;  $5 - 3 \times (3/4) = 11/4$ .

Число  $z_1 - c_1$  в 4-й строке столбца вектора  $P_1$  табл. 7 можно найти двумя способами:

1) по формуле  $Z_1 - C_1 = (C, P_1) - C_1$  имеем  $0 \cdot 9 + 16 \cdot 3/4 + 0 \cdot 11/4 - 9 = 3$ ;

2) по правилу треугольника получим  $-9 - (-16) \cdot (3/4) = 3$ .

Аналогично находим элементы столбца вектора  $P_2$ .

Элементы столбца вектора  $P_5$  вычисляем по правилу треугольника. Однако построенные для определения этих элементов треугольники выглядят иначе.

При вычислении элемента 1-й строки указанного столбца получается треугольник, образованный числами 0, 12 и  $1/8$ . Следовательно, искомый элемент равен  $0 - 12 \times (1/8) = -3/2$ . Элемент, стоящий в 3-й строке данного столбца, равен  $0 - 3 \times (1/8) = -3/8$ .

По окончании расчета всех элементов табл. 7 в ней получены новый опорный план и коэффициенты разложения векторов  $P \{j = \overline{1, 6}\}$  через базисные векторы  $P_4, P_3, P_6$  и значения  $\Delta'_j$  и  $F'_0$ . Как видно из этой таблицы, новым опорным планом задачи является план  $X = (0; 0; 24; 72; 0; 108)$ . При данном плане производства изготавливается 24 изделия  $C$  и остается неиспользованным 72 кг сырья I вида и 108 кг сырья III вида. Стоимость всей производимой при этом плане продукции равна 384 руб. Указанные числа записаны в столбце вектора  $P_0$  табл. 7. Как видно, данные этого

Прикладная математика

столбца по-прежнему представляют собой параметры рассматриваемой задачи, хотя они претерпели значительные изменения. Изменились данные и других столбцов, а их экономическое содержание стало более сложным. Так, например, возьмем данные столбца вектора  $P_2$ . Число  $1/2$  во 2-й строке этого столбца показывает, на сколько следует уменьшить изготовление изделий  $C$ , если запланировать выпуск одного изделия  $B$ . Числа  $9$  и  $3/2$  в 1-й и 3-й строках вектора  $P_2$  показывают соответственно, сколько потребуется сырья I и II вида при включении в план производства одного изделия  $B$ , а число  $-2$  в 4-й строке показывает, что если будет запланирован выпуск одного изделия  $B$ , то это обеспечит увеличение выпуска продукции в стоимостном выражении на 2 руб. Иными словами, если включить в план производства продукции одно изделие  $B$ , то это потребует уменьшения выпуска изделия  $C$  на  $1/2$  ед. и потребует дополнительных затрат 9 кг сырья I вида и  $3/2$  кг сырья III вида, а общая стоимость изготавливаемой продукции в соответствии с новым оптимальным планом возрастет на 2 руб. Таким образом, числа  $9$  и  $3/2$  выступают как бы новыми “нормами” затрат сырья I и III вида на изготовление одного изделия  $B$  (как видно из табл. 6, ранее они были равны 15 и 3), что объясняется уменьшением выпуска изделий  $C$ .

Такой же экономический смысл имеют и данные столбца вектора  $P_1$  табл. 7. Несколько иное экономическое содержание имеют числа, записанные в столбце вектора  $P_5$ . Число  $1/8$  во 2-й строке этого столбца, показывает, что увеличение объемов сырья II вида на 1 кг позволило бы увеличить выпуск изделий  $C$  на  $1/8$  ед. Одновременно потребовалось бы дополнительно  $3/2$  кг сырья I вида и  $3/8$  кг сырья III вида. Увеличение выпуска изделий  $C$  на  $1/8$  ед. приведет к росту выпуска продукции на 2 руб.

Из изложенного выше экономического содержания данных табл. 7 следует, что найденный на II итерации план задачи не является оптимальным. Это видно и из 4-й строки табл. 7, поскольку в столбце вектора  $P_2$  этой строки стоит отрицательное число  $-2$ . Значит, в базис следует ввести вектор  $P_2$ , т. е. в новом плане следует предусмотреть выпуск изделий  $B$ . При определении возможного числа изготовления изделий  $B$  следует учитывать имеющееся количество сырья каждого вида, а именно:

возможный выпуск изделий  $B$  определяется  $\min\left(\frac{b'_i}{a'_{i2}}\right)$  для  $a'_{i2} > 0$ , т. е. находим

$$\Theta_0 = \min\left(\frac{72}{9}; \frac{24 \cdot 2}{1}; \frac{108 \cdot 2}{3}\right) = \frac{72}{9} = 8.$$

Следовательно, исключению из базиса подлежит вектор  $P_4$  иными словами, выпуск изделий  $B$  ограничен имеющимся в распоряжении предприятия сырьем I вида. С учетом имеющихся объемов этого сырья предприятию следует изготовить 8 изделий  $B$ . Число 9 является

Прикладная математика

разрешающим элементом, а столбец вектора  $P_2$  и 1-я строка табл. 7 являются направляющими. Составляем таблицу для III итерации (табл. 8).

Таблица 8

$i$	Базис	$C_b$	$P_0$	9	1	1	0	0	0
				$P_1$	$P_2$	$P_3$	$p_4$	$p_5$	$P_6$
1	$P_2$	1	8	1	1	0	1/9	-	0
2	$P_3$	0	20	1/4	0	1	-	5/2	0
3	$P_6$	1	96	5/4	0	0	1/18	4	1
4		0	40	4	0	0	-	-	0
			0	5			1/6	1/8	
							2/9	5/3	

В табл. 8 сначала заполняем элементы 1-й строки, которая представляет собой строку вновь вводимого в базис вектора  $P_2$ . Элементы этой строки получаем из элементов 1-й строки табл. 7 делением последних на разрешающий элемент (т.е. на 9). При этом в столбце  $C_b$  данной строки записываем  $C_2 = 10$ .

Затем заполняем элементы столбцов векторов базиса и по правилу треугольника вычисляем элементы остальных столбцов. В результате в табл. 8 получаем новый опорный план  $X=(0; 8; 20; 0; 0; 96)$  и коэффициенты разложения векторов  $P_j (j = \overline{1, 6})$  через базисные векторы  $(j = \overline{1, n})$  и соответствующие значения  $\Delta_j^*$  и  $F_0^*$ .

Проверяем, является ли данный опорный план оптимальным или нет. Для этого рассмотрим 4-ю строку, табл. 8. В этой строке среди чисел  $\Delta_j^*$  нет отрицательных. Это означает, что найденный опорный план является оптимальным и  $F_{\max}^* = 400$ .

Следовательно, план выпуска продукции, включающий изготовление 8 изделий  $B$  и 20 изделий  $C$ , является оптимальным. При данном плане выпуска изделий полностью используется сырье I и II видов и остается

Прикладная математика

неиспользованным 96 кг сырья III вида, а стоимость производимой продукции равна 400 руб.

Оптимальным планом производства продукции не предусматривается изготовление изделий *A*. Введение в план выпуска продукции изделий вида *A* привело бы к уменьшению указанной общей стоимости. Это видно из 4-й строки столбца вектора  $P_1$ , где число 5 показывает, что при данном плане включение в него выпуска единицы изделия *A* приводит лишь к уменьшению общей величины стоимости на 5 руб.

Решение данного примера симплексным методом можно было бы проводить, используя лишь одну таблицу (табл. 9). В этой таблице последовательно записаны одна за другой все три итерации вычислительного процесса.

Таблица 9

$i$	Баз ис	$C$ $b$	$P_0$	9	1 0	1 6	0	0	0
				$P_1$	$P_2$	$P_3$	$p_4$	$p_5$	$P_6$
1	$P_4$	0	36	18	1	1	1	0	0
2	$p_5$	0	0	6	5	2	0	1	0
3	$p_6$	0	19	5	4	8	0	0	1
4	$P_4$	0	2	-9	3	3	0	0	0
1	$p_3$	1	0	9	-	-	1	-	0
2	$p_6$	6	0	3/4	10	16	0	3/2	0
3	$P_2$	0	72	11/4	1/2	1	0	1/8	1
4	$p_3$	0	24	4	2	0	0	-	0
1	$p_6$	1	10	3	3/2	0	0	3/8	0
2		6	8	1	2	0	1/9	2	0
3		0	38	1/4	-2	0	9	-	0
4			4	5/4	1	1	1/18	1/6	1
			8						0

Прикладная математика

			20	5	0	0	-	5/2
			96		0	0	1/6	4
			40		0		2/9	-
			0					1/8
								5/3

**Пример 10.**

Найти максимум функции  $F = 2x_1 - 6x_2 + 5x_5$  при условиях

$$\begin{cases} -2x_1 + x_2 + x_3 + x_5 = 20, \\ -x_1 - 2x_1 + x_4 + 3x_5 = 24, \\ 3x_1 - x_2 - 12x_5 + x_6 = 18, \\ x_j \geq 0 \ (j = \overline{1, 6}). \end{cases}$$

**Решение.** Систему уравнений задачи запишем в векторной форме:

$$x_1 P_1 + x_2 P_2 + x_3 P_3 + x_4 P_4 + x_5 P_5 + x_6 P_6 = P_0,$$

где

$$P_1 = \begin{pmatrix} -2 \\ -1 \\ 3 \end{pmatrix}; P_2 = \begin{pmatrix} 1 \\ -2 \\ -1 \end{pmatrix}; P_3 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}; P_4 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}; P_5 = \begin{pmatrix} 1 \\ 3 \\ -12 \end{pmatrix}; P_6 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}; P_0 = \begin{pmatrix} 20 \\ 24 \\ 18 \end{pmatrix}.$$

Так как среди векторов  $P_1, P_2, P_3, P_4, P_5, P_6$  имеется три единичных вектора, то для данной задачи можно непосредственно найти опорный план. Таковым является план  $X = (0, 0, 20, 24; 0; 18)$ . Составляем симплексную таблицу (табл. 10) и проверяем, является ли данный опорный план оптимальным.

**Таблица 10**

$i$	Бази с	$C$ $b$	$P$ 0	2	-6	0	0	5	0
				$P$ 1	$P$ 2	$P$ 3	$P$ 4	$P$ 5	$P$ 6

Прикладная математика

1	$p_3$	0	2	-2	1	1	0	1	0
2	$P_4$	0	0 24 18	-1	-2	0	1	3	0
3	$p_6$	0	0	3	-1	0	0	-	1
4				-2	6	0	0	12 - 5	0

Как видно из табл. 10, исходный опорный план не является оптимальным. Поэтому переходим к новому опорному плану. Это можно сделать, так как в столбцах векторов  $P_1$  и  $p_5$ , 4-я строка которых содержит отрицательные числа, имеются положительные элементы. Для перехода к новому опорному плану введем в базис вектор  $p_5$  и исключим из базиса вектор  $p_4$ . Составляем таблицу II итерации.

Таблица 11

$i$	Бази с	$C$ $b$	$P_0$	2	-6	0	0	5	0
				$P$ 1	$P_2$	$P$ 3	$p_4$	$p$ 5	$P$ 6
1	$p_3$	0	12	-	5/	1	-	0	0
2	$P_5$	5	8	5/3	3	0	1/3	1	0
3	$p_6$	0	11	-	-	0	1/	0	1
			4	1/3	2/3	0	3	0	0
			40	-	-9	0	4	0	0
				1	8/		5/		
				-	3		3		
				11/3					

Как видно из табл. 11, новый опорный план задачи не является оптимальным, так как в 4-й строке столбца вектора  $P_1$  стоит отрицательное число  $-11/3$ . Поскольку в столбце этого вектора нет положительных элементов, данная задача не имеет оптимального плана.

2.2. МЕТОДЫ НАХОЖДЕНИЯ ОПОРНЫХ ПЛАНОВ.

Прикладная математика

Цель: показать значимость нахождения оптимальных путей транспортировки груза; разобрать методы нахождения опорного плана транспортной задачи и разобрать их преимущества и недостатки.

Понимать смысл поставленной задачи.

Научиться схематизировать. Важным частным случаем задачи линейного программирования является так называемая транспортная задача.

Это одна из самых распространённых проблем во всех областях экономики, транспортировка груза или товара с минимальными денежными и временными затратами. Так как огромное количество возможных вариантов перевозок затрудняет получение самого экономичного плана эмпирическим или экспертным путём, то появилась необходимость разработки специальной теории, позволяющей быстро решать подобные задачи с помощью алгоритмизации. Применение математических методов в планировании перевозок даёт большой экономический эффект. Алгоритм и методы решения транспортной задачи могут быть использованы при решении некоторых экономических задач, не имеющих ничего общего с транспортировкой груза. В этом случае величины тарифов  $C_{ij}$  имеют различный смысл в зависимости от конкретной экономической задачи.

К таким задачам относятся следующие:

1. оптимальное закрепление за станками операций по обработке деталей.

В них  $C_{ij}$  является таким экономическим показателем, как производительность. Задача позволяет определить, сколько времени и на какой операции нужно использовать каждый из станков, чтобы обработать максимальное количество деталей. Так как транспортная задача требует нахождения минимума, то значения берутся с отрицательным знаком;

2. оптимальные назначения, или проблема выбора. Имеется  $m$  механизмов, которые могут выполнять  $n$  различных работ с производительностью  $C_{ij}$ . Задача позволяет определить, какой механизм и на какую работу надо назначить, чтобы добиться максимальной производительности;

3. задача о сокращении производства с учетом суммарных расходов на изготовление и транспортировку продукции;

4. увеличение производительности автомобильного транспорта за счет минимизации порожнего пробега. Уменьшение порожнего пробега сократит количество автомобилей для перевозок, увеличив их производительность;

Прикладная математика

5. решение задач с помощью метода запрещения перевозок. Используются в том случае, если груз от некоторого поставщика по каким-то причинам не может быть направлен одному из потребителей. Данное ограничение можно учесть, присвоив соответствующей клетке достаточно большое значение стоимости, тем самым в эту клетку не будут производиться перевозки.

Общая постановка транспортной задачи состоит в определении оптимального плана перевозок некоторого однородного груза из  $m$  пунктов отправления,  $A_1, A_2, \dots, A_m$  в  $n$  пунктов назначения  $B_1, B_2, \dots, B_n$ . При этом в качестве критерия оптимальности обычно берётся либо минимальная стоимость перевозок всего груза, либо минимальное время его доставки. Рассмотрим транспортную задачу, в качестве критерия оптимальности которой взята минимальная стоимость перевозок всего груза.

Обозначим через  $C_{ij}$  тарифы перевозки единицы груза из  $i$ -ого пункта отправления в  $j$ -ый пункт назначения, через  $a_j$ - запасы груза в  $i$ -м пункте отправления, через  $b_j$ - потребности в грузе в  $j$ -м пункте назначения, а через  $x_{ij}$ -количество единиц груза, перевозимого из  $i$ -го пункта отправления в  $j$ -й пункт назначения.

Тогда математическая постановка задачи состоит в определении минимального значения функции.

$$F = \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij}$$

При условиях

$$\sum_{i=1}^m x_{ij} = b_j \quad (j = 1, n)$$

$$\sum_{j=1}^n x_{ij} = a_i \quad (i = 1, m),$$

$$x_{ij} > 0 \quad (i = 1, m; j = 1, n)$$

Всякое неотрицательное решение систем линейных уравнений, определяемое матрицей

$$X = (x_{ij}) \quad (i = 1, m, j = 1, n)$$

План,  $X = (x_{ij}) \quad (i = 1, m, j = 1, n)$  при котором функция принимает своё минимальное значение,

Прикладная математика

называется оптимальным планом транспортной задачи.

Таблица 2.1

Пункты отправления	Пункты назначения					Запасы
	$B_1$	...	$B_j$	...	$B_n$	
$A_1$	$c_{11}$ $x_{11}$	...	$c_{1j}$ $x_{1j}$	...	$c_{1n}$ $x_{1n}$	$a_1$
...	...	...	...	...	...	...
$A_i$	$c_{i1}$ $x_{i1}$	...	$c_{ij}$ $x_{ij}$	...	$c_{in}$ $x_{in}$	$a_i$
...	...	...	...	...	...	...
$A_m$	$c_{m1}$ $x_{m1}$	...	$c_{mj}$ $x_{mj}$	...	$c_{mn}$ $x_{mn}$	$a_m$
Потребности	$b_1$	...	$b_j$	...	$b_n$	

Очевидно, общее наличие груза у поставщиков равно  $\sum_{i=1}^m a_i$ , а общая потребность в грузе в пунктах назначения равна  $\sum_{j=1}^n b_j$  единиц. Если общая потребность в грузе в пунктах назначения равна запасу груза в пунктах отправления, т.е.

$$\sum_{i=1}^m a_i = \sum_{j=1}^n b_j$$

то модель такой транспортной задачи называется закрытой, если же указанное условие не выполняется, то модель называется открытой. Для разрешимости транспортной задачи необходимо и достаточно, чтобы запасы груза в пунктах отправления были равны потребностям в грузе в пунктах назначения.

В случае превышения запаса над потребностью, т.е.  $\sum_{i=1}^m a_i > \sum_{j=1}^n b_j$  вводится фиктивный (n+1)-й

Прикладная математика

пункт назначения с потребностью  $b_{n+1} = \sum_{i=1}^m a_i - \sum_{j=1}^n b_j$  и соответствующие тарифы считаются равными нулю:

$c_{in+1} = 0 (i = 1, m)$  Полученная задача является транспортной задачей, для которой выполняется

равенство закрытой транспортной задачи. Аналогично, при  $\sum_{i=1}^m a_i = \sum_{j=1}^n b_j$  вводится фиктивный

$(m+1)$ -й пункт отправления с запасом груза и тарифы полагаются равными нулю  $c_{m+1j} = 0 (j = 1, n)$

Этим задача сводится к обычной транспортной задаче, из оптимального плана которой получается оптимальный план исходной задачи.

Для определения опорного плана существует несколько методов. Три из них используются

наиболее часто, это метод северо-западного угла, метод аппроксимации Фогеля и метод минимального элемента.

Для определения оптимального плана транспортной задачи разработаны специальные методы.

Один из них – метод потенциалов.

Определение опорного плана транспортной задачи.

Определение опорного плана транспортной задачи.

Определение оптимального плана транспортной задачи начинают, прежде всего,

с нахождения какого-нибудь её опорного плана. Этот план может быть найден одним из методов – метод северо-западного угла, метод

Прикладная математика

аппроксимации Фогеля либо метод минимального элемента. Сущность этих методов заключается в том, что опорный план

находят последовательно за  $n+m-1$  шагов, на каждом из которых в таблице условий

заполняют одну клетку, которую называют занятой. Заполнение одной из клеток

обеспечивает полностью либо удовлетворения потребности в грузе одного из пунктов

назначения, либо вывоз груза из одного пункта отправления. После того как проделаны  $m+n-2$  описанных выше шагов, получают задачу с одним пунктом отправления и одним пунктом назначения. При этом останется свободной только одна клетка, а запасы оставшегося пункта отправления будут равны потребностям оставшегося пункта

назначения. Заполнив эту клетку, тем самым делают последний шаг ( $m+n-1$ ) и получают искомый опорный план Метод северо-западного угла.

При нахождении опорного плана транспортной задачи методом северо-западного угла на каждом шаге рассматривают первый из оставшихся пунктов отправления и первый из оставшихся пунктов назначения. Заполнение клеток таблицы условий начинается с левой верхней клетки для неизвестного  $x_{mn}$ , т.е. идет как бы по диагонали таблицы.

Метод аппроксимации Фогеля.

При определении оптимального плана транспортной задачи методом аппроксимации Фогеля на каждой итерации по всем столбцам и по всем строкам находят разность между двумя записанными в них минимальными тарифами. Эти разности записывают в специально отведенных для этого строке

и столбцов таблице условий задачи. Среди указанных разностей выбирают минимальную. В строке, которой данная разность соответствует, определяют минимальный тариф. Клетку, в которой он записан, заполняют на данной итерации. Если минимальный тариф одинаков для нескольких клеток данной строки, то для заполнения выбирают ту клетку, которая расположена в столбце, соответствующем наибольшей разности между двумя минимальными тарифами, находящимися в данном столбце(строке).

Метод минимального элемента.

Прикладная математика

В методе северо-западного угла на каждом шаге потребности первого из оставшихся пунктов назначения удовлетворялись за счет запасов первого из оставшихся пунктов отправления. Очевидно, выбор пунктов назначения и отправления целесообразно производить, ориентируясь

на тарифы перевозок, а именно: на каждом шаге следует выбирать какую-нибудь клетку, отвечающую минимальному тарифу, и рассмотреть пункты назначения и отправления, соответствующие выбранной клетке. Сущность этого метода и состоит в выборе клетки с минимальным тарифом. Следует отметить, что этот метод, как правило, позволяет найти опорный план транспортной задачи, при котором общая стоимость перевозок груза меньше, чем общая стоимость перевозок при плане, найденном для задачи с помощью северо-западного угла. Поэтому наиболее целесообразно находить опорный план методом минимального элемента.

Все рассмотренные методы разберем на примере.

Задача №1.

Дано:

Три предприятия данного экономического района могут производить некоторую однородную продукцию в количествах, соответственно равных 180, 350 и 20 ед. Эта продукция должна быть поставлена пяти потребителям в количествах, соответственно равных 110, 90, 120, 80 и 150 ед. Затраты, связанные с производством и доставкой единицы продукции, задаются матрицей

$$C = \begin{pmatrix} 7 & 12 & 4 & 6 & 5 \\ 1 & 8 & 6 & 5 & 3 \\ 6 & 13 & 8 & 7 & 4 \end{pmatrix}$$

Составить такой опорный план прикрепления потребителей к поставщикам, при котором общие затраты являются минимальными.

Решение:

Прикладная математика

	$B_1$	$B_2$	$B_3$	$B_4$	$B_5$	Запасы
$A_1$	7	12	4	6	5	180
$A_2$	1	8	6	5	3	350
$A_3$	6	13	8	7	4	20
Потребности	110	90	120	80	150	550

Метод северо-западного угла:

	$B_1$	$B_2$	$B_3$	$B_4$	$B_5$	Запасы
$A_1$	7 110	12 70	4 -	6 -	5 -	180
$A_2$	1 -	8 20	6 120	5 80	3 130	350
$A_3$	6 -	13 -	8 -	7 -	4 20	20
Потребности	110	90	120	80	150	550

$$F=7*110+12*70+8*20+6*120+5*80+3*130+4*20=3360$$

Метод минимального элемента:

	$B_1$	$B_2$	$B_3$	$B_4$	$B_5$	Запасы
$A_1$	7 -	12 60	4 120	6 -	5 -	180
$A_2$	1 110	8 10	6 -	5 80	3 150	350
$A_3$	6 -	13 20	8 -	7 -	4 -	20
Потребности	110	90	120	80	150	550

$$F=12*60+4*120+1*110+8*10+5*80+3*150+13*20=2420$$

Метод аппроксимации Фогеля:

Прикладная математика

	$B_1$	$B_2$	$B_3$	$B_4$	$B_5$	Запасы	Разности по строкам					
$A_1$	7 -	12 -	4 120	6 60	5 -	180	1	1	1	1	2	6
$A_2$	1 110	8 90	6 -	5 20	3 130	160	2	2	2	2	1	5
$A_3$	6 -	13 -	8 -	7 -	4 20	140	2	3	3	-	-	-
Потребности	150	250	120	180	150	700						
Ра зн ос ти по ст ол бц ам	5	4	2	1	1							
	-	4	2	1	1							
	-	-	2	1	1							
	-	-	2	1	2							
	-	-	2	1	-							
	-	-	-	1	-							

$$F=4*120+6*60+1*110+8*90+5*20+3*130+4*20=2240.$$

### 3.2 ОПТИМИЗАЦИИ ПРИ НАЛИЧИИ ОГРАНИЧЕНИЙ

Как правило, общий подход к решению задач оптимизации при наличии ограничений состоит в замене исходной задачи с ограничениями на другую более легко реализуемую задачу и которая в последующем используется как основа для некоторых итерационных процессов. В качестве основной особенности первоначально предложенных методов можно отметить следующее: исходная задача с ограничениями заменяется на задачу без ограничений, но с применением метода штрафных функций вблизи или около налагаемых значений для ограничений. При таком подходе задача оптимизации при наличии ограничений решалась через введение некой последовательности параметризованных задач оптимизации без наложения ограничений, которые в пределе (выбранной последовательности) сходились к искомой задаче с ограничениями. В настоящее время такой подход считается относительно малоэффективным и, соответственно, был заменен на методы решения, основанными на формулировке и последующем решении так называемых уравнений Куна-Такера (КТ). В уравнениях КТ вводятся дополнительные предположения о характере ограничений и понятии оптимальности для задачи оптимизации при наличии ограничений. Если поставленная задача является так называемой задачей выпуклого программирования, то есть  $f(x)$  и  $G_i(x), i = 1, \dots, m$  являются выпуклыми

функциями, то уравнения КТ являются необходимыми и достаточными условиями для общей постановки задачи.

Применительно к уравнению (3-1) (GP) уравнения Куна-Таккера записываются в виде

$$\begin{aligned} \nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \cdot \nabla G_i(x^*) &= 0 \\ \lambda_i^* \cdot G_i(x^*) &= 0 \quad i = 1, \dots, m \\ \lambda_i^* &\geq 0 \quad i = m_e + 1, \dots, m \end{aligned} \quad (3-26)$$

Первые уравнения представляют собой описание процесса исчезновения градиента между целевой функцией и активными ограничениями в точке решения. Поскольку градиенты подлежат выходу на нулевые значения, то множители Лагранжа ( $\lambda_i, i = 1, \dots, m$ ) будут необходимы для того, что бы уравновесить отклонения по величине данной целевой функции и градиентов ограничений. Поскольку только активные ограничения вовлечены в данную процедуру оптимизации, то ограничения не активны и не должны подвергаться данной процедуре и поэтому соответствующие множители Лагранжа будут равны нулю. Это обстоятельство неявно выражено в двух последних уравнениях 3-26.

Подобное решение уравнений Куна-Таккера служит основой для большинства алгоритмов нелинейного программирования. В этих алгоритмах часто используется прямое вычисление множителей Лагранжа. Квазиньютоновские методы обеспечивают сверхлинейную сходимость путем накопления информации второго порядка относительно уравнений Куна-Таккера, использующих процедуры квазиньютоновской корректировки. В общем случае эти методы можно отнести к задачам Последовательного квадратичного программирования (SQP), поскольку проблема QP решается на каждой главной итерации (иногда их еще называют методами Итерационного квадратичного программирования, Рекурсивного квадратичного программирования или Переменной метрики при наличии ограничений).

### Последовательное квадратичное программирование SQP

SQP метод является одним из самых современных методов в области нелинейного программирования. Шитковский [38], к примеру, успешно реализовал и провел тестовые расчеты по данной версии оптимизации и получил всестороннее превосходство, по сравнению с другими тестовыми методами, в части эффективности, точности и процента успешного решения задачи для большого числа тестовых задач.

Основанный на работах Бигса [1], Хана [24] и Пауэлла ([34], [35]) данный метод позволяет достаточно точно имитировать метод Ньютона для оптимизации при наличии ограничений, как это сделано для оптимизации без наличия ограничений. На каждой основной итерации осуществляется аппроксимация Гессиана для функций Лагранжа при помощи квазиньютоновского модифицированного метода. Такой подход далее будет востребован для постановки подзадачи QR, решение которой далее уже используется для формирования направления поиска в процедуре линейного поиска. Обзор методов SQP можно найти в работах Флетчера [15], Гиль и др. [21], Пауэлла [37] и Шитковского [25]. Тем не менее, далее приводится описание обобщенного метода.

Согласно описанию задачу метода GP (уравнение 3-1) основная идея постановки подзадачи QR заключается в квадратичной аппроксимации функции Лагранжа.

$$L(x, \lambda) = f(x) + \sum_{i=1}^m \lambda_i \cdot g_i(x) \quad (3-27)$$

Последнее представляет собой упрощение уравнения 3-1 при предположении, что связанные ограничения могут быть представлены через ограничения в виде неравенств. Посредством линеаризации нелинейных ограничений можно получить подзадачу QR.

### Подзадача квадратичного программирования (QR)

$$\begin{aligned} \underset{d \in \mathcal{R}^n}{\text{minimize}} \quad & \frac{1}{2} d^T H_k d + \nabla f(x_k)^T d \\ \nabla g_i(x_k)^T d + g_i(x_k) = 0 \quad & i = 1, \dots, m_e \\ \nabla g_i(x_k)^T d + g_i(x_k) \leq 0 \quad & i = m_e + 1, \dots, m \end{aligned} \quad (3-27)$$

Данная подзадача может быть решена посредством применения алгоритма QR (см., например, раздел Решение квадратичного программирования). Такое решение основано на формировании новой итерации следующего вида

$$x_{k+1} = x_k + \alpha_k d_k$$

Параметр при длине шага  $\alpha_k$  определяется из соответствующей процедуры линейного поиска, которая обеспечивает приемлемое уменьшение получаемой функции выгоды (см. раздел Корректировка матрицы Гессе). Матрица  $H_k$  является положительно определенной аппроксимацией матрицей

Гессе для Лагранжевой функции. (уравнение 3-27).  $H_k$  может быть корректироваться посредством любого из квазиньютоновских методов, хотя метод BFGS (смотри раздел Корректировка матрицы Гессе), скорее всего, является наиболее популярным.

В отличие от решения методом SQP для задач без ограничений, нелинейные задачи при наличии ограничений решаются за некоторое число итераций. Одной из причин такого факта является то, что, вследствие наличия пределов на обозримые области, оптимизатор может принимать осознанные решения относительно направлений поиска и размера шага.

Рассмотрим функцию Розенброка (уравнение 3-2) при наличии дополнительных нелинейных ограничений в виде неравенств,  $g(x)$ ,

$$x_1^2 + x_2^2 - 1.5 \leq 0 \quad (3-29)$$

При применении метода SQP эта задача решается за 96 итераций по сравнению с 140 итерациями для задач без ограничений. На рисунке 3-6 представлен путь к точке решения  $x = [0.9072, 0.8228]$  начиная со стартовой точки  $x = [-1.9, 2]$ .

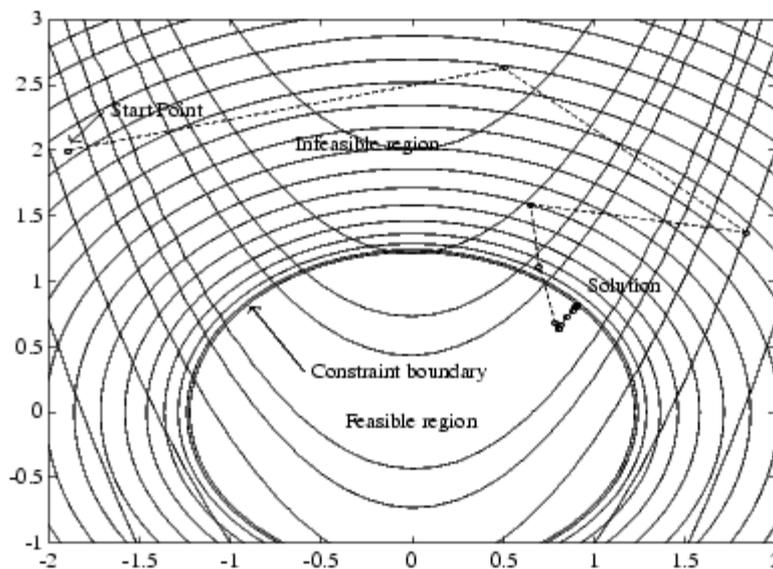


Рис. 3-6. Применение метода SQP для функции Розенброка (уравнение 3-2) с линеаризованными нелинейными ограничениями

### Реализация метода SQP

Реализация метода SQP состоит из трех основных стадий, которые далее кратко обсуждаются в следующих подразделах

Прикладная математика

- Корректировка матрицы Гессе для Лагранжевой функции.
- Решение задачи квадратичного программирования.
- Вычисление линейного поиска и функции выгоды

Корректировка матрицы Гессе

На каждой главной итерации положительно определенная квазиньютоновская аппроксимация для функции Лагранжа,  $H$ , рассчитывается с помощью метода BFGS, где  $\lambda_i (i = 1, \dots, m)$  представляют собой оценку множителей Лагранжа.

$$H_{k+1} = H_k + \frac{q_k q_k^T}{q_k^T s_k} - \frac{H_k^T H_k}{s_k^T H_k s_k}, \text{ где} \tag{3-30}$$

$$s_k = x_{k+1} - x_k$$

$$q_k = \nabla f(x_{k+1}) + \sum_{i=1}^n \lambda_i \cdot \nabla g_i(x_{k+1}) - \left( \nabla f(x_k) + \sum_{i=1}^n \lambda_i \cdot \nabla g_i(x_k) \right)$$

Пауэл вообще [35] рекомендует поддерживать значения матрицы Гессе положительно определенными, даже не смотря на то, что в точке эти решения могут и не иметь положительные решения. Положительные значения матрицы Гессе поддерживаются в том случае, если величина  $q_k^T s_k$  будет больше нуля для каждой корректировки и, что  $H$  инициализируется как положительно определенная матрица. Когда величина  $q_k^T s_k$  не является положительной, то параметр  $q_k$  модифицируется поэлементно, шаг за шагом, так что бы выполнялось условие  $q_k^T s_k > 0$ . Обобщенная цель такой модификации заключается в том, что бы несколько изменить элементы  $q_k$ , которые составляют основной вклад в положительно определенную корректировку, как можно меньше. Таким образом, на начальной стадии принятой модификации, отрицательный наибольший элемент их набора  $q_k \cdot s_k$  последовательно уменьшается на половину. Такая процедура продолжается до тех пор, пока  $q_k^T s_k$  больше или равна  $1e-5$ . Если после такой процедуры  $q_k^T s_k$  все равно остается положительной, то производится модификация у  $q_k$  путем добавления некого вектора  $v$ , умноженного на некую скалярную постоянную  $w$ , а именно,

$$q_k = q_k + wv \tag{3-31}$$

где

Прикладная математика

$$v_i = \nabla g_i(x_{k+1}) \cdot g_i(x_{k+1}) - \nabla g_i(x_k) \cdot g_i(x_k),$$

если  $(q_k)_i \cdot w < 0$  и

$$(q_k)_i \cdot (s_k)_i < 0 (i = 1, \dots, m)$$

$$v_i = 0$$

или же в противном случае методично увеличивают  $w$  до тех пор, пока член  $q_k^T s_k$  станет положительным.

Программы `fmincom`, `fvinmax`, `fgoalattain` и `fseminf` реализованы на основе метода SQP. Если опционный параметр `Display` установлен как отображение информации на каждой итерации 'iter', то выводится различная информация, включая значение функции и максимальное нарушение поставленных ограничений. Также отображается модифицированный Гессиан, в случае, если он претерпевает некую коррекцию на первоначальной стадии выполняемой процедуры с целью поддержания его положительного значения. Если Гессиан подвергается вторичной корректировке на второй стадии приведенного выше подхода, то отображается его вторичное значение. Если подзадача QP не является выполнимой, то этот факт так же отображается. Такая отображаемая информация, как правило, не представляет особого интереса, но указывает на то, что данная задача является существенно нелинейной, а сходимость может быть более длительной, чем обычно. Иногда отображаемое сообщение не изменяется, что указывает, что величина  $q_k^T s_k$  будет порядка нуля. Это может указывать, что постановка задачи является неправильной или проводится оптимизация прерывистой функции.

### Решение задачи квадратичного программирования.

На каждой основной итерации метода SQP решается задача QP в следующей форме, где  $A_i$  относится к  $i$ -той строке матрицы  $A$  размерностью  $m \times n$ :

$$\begin{aligned} \underset{d \in \mathbb{R}^n}{\text{minimize}} \quad & q(d) = \frac{1}{2} d^T H d + c^T d \\ & A_i d = b_i \quad i = 1, \dots, m_e \\ & A_i d \leq b_i \quad i = m_e + 1, \dots, m \end{aligned} \tag{3-32}$$

Используемый в тулбоксе Оптимизация метод основан на стратегии активных наборов (более известен как метод проекций) и аналогичен методу, описанному в работе Гиля и др., [20] [19]. Этот метод был модифицирован

применительно для задач Линейного программирования (LP) и Квадратичного программирования (QP).

Процедура решения включает в себя две фазы. Первая фаза представляет собой расчет наиболее вероятной точки. Вторая фаза включает в себя генерацию некой итеративной последовательности наиболее вероятных точек, которая уже сходится к требуемому решению. В таком методе утверждается, что активный набор,  $\bar{A}_k$ , является некой оценкой активных ограничений (т.е. то, что представляет собой ограничительные пределы) в данной точке решения. Фактически все алгоритмы QP являются методами активных наборов. Это следует особо подчеркнуть, поскольку существует множество различных методов, которые чрезвычайно схожи по своей сути, но представлены в виде различных способов своего представления.

$\bar{A}_k$  корректируется на каждой итерации и  $k$  используется для построения основы для направления поиска  $\hat{d}_k$ . Ограничения в виде равенств всегда остаются в данном активном наборе  $\bar{A}_k$ . Обозначение для переменной в виде  $\hat{d}_k$  в данном случае используется для того, что бы различать эти переменные от  $d_k$ , используемых на главных итерациях метода SQP. Направление поиска  $\hat{d}_k$  рассчитывается и далее используется для минимизации целевой функции, несмотря на то, что остаются в силе все возможные границы активных ограничений. Все возможное пространство для переменных  $\hat{d}_k$  образуется из базиса  $Z_k$ , чьи колонки являются ортогональными относительно полученных результатов расчета активного набора  $\bar{A}_k$  (т.е.,  $\bar{A}_k Z_k = 0$ ). Таким образом, направление поиска, которое формируется путем суммирования всех возможных сочетаний колонок  $Z_k$ , обеспечивает сохранение границы активных ограничений.

Принятая матрица  $Z_k$  формируется из последних  $m-l$  колонок разложения QR матрицы  $\bar{A}_k^T$ , где  $l$  есть число активных ограничений, а так же справедливо  $l < m$ . Т.е.  $Z_k$  определяется как

$$Z_k = Q[:, l+1:m] \quad (3-33)$$

где

$$Q^T \bar{A}_k^T = \begin{bmatrix} R \\ 0 \end{bmatrix}$$

Прикладная математика

Как только  $Z_k$  будет найдено, то далее ищется новое направление поиска  $\hat{d}_k$ , которое в свою очередь приводит к минимуму  $q(d)$ , где  $\hat{d}_k$  есть нулевое пространство активных ограничений. Т.е.  $\hat{d}_k$  есть линейная комбинация столбцов  $Z_k$ :  $\hat{d}_k = Z_k p$  для некоторого вектора  $p$ . На следующем этапе, если при помощи соответствующей постановки для  $\hat{d}_k$  ввести квадратное уравнение как некую функции от  $p$ , то получим:

$$q(p) = \frac{1}{2} p^T Z_k^T H Z_k p + c^T Z_k p \quad (3-34)$$

После дифференцирования по  $p$  получим

$$\nabla q(p) = Z_k^T H Z_k p + Z_k^T c \quad (3-35)$$

Член  $\nabla q(p)$  имеет отношение к проекции градиента квадратичной функции, поскольку он является проекцией градиента в подпространстве функций  $Z_k$ . Член  $Z_k^T H Z_k$  есть так называемая проекция Гессеана. Полагая, что матрица Гессе  $H$  является положительно определенной (что справедливо применительно к реализации метода SQP), то минимум функции  $q(p)$  в подпространстве функций  $Z_k$  будет определяться условием  $\nabla q(p) = 0$ , и, следовательно, для поиска минимума функции необходимо решение системы линейных уравнений.

$$Z_k^T H Z_k p = -Z_k^T c \quad (3-36)$$

А шаг в направлении минимума будет в следующем виде

$$x_{k+1} = x_k + \alpha \hat{d}_k \quad \text{where } \hat{d}_k = Z_k^T p \quad (3-37)$$

Вследствие квадратичной природы целевой функции на каждой итерации следует, что существует только одно возможное значения размера шага  $\alpha$ . Единичный шаг в направлении  $\hat{d}_k$  как раз и есть шаг в направлении минимума функции, ограниченной пределами нулевого пространства  $\bar{A}_k$ . Если будет возможно принять такой шаг, без нарушения принятых ограничений, то это как раз и будет решение QP (см. уравнение 3-33). В противном случае, шаг вдоль направления  $\hat{d}_k$  в сторону ближайшего ограничения будет меньше единицы и в активный набор на следующей итерации будут включены новые ограничения. Расстояние до границ ограничений в любом направлении  $\hat{d}_k$  можно представить как

$$\alpha = \min_i \left\{ \frac{-(A_i x_k - b_i)}{A_i \hat{d}_k} \right\} \quad (i = 1, \dots, m) \quad (3-38)$$

Которое определено для ограничений не из активного набора и где направление  $\hat{d}_k$  показывает направление к границам ограничений, т.е.  $A_i \hat{d}_k > 0, i = 1, \dots, m$ .

В случае включения  $n$  независимых ограничений в активный набор, без локализации минимум, то для соответствия невырожденной системе линейных уравнений

$$\bar{A}_k^T \lambda_k = c \quad (3-39)$$

следует рассчитывать множители Лагранжа.

Если все элементы  $\lambda_k$  больше нуля, то  $x_k$  является решением для точки оптимума задачи (уравнение 3-33). Однако, если какой-нибудь из элементов  $\lambda_k$  будут меньше нуля, то эти компоненты не будут соответствовать ограничениям в виде равенств и, следовательно, соответствующий элемент необходимо определять из активного набора, так же требуется обращение к новой итерации.

**Инициализация.** В данном алгоритме для успешного начала работы требуется некая допустимая стартовая точка. Если текущая точка из метода SQP не является допустимой, то тогда некую точку можно найти из решения задачи линейного программирования.

$$\begin{aligned} & \underset{\gamma \in \mathbb{R}, x \in \mathbb{R}^n}{\text{minimize}} \quad \gamma \\ & A_i x = b_i \quad i = 1, \dots, m_e \\ & A_i x - \gamma \leq b_i \quad i = m_e + 1, \dots, m \end{aligned} \quad (3-40)$$

Обозначение  $A_i$  указывает на  $i$ -ую точку матрицы. Допустимую точку для уравнения 3-40 можно найти (если таковая существует) путем установки  $x$  в некое удовлетворяющее ограничениям типа равенство значение. Такое значение можно определить из решения недо- или переопределенной системы линейных уравнений, полученной из системы ограничений типа равенств. Если решение такой задачи существует, то фиктивная переменная  $\gamma$  устанавливается в данном случае как некое максимальное расхождение.

Упомянутый выше алгоритм QR можно модифицировать применительно к задачам LP путем установки на каждой итерации направления поиска в

направление наискорейшего спуска, где  $g_k$  является градиентом взятой целевой функции (равенство коэффициентов линейной целевой функции).

$$\hat{d}_k = -Z_k Z_k^T g_k \quad (3-41)$$

Если после использования ранее упомянутого метода LP найдена некая допустимая точка, то, далее, запускается основная фаза метода QR. Направление поиска  $\hat{d}_k$  инициализируется совместно с направлением поиска  $\hat{d}_1$ , полученным из решения системы линейных уравнений

$$H\hat{d}_1 = -g_k \quad (3-42)$$

Где  $g_k$  есть градиент целевой функции для текущей итерации  $x_k$  (т.е.  $Hx_k + c$ ).

Если для задачи QR невозможно определить допустимую точку, то направление поиска для основной подпрограммы SQP  $\hat{d}_k$  принимается из условия минимизации величины  $\gamma$ .

### Линейный поиск и функция выгоды

Решение подзадачи QR приводит к формированию вектора  $d_k$ , который, в свою очередь, используется при формировании новой итерации типа

$$x_{k+1} = x_k + \alpha d_k \quad (3-43)$$

Для того, что бы получить достаточное уменьшение функции выгоды оценивается параметр длины шага  $\alpha_k$ . Реализованная в данном алгоритме функция выгоды была ранее использована Наном [24] и Поуэлом [35] и имеет следующий вид.

$$\Psi(x) = f(x) + \sum_{i=1}^{m_r} r_i \cdot g_i(x) + \sum_{i=m_r+1}^m r_i \cdot \max\{0, g_i(x)\} \quad (3-44)$$

Поуэл рекомендует введение следующего штрафного параметра:

$$r_i = (r_{k+1})_i = \max_i \left\{ \lambda_i, \frac{1}{2}((r_k)_i + \lambda_i) \right\}, \quad i = 1, \dots, m \quad (3-45)$$

Данный подход дает положительный вклад от принятых ограничений, но которые являются неактивными для решения задачи QR, хотя ранее это были

Прикладная математика

активные значения. В принятом методе реализации параметр штрафа в качестве исходного параметра имеет вид:

$$r_i = \frac{\|\nabla f(x)\|}{\|\nabla g_i(x)\|} \quad (3-46)$$

Где  $\|\cdot\|$  представляет собой эвклидову норму.

Такой подход обеспечивает существенный вклад от ограничения с небольшими параметрами в значения штрафных параметров, что особенно актуально для активных ограничений вблизи точки решения.

**Алгоритм симплексного метода**

Предложенный еще Георгом Дантзигом в 1947 г. алгоритм симплексного метода является одним из самых ранних и наиболее известных методов оптимизации. Данный алгоритм решает следующую задачу линейного программирования

$$\begin{aligned} \min_x f^T x \quad \text{при условии, что} \quad & A \cdot x \leq b \\ & Aeq \cdot x = beq \\ & lb \leq x \leq ub \end{aligned}$$

Данный алгоритм представляет собой продвижение вдоль краев многогранника, определенного принятыми ограничениями, от одной вершины к другой при снижении величины целевой функции,  $f^T x$ , на каждом шаге. Далее приводится описание улучшенной версии исходного метода симплексного алгоритма, который так же приводит к оптимальным решениям, находящимся на вершинах многоугольника.

**Основной алгоритм.**

Алгоритм симплексного метода включает в себя две фазы:

Фаза 1. - Расчет начальной основной допустимой точки.

Фаза 2. - Расчет оптимального решения согласно поставленной задаче.

**Примечание.** Применительно к программе linprog нельзя принимать исходную точку  $x_0$  для расчета по алгоритму симплексного метода. Если все же  $x_0$  будет передано в качестве входного аргумента, то программа linprog игнорирует эту точку  $x_0$  и рассчитывает свою собственную начальную точку.

**Фаза 1.** При выполнении фазы 1 в данном алгоритме определяется начальное основное допустимое решение (см. определения из раздела

**Основные и неосновные переменные)** посредством решения кусочно-правильной вспомогательной задачи линейного программирования. Целевая функция этой вспомогательной задачи является *линейной штрафной функцией*

$$P = \sum_j P_j(x_j),$$

где члены  $P_j(x_j)$  определены как

$$P_j(x_j) = \begin{cases} x_j - u_j & \text{if } x_j > u_j \\ 0 & \text{if } l_j \leq x_j \leq u_j \\ l_j - x_j & \text{if } l_j > x_j \end{cases}$$

$P(x)$  служит мерой того, как много точка  $x$  переступает нижний и верхний граничные условия. А вспомогательная задача будет

$$\min_x \sum_j \quad \text{при условии, что} \quad \begin{array}{l} A \cdot x \leq b \\ A_{eq} \cdot x = b_{eq} \end{array}$$

Исходная задача имеет допустимую базисную точку тогда и только тогда, когда вспомогательная задача имеет равное нулю минимальное значение.

В данном алгоритме начальная точка для вспомогательной задачи определяется с помощью эвристического метода, который по мере необходимости добавляет неактивные и искусственные переменные. Далее согласно выбранному алгоритму эта начальная точка используется совместно с симплексным алгоритмом для решения вспомогательной задачи. Такое оптимальное решение, в свою очередь, является начальной точкой для фазы 2 основного алгоритма.

**Фаза 2.** Согласно фазе 2, для решения исходной поставленной задачи в данном алгоритме используется симплексный алгоритм, начиная уже с начальной стартовой точки фазы 1. На каждой итерации в данном алгоритме проводится проверка условия оптимальности и задача останавливается в случае, если текущие решения оказываются оптимальным. Если текущее решение не оптимально, то согласно данному алгоритму проводятся следующие действия:

- Выбирается одна переменная, называемая как *вводимая переменная*, из небазисных переменных и добавляется соответствующая колонка из небазисного вида в рассматриваемый базисный вид (в соответствии с принятыми определениями см. раздел **Базисные и небазисные переменные**).

Прикладная математика

- Выбирается некая переменная, (так называемая *исключаемая переменная*) из набора базисных переменных и из базисного набора удаляется соответствующая колонка.
- Проводится корректировка текущего решения и значения текущей цели.

Выбор вводимых и исключаемых переменных производится путем решения двух систем линейных уравнений, при сохранении свойства допустимости решения.

**Предварительная подготовка.**

В алгоритме симплексного метода используются те же самые подготовительные операции, что и в решателе задачи линейного программирования для алгоритмов большой размерности, как это отмечено в одноименном разделе. Дополнительно, в данном алгоритме используются еще две операции:

Исключаются колонки, которые имеют только один ненулевой элемент, а так же исключаются их соответствующие строчки.

Для каждого уравнения из числа ограничений  $a \cdot x = b$ , где  $a$  есть строчные элементы для  $Aeq$ , в данном алгоритме для линейной комбинации  $a \cdot x$  в виде  $rlb$  и  $rub$ , рассчитываются нижняя и верхняя границы. Такой подход является справедливым в случае конечных значений нижней и верхней границ. Если или  $rlb$  или  $rub$  равно  $b$ , то эта константа называется *принудительным ограничением*. В данном алгоритме каждый раз устанавливается некая переменная соответствующая ненулевой коэффициенту из  $a \cdot x$ , равному его верхней или нижней границе в зависимости от принудительного ограничения. Далее согласно алгоритму удаляются колонки, соответствующие этим переменным, а так же удаляются соответствующие принудительным ограничениям строчки.

**Использование симплексного алгоритма.**

Для того, что бы установить использование алгоритма симплексного метода, следует установить опционные параметры следующего типа

'LargeScale' в 'off' и 'Simplex' в 'on'.

```
options = optimset('LargeScale', 'off', 'Simplex', 'on')
```

Далее вызывается функция `linprog` с соответствующими входными опциями и аргументами. Детали использования функции `linprog` приведены соответствующем разделе.

Прикладная математика

В ситуации если программа определит недопустимость или наличие неограниченности при выполнении процедуры предварительной подготовки данных, то Linprog возвращает пустые выходные аргументы на месте  $x$  и  $fval$ . Linprog возвращает значение текущей точки в ситуациях, когда

- Есть превышение максимального числа итераций.

<P

- Будет определено, что поставленная задача является невыполнимой или не имеющей границ относительно фаз 1 и 2.
- Когда задача не имеет предельных границ, то linprog возвращает значения  $x$  и  $fval$  относительно направления отсутствия предельных границ.

**Основные и неосновные переменные.**

В данном разделе дается определение таких терминов, как базис, небазис и базисное допустимое решение для задачи линейного программирования. Для данных определений примем, что поставленная задача имеет следующий стандартный вид:

$$\min_x f^T x \quad \text{при условии} \quad \begin{aligned} A \cdot x &= b \\ lb &\leq x \leq ub \end{aligned}$$

(Отметим, что  $A$  и  $b$  не являются некими матрицей или вектором, определяющими неравенства в исходной задаче). Предположим, что  $A$  есть матрица размерностью  $m \times n$  с рангом  $m < n$  и соответствующими колонками  $\{a_1, a_2, \dots, a_n\}$ . Примем, что  $\{a_{i_1}, a_{i_2}, \dots, a_{i_m}\}$  является пространством колонок  $A$ , с набором индексов  $B = \{i_1, i_2, \dots, i_m\}$  и  $N = \{1, 2, \dots, n\}$ .  $N$  является дополнением  $B$ . Субматрица  $A_B$  есть так называемый *базис*, а дополнительная субматрица  $A_N$  есть так называемый *небазис*.  $x_B$  есть вектор *базисных переменных* и  $x_N$  есть вектор *небазисных переменных*. При выполнении каждой итерации фазы 2 данного алгоритма производится замена одной колонки текущего базиса на колонку небазиса и, соответственно, корректируются переменные  $x_B$  и  $x_N$ .

Если  $x$  есть решение системы  $A \cdot x = b$  и все небазисные переменные для  $x_N$  равны или их нижней или верхней границе, то  $x$  называется *базисным решением*.

## 4.1 РАСПЛЫВЧАТЫЕ ЦЕЛИ.

### ФУНКЦИИ ПРИНАДЛЕЖНОСТИ И ПРИНЯТИЕ РЕШЕНИЙ

В этой главе излагается метод построения функций принадлежности нечетких множеств и его применение для сравнения вариантов решений по многим критериям. В основу многокритериального анализа вариантов положена процедура слияния нечётких целей и ограничений, известная как принцип Беллмана-Заде [5]. Глава базируется на работах [44,46,75].

#### 7.1. Принцип слияния целей и ограничений

В общепринятом подходе главными элементами процесса принятия решения являются: а) множество альтернатив; б) множество ограничений, которые необходимо учитывать при выборе между различными альтернативами и в) функция предпочтительности, ставящая каждой альтернативе в соответствие выигрыш (или проигрыш), который будет получен в результате выбора этой альтернативы.

При рассмотрении этого процесса с более общих позиций принятия решений в нечётких условиях естественной представляется другая логическая схема, важнейшей чертой которой является симметрия по отношению к целям и ограничениям. Эта симметрия устраняет различия между целями и ограничениями и позволяет довольно просто сформировать на их основе решение.

##### 7.1.1. Нечёткие цели и ограничения

Пусть  $X = \{x\}$  - заданное множество альтернатив. Тогда нечёткая цель, или просто цель,  $G$  будет отождествляться с фиксированным нечётким множеством  $G$  в  $X$ . Например, если  $X = R^1$  (действительная прямая), а нечёткая цель формулируется как " $x$  должно быть значительно больше 10", то ее можно представить как нечёткое множество в  $R^1$  с функцией принадлежности, имеющей, скажем, такой вид:

$$\mu_G(x) = \begin{cases} 0, & x < 10, \\ \left(1 + (x - 10)^2\right)^{-1}, & x \geq 10. \end{cases} \quad (7.1)$$

Прикладная математика

Аналогично цели " $x$  должно быть в окрестности 15" может быть поставлено в соответствие нечёткое множество с функцией принадлежности:

$$\mu_G(x) = \left(1 + (x - 15)^4\right)^{-1}, \quad (7.2)$$

Подобным же образом нечёткое ограничение, или просто ограничение,  $C$  в пространстве  $X$  определяется как некоторое расплывчатое множество в  $X$ . Например, в случае  $X = \mathbb{R}^1$  ограничение " $x$  должно находиться в диапазоне 2-10" может быть представлено нечётким множеством с функцией принадлежности, скажем, вида:

$$\mu_C(x) = \left(1 + a(x - b)^m\right)^{-1},$$

где  $a$  - положительное число и  $m$  - четное положительное число, выбираемое так, чтобы передать смысл, в котором следует понимать "приближение" к заданному интервалу.

Важным аспектом приведенных выше определений является то, что и цель и ограничения рассматриваются как нечёткие множества в пространстве альтернатив; это дает возможность не делать между ними различия при формировании решения.

Действительно, предположим, например, что нечёткая цель  $G$  и нечёткое ограничение  $C$  заданы следующим образом:

$G$ :  $x$  должно быть значительно больше 10 и

$C$ :  $x$  должно быть в окрестности 15.

$[\mu_G(x)$  и  $\mu_C(x)$  задаются соответственно формулами (7.1) и (7.2)].

Заметим, что цель  $G$  и ограничения  $C$  соединены между собою союзом "И", причём "И" соответствует пересечению нечётких множеств. Это означает, что в рассматриваемом примере совокупное влияние нечёткой цели  $G$  и нечёткого ограничения  $C$  на выбор альтернатив может быть представлено пересечением  $G \cap C$ . Функция принадлежности для пересечения задается соотношением

$$\mu_{G \cap C}(x) = \mu_G(x) \wedge \mu_C(x)$$

или, в развернутой форме

$$\mu_{G \cap C}(x) = \begin{cases} \text{Min} \left( \left(1 + (x - 10)^{-2}\right)^{-1}, \left(1 + (x - 15)^4\right)^{-1} \right) & \text{для } x \geq 10, \\ 0, & \text{для } x < 10. \end{cases}$$

Отметим, что в силу выпуклости расплывчатых множеств  $G$  и  $C$  множество  $G \cap C$  также является выпуклым.

### 7.1.2. Нечёткие решения

Обратимся теперь к понятию *решения*. Интуитивно ясно, что *решение* - это по существу выбор одной или нескольких из имеющихся альтернатив. Предыдущий пример наводит на мысль, что расплывчатое решение, или просто решение, следует определить как нечёткое множество в пространстве альтернатив, получающееся в результате пересечения заданных целей и ограничений. Следующее определение уточняет эту мысль.

**Определение 7.1.** Пусть в пространстве альтернатив  $X$  заданы нечёткая цель  $G$  и нечёткое ограничение  $C$ . Тогда нечёткое множество  $D$ , образуемое пересечением  $G$  и  $C$ , называется решением. В символической форме

$$D = G \cap C \tag{7.3}$$

и соответственно  $\mu_D = \mu_G \cap \mu_C$ . Взаимосвязь между  $G$  и  $C$  показана на рис.7.1

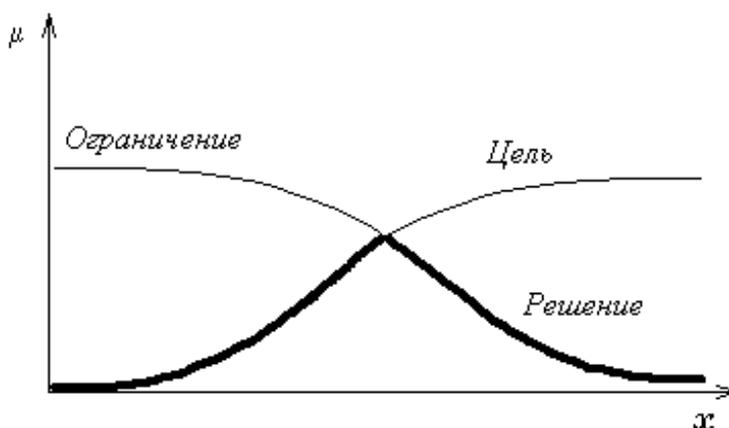


Рис.7.1. Функции ограничения, цели и решения.

Прикладная математика

В более общем случае, если имеется  $n$  целей и  $m$  ограничений, то результирующее решение определяется пересечением всех заданных целей и ограничений, т.е.

$$D = G_1 \cap G_2 \cap \dots \cap G_n \cap C_1 \cap C_2 \cap \dots \cap C_m \quad (7.4)$$

и соответственно

$$\mu_D = \mu_{G_1} \cap \mu_{G_2} \cap \dots \cap \mu_{G_n} \cap \mu_{C_1} \cap \mu_{C_2} \cap \dots \cap \mu_{C_m} \quad (7.5)$$

Заметим, что в приведенном определении нечёткого решения цели и ограничения входят в выражение для  $D$  совершенно одинаковым образом, что и соответствует тождественности целей и ограничений в логической схеме процесса принятия решений в нечётких условиях.

Коротко обобщенное определение решения можно сформулировать следующим образом:

*Решение = Слияние целей и ограничений.*

В качестве иллюстрации к соотношению (7.5) рассмотрим простой пример, в котором  $X = \{1, 2, \dots, 10\}$ , а  $G_1$ ,  $G_2$ ,  $C_1$  и  $C_2$  определяются табл.

7.1. Образуя конъюнкцию  $\mu_{G_1}$ ,  $\mu_{G_2}$ ,  $\mu_{C_1}$  и  $\mu_{C_2}$ , получим таблицу значений для  $\mu_D(x)$  (табл.7.2). Решение в этом случае есть расплывчатое множество

$$D = \{(2; 0,1), (3; 0,4), (4; 0,7), (5; 0,8), (6; 0,6), (7; 0,4), (8; 0,2)\}.$$

Заметим, что ни одно  $x$  из  $X$  не принадлежит решению  $D$  полностью (т.е. со степенью принадлежности, равной 1). Это, конечно, является следствием того, что заданные цели и ограничения вступают в конфликт друг с другом, исключая тем самым возможность существования альтернативы, которая бы полностью им всем удовлетворяла.

Понятие решения как расплывчатого множества в пространстве альтернатив может по началу показаться несколько искусственным. На самом деле оно совершенно естественно, поскольку расплывчатое решение может рассматриваться как некоторая "инструкция", расплывчатость которой является следствием неточности формулировки поставленных целей и ограничений. Так, в приведенном примере  $G_1$ ,  $G_2$ ,  $C_1$  и  $C_2$  могли бы быть выражены следующими фразами: " $x$  следует взять близким к 5", " $x$  следует взять близким к 3", " $x$  следует взять близким к 4", " $x$  следует

Прикладная математика

взять близким к 6". Тогда решение состоит в том, что "следует взять" "X близкое к 5". При этом точное значение слова "близко" определяется в каждом случае значением соответствующей функции принадлежности.

Таблица 7.1. Определение нечётких целей и ограничений

x	1	2	3	4	5	6	7	8	9	10	11
$\mu_{C_1}$	0	.1	.4	.8	.0	.7	.4	.2	0	0	0
$\mu_{C_2}$	.1	.6	.0	.9	.8	.6	.5	.3	0	0	0
$\mu_{C_3}$	0	0	0	1	0	0	0	0	0	0	0
$\mu_{C_4}$	.3	.6	.9	.0	.8	.7	.5	.3	.2	.1	0
$\mu_{C_5}$	0	0	0	0	0	1	0	0	0	0	0

Таблица 7.2. Определение нечёткого решения

x	1	2	3	4	5	6	7	8	9	10	11
$\mu_D$	0	.1	.4	.7	.8	.6	.4	.2	0	0	0

Как следует выполнять расплывчатые инструкции типа "X следует взять близким к 5"? Хотя на вопросы такого типа не представляется возможным дать универсальный ответ, во многих случаях все же разумно выбрать те альтернативы, которые имеют максимальную степень принадлежности к D. В нашем примере этому соответствует  $x = 5$ .

В определении нечёткого решения D как пересечения или, в более общем смысле, как слияния цели и ограничений подразумевается, что все входящие в D цели и ограничения имеют в некотором смысле одинаковую важность. Однако встречаются ситуации, в которых некоторые цели и, возможно, некоторые ограничения являются более важными, чем остальные. В таких случаях решение D может быть выражено выпуклой комбинацией целей и ограничений с весовыми коэффициентами, характеризующими относительную важность составляющих элементов. Таким образом,  $\mu_D(x)$  может быть записано в виде:

$$\mu_D(x) = \sum_{i=1}^n a_i(x) \mu_{G_i}(x) + \sum_{j=1}^m \beta_j(x) \mu_{C_j}(x) \quad (7.6)$$

где  $a_i$  и  $\beta_j$  - функции принадлежности, такие, что

$$\sum_{i=1}^n a_i(x) + \sum_{j=1}^m \beta_j(x) \equiv 1$$

С учетом этого ограничения  $a_i(x)$  и  $\beta_j(x)$  могут быть подобраны так, чтобы передавать относительную важность целей  $G_1, G_2, \dots, G_n$  и ограничений  $C_1, C_2, \dots, C_m$ . В частности, если  $m = n = 1$ , получится любое расплывчатое множество, содержащееся в  $G \cup C$  и включающее  $G \cap C$ . Отметим, что формула (7.6) напоминает известный способ сведения векторного критерия к скалярному с помощью образования линейной комбинации компонент векторной функции цели.

## 5.1 КОНКУРЕНТНЫЕ МОДЕЛИ . МОДЕЛИ ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ.

**Динамические системы** — системы, под действием внешних и внутренних сил изменяющие во времени свои состояния. Представления о динамических системах возникли как обобщение понятия механической системы, поведение которой описывается законами динамики Ньютона. В современной науке понятие динамической системы охватывает системы практически любой природы—физические, химические, биологические, экономические, социальные и др. При этом системы характеризуются различной внутренней организацией—жестко-детерминированные, стохастические, нелинейные, системы с элементами самоорганизации, самоорганизующиеся.

Важнейшим свойством динамических систем является их устойчивость, т. е. сохранение системой своей базовой структуры и основных выполняемых функций в течение определенного времени и при относительно небольших и разнообразных внешних воздействиях и внутренних возмущениях. Устойчивость есть внутреннее свойство систем, а не результат

### Прикладная математика

внешнего воздействия. Представления же о развитии этих систем отражают такие изменения их структурной организации, которые ведут к более эффективному выполнению системой своих основных функций. Качественные перестройки систем анализируются в теории катастроф, которая рассматривается как ветвь общей теории динамических систем.

Развитие представлений о динамических системах связано с переходом к познанию все более сложных систем. При этом особую роль приобретает изучение динамики внутренних свойств систем. В случае механических систем действие внутренних факторов сводилось к силам инерции. По мере усложнения систем возрастает значение внутренних факторов. На первый план выходят проблемы изучения источников внутренней активности систем и природы их целенаправленного функционирования и поведения.

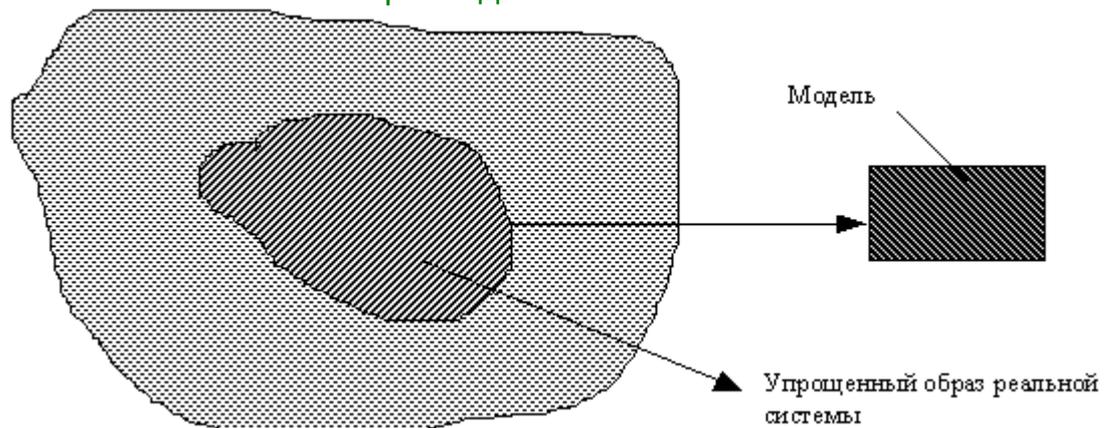
### **. Моделирование экономических процессов.**

Моделирование представляет собой построение математической модели. Для этого необходимо иметь строгое представление о цели функционирования исследуемой экономической системы и располагать информацией об ограничениях, которые определяют область допустимых значений управляемых переменных. Цель и ограничения должны быть представлены в виде функций от управляемых переменных. Анализ модели должен привести к определению наилучшего управляющего воздействия на объект управления при выполнении всех установленных ограничений.

Сложность реальных систем может сильно затруднить представление цели и ограничений в аналитическом виде. Несмотря на слишком большое число переменных и ограничений, которые на первый взгляд необходимо учитывать при анализе реальных ситуаций, лишь небольшая их часть оказывается существенной для описания исследуемых систем. Поэтому при моделировании систем прежде всего следует идентифицировать доминирующие переменные, параметры и ограничения.

Изобразим схематически уровни абстракции, соответствующие переходу от системы-оригинала к ее модели.

Прикладная математика



Упрощенный образ реальной системы отличается от системы-оригинала тем, что в нем находят отражение только доминирующие факторы (переменные, ограничения, параметры), определяющие основную линию поведения реальной системы.

Модель, будучи дальнейшим упрощением образа системы-оригинала, представляет собой наиболее существенные для описания системы соотношения в виде целевой функции и совокупности ограничений.

Рассмотрим пример, чтобы пояснить различные уровни абстракции.

Процесс создания конечного продукта материального производства обычно состоит из нескольких этапов, которые можно представить в определенной хронологической последовательности - от замысла проектировщика(конструктора) до поставки потребителю.

- После утверждения соответствующей проектно-технологической документации в производственный отдел предприятия направляется наряд-заказ на изготовление данной продукции.
- В свою очередь производственный отдел направляет на склад заявки на необходимое сырье и материалы. Склад либо удовлетворяет эти требования поставки нужного сырья и материалов со складов, либо ставит перед отделом закупок (маркетинга) вопрос об организации снабжения нужными материалами со стороны.
- После того как продукция изготовлена, отдел сбыта(маркетинга) реализует функции, связанные со сбытом и доставкой готовой продукции потребителю.

Теперь предположим, что задача исследования состоит в определении “наиболее выгодного” для предприятия-изготовителя объема производства данной продукции. Рассматривая исследуемую производственную систему как единое целое, легко убедиться в том, что на объем производства может

Прикладная математика

влиять большое число факторов. Приведем несколько примеров таких факторов, группируя их по соответствующим подразделениям предприятия.

1. Производственный отдел. Возможное время загрузки оборудования, последовательность технологических операций, выполняемых на оборудовании, объем незавершенного производства, количество бракованных изделий и производительность службы технического контроля.

2. Склад. Имеющиеся запасы сырья и материалов, ограничения запасов готовой продукции.

3. Отдел сбыта (маркетинга). Прогноз объемов реализации продукции, активность рекламной деятельности, возможности организационно-технической базы сбыта продукции, влияние конкуренции.

Учесть все эти факторы в модели, предназначенной для определения оптимального объема производства, довольно трудно. В этом случае пришлось бы ввести такие переменные, как распределение времени использования оборудования и рабочей силы, производительность службы технического контроля, а в качестве ограничений использовать данные, характеризующие мощность оборудования, фонд рабочего времени, предельно допустимый объем незавершенного производства, а также ограничения, связанные со сбытом продукции и хранением готовой продукции на складах. Пересечение всех этих факторов показывает, насколько сложны соотношения, которые пришлось бы использовать для того, чтобы выразить объем производства в виде функции от такого количества переменных.

Для формирования упрощенного абстрактного образа системы-оригинала следует рассматривать систему как единый объект, а не заниматься частными деталями анализируемой проблемы. По существу систему как единое целое можно рассматривать с точек зрения изготовителя и потребителя. В первом случае систему можно характеризовать ее *производительностью*, а во втором - *объемом реализации продукции*.

*Производительность* системы зависит от таких факторов, как располагаемое время использования рабочей силы и оборудования, последовательность технологических операций, наличие сырья.

*Объем реализации продукции* определяется ограничениями, присущими системе распределения готовой продукции, и прогнозом сбыта.

Прикладная математика

Упрощения, осуществляемые при переходе от системы оригинала к ее упрощенному образу, достигается за счет “объединения” нескольких первичных факторов в один фактор. Однако, следует помнить, что степень сложности модели всегда находится в обратной зависимости от степени упрощения реальной системы. Можно, например, принять допущение о том, что производительность системы и объем сбыта либо постоянны, либо зависят от времени. Очевидно, что в последнем случае необходимо разрабатывать более сложную модель.

Правил, определяющих переход от реальной системы к модели не существует. Сведение множества факторов, управляющих поведением системы, к относительно небольшому количеству доминирующих факторов и переход от упрощенного образа системы-оригинала к модели - в большей мере искусство, чем наука. Степень адекватности построенной модели реальной системе зависит прежде всего от творческих способностей и интуиции разработчика. Ясно, что проявление этих чисто индивидуальных качеств нельзя отразить в рамках формализованных правил построения моделей.

Несмотря на то что строгие предписания о том, как следует разрабатывать модель, сформулировать невозможно, все же полезно иметь некоторое представление о возможных типах моделей, их общей структуре и характеристиках.

**Классификация моделей.**

- Математические модели.

В основе их построения лежит допущение о том, что все релевантные(переставляемые) переменные, параметры и ограничения, а также целевая функция количественно измеримы. Поэтому, если  $x_j, j=1,2,\dots,n$  представляют собой  $n$  управляемых переменных и условия функционирования исследуемой экономической системы характеризуются  $m$  ограничениями, то математическая модель может быть записана в следующем виде:

найти оптимум  $z=f(x_1,x_2,\dots,x_n)$  (целевая функция)

при

$g_i(x_1,x_2,\dots,x_n) \leq b_i, i=1,2,\dots, m$  (ограничения).

$x_1,x_2,\dots,x_n \geq 0$

### Прикладная математика

Ограничения  $x_1, x_2, \dots, x_n \geq 0$  называются условиями неотрицательности. Эти условия требуют, чтобы переменные принимали только положительное или нулевое значение. В большинстве практических случаев такое требование вполне естественно. Нахождение оптимума осуществляется для определения наилучшего значения целевой функции, например, максимума прибыли или минимума затрат.

Общие классы математических моделей: модели линейного, целочисленного и динамического программирования, вероятностные модели, модели нелинейного программирования.

- Имитационные модели.

Воспроизводят поведение системы на протяжении некоторого промежутка времени. Это достигается путем идентификации ряда событий, распределение которых во времени дает важную информацию о поведении системы. После того как такие события определены, требуемые характеристики системы необходимо регистрировать только в моменты реализации этих событий. Информация характеристиках системы накапливается в виде статистических данных таких наблюдений. Эта информация обновляется всякий раз при наступлении каждого из интересующих нас событий.

Для построения имитационных моделей не требуется использования математических функций, явным образом связывающих те или иные переменные. Эти модели позволяют имитировать поведение сложных систем, для которых построение математических моделей и получение решений невозможно. Более того, гибкость, присущая имитационным моделям, позволяет добиться более точного представления системы.

Основной недостаток имитационного моделирования заключается в том, что его реализация эквивалентна проведению множества экспериментов, а это неизбежно обуславливает наличие экспериментальных ошибок. Кроме того, сам процесс оптимизации также вызывает затруднения.

- Эвристические методы.

Базируются на интуитивно или эмпирически выбираемых правилах, которые позволяют улучшить уже имеющееся решение. Используются в том случае, когда соответствующие математические построения оказываются настолько сложными, точное решение сформулированной задачи найти нельзя.

### Прикладная математика

По существу эвристические методы представляют собой процедуры поиска разумного перехода от одной точки пространства решений к некоторой другой точке с целью улучшения текущего значения целевой функции модели. Когда дальнейшего приближения к оптимуму добиться невозможно, лучшее из полученных решений принимается в качестве приближенного решения оптимизационной задачи.

#### **Оптимизация заданной целевой функции.**

Модели экономических процессов разрабатываются с целью оптимизации заданной целевой функции при некоторой совокупности ограничений. Термин “оптимизация” обычно используется для обозначения процессов максимизации или минимизации целевой функции. Поэтому для одной и той же задачи можно предложить две различные модели с различными критериями оптимизации. Например, мы можем предпочесть максимизировать прибыль, или с неменьшим основанием исходить из другой целевой установки - минимизации затрат. Эти критерии не эквивалентны, так как величина затрат может быть функцией переменных, находящихся под контролем данной фирмы, тогда как величина прибыли зависит от внешних неуправляемых факторов, например от ситуации на рынке сбыта, складывающейся под влиянием конкурентов. Использование соответствующих этим критериям оптимизационных моделей при одинаковых ограничениях не обязательно приведет к получению одинаковых оптимальных решений.

Основной вывод, который следует из вышеизложенного, заключается в том, что полученное с помощью некоторой модели конкретное оптимальное решение является наилучшим только в рамках использования именно этой модели. Другими словами, оно является наилучшим из всех возможных только тогда, когда выбранный критерий оптимизации можно считать полностью адекватным целям организации, в которой возникла исследуемая проблемная ситуация.

#### **Проблема информационного обеспечения моделей.**

Любая модель экономической системы независимо от ее сложности и адекватности системе-оригиналу принесет мало пользы при отсутствии необходимой информации.

Предположим, например, что некоторое предприятие выпускает продукцию двух видов, изготавливаемую из одного и того же сырья, имеющегося в ограниченном количестве. Пусть расход сырья на

Прикладная математика

изготовление единицы продукции вида 1 равен  $a_1$ , а продукции вида 2 -  $a_2$ . Если через  $b$  обозначить имеющийся запас сырья, то при объемах производства каждого вида продукции, равных  $x_1$  и  $x_2$  соответственно, рассматриваемый производственный процесс характеризуется ограничением  $a_1 x_1 + a_2 x_2 \leq b$ . Приступая к решению такой задачи мы должны найти значения  $a_1$ ,  $a_2$ , и  $b$ . Определение этих параметров может оказаться затруднительным и потребовать тщательного анализа большого объема данных, характеризующих работу всего предприятия.

**Этапы исследования экономических процессов.**

Работа, выполняемая в процессе исследования, состоит из следующих этапов:

- 1) идентификации проблемы;
- 2) построения модели;
- 3) решения поставленной задачи с помощью модели;
- 4) проверки адекватности модели;
- 5) реализации результатов исследования.

Хотя эта последовательность не обязательна, ее считают общепринятой.

За исключением этапа, связанного с получением решения на основе разработанной модели, когда используются формализованные методы (линейное программирование, управление запасами, теория массового обслуживания, календарное планирование и т.д.), все остальные этапы исследования выполняются без строгой ориентации на какие-либо регламентирующие правила.

• На первом этапе задача исследования заключается в *идентификации проблемы*. Здесь можно выделить следующие основные стадии:

1. формулировка задачи или цели исследования,
2. выявление возможных альтернатив решения применительно к исследуемой ситуации,
3. определение присущих исследуемой системе требований, условий и ограничений.

• Второй этап связан с *построением модели*. На этом этапе выбирается модель, наиболее подходящая для адекватного описания

### Прикладная математика

исследуемой системы. При построении такой модели должны быть установлены количественные соотношения для выражения целевой функции и ограничений в виде функций от управляемых переменных. Если разработанная модель соответствует некоторому общему классу математических моделей экономических процессов (например, моделям линейного программирования или календарного программирования), то для получения решения нужно воспользоваться известными математическими методами. Если же математические соотношения слишком сложны и не позволяют получить аналитического решения задачи, более подходящей для исследования может оказаться имитационная модель. В некоторых случаях возникает необходимость совместного использования математических, имитационных и эвристических моделей. Это все зависит от характерных особенностей и сложности исследуемой задачи.

- На третьем этапе осуществляется *решение сформулированной задачи*. При использовании математической модели решение получают с помощью апробированных оптимизационных методов; при этом модель приводит к оптимальному решению задачи. В случае применения имитационных или эвристических моделей понятие оптимальности становится менее определенным и получаемое решение соответствует лишь приближенным оценкам критериев оптимальности функционирования экономической системы.

На данном этапе кроме нахождения решения всякий раз, когда это возможно, должно быть обеспечено также получение дополнительной информации о возможных изменениях решения при изменении параметров системы. Эту часть исследования называют анализом модели на чувствительность. Он необходим, например, в тех случаях, когда некоторые характеристики исследуемой системы не поддаются точной оценке. В такой ситуации весьма важно исследовать возможные изменения оптимального решения в зависимости от соответствующих параметров системы в некоторых интервалах их количественных значений.

- Четвертый этап заключается в *проверке адекватности модели*. Модель можно считать адекватной, если, несмотря на некоторые неточности отображения системы-оригинала, она способна обеспечить достаточно надежное предсказание поведения системы. Общий метод проверки адекватности модели состоит в сопоставлении получаемых результатов с характеристиками системы. Если при аналогичных входных параметрах модель достаточно точно воспроизводит поведение системы-оригинала, то она считается адекватной. Однако такое сопоставление не дает полной уверенности в том, что поведение системы в предстоящем периоде будет таким же, как в прошлом. А поскольку построение модели осуществляется с использованием

### Прикладная математика

ретроспективных данных, то благоприятный исход такого сравнения во многом предопределен. В отдельных случаях, когда система-оригинал исследуется с помощью математической модели, допустима параллельная разработка имитационной модели, предназначенной для проверки основной математической модели.

- *Заключительный пятый этап связан с реализацией полученных результатов.* На данном этапе необходимо оформить конечные результаты исследования в виде детальных инструкций, которые должны быть составлены таким образом, чтобы они легко воспринимались лицами, ответственными за управление экономической системой (службой) и обеспечение ее функционирования.

А теперь познакомимся с одной моделью из математического аппарата исследования экономических процессов - сетевые модели планирования и управления.

#### **Календарное планирование процессов сетевыми методами.**

Календарное планирование - планирование во времени. Здесь общепринята следующая терминология.

**Программа** - определяет совокупность взаимосвязанных операций, которые необходимо выполнить в определенном порядке, чтобы достигнуть поставленной в программе цели.

**Операция** программы обычно рассматривается как работа, для выполнения которой требуются затраты времени и ресурсов. Операции логически упорядочены в том смысле, что одни операции нельзя начать, прежде чем не будут завершены другие. Как правило, совокупность операций программы не повторяется.

Задача календарного планирования заключается в минимизации продолжительности выполнения программы с учетом экономических факторов использования имеющихся ресурсов.

Решается задача календарного планирования с помощью двух аналитических методов структурного и календарного планирования и оперативного управления программами. Эти два метода получили название

- метод критического пути (МКП), предложен фирмой E.I. du Pont de Nemours & Company для управления программами строительства,

Прикладная математика

- метод оценки и пересмотра программ(ПЕРТ), разработан консультативной фирмой по заказу военно-морского министерства США для календарного планирования научно-исследовательских и опытно-конструкторских работ программы создания ракет “Поларис”.

Оба метода основное внимание уделяют временному аспекту выполнения программы и в конечном счете определяют календарный план программы. Они очень схожи между собой. Самым существенным различием является то, что в методе оценке МКП продолжительности операций предполагаются детерминированными величинами, а в методе ПЕРТ - случайными. В настоящее время оба метода составляют единый метод сетевого планирования и управления (СПУ) программами.

Сетевое планирование и управление программами включает три основных этапа: структурное планирование, календарное планирование и оперативное планирование.

Этап *структурного планирования* начинается с разбиения программы на четко определенные операции. Затем определяется оценка продолжительности операций и строится сетевая модель (сетевой график, стрелочная диаграмма), каждая дуга (стрелка) которой отображает работу. Вся сетевая модель в целом является графическим представлением взаимосвязей операций программы. Построение сетевой модели на этапе структурного планирования позволяет детально проанализировать все операции и внести улучшения в структуру программы еще до начала ее реализации.

Конечной целью этапа календарного планирования является построение календарного графика, определяющего моменты начала и окончания каждой операции, а также ее взаимосвязи с другими операциями программы. Кроме того, календарный график дает возможность выявлять критические операции (с точки зрения времени), которым необходимо уделять особое внимание, чтобы закончить программу в назначенный срок. Что касается некритических операций, то календарный план позволяет определять их резервы времени, которые можно выгодно использовать при задержке выполнения таких операций или с позиций эффективного использования ресурсов.

Заключительным этапом является оперативное управление процессом реализации программы. Этот этап включает использование сетевой модели и календарного графика для составления периодических отчетов о ходе выполнения программы. Сетевая модель подвергается анализу и в случае необходимости корректируется. В этом случае разрабатывается новый календарный план выполнения остальной части программы.

Сетевая модель отображает взаимосвязи между операциями и порядок их выполнения.

Для представления операции используется **стрелка**, направление которой соответствует процессу реализации программы во времени.

Отношение упорядочения между операциями задается с помощью событий. **Событие** определяется как момент времени, когда завершаются одни операции и начинаются другие.

Начальная и конечная точка любой операции описываются, таким образом, парой событий, которые обычно называют *начальным* событием и *конечным* событием.

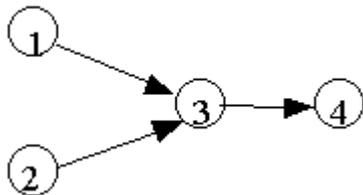
Операции, выходящие из некоторого события, не могут начинаться, пока не будут завершены все операции, входящие в это событие.

Каждое событие - это узел(вершина), а каждая операция - это ориентированная дуга.

Типичный пример графического изображения операции  $i, j$  с начальным событием  $i$  и конечным событием  $j$ .



Другой пример, из которого видно, что для возможности начала операции (3,4) требуется завершение операций (1,3) и (2,3).



Протекание операций во времени задается путем нумерации событий, причем номер начального события всегда меньше номера конечного.

### Правила построения сетевой модели.

Правило 1. Каждая операция в сети представляется одной и только одной дугой(стрелкой).

Ни одна из операций не должна появляться в модели дважды. При этом следует различать случай, когда какая-либо операция разбивается на части;

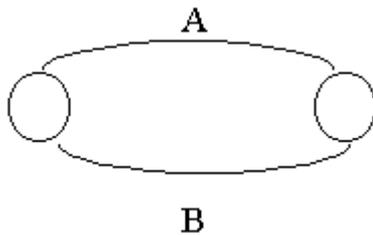
Прикладная математика

тогда каждая часть изображается отдельной дугой. Так, например, прокладку трубопровода можно расчленить на прокладку отдельных секций и рассматривать прокладку каждой секции как самостоятельную операцию.

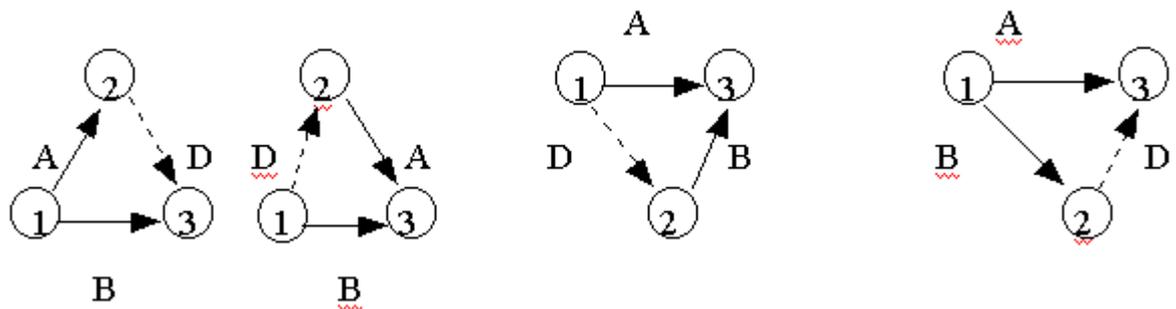
Правило 2. Ни одна пара операций не должна определяться одинаковыми начальными и конечными событиями.

Возможность неоднозначного определения операций через события появляется в случае, когда две или большее число операций допустимо выполнять одновременно.

Пример, когда операции А и В имеют одинаковые начальное и конечное события.



Чтобы исключить такую ошибку между А и конечным(начальным) событием или между В и конечным(начальным) событием вводится фиктивная операция D.

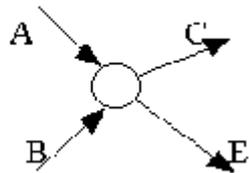


В результате операции А и В определяются теперь однозначно парой событий, отличающихся либо номером начального, либо номером конечного события.

Фиктивные операции не требуют затрат ни времени, ни ресурсов.

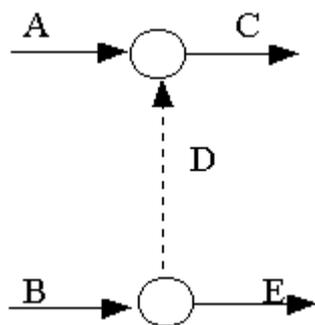
Они также позволяют правильно отражать логические связи. Предположим, что в некоторой программе операции А и В должны непосредственно предшествовать С, а операции Е непосредственно предшествует только В.

Неправильное отражение этих условий.



Упорядочения между A, B и C показаны правильно, но операции E непосредственно предшествуют обе операции A и B.

Правильное представление указанных условий достигается через фиктивную операцию D.



Правило 3. При включении каждой операции в сетевую модель для обеспечения правильного упорядочения необходимо дать ответ на следующие вопросы.

а) Какие операции необходимо завершить непосредственно перед началом рассматриваемой операции?

б) Какие операции должны непосредственно следовать после завершения данной операции?

в) Какие операции могут выполняться одновременно с рассматриваемой?

### Пример.

Постройте сетевую модель, включающую операции A, B, C, ... , L, которая отображает следующие отношения упорядочения.

1. A, B и C - исходные операции, которые можно начинать одновременно.

2. A и B предшествуют D.

Прикладная математика

3. В предшествует E, F и H.
4. F и C предшествуют G.
5. E и H предшествуют I и J.
6. C, D, F, J предшествуют K.
7. K предшествует L.
8. I, G, L - завершение операции программы.

События сети пронумеровать таким образом, чтобы возрастание номеров соответствовало ходу выполнения программы. Использовать три фиктивные операции:

D1 и D2 - для того, чтобы правильно отразить отношения следования.

D3 - для однозначного определения операций E и H по конечным событиям.

Рассмотрим пример.

А) Укажите, какое влияние оказывает добавление каждой из следующих операций на отношения упорядочения в сети. Все случаи рассмотрите независимо.

- 1) фиктивная операция (3,5) [Ответ. A предшествует I, J ]
- 2) фиктивная операция (3,4) [Ответ. A предшествует I, J ]
- 3) фиктивная операция (5,6) [Ответ. E, H предшествуют G ]
- 4) фиктивная операция (3,6) [Ответ. A предшествует G ]

б) Укажите, как можно ввести каждое из следующих отношений упорядочения в сеть.

1) Операции A и B предшествуют G [Ввести фиктивную операцию (3,6) ]

2) Операция D предшествует G [Ввести фиктивную операцию между конечным событием операции D и событием 7, а затем соединить конечное событие операции D и событие 6 фиктивной операцией ]

Прикладная математика

3) Операция С предшествует D. [Ввести фиктивную операцию между конечным событием операции С и событием б, а затем соединить конечное событие операции С и событие 3 фиктивной операцией.]

**Задачи.**

1. Постройте сетевую модель, включающую операции А, В, С, ... , Р и удовлетворяющую указанным ниже отношениям упорядочения.

- 1) А, В, С - начальные операции, которые можно начинать одновременно.
- 2) операции D, E, F начинаются сразу по окончании операции А.
- 3) I, G начинаются после завершения как В, так и D.
- 4) H начинается после окончания С, G.
- 5) К, L следуют за I.
- 6) J следует за E, H.
- 7) M, N следуют за F, но не могут начаться, пока не завершены E, H.
- 8) O следует за M и L.
- 9) P следует за J, L, O.
- 10) Операции К, N, P - завершающие.

2. Фундамент здания можно построить в виде четырех последовательных секций. Операции по каждой секции включают рытье котлована, монтаж металлоконструкций и укладку бетона. Отрывку котлована одной секции нельзя начинать, пока не закончено рытье котлована предыдущей секции. То же относится к заливке бетона. Постройте сетевую модель.

3. Программа опроса общественного мнения включает разработку и распечатку анкет, прием на работу и обучение персонала, выбор опрашиваемых лиц, рассылку им анкет и анализ полученных данных. Постройте сетевую модель программы, указав все принятые допущения относительно порядка выполнения операций.

## 7.1 ДЕТЕРМИНИРОВАННЫЕ ВРЕМЕННЫЕ РЯДЫ.

Группировка первичного материала и вычисление обобщающих показателей позволяют на этапе статистического исследования решать задачи анализа развития изучаемых явлений, во-первых, во времени и, во-вторых, во взаимосвязи.

Первая задача решается путем построения и обработки динамических (временных) рядов, которые представляют собой совокупности последовательно расположенных показателей, характеризующих изменение какого-либо явления во времени.

При изучении временных рядов ставятся следующие цели: краткое описание характерных особенностей ряда; подбор статистических моделей, описывающих временной ряд; предсказание будущих значений на основе прошлых наблюдений; управление процессом, порождающим временной ряд.

На практике эти и подобные цели достижимы далеко не всегда и далеко не в полной мере. Этому, препятствует, во-первых, недостаточный объем наблюдений и, во-вторых, изменяющаяся с течением времени статистическая структура временного ряда. Из-за этих факторов значение прошлых наблюдений обесценивается, и они уже не помогают предвидеть будущее.

При практическом анализе временных рядов последовательно проходятся следующие этапы: графическое представление и описание поведения временного ряда; выделение и удаление закономерных составляющих временного ряда, зависящих от времени: тренда, сезонных и циклических составляющих; выделение и удаление низко- или высокочастотных составляющих процесса; исследование случайной составляющей временного ряда, оставшейся после удаления перечисленных выше составляющих; построение математической модели для описания случайной составляющей; прогнозирование будущего развития процесса, представленного временным рядом; исследование взаимодействий между различными временными рядами. С учетом необходимости решения указанных выше задач исследователями используется большое количество методов: сглаживание и фильтрация - для преобразования временных рядов с целью удаления из них высокочастотных или сезонных колебаний; модели скользящего среднего и регрессии - для описания и прогнозирования процессов, проявляющих однородные колебания вокруг среднего значения;

экстраполяция и прогнозирование - для предсказания будущих значений временного ряда на основе подобранной модели поведения; корреляционный анализ - для выявления существенных зависимостей и их лагов внутри одного процесса или между несколькими процессами; спектральный анализ - для

нахождения периодических и квазипериодических составляющих временного ряда.

Построение динамического ряда является простым и наглядным приемом отображения изменения какого-либо явления во времени. В одних случаях построение ряда без дополнительной обработки выявляет тенденцию развития явления. Но бывают случаи, когда тенденция развития явно не просматривается, так как показатели ряда колеблются. В этом случае для выявления тенденций требуется определенная аналитическая обработка динамического ряда.

При анализе временного ряда его видимую изменчивость стараются разделить на закономерную и случайную составляющие. Закономерные изменения членов временного ряда следуют определенному правилу и поэтому предсказуемы. Эта составляющая  $U_t$  может быть вычислена при каждом  $t$  как некоторая функция от текущего момента  $t$ , причем эта функция может зависеть от некоторого набора параметров. Когда параметры неизвестны, их приходится оценивать по имеющимся наблюдениям - так бывает в случае использования регрессии.

Изменчивость, оставшаяся необъяснимой, иррегулярна и хаотична. Для ее описания необходим статистический подход.

Под закономерной {детерминированной} составляющей  $U_t$  временного ряда  $U_1, U_2, \dots, U_n$  понимается числовая последовательность  $D_1, D_2, \dots, D_n$ , элементы которой  $D_t$  вычисляются по определенному правилу как функция времени  $t$ .

Детерминированная составляющая отражает действие каких-либо определенных причин или факторов. Для рядов в криминологии и социальных науках причины, порождающие закономерные составляющие, могут быть неочевидными. Тем не менее, их совокупное влияние может быть устойчивым в течение достаточно длительных промежутков времени. Это обеспечивает возможность прогноза для временных рядов.

Если удастся полностью выявить закономерную составляющую в поведении временного ряда, то оставшаяся часть выглядит хаотично и непредсказуемо. Ее обычно называют случайным компонентом и обозначают

Формы разложения временного ряда на детерминированный и случайный компоненты могут различаться.

Прикладная математика

Аддитивной моделью временного ряда называется представление ряда в виде суммы детерминированного и случайного компонентов:

Мультипликативной моделью временного ряда называется представление ряда в виде произведения детерминированного и случайного компонентов:

При переходе к логарифмам в мультипликативной модели вновь получаем аддитивную модель, но уже для логарифмов .

Способы описания детерминированных компонентов временного ряда зависят от области приложений.

В детерминированном компоненте временного ряда, обычно выделяют три составляющие: тренд  $TR_t$ , сезонный компонент  $St$  и циклический компонент  $Ct$ . При этом термины «тренд», «сезонный компонент» и «циклический компонент» не имеют однозначных общепринятых определений.

Анализ временного ряда обычно начинается с выделения компонента, называемого в правовой статистике, трендом: его присутствие или отсутствие наглядно показывает график временного ряда.

Трендом временного ряда  $TR_t$ , при  $t=1, \dots, n$  называют плавно изменяющийся, не циклический компонент, описывающий чистое влияние долговременных факторов, эффект которых сказывается постепенно.

Сезонный компонент отражает повторяемость процессов во времени.

Сезонный компонент  $St$  временного ряда при  $t = 1, \dots, n$  описывает поведение, изменяющееся регулярно в течение заданного периода (года, месяца, недели, дня и т.п.). Он состоит из последовательности почти повторяющихся циклов.

Сезонные эффекты присущи многим сферам человеческой активности: многие виды продукции имеют сезонный характер производства, некоторые виды преступлений (кражи, изнасилования) также имеют ярко выраженную сезонность.

Сезонный компонент может иметь плавающий, или изменяющийся характер. Идея подхода к анализу сезонных компонентов заключается в переходе к сравнению значений через определенный период времени. Так, при изучении динамики месячных количеств краж за несколько лет данные

декабря одного года обычно сравнивают с данными декабря предыдущего года.

Циклический компонент занимает промежуточное положение между закономерной и случайной составляющими временного ряда. Если тренд - это плавные изменения, проявляющиеся на больших временных промежутках, если сезонный компонент - это периодическая функция времени, ясно видимая, когда ее период много меньше общего времени наблюдений, то под циклическим компонентом обычно подразумевают изменения временного ряда, достаточно плавные и заметные для того, чтобы не включать их в случайную составляющую, и которые нельзя отнести ни к тренду, ни к сезонному, периодическому компоненту.

Циклический компонент  $S_t$  временного ряда описывает длительные периоды относительного подъема и спада. Он состоит из циклов, которые меняются по амплитуде и протяженности.

Числовые значения показателей динамического ряда называются уровнями ряда, которые могут быть выражены абсолютными показателями, относительными и средними величинами.

В тех случаях, когда уровни выражены абсолютными показателями, различают интервальные и моментные ряды.

Интервальным называется такой ряд, абсолютные уровни которого характеризуют величину изучаемого явления, полученную в итоге за определенный период времени.

Моментным называется ряд, абсолютные уровни которого характеризуют величину явления по состоянию на определенные моменты времени.

Таким образом, в интервальном ряду интервал - это промежуток времени, за который обобщены приводимые сведения, а в моментном ряду интервал - это промежуток времени между датами, на которые приведены сведения. В связи с этим значения показателей интервального ряда можно складывать. Значения показателей же моментного ряда складывать нельзя.

Основным требованием, предъявляемым к построению любых динамических рядов, является сопоставимость их уровней. Это требование должно обеспечиваться выбором одинакового интервала, единством методологии учета и расчета показателей.

Простейшими характеристиками динамического ряда являются: уровни ряда; абсолютный прирост; средний уровень ряда; темп роста; темп прироста.

Прикладная математика

Динамика зарегистрированных преступлений характеризуется следующими данными: 2006г. - 15 600, 2007г. - 17 300, 2008г. - 18 100, 2009г. - 19 500, 2010г. - 20 000. Абсолютный прирост преступлений в 2007г. по сравнению с 2006г. составляет 1700 (17 300 - 15 600), в 2008г. по сравнению с 2007г. - 800 (18 100 - 17 300) и т.д.

Средний уровень ряда, будучи обобщающим показателем для интервальных рядов и рядов средних величин, рассчитывается как среднее арифметическое из отдельных уровней.

Этот показатель характеризует среднегодовое значение уровней ряда за какой-то рассматриваемый период.

Для моментных рядов средний уровень рассчитывается несколько иначе.

Темпом роста называется отношение данного уровня к сравниваемому. Это относительный показатель, выраженный в процентах; если этот показатель исчисляется в долях, то он называется коэффициентом роста.

Темп роста показывает, во сколько раз уровень данного периода больше или меньше предыдущего или базисного. Темпы роста, как уже отмечалось, могут вычисляться цепным способом, когда каждый последующий уровень делится на каждый предыдущий, и базисным способом, когда каждый уровень делится на один и тот же уровень, взятый за базу.

Темп прироста - это относительный показатель, характеризующий, на сколько процентов один уровень больше или (меньше) другого уровня. Он может быть получен путем вычитания 100% из темпа роста. В нашем примере темп прироста будет равен 7% (107%-100%). Темп прироста можно получить также путем деления абсолютного прироста на базисный уровень, по сравнению с которым рассчитан абсолютный прирост. В данном случае находим абсолютный прирост  $1906988 - 1769565 = 137443$  и делим его на базисный уровень, т.е.

Построение динамического ряда в отдельных случаях, когда показатели уровней последовательно увеличиваются или уменьшаются, может сразу обнаружить тенденцию развития явления во времени. Но такие случаи бывают редко. Чаще всего уровни ряда на протяжении определенного времени колеблются. Эти колебания вызываются одновременным действием случайных и систематических, краткосрочных и долгосрочных факторов. Задача состоит в том, чтобы, отбросив случайные факторы, выявить общую тенденцию в изменении уровней ряда.

### Прикладная математика

Такая задача решается путем применения методов выравнивания динамических рядов.

Методы выравнивания динамических рядов можно разделить на две группы: эмпирические и аналитические.

Одним из самых простых эмпирических методов выравнивания динамических рядов является метод укрупнения интервалов. Суть его состоит в том, что в результате анализа ряда выбирается соответствующий укрупненный интервал и в пределах этого интервала складываются показатели уровней имеющегося ряда, в результате чего получается новый выровненный ряд. Рассмотрим пример.

Для выравнивания ряда выбираем интервал в три месяца и, сложив в пределах этих месяцев показатели уровней ряда, получим выровненный ряд:

Следует иметь в виду, что данный способ применим только к интервальным рядам. Для моментных рядов и рядов средних величин рассчитывается средний уровень по новым укрупненным интервалам.

Другим эмпирическим методом выравнивания динамических рядов является метод скользящей средней. Суть метода заключается в замене фактических уровней ряда скользящими средними, взятыми в пределах последовательно сдвигаемых интервалов. При этом способе по каждому укрупненному интервалу берется не сумма показателей, а их средняя арифметическая, причем после вычисления первой средней интервал переносится на один шаг вправо.

Для расчета скользящей средней можно, например, взять сумму пяти уровней и вычислить среднюю арифметическую этой суммы, разделив ее на пять. Далее, переходя на один интервал, подсчитывают среднюю для следующих пяти членов (начиная со второго) и так до конца.

В результате получается динамический ряд скользящих средних, который помогает более отчетливо выявить тенденции в развитии явлений.

Интервал в пять членов ряда выбран в данном случае произвольно. Можно производить сглаживание скользящими средними за 2, 3, 4 периода и т.д. в зависимости от характера динамического ряда.

Чем больше интервал, за который исчисляется средняя, тем более сглаженный ряд приближается к фактическому.

Помимо эмпирических методов существуют более сложные аналитические методы обработки динамических рядов. Сущность этих

### Прикладная математика

методов заключается в том, что на основе фактических данных подбираются подходящие для отражения тенденций развития явления математические уравнения, по которым рассчитываются теоретические значения уровней ряда. Используя математические критерии сравнения, среди уравнений выбирается такое, которое наилучшим образом описывает экспериментальные данные. Затем с его помощью находят расчетные уровни ряда динамики, близкие к фактическим и выявляющие тенденции развития явления, нашедшего отражение в форме исходного динамического ряда. В качестве зависимостей для аналитической обработки динамических рядов могут выбираться прямолинейная, параболическая, гиперболическая, экспоненциальная, логарифмическая и др.

Простейшей математической формулой, выражающей тенденции развития, является формула прямой линии. Прямая линия характеризует равномерное изменение динамики. Выравнивание по прямой осуществляется методом наименьших квадратов. Этот метод обеспечивает минимальную разность между фактическими и теоретическими уровнями.

Выравнивание, как особый способ обработки динамических рядов, решает задачу выявления тенденций развития того или иного явления к настоящему моменту. Но при исследовании может возникнуть и другая задача, а именно: как данное явление будет развиваться в будущем.

Нахождение по известным значениям недостающих уровней внутри динамического ряда называется интерполяцией. Таким образом, экстраполяция и интерполяция - это распространение выводов, полученных из наблюдения над одной частью явления на другую его часть.

Метод экстраполяции заключается в нахождении значений, лежащих за пределами данного статистического ряда: по известным значениям статистического ряда находят другие значения, лежащие за пределами этого ряда.

При экстраполяции выводы, которые сделаны при изучении тенденций развития явления в прошлом и настоящем, переносятся на будущее, т.е. в основе метода лежит предположение об определенной стабильности факторных признаков, влияющих на развитие данного явления.

Ввиду такого предположения экстраполяцию можно применять при прогнозировании на короткий срок, причем прогнозируемый период должен быть значительно меньше периода, за который выявлена тенденция изучаемого явления.

Прикладная математика

При экстраполяции используется следующая терминология и понятия :  $t_1$  — глубина ретроспекции;  $t_2$  — момент прогнозирования;  $t_3$  — прогнозный горизонт;  $t_2 - t_1$ -интервал наблюдения (промежуток времени, на базе которого исследуется история развития объекта прогнозирования);  $t_3 - t_2$  — интервал упреждения (промежуток времени, на который разрабатывается прогноз).

Чем более устойчивый характер носят прогнозируемые процессы и тенденции, тем дальше может быть отодвинут горизонт прогнозирования. Как показывает практика, интервал наблюдения должен быть в три и более раза длиннее интервала упреждения. Как правило, этот период довольно короткий: до 1-го года. Метод экстраполяции не работает и при скачкообразных процессах.

Метод экстраполяции легко реализуется на персональных компьютерах. В частности, он особенно оперативен в реализации при использовании табличного процессора MS Excel, который можно установить практически на все современные персональные компьютеры.

Однако следует заметить, что при этом необходимо быть внимательным при выборе вида нелинейной функции линии тренда. Чем больше значение коэффициента достоверности  $R^2$ , тем точнее сглаживающая кривая описывает эмпирические данные, но долгосрочная тенденция отражается тогда в меньшей степени. Поэтому при интерпретации полученной линии тренда, когда она неограниченно растет или резко уменьшается, надо учитывать также и физический смысл.

Следует иметь в виду и то обстоятельство, что экстраполяция - это не конечный результат, а отправной момент прогнозирования. Прогноз разрабатывается с учетом результатов экстраполяции, но с привлечением дополнительной информации, не содержащейся в самом динамическом ряду.

Если значения уровней динамического ряда последовательно увеличиваются или последовательно уменьшаются, то экстраполяцию можно осуществлять методом среднего геометрического. В этом случае находится средний темп роста, а он, как нам уже известно, определяется с помощью среднего геометрического. Умножая последний известный уровень ряда на средний темп, находим первый расчетный уровень; умножая его на средний темп роста, находим второй расчетный уровень и т.д.

Для динамических рядов, значения показателей уровней которых колеблются, можно применять аналитические методы выравнивания.

Аналитическая обработка интервальных рядов динамики в целях выделения из их уровней случайных компонентов и установления тенденций

### Прикладная математика

развития осуществляется с помощью ряда способов, одним из наиболее часто применяемых и теоретически обоснованных среди которых является метод наименьших квадратов.

Посредством выравнивания по способу наименьших квадратов не только устанавливается общая тенденция развития явления, но и дается количественная характеристика изменения уровней ряда.

Выравнивание может быть произведено по прямой или какой-либо другой линии, выражающей функциональную зависимость уровня динамического ряда от времени:

Выравниванию предшествует теоретический анализ динамического ряда в целях познания сущности исследуемого явления и законов его развития, на основе чего устанавливаются характер динамики и тип необходимой кривой. Теоретический, качественный анализ является основой и в дальнейшем преобразовании ряда. Выравнивание же выступает лишь в качестве технического приема, инструмента, который способствует теоретическому анализу. Если явление развивается с относительно стабильными абсолютными приростами  $\Delta Y$ , то на практике чаще всего применяется выравнивание по прямой линии.

При выравнивании ряда динамики по прямой линии фактические уровни  $Y$  ряда заменяют теоретическими  $Y_t$ , которые равномерно возрастают или убывают, так как прямой линией характеризуется равномерное изменение динамики.

Метод наименьших квадратов предполагает в этом случае, что расчетные уровни ряда должны лежать на прямой линии, которая ближе всего подходит к линии фактических уровней ряда и наиболее точно отражает тенденции изменения ряда. Найденная прямая обладает следующим свойством: сумма квадратов отклонений ее ординат от соответствующих ординат исходного ряда, т.е. разность между экспериментальными и теоретическими уровнями при одних и тех же абсциссах является наименьшей:

В приведенной формуле значения  $t$  всегда известны, а для нахождения нужно определить параметры прямой  $a$  и  $b$ .

Для нахождения параметров линейного уравнения получается система линейных уравнений:

В целях облегчения нахождения параметров  $a$  и  $b$  систему упрощают, придавая условно показателям времени  $t$  такие значения, при которых их сумма становится равной нулю. Для этого в рядах с нечетным числом членов

Прикладная математика

серединный член обозначается нулевым интервалом  $t = 0$ , а другие члены получают условные номера  $-1, -2, -3$  и т.д. и  $+1, +2, +3$  и т.д. Если число членов ряда четное, то два серединных члена обозначаются нулевыми интервалами, а другие члены, как и ранее,  $-1, -3, -5, -7$  и т.д.,  $+1, +3, +5, +7$  и т.д.

При таком подходе рассмотренная выше система уравнений упрощается и принимает следующий вид:

Определив параметры  $a$  и  $b$ , легко вычислить теоретические (расчетные) уровни, т.е. ординаты точек искомой прямой  $U_t$ .

Аналитическая обработка динамического ряда позволяет за колебаниями уровней исходного ряда обнаружить определенную тенденцию (тренд), количественным выражением которой будет значение параметра  $b$ , и экстраполировать динамический ряд в будущее. При этом следует иметь в виду, что интервал прогноза должен быть гораздо короче того периода, за который выявлена тенденция прогнозируемого процесса.