Министерство образования Российской Федерации
Уральский государственный университет
им. А. М. Горького

УЧЕБНОЕ ПОСОБИЕ

Екатеринбург
Издательство Уральского университета
2001

ВВЕДЕНИЕ В ТЕОРИЮ, МЕТОДЫ И ЭКОНОМИЧЕСКИЕ ПРИЛОЖЕНИЯ ЗАДАЧ О ДОПОЛНИТЕЛЬНОСТИ

Л.Д.ПОПОВ

2001 год

УДК 519.863
ББК 22.18
П 58

Печатается по решению
редакционно-издательского совета
Уральского государственного университета
им. А.М.Горького

Научный редактор профессор Вл.Д.Мазуров

Рецензенты: кафедра анализа систем и принятия решений Уральского государственного технического университета; доктор физ.-мат. наук, профессор Н. Н. Астафьев

Попов Л.Д.

П58 Введение в теорию, методы и экономические приложения задач о дополнительности: Учеб. пособие. Екатеринбург: Изд-во Урал. ун-та, 2001. 124 с.

ISBN 5–7996–010–6

Пособие содержит вводный материал и упражнения по теории, методам и экономическим приложениям линейных и нелинейных задач о дополнительности, а также конечномерным вариационным неравенствам. Может быть полезно студентам математических и экономических специальностей вузов, аспирантам и специалистам в области вычислительной математики и математической экономики, а также всем, кто сталкивается с математическим моделированием экономических процессов и решением сложных нелинейных задач.

УДК 519.863
ББК 22.18

ISBN 5–7996–010–6

Оглавление

ПРЕДИСЛОВИЕ

Вариационные неравенства и линейные и нелинейные задачи о дополнительности стали изучаться с середины 1960-х годов, начиная с работ Дж.Стампаккьи, К.Лемке, Дж.Данцига, Р.Коттла и др. Эти постановки сыграли значительную роль в математическом программировании и являются центральными в моделировании, численном решении и анализе многих задач в экономических, физических, технических и социальных исследованиях.

В последние десятилетия значительно выросла активность исследований в области конечномерных вариационных неравенств и задач о дополнительности. Эта тенденция вызвана обновленным интересом к применению этих постановок в качестве унифицированного математического аппарата изучения экономических, транспортных и теоретико-игровых задач о равновесии — аппарата, который позволил развить новые, высокоэффективные алгоритмы численного определения такого сорта равновесия. Вместе с тем в отечественной учебной литературе эта научная область представлена недостаточно.

Предлагаемое пособие написано на основе зарубежных и отечественных журнальных публикаций, а также оригинального материала и легло в основу лекций, прочитанных автором студентам 3–4 курсов математико-механического факультета Уральского госуниверситета, специализирующимся на применении математических методов и информатики в экономике. В пособие включены основные факты из качественной теории конечномерных вариационных неравенств и задач о дополнительности, а также краткий обзор методов их решения, главным образом тех, что используют свойства монотонности входящих в постановки отображений. Отдельно разобран случай линейной задачи о дополнительности и рассмотрены конечные методы ее решения. Приведены разнообразные экономические приложения, в том числе модели равновесия в транспортных сетях и модель общего экономического равновесия Вальраса.

Для понимания изложенного достаточно начальных сведений по математическому программированию и теории игр, выпуклому анализу, микро- и макроэкономике.

Список обозначений

$\mathbb{R}^n$ — n-мерное евклидово пространство векторов.

$x = (x_1, \ldots, x_n),\ y = (y_j)_m,\ z, \ldots$ — векторы (в детальной и краткой записи), в матричных вычислениях — векторы-столбцы.

$x_i,\ y_j,\ z_k, \ldots$ — их координаты с номерами i, j и k соответственно.

$\mathbb{R}^n_+ = \{x \in \mathbb{R}^n :\ x_i \geq 0, \quad i = 1, \ldots, n\}$ — неотрицательный ортант евклидова пространства.

$x^\top y = \sum_{i=1}^n x_i y_i$ — скалярное произведение векторов x, y.

$\|x\| = \sqrt{x^\top x}$ — длина (или норма) вектора x.

$x * y = (x_1 y_1, \ldots, x_n y_n)$ — векторное произведение Адамара.

$A = (a_{ij})_{n\times m},\ B = (b_{ij})_{n\times n},\ C = (c_{ij}), \ldots$ — матрицы (с указанием и без указания их размерности).

$A_i,\ B_j,\ C_k, \ldots$ — их столбцы с номерами i, j и k соответственно.

$A^\top$ — матрица, транспонированная к A.

$diag(a_1, \ldots, a_n)$ — диагональная матрица с диагональными элементами $a_1, \ldots, a_n$.

E — единичная матрица (по диагонали единицы).

A^{-1} — матрица, обратная к A, т.е. $A^{-1}A = AA^{-1} = E$.

$det(A)$ — определитель матрицы A.

$F : \mathbb{R}^n \to \mathbb{R}^n$ — однозначное отображение $\mathbb{R}^n$ в себя.

$\nabla F(x) = (\partial F_i(x)/\partial x_j)_{n\times n}$ — матрица Якоби для отображения $F(x)$, вычисленная в точке x.

$f : \mathbb{R}^n \to \mathbb{R}$ — числовая функция векторного аргумента.

$\nabla f(x)$ — градиент числовой функции $f(x)$ в точке x.

$\nabla_x f(x, y)$ — частный градиент числовой функции $f(x, y)$ по переменной x в точке (x, y).

$\min\{f(x) :\ g_i(x) \leq 0,\ i = 1, \ldots, m,\ x \in X\}$ — задача минимизации числовой функции $f(x)$ на множестве, заданном системой неравенств и включений.

$\operatorname{Arg}\min\limits_{x\in X} f(x)$ — точки допустимого множества X, в которых достигается наименьшее значение минимизируемой функции; вообще, если $\mathcal{P}$ — некоторая математическая задача, то $\operatorname{Arg}\,\mathcal{P}$ — множество ее решений.

Глава 1

ЛИНЕЙНЫЕ ЗАДАЧИ О ДОПОЛНИТЕЛЬНОСТИ

Линейная задача о дополнительности возникает как обобщение классических постановок из линейного и квадратичного программирования и теории матричных игр и является фундаментальной математической проблемой.

§ 1. Постановка задачи и истоки

Пусть заданы вектор $q \in \mathbb{R}^n$ и вещественная $(n \times n)$-матрица M. *Линейной задачей о дополнительности* называется задача решения следующей смешанной системы неравенств и уравнений, выписанных относительно векторов переменных $w \in \mathbb{R}^n$ и $z \in \mathbb{R}^n$:

$$w = q + Mz, \quad w \geq 0, \ z \geq 0, \tag{1.1}$$

$$w^\top z = 0. \tag{1.2}$$

Нетрудно видеть, что, в силу неотрицательности w и z, множество решений выписанной системы не изменится, если условие (1.2) заменить требованием

$$w_i \, z_i = 0 \ \text{ при всех } i = 1, \, \ldots, \, n. \tag{1.2a}$$

Будем обозначать сформулированную задачу $LCP(q, M)$.

Геометрически задача (1.1), (1.2) состоит в поиске неотрицательного вектора, образ которого при заданном аффинном преобразовании также неотрицателен и ортогонален ему.

Остановимся на некоторых источниках возникновения постановки (1.1), (1.2). В качестве первого из них укажем на задачи *линейного программирования*, широкие прикладные возможности которых при решении технических, экономических, социальных проблем хорошо известны. Выпишем задачу линейного программирования

$$\min\{ \, c^\top x : Ax \geq b, \quad x \geq 0 \, \} \tag{1.3}$$

и двойственную к ней задачу

$$\max\{ \, b^\top y : A^\top y \leq c, \quad y \geq 0 \, \}; \tag{1.4}$$

здесь $x, \, c \in \mathbb{R}^n$; $b, \, y \in \mathbb{R}^m$; A — вещественная $(m \times n)$-матрица. В соответствии с теорией двойственности задачи (1.3), (1.4) одновременно разрешимы или не разрешимы и, если они разрешимы, их оптимальные значения совпадают. Последнее имеет место только в случае совместности систем ограничений в (1.3), (1.4). Поскольку для всех допустимых x и y верно

$$y^\top b \le y^\top Ax \le c^\top x,$$

то решение исходной пары задач сводится к поиску таких допустимых x и y, что

$$y^\top b = c^\top x. \tag{1.5}$$

Ограничения-неравенства в задачах (1.3), (1.4) путем введения дополнительных неизвестных можно конвертировать в эквивалентные системы уравнений с неотрицательными неизвестными. Эти системы вкупе эквивалентны системе

$$\begin{aligned} &Ax - v = b, \quad\ \ x \ge 0,\ v \ge 0, \\ &A^\top y + u = c,\ \ y \ge 0,\ u \ge 0, \end{aligned} \tag{1.6}$$

а сама задача линейного программирования — поиску векторов u, v, x, y таких, что

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} c \\ -b \end{pmatrix} + \begin{pmatrix} 0 & -A^\top \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \quad \begin{matrix} u \ge 0,\ v \ge 0, \\ x \ge 0,\ y \ge 0, \end{matrix} \tag{1.7}$$

и, в силу (1.5),

$$x^\top u + y^\top v = 0. \tag{1.8}$$

Нетрудно заметить, что задача (1.7), (1.8) имеет форму линейной задачи о дополнительности (1.1), (1.2), где

$$w = \begin{pmatrix} u \\ v \end{pmatrix},\ q = \begin{pmatrix} c \\ -b \end{pmatrix},\ M = \begin{pmatrix} 0 & -A^\top \\ A & 0 \end{pmatrix},\ z = \begin{pmatrix} x \\ y \end{pmatrix}. \tag{1.9}$$

Другим источником постановок вида (1.1)–(1.2) являются задачи *квадратичного программирования*. Пусть задача квадратичного программирования записана в виде

$$\min\{\, z(x) = c^\top x + \frac{1}{2}\, x^\top Dx :\ Ax \ge b,\ x \ge 0 \,\}; \tag{1.10}$$

здесь $x, c \in \mathbb{R}^n$; $b \in \mathbb{R}^m$; A — вещественная $(m \times n)$-матрица , D — симметричная вещественная $(n \times n)$-матрица. Как известно, целевая функция в (1.10) выпукла, если только $x^\top Dx \ge 0 \quad \forall x$, т. е. матрица D положительно полуопределена. В этом случае говорят о задаче выпуклого квадратичного программирования. Ясно, что когда D – нуль-матрица, (1.10) реформируется в задачу линейного программирования (1.3).

Для задачи (1.10) определим u и v как

$$u = Dx - A^\top y + c, \quad v = Ax - b. \tag{1.11}$$

В соответствии с теоремой Куна—Таккера вектор $\hat{x}$ доставляет минимум функции $z(x)$, только если существуют вектор $\hat{y}$ и векторы $\hat{u}$, $\hat{v}$ такие, что

$$\begin{gathered} \hat{u} = D\hat{x} - A^\top \hat{y} + c \ge 0, \quad \hat{v} = A\hat{x} - b \ge 0,\ \hat{x} \ge 0, \quad \hat{y} \ge 0, \\ \hat{x}^\top \hat{u} = 0, \quad \hat{y}^\top \hat{v} = 0. \end{gathered} \tag{1.12}$$

В случае задачи выпуклого квадратичного программирования эти условия являются и достаточными, так как

$$z(x) - z(\hat{x}) = c^\top (x - \hat{x}) + \frac{1}{2}\, x^\top Dx - \frac{1}{2}\, \hat{x}^\top D\hat{x} =$$

$$= \hat{u}^\top(x-\hat{x}) + \hat{y}^\top(v-\hat{v}) + \frac{1}{2}\,(x-\hat{x})^\top D(x-\hat{x}) =$$

$$= \hat{u}^\top x + \hat{y}^\top v + \frac{1}{2}\,(x-\hat{x})^\top D(x-\hat{x}) \geq 0$$

для любого допустимого вектора x.

Таким образом, задача квадратичного программирования приводит к поиску решения системы

$$\begin{aligned} u &= Dx - A^\top y + c, \qquad & x \geq 0,\ y \geq 0, \\ v &= Ax - b, & u \geq 0,\ v \geq 0, \end{aligned} \tag{1.13}$$

$$x^\top u + y^\top v = 0. \tag{1.14}$$

Очевидно, задача (1.13), (1.14) также имеет форму (1.1), (1.2), где

$$w = \begin{pmatrix} u \\ v \end{pmatrix},\ q = \begin{pmatrix} c \\ -b \end{pmatrix},\ M = \begin{pmatrix} D & -A^\top \\ A & 0 \end{pmatrix},\ z = \begin{pmatrix} x \\ y \end{pmatrix}. \tag{1.15}$$

Соотношения (1.13)–(1.15) подводят к рассмотрению матриц более общего вида

$$M = \begin{pmatrix} D & -A^\top \\ A & G \end{pmatrix},$$

где подматрица G, подобно D, положительно полуопределена и симметрична. Матрицы такого вида возникают в связи с так называемой симметричной задачей квадратичного программирования

$$\min\{\, c^\top x + \frac{1}{2}\,\left(x^\top Dx + y^\top Gy\right) : Ax + Gy \geq b,\quad x \geq 0,\ y \geq 0\,\}, \tag{1.16}$$

и двойственной к ней задачей

$$\max\{\, b^\top y - \frac{1}{2}\,\left(x^\top Dx + y^\top Gy\right) : -Dx + A^\top y \leq c,\ x \geq 0,\ y \geq 0\,\}. \tag{1.17}$$

Все результаты по двойственности в линейном программировании распространяются и на квадратичный случай.

В качестве последнего примера математических постановок, приводящих к линейным задачам о дополнительности, рассмотрим *биматричную игру* $\Gamma(A,B)$ (игру двух лиц с ненулевой суммой, задаваемую парой вещественных $(m \times n)$ матриц A и B). Каждая из сторон обладает конечным набором согласованных с *правилами игры* способов поведения или *чистых* стратегий, которые применяет в конкретной *партии* (реализации игры) втайне от другой стороны. Предполагается, что результат партии полностью определяется выбором чистых стратегий, а именно, если первый игрок применил свою чистую стратегию с номером i, а второй — чистую стратегию с номером j, их ожидаемые *потери,* которые игроки стремятся минимизировать, равны величинам a_{ij} и b_{ij} соответственно. При проведении бесконечной серии партий с применением игроками *смешанных* стратегий $x = (x_1, \ldots, x_m) \geq 0$, $y = (y_1, \ldots, y_n) \geq 0$, где $\|x\|_1 = \sum_{i=1}^m x_i = 1$, $\|y\|_1 = \sum_{j=1}^n y_j = 1$, ожидаемые средние потери игроков составят соответственно $x^\top Ay$ и $x^\top By$ (компоненты смешанных стратегий выступают как вероятности, с которыми игроки выбирают соответствующие им чистые стратегии в той или иной конкретной партии).

Пара смешанных стратегий $(\hat{x}, \hat{y})$ называется *точкой равновесия по Нэшу* в игре $\Gamma(A,B)$, если $\hat{x}^\top A\hat{y} \leq x^\top A\hat{y}$ при всех смешанных стратегиях x первого игрока и $\hat{x}^\top B\hat{y} \le \hat{x}^\top By$ при всех смешанных стратегиях y второго игрока. Такие пары интересны тем, что игрок, придерживающийся своей равновесной стратегии (например, $\hat{x}$), фактически вынуждает другого игрока придерживаться его равновесной стратегии (соответственно $\hat{y}$).

Точка равновесия в игре $\Gamma(A,B)$, очевидно, не изменится, если A и B заменить на $A' = (a_{ij} + a_0)$ и $B' = (b_{ij} + b_0)$ соответственно, где a_0 и b_0 — произвольные числа. Поэтому без ограничения общности можно считать $A > 0$, $B > 0$, что и будет сделано ниже.

Обозначим через e_k вектор размерности k, все компоненты которого равны 1. Нетрудно видеть, что смешанные стратегии $\hat{x}$, $\hat{y}$ образуют точку равновесия в игре $\Gamma(A,B)$, если и только если они предпочтительнее любой из чистых стратегий, т. е. если

$$(\hat{x}^\top A\hat{y})\ e_m \le A\hat{y} \qquad (A > 0), \tag{1.18}$$

$$(\hat{x}^\top B\hat{y})\ e_n \le B^\top\hat{x} \qquad (B > 0). \tag{1.19}$$

Эта простая характеризация точки равновесия приводит к теореме, связывающей задачу поиска равновесной пары стратегий с решением системы вида (1.1), (1.2).

Теорема 1.1. *Пусть* $A > 0$, $B > 0$. *Если точки* x, y *удовлетворяют системе соотношений*

$$u = Ay - e_m,\ u \ge 0,\ y \ge 0, \tag{1.20}$$

$$v = B^\top x - e_n,\ v \ge 0,\ x \ge 0, \tag{1.21}$$

$$x^\top u + y^\top v = 0, \tag{1.22}$$

то точка

$$(\hat{x},\hat{y}) = \left(\frac{x}{\|x\|_1}\,;\ \frac{y}{\|y\|_1}\right) \tag{1.23}$$

есть точка равновесия в игре $\Gamma(A,B)$. *Обратно, если* $(\hat{x},\hat{y})$ *есть точка равновесия в игре* $\Gamma(A,B)$, *то точки*

$$x = \frac{\hat{x}}{\hat{x}^\top B\hat{y}},\quad y = \frac{\hat{y}}{\hat{y}^\top A\hat{y}}$$

удовлетворяют соотношениям (1.20) – (1.22).

Д о к а з а т е л ь с т в о носит выкладочный характер. Пусть соотношения (1.20)–(1.22) верны. Умножая первые два уравнения на x и y соответственно и учитывая, что (1.22) влечет равенства $x^\top u = \ = y^\top v = 0$, получаем

$$x^\top Ay = x^\top e_m = \|x\|_1, \qquad x^\top By = y^\top e_n = \|y\|_1.$$

Заметим, что величины $\|x\|_1$ и $\|y\|_1$ отличны от нуля, так как иначе u и v из соотношений (1.20), (1.21) оказались бы отрицательными. С учетом этого для пары смешанных стратегий из (1.23) имеем

$$A\hat{y} - (\hat{x}^\top A\hat{y})\,e_m =$$
$$= \frac{1}{\|y\|_1}\left(Ay - \frac{x^\top Ay}{\|x\|_1}\ e_m\right) = \frac{1}{\|y\|_1}\,(Ay - e_m) = \frac{u}{\|y\|_1} \ge 0$$

и

$$B^\top\hat{x} - (\hat{x}^\top B\hat{y})\,e_n =$$

$$= \frac{1}{\|x\|_1}\left(B^\top x - \frac{x^\top B y}{\|y\|_1}\ e_n\right) = \frac{1}{\|x\|_1}\left(B^\top x - e_n\right) = \frac{v}{\|x\|_1} \ge 0,$$

т. е. эти стратегии удовлетворяют условиям (1.18), (1.19).

Обратно, пусть пара смешанных стратегий $(\hat{x};\ \hat{y})$ удовлетворяет условиям (1.18), (1.19). В силу положительности платежных матриц A, B имеем $\hat{x}^\top A\hat{y} > 0$, $\hat{x}^\top B\hat{y} > 0$. Деля обе части неравенств (1.18), (1.19) на $\hat{x}^\top A\hat{y}$ и $\hat{x}^\top B\hat{y}$ соответственно, убеждаемся в справедливости соотношений (1.20)–(1.21). Остается проверить, что

$$x^\top u + y^\top v = x^\top(Ay - e_m) + y^\top(B^\top x - e_n) =$$

$$= \frac{1}{\hat{x}^\top B\hat{y}}(1 - \hat{x}^\top e_m) + \frac{1}{\hat{x}^\top A\hat{y}}(1 - \hat{y}^\top e_n) = 0,$$

т. е. равенство (1.22) также соблюдено.

Очевидно, что система (1.20)–(1.22) имеет форму (1.1), (1.2). Достаточно определить

$$w = \begin{pmatrix} u \\ v \end{pmatrix},\ q = \begin{pmatrix} -e_m \\ -e_n \end{pmatrix},\ M = \begin{pmatrix} 0 & A \\ B^\top & 0 \end{pmatrix},\ z = \begin{pmatrix} x \\ y \end{pmatrix}.$$

Заметим, что неравенства $A > 0, B > 0$ исключают для матрицы M возможность быть положительно полуопределенной.

Упражнения

1. Для симметричных задач квадратичного программирования (1.16), (1.17) докажите слабую теорему двойственности:

$$c^\top x + \frac{1}{2}\,\left(x^\top Dx + y^\top Gy\right) \geq b^\top u - \frac{1}{2}\,\left(v^\top Dv + u^\top Gu\right)$$

для произвольных допустимых пар (x, y) и (u, v) задач (1.16) и (1.17) соответственно. Воспользуйтесь для этого неравенствами $2x^\top Dv \leq\ \leq x^\top Dx + v^\top Dv,\ 2u^\top Gy \leq u^\top Gu + y^\top Gy$.

2. Докажите, что из разрешимости задачи (1.16) вытекает разрешимость задачи (1.17) и совпадение их оптимальных значений. Для этого проведите линеаризацию прямой задачи в оптимальной точке и включите в рассмотрение решение задачи, двойственной к линеаризованной.

3. Покажите, что если $(\bar{x}, \bar{y})$ есть решение задачи (1.16), то найдется такой вектор $\bar{v}$, что пара $(\bar{x}, \bar{v})$ даст решение двойственной задачи (1.17). Симметрично, если $(\bar{u}, \bar{v})$ — решение двойственной задачи, то найдется такой вектор $\bar{x}$, что пара $(\bar{x}, \bar{v})$ будет решением задачи (1.16).

4. Покажите, что если задачи (1.16), (1.17) разрешимы, то существуют векторы $\bar{x}$ и $\bar{v}$, образующие оптимальную пару $(\bar{x}, \bar{v})$ одновременно как для прямой, так и для двойственной задач.

§ 2. Основные матричные классы

В связи с задачей (1.1)–(1.2) выделяют ряд матричных классов. Перечислим важнейшие из них (все матрицы ниже предполагаются квадратными):

P — класс матриц, все главные миноры которых положительны;

P_0 — класс матриц с неотрицательными главными минорами;

PD — класс положительно определенных матриц, т. е. матриц M таких, что $x^\top Mx > 0 \quad \forall x \neq 0$;

PSD — класс положительно полуопределенных матриц, т. е. матриц M таких, что $x^\top Mx \geq 0 \ \ \forall x$;

CP — класс коположительных матриц; матрица M коположительна, если $x^\top Mx \geq 0 \ \ \forall x \geq 0$;

SCP — класс строго коположительных матриц; матрица M строго коположительна, если $x^\top Mx > 0 \quad \forall x \geq 0,\ x \neq 0$;

BG — класс матриц вида

$$M = \begin{pmatrix} 0 & C \\ B & 0 \end{pmatrix},$$

где подматрицы $B > 0,\ C > 0,$ возникает в биматричных играх;

BS — класс бисимметричных матриц, т. е. матриц вида

$$M = \begin{pmatrix} D & G \\ -G^\top & F \end{pmatrix},$$

где подматрицы D и F симметричны и положительно полуопределены;

SS — класс кососимметричных матриц; матрица M кососимметрична, если $M^\top = -M$;

RS — класс строчно-достаточных матриц; матрица M называется строчно-достаточной, если $XM^\top x \leq 0 \Rightarrow XM^\top x = 0,$ где $X = \operatorname{diag}\,(x_1, \ldots, x_n)$;

CS — класс столбцово-достаточных матриц; матрица M называется столбцово-достаточной, если $XMx \le 0 \Rightarrow XMx = 0$;

S — класс достаточных матриц; представляет собой пересечение классов RS и CS.

Связь перечисленных классов по включению показана на рисунке.

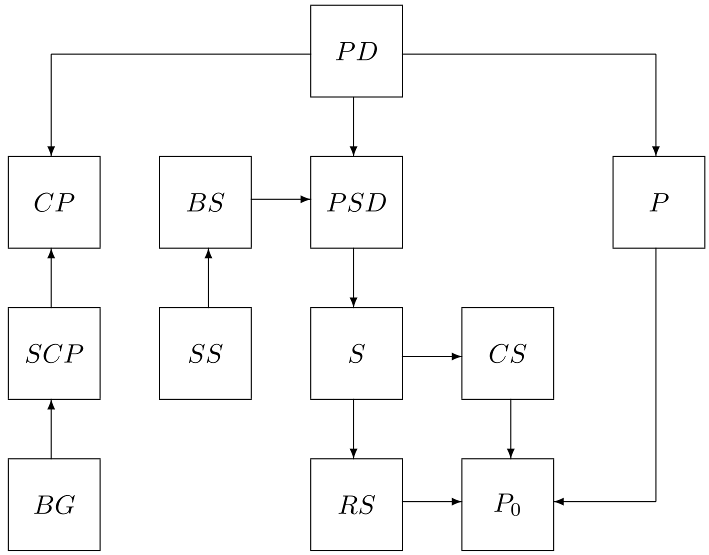
Важным (хотя и простым) свойством классов P, PSD, SS, RS, CS и S является их *полнота* в следующем смысле: перечисленные классы содержат все главные подматрицы своих членов. Напомним также, что для симметричных матриц эквивалентны условия:

1) матрица положительно (полу)определена;

2) все ее главные миноры положительны (неотрицательны);

3) все ее собственные значения положительны (неотрицательны).

Поясним включение PD в P и RS и CS в P_0. В силу полноты этих классов достаточно установить, что $M \in PD \Rightarrow \det\ (M) > 0$ и $M \in RS(CS) \Rightarrow \det\ (M) \ge 0$. Пусть, от противного, $M \in PD$ и $\det\ (M) \le 0$. Поскольку определитель произвольной матрицы равен произведению ее собственных значений, среди последних найдется $\lambda \le 0$ — вещественное и неположительное. Для соответствующего собственного вектора $x(\lambda) \ne 0$ верно $x(\lambda)^\top Mx(\lambda) = \lambda\|x(\lambda)\|^2 \le 0$, что противоречит положительной определенности M.



Связь основных матричных классов

Аналогично пусть $M \in CS$ и $\det\ (M) < 0$. Теперь среди собственных значений матрицы M найдется $\lambda < 0$ — вещественное и отрицательное. Для соответствующего собственного вектора $x(\lambda) \neq 0$ имеем

$$X(\lambda)Mx(\lambda) = \lambda\,\mathrm{diag}\ (x(\lambda)_1^2, \ldots, x(\lambda)_n^2) \leq 0,$$

причем хотя бы одна координата этого вектора отрицательна, что противоречит определению достаточности M по столбцам.

Наконец, пусть $M \in RS$. Тогда $M^\top \in CS$ и по доказанному имеем $M^\top \in P_0$. Но определитель матрицы при транспонировании не меняется. Поэтому также $M \in P_0$.

Перечисленные классы матриц и их свойства применяют при выводе теорем существования и единственности решения линейной задачи о дополнительности. Часть этих теорем будет представлена как следствие более общих результатов из раздела, посвященного нелинейным задачам. Другая часть будет получена ниже конструктивным образом из анализа метода Лемке—Хаусона и различных его обобщений.

В связи с линейными задачами о дополнительности рассматривают еще два важных матричных класса, определяемых не конструктивно. Это класс Q, состоящий из матриц, порождающих разрешимые задачи о дополнительности вне зависимости от выбора вектора свободных членов, и класс Q_0 матриц, для которых свойство разрешимости задачи $LCP(q, M)$ эквивалентно свойству совместности системы ее ограничений.

Рассмотрим класс Q_0 подробнее. Выпишем условия линейной задачи о дополнительности

$$w = q + Mz, \quad w \geq 0,\ z \geq 0, \tag{2.1}$$

$$w^\top z = 0, \tag{2.2}$$

где матрица M фиксирована, а вектор q меняется. Определим множества

$$K(M) \ = \{q: \ \text{ задача (2.1)–(2.2) имеет решение}\},$$

$$K_0(M) = \{q: \ \text{ система условий (2.1) совместна }\}.$$

Эти множества, очевидно, не пусты (содержат неотрицательный ортант пространства перемены) и являются конусами. При этом $K_0(M)$ представляет собой конус, *порожденный векторами-столбцами* расширенной матрицы

$$\bar{M} = \left(E \mid -M \right)_{n\times 2n},$$

составленной из матриц E и $-M$, так как

$$q \in K_0(M) \Longleftrightarrow q = w - Mz, \text{ где } w \geq 0,\ z \geq 0.$$

Множество $K(M)$ устроено чуть сложнее. Поскольку в любом решении линейной задачи о дополнительности в каждой паре дополнительных переменных (w_i, z_i) по крайней мере одна из переменных обращается в нуль, то вектор свободных членов такой задачи лежит в конусе K'_S, порожденном векторами набора

$$\{E_i:\ i \in S\} \bigcup \{-M_i:\ i \notin S\},$$

где $S = \{i \in \{1, \ldots, n\} :\ w_i \neq 0\}$. При этом само множество $K(M)$ представимо как объединение таких конусов K'_S по всем подмножествам S индексного множества $\{1, \ldots, n\}$, т. е.

$$K(M) = \bigcup_{S\subseteq\{1,\ldots,n\}} K'_S.$$

Разумеется, $K_0(M) \supseteq K(M)$, так как разрешимая задача всегда допустима. Обратное, вообще говоря, не верно.

Теорема 2.1. *Множества $K(M)$ и $K_0(M)$ совпадают в том и только в том случае, когда множество $K(M)$ выпукло.*

Д о к а з а т е л ь с т в о. Необходимость вытекает из очевидной выпуклости конуса $K_0(M)$. Обратимся к достаточности. Пусть множество $K(M)$ выпукло. Будучи объединением конусов, оно само является конусом и по определению выпуклого конуса содержит все неотрицательные линейные комбинации своих элементов, т. е. содержит все векторы вида

$$q = w - Mz, \text{ где } z \geq 0,\ w \geq 0,$$

так как содержит слагаемые $w \in K'_{\{1,\ldots,n\}}$ и $-Mz \in K'_{\{\emptyset\}}$.

Свойства других матричных классов будут рассматриваться по мере необходимости.

Упражнения

1. Приведите примеры P-матриц, не являющихся положительно определенными, и примеры P_0-матриц, не являющихся положительно полуопределенными.

2. Приведите примеры SCP-матриц, не являющихся положительно определенными.

3. Матрица M называется *полумонотонной* (SM-матрицей), если для любого $0 \neq x \geq 0$ найдется индекс k такой, что $x_k > 0$ и $M_k x \geq 0$. Покажите, что CP- и P_o-матрицы являются SM-матрицами.

4. Матрица M называется *коположительной-плюс*, если

$$x^\top Mx \geq 0 \text{ при всех } x \geq 0,$$

$$\big(x \geq 0,\ x^\top M x = 0\big) \Rightarrow (M + M^\top)x = 0.$$

Покажите, что матрица

$$M = \begin{pmatrix} M_1 & -A^\top \\ A & M_2 \end{pmatrix}$$

будет коположительной-плюс тогда и только тогда, когда такими будут матрицы M_1 и M_2.

5. Покажите, что класс коположительных-плюс матриц включает в себя: а) класс строго коположительных матриц, б) класс положительно полуопределенных матриц, в) класс положительных матриц.

6. Покажите, что матрица

$$M = \begin{pmatrix} 0 & -1 & 2 \\ 2 & 0 & -2 \\ -1 & 1 & 0 \end{pmatrix}$$

строчно-достаточна, но не является ни P- ни PD-, ни даже PSD-матрицей.

7. Приведите примеры P_0-матриц, не являющихся строчно(столбцово)-достаточными.

§ 3. Метод Лемке—Хаусона

Рассмотрим итеративную технику Лемке—Хаусона, первоначально разработанную для нахождения точек равновесия в биматричных играх и в дальнейшем распространенную Лемке на общий случай линейной задачи о дополнительности.

Выпишем отдельно систему линейных уравнений

$$w = q + Mz, \tag{3.1}$$

здесь, как и ранее, n-вектор q и $(n \times n)$-матрица M заданы, z и w — n-векторы неизвестных. Неизвестные z_i и w_i принято называть *дополнительными* друг к другу. Решение уравнений (3.1) удовлетворяет *условиям дополнительности*, если

$$z_i w_i = 0, \quad i = 1, \ldots, n. \tag{3.2}$$

Таким образом, решение задачи $LCP(q, M)$ есть просто неотрицательное решение системы (3.1), удовлетворяющее условиям (3.2).

Будем говорить, что решение системы (3.1) *почти удовлетворяет условиям дополнительности*, если $z_i w_i \neq 0$ только для одного значения индекса i, скажем $i = i_0$ (т. е. $z_{i_0} \neq 0,\ w_{i_0} \neq 0$).

Рассмотрим выпуклое многогранное множество

$$\mathcal{P} = \{(w, z) : w = q + Mz, \quad z \geq 0,\ w \geq 0\}.$$

Напомним, что его *крайними точками* (вершинами) называют точки, не являющиеся внутренними ни для одного из лежащих в нем нетривиальных отрезков. Крайние точки множества $\mathcal{P}$ тесно связаны с допустимыми *базисными решениями* системы (3.1), т. е. решениями, ненулевым компонентам которых соответствуют линейно независимые столбцы матрицы ее коэффициентов. Понятно, что такие решения не могут содержать более n ненулевых компонент. Если число ненулевых компонент равно n, базисное решение принято называть *невырожденным*. *Невырожденной* называется и система уравнений, все базисные решения которой не вырождены. При предположении о невырожденности системы (3.1) крайние точки множества $\mathcal{P}$ находятся с допустимыми базисными решениями системы (3.1) во взаимно-однозначном соответствии.

Процедура Лемке основана на переборе крайних точек множества $\mathcal{P}$. Она отталкивается от начала некоторого луча (неограниченного ребра множества $\mathcal{P}$), состоящего из точек $\eta = (w, z)$, почти удовлетворяющих условиям дополнительности, т. е. точек, у которых $z_i\,w_i = 0$ для всех $i = 1, \ldots, n$ за исключением некоторого $i = i_0$. Для произвольной матрицы M найти такую стартовую точку не просто. Исключение составляют, пожалуй, только два важных случая: первый связан с биматричными играми (они будут рассмотрены ниже), второй — с наличием в матрице M положительного столбца. Этот положительный столбец, впрочем, всегда можно ввести в задачу искусственно, что является полезным приемом при инициировании процедуры Лемке и ей подобных в общем случае.

Каждая итерация метода Лемке соответствует (как и в обычном симплекс-методе в линейном программировании) движению из крайней точки $\eta^{(k)}$ многогранного множества $\mathcal{P}$ вдоль некоторого его ребра, почти удовлетворяющего условиям дополнительности. Если это ребро ограничено, такое движение заканчивается в смежной крайней точке $\eta^{(k+1)}$, которая или также почти удовлетворяет условиям дополнительности, или удовлетворяет им точно. Процесс прекращается, если текущее ребро не ограничено, или точка $\eta^{(k+1)}$ уже генерировалась ранее, или она удовлетворяет условиям дополнительности точно (т. е. является решением исходной задачи).

Как уже оговаривалось, при стандартном предположении о невырожденности системы (3.1) крайние точки множества $\mathcal{P}$ находятся во взаимно-однозначном соответствии с допустимыми базисными решениями (3.1). При этом же предположении допустимое базисное решение будет удовлетворять условиям (3.2), только если из каждой дополнительной пары неизвестных (z_i, w_i) ровно одна является небазисной (свободной). Цель алгоритма — получить допустимое базисное решение с указанным свойством.

В невырожденном допустимом базисном решении, почти удовлетворяющем условиям дополнительности, имеется ровно одна пара неизвестных, скажем z_{i_0} и w_{i_0}, из которых обе являются базисными. Будем называть эту пару *базисной*. Соответственно существует ровно одна дополнительная пара (z_{j_0}, w_{j_0}) такая, что обе ее неизвестные свободны. Будем называть ее *небазисной (свободной) парой* неизвестных. Ребро, точки которого почти удовлетворяют условиям дополнительности, получается из текущего допустимого базисного решения путем *наращивания* значения одной из неизвестных, входящих в небазисную пару. Все прочие небазисные неизвестные остаются на нулевом уровне. Поэтому для каждой крайней точки, связанной с базисным решением, почти удовлетворяющим условиям дополнительности, существует *ровно два ребра*, исходящих из нее и обладающих тем же свойством [т. е. почти удовлетворяющих условиям (3.2)].

Пусть z_{j_0} — наращиваемая неизвестная небазисной пары. Вместе с ее ростом будут линейно изменяться и значения всех базисных неизвестных. При достаточно малых положительных значениях z_{j_0} свойство допустимости решения будет сохраняться. Это является следствием предположения о невырожденности системы (3.1). Однако при значительном росте z_{j_0} допустимость может быть утеряна.

Если значение z_{j_0} может быть сделано сколь угодно большим без того, чтобы какая-либо базисная неизвестная приняла отрицательное значение, перед нами луч — неограниченное ребро множества $\mathcal{P}$. В этом случае процесс завершается (без получения решения исходной задачи). Если же рост неизвестной z_{j_0} *блокирует-*

ся некоторой базисной неизвестной (той, что в процессе убывания первой достигает нулевого уровня), мы приходим к новому базисному решению. Если неизвестная, покинувшая базис, принадлежит базисной паре, перед нами решение исходной задачи. В противном случае имеем очередную крайнюю точку множества $\mathcal{P}$, почти удовлетворяющую условиям дополнительности. В этой точке наращиваемая и блокирующая неизвестные поменялись ролями: первая вошла в число базисных, а вторая — в число свободных неизвестных.

Главное *правило метода Лемке* касается выбора очередной наращиваемой (вводимой в базис) неизвестной. А именно, наращивается значение той неизвестной, которая является дополнительной к неизвестной, покинувшей базис на предыдущем шаге (итерации).

Генерируемая последовательность крайних точек, почти удовлетворяющих условиям (3.2), образует своего рода путь (*путь Лемке*) по границе множества $\mathcal{P}$.

Теорема 3.1. *Вдоль пути, почти удовлетворяющего условиям дополнительности, только исходное базисное решение может встретиться дважды.*

Доказательство с очевидностью следует из того, что каждое допустимое базисное решение, почти удовлетворяющее условиям дополнительности, связано не более чем с двумя другими такими точками. Поэтому, войдя и выйдя из такой точки, мы уже не можем в нее вернуться.

Следствие 3.1. *Если путь Лемке был начат из конца луча, почти удовлетворяющего условиям дополнительности, то метод прервется или на другом таком луче (называемом альтернативным), или на решении исходной задачи.*

Пример. Рассмотрим задачу $LCP(q, M)$, где

$$q = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, \ M = \begin{pmatrix} 1 & 2 & 1 \\ 1 & 1 & 2 \\ 2 & 1 & 1 \end{pmatrix}.$$

Ее уравнения имеют вид

$$\begin{aligned} w_1 &= 1 + z_1 + 2z_2 + z_3, \\ w_2 &= -1 + z_1 + z_2 + 2z_3, \\ w_3 &= 1 + 2z_1 + z_2 + z_3 \end{aligned}$$

или в табличной форме

		z_1	z_2	z_3	
w_1	1	1	2	1	2
w_2	−1	1	1	2	0
w_3	1	2	1	1	3
		1	0	0	

Начальный луч отвечает зафиксированному значению переменных $z_2 = z_3 = 0$ и свободному изменению переменной z_1 в диапазоне от 1 до $+\infty$. Значение $z_1 = 1$ является минимальным, при котором базисные переменные w_1, w_2, w_3 неотрицательны, причем w_2 обращается в нуль. Меняем ролями переменные z_1 и w_2. Первая из них становится базисной, вторая — свободной. Разрешая исходную систему относительно переменных w_1, z_1, w_3, приходим к эквивалентной системе

$$\begin{aligned} w_1 &= 2 + w_2 + z_2 - z_3, \\ z_1 &= 1 + w_2 - z_2 - 2z_3, \\ w_3 &= 3 + 2w_2 - z_2 - 3z_3 \end{aligned}$$

или в табличной форме

		w_2	z_2	z_3	
w_1	2	1	2	−1	2
z_1	1	1	−1	−2	1
w_3	3	2	−1	−3	3
		0	0	0	

Далее вводим в базис переменную z_2, которая является дополнительной к переменной w_2, покинувшей базис на предыдущем шаге. Рост значения переменной z_2 блокируется базисной переменной z_1, которая, убывая, раньше других базисных переменных обращается в нуль. Поэтому переменная $z1$ покидает базис, а исходная система приобретает вид

$$\begin{aligned} w_1 &= 3 + 2w_2 - z_1 - 3z_3, \\ z_2 &= 1 + w_2 - z_1 - 2z_3, \\ w_3 &= 2 + w_2 + z_1 - z_3 \end{aligned}$$

или в табличной форме

		w_2	z_1	z_3	
w_1	3	2	−1	−3	3
z_2	1	1	−1	−2	1
w_3	2	1	1	−1	2
		0	0	0	

Полученное базисное решение $(w, z) = (3, 0, 2; 0, 1, 0)$ удовлетворяет условиям дополнительности, т. е. является решением исходной задачи.

В принципе, если в качестве начальной точки брать не конец некоторого луча, в методе Лемке можно столкнуться с явлением зацикливания. Зацикливание означает, что после некоторого числа шагов метод возвращается к исходной точке. Вместе с тем даже если исходная задача разрешима, метод с правильно выбранной точки может тем не менее завершиться на альтернативном луче.

§ 4. Сходимость метода для матричных игр

Дадим алгебраическую интерпретацию завершению метода на альтернативном луче.

Лемма 4.1. *Пусть* (w, z) *— допустимое базисное решение системы* (3.1), *почти удовлетворяющее условиям дополнительности. Если оно инцидентно лучу, точки которого также почти удовлетворяют условию* (3.2), *то существуют* n*-векторы* $\bar{w}$ *и* $\bar{z}$ *такие, что*

$$\bar{w} = M\bar{z}, \quad \bar{w} \geq 0, \; \bar{z} \geq 0, \; \bar{z} \neq 0, \tag{4.1}$$

причем точки луча имеют вид

$$(w + \lambda\bar{w},\ z + \lambda\bar{z}), \quad \lambda \geq 0, \tag{4.2}$$

и удовлетворяют условию

$$(w_i + \lambda\bar{w}_i)(z_i + \lambda\bar{z}_i) = 0 \ \text{ при всех } \ i = 1, \ldots, n,\ i \neq i_0. \tag{4.3}$$

Доказательство очевидно. Подчеркнем, что индекс i_0 не меняется вдоль всего пути Лемке. Смена его означала бы, что одна из неизвестных исходной базисной пары покинула базис, что является признаком получения искомого решения и завершения работы алгоритма в целом.

Лемма 4.1 позволяет сформулировать первый результат о существовании решения линейной задачи о дополнительности.

Теорема 4.1. *Если $M > 0$, то система (3.1) при любом q имеет хотя бы одно допустимое базисное решение, удовлетворяющее условиям дополнительности.*

Доказательство. В качестве начальных базисных неизвестных выберем $w_1, \ldots, w_n$. Если $q \geq 0$, этот базис немедленно дает решение исходной задачи, а именно вектор $(q; 0)$. В противном случае присоединим к множеству ненулевых неизвестных неизвестную z_1. Поскольку $M > 0$, то при достаточно большом значении $z_1 > 0$ все w_i, где $i = 1, \ldots, n$, также будут положительны. Постепенно снижая значение z_1, остановимся на таком его значении, при котором одна из базисных неизвестных, единственная в силу предположения о невырожденности системы (3.1), обратится в нуль. Стартовая крайняя точка пути Лемке получена.

Строим путь Лемке, применяя описанные выше правила. По следствию 3.1 из теоремы 3.1 процесс прервется или по достижении решения исходной задачи, или по выходу на неограниченное ребро (луч) множества $\mathcal{P}$. Покажем, что последнее невозможно. В самом деле, если бы это было так, мы столкнулись бы с выполнением условий (4.1)—(4.3), где $i_0 = 1$. Поскольку $M > 0$ и $\bar{z} \geq 0$, то $\bar{w} > 0$. Отсюда, в силу (4.3), $z_i = \bar{z}_i = 0$ при всех $i \neq 1$. Следовательно, z_1 — единственная изменяющаяся вместе с λ неизвестная (наряду с компонентами $w_1, \ldots, w_n$). Но это значит, что конечный луч совпадает с начальным, что по следствию 3.1 невозможно.

Обоснование следующего факта требует модификации схемы выбора начальной точки.

Теорема 4.2. *Биматричная игра $\Gamma(A, B)$ разрешима, и ее решение можно получить методом Лемке.*

Доказательство. Инициируем алгоритм, выбрав наименьшее положительное значение неизвестной z_1, при котором

$$v = -e_n + B_1^\top z_1 \geq 0, \tag{4.4}$$

здесь $B_1^\top$ — первый столбец матрицы $B^\top$. Из стандартного предположения о невырожденности системы уравнений задачи о дополнительности, ассоциированной с игрой $\Gamma(A, B)$, следует, что вектор v имеет ровно одну нулевую компоненту, скажем v_r. Требуемый луч формируется путем ввода в базис неизвестной z_1 и всех дополнительных неизвестных u и v, кроме v_r. Дополнительная к v_r неизвестная w_r

выбирается в качестве той небазисной неизвестной, которая может расти неограниченно. Очевидно, что при достаточно большом ее значении все базисные неизвестные также положительны, причем точки луча почти удовлетворяют условиям оптимальности. Единственная пара неизвестных, которая может дать ненулевое произведение, — это (z_1, u_1). Понижая постепенно значение неизвестной w_r, при некотором ее положительном значении получаем начальную крайнюю точку пути Лемке.

Покажем, что реализация шагов рассматриваемого алгоритма не может завершиться на альтернативном луче. В самом деле, если бы это было так, мы получили бы точки вида

$$\begin{pmatrix} u+\lambda\bar{u} \\ v+\lambda\bar{v} \end{pmatrix} = \begin{pmatrix} -e_m \\ -e_n \end{pmatrix} + \begin{pmatrix} 0 & A \\ B^\top & 0 \end{pmatrix} \begin{pmatrix} z+\lambda\bar{z} \\ w+\lambda\bar{w} \end{pmatrix}, \tag{4.5}$$

где $\lambda \geq 0$ — любое и

$$(u_i+\lambda\bar{u}_i)(z_i+\lambda\bar{z}_i) = 0 \quad \text{при всех } i \neq 1, \tag{4.6}$$

$$(v_j+\lambda\bar{v}_j)(w_j+\lambda\bar{w}_j) = 0 \quad \text{при всех } j. \tag{4.7}$$

Предположим, что $\bar{z} \neq 0$. Тогда $\bar{v} = B^\top\bar{z} > 0$. В силу (4.7), $w_j + \lambda\bar{w} = 0$ при всех j и всех $\lambda \geq 0$. Но тогда $u + \lambda\bar{u} = -e_m < 0$, чего не может быть. Поэтому $\bar{z} = 0$. Пусть далее $\bar{w} \neq 0$. Тогда $\bar{u} = A\bar{w} > 0$. В силу (4.6), $z_i = 0$ при всех $i \neq 1$ и $\bar{z}_i = 0$ при всех i. Отсюда $\bar{v} = B^\top\bar{z} = 0$ и v есть то же самое, что и в (4.4), поскольку начальное значение z_1 выбиралось наименьшим среди тех, для которых (u, v, z, w) — крайняя точка. По предположению невырожденности только $v_r = 0$, а все прочие $v_j > 0, j \neq r$. Следовательно, условие (4.7) влечет равенство $w_j + \lambda\bar{w}_j = 0$ при всех $j \neq r$. В результате мы установили, что конечный луч совпадает с начальным, что по следствию 3.1 невозможно. Значит, алгоритм не может завершиться на луче. Он завершается на решении исходной задачи.

Пример. Рассмотрим задачу $LCP(q, M)$, где

$$q = \begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix}, \; M = \begin{pmatrix} 0 & 1 & 2 \\ 0 & 2 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

Ее уравнения имеют вид

$$\begin{aligned} w_1 &= -1 \phantom{{}+z_1} + z_2 + 2z_3, \\ w_2 &= -1 \phantom{{}+z_1} + 2z_2 + z_3, \\ w_3 &= -1 + z_1 \end{aligned}$$

или в табличной форме

$$\begin{array}{c|c|ccc|c} \multicolumn{1}{c}{} & \multicolumn{1}{c}{} & z_1 & z_2 & \multicolumn{1}{c}{z_3} & \\ \cline{2-5} w_1 & -1 & 0 & 1 & 2 & 1 \\ w_2 & -1 & 0 & 2 & 1 & 0 \\ w_3 & -1 & 1 & 0 & 0 & 0 \\ \cline{2-5} \multicolumn{1}{c}{} & \multicolumn{1}{c}{} & 1 & 0 & \multicolumn{1}{c}{1} & \end{array}$$

Устанавливая свободные переменные z_1, z_2, z_3 в нуль, получаем отрицательные значения базисных переменных $w_1 = w_2 = w_3 = -1$. Наращивая далее z_1 до значения $z_1 = 1$, обеспечиваем неотрицательность переменной w_3. Далее, поскольку

$w_3 = 0$, следует наращивать дополнительную к ней переменную z_3. Наращиваем ее до значения $z_3 = 1$, при котором остальные базисные переменные w_1, w_2 становятся также неотрицательными. Поскольку вторая из них фактически обращается в нуль, она заменяется в базисе на переменную z_3. Разрешая исходные уравнения относительно нового набора базисных переменных w_1, z_1, z_3, получаем эквивалентную систему

$$\begin{aligned} w_1 &= 1 + 2w_2 - 3z_2, \\ z_3 &= 1 + w_2 - 2z_2, \\ z_1 &= 1 \qquad\qquad\qquad\quad + w_3 \end{aligned}$$

или в табличной форме

$$\begin{array}{c|c|ccc|c} & & w_2 & z_2 & w_3 & \\ \hline w_1 & 1 & 2 & -3 & 0 & 1 \\ z_3 & 1 & 1 & -2 & 0 & 1 \\ z_1 & 1 & 0 & 0 & 1 & 1 \\ \hline & & 0 & 0 & 0 & \end{array}$$

Вводим в базис переменную z_2, дополнительную к последней покинувшей базис переменной w_2. Рост значения переменной z_2 блокируется базисной переменной w_1, которая, убывая, первой из базисных переменных обращается в нуль. Разрешая уравнения задачи относительно нового набора базисных уравнений, получаем систему

$$\begin{aligned} z_2 &= \tfrac{1}{3} + \tfrac{2}{3}w_2 - \tfrac{1}{3}w_1, \\ z_3 &= \tfrac{1}{3} - \tfrac{1}{3}w_2 + \tfrac{2}{3}w_1, \\ z_1 &= 1 \qquad\qquad\qquad\quad + w_3 \end{aligned}$$

или в табличной форме

$$\begin{array}{c|c|ccc|c} & & w_2 & w_1 & w_3 & \\ \hline z_2 & \frac{1}{3} & \frac{2}{3} & -\frac{1}{3} & 0 & \frac{1}{3} \\ z_3 & \frac{1}{3} & -\frac{1}{3} & \frac{2}{3} & 0 & \frac{1}{3} \\ z_1 & 1 & 0 & 0 & 1 & 1 \\ \hline & & 0 & 0 & 0 & \end{array}$$

Эта таблица последняя, так как соответствующее ей базисное решение $(w, z) = (0, 0, 0; \frac{1}{3}, \frac{1}{3}, 1)$ допустимо и удовлетворяет условиям дополнительности.

При доказательстве приведенных выше утверждений использовалось предположение о невырожденности системы уравнений в (3.1). Это предположение можно обойти за счет лексикографического правила выбора неизвестной, покидающей базис, так же, как это делается в линейном программировании при обосновании конечности симплекс-метода

§ 5. Применение искусственной переменной

Перейдем к изучению линейной задачи о дополнительности с матрицами более общего вида. Нам придется ввести в систему ее уравнений дополнительную переменную z_0:

$$w = q + e_n z_0 + Mz, \quad \text{где } e_n = (\underbrace{1, 1, \ldots, 1}_{n})^\top. \tag{5.1}$$

Переменная z_0 называется *искусственной* и призвана обеспечить наличие стартового допустимого базисного решения. Она играет роль своеобразного параметра, начальное положительное значение которого в ходе вычислений будет доведено до нуля. Решение параметризованной системы (5.1) будем называть *почти удовлетворяющим условиям дополнительности*, если $z_i w_i = 0$ при всех $i = 1, \ldots, n$, и *удовлетворяющим условиям дополнительности точно*, если также $z_0 = 0$. Обозначим

$$\mathcal{P}_0 = \{(w, z_0, z) : \ w = q + e_n z_0 + Mz \ \ z_0 \geq 0, \ z \geq 0, \ w \geq 0\}.$$

Пусть вначале неизвестные $w_1, \ldots, w_n$ составляют базис, а прочие неизвестные $z_0, z_1, \ldots, z_n$ остаются свободными Придавая неизвестной z_0 достаточно большое положительное значение ζ, можно добиться того, чтобы выполнялось неравенство

$$w = q + e_n z_0 > 0.$$

При снижении значения z_0 значения неизвестных w_i также будут понижаться. Начальная для метода Лемке крайняя точка множества $\mathcal{P}_0$ получится при минимальном значении $z_0 \geq 0$, обеспечивающем неравенство $w = q + e_n z_0 \geq 0$. Если $z_0 = 0$, то $q \geq 0$ и исходная задача имеет очевидное решение — вектор $(q; 0)$. В противном случае $z_0 > 0$ и среди компонент вектора w ровно одна, скажем w_r, обращена в нуль. Вводим z_0 в базис на место неизвестной w_r, которая переходит в разряд свободных.

Далее применяем правило Лемке выбора очередной вводимой в базис неизвестной, т. е. на очередном шаге вводим в базис неизвестную, которая является дополнительной к неизвестной, покинувшей базис на предыдущем шаге. Покидающая базис неизвестная определяется стандартным образом как неизвестная, которая, убывая, первой достигает нулевого уровня. Если все старые базисные неизвестные при наращивании вводимой в базис свободной неизвестной не убывают, алгоритм завершает работу на альтернативном луче без получения решения исходной задачи. Если покидающей базис неизвестной оказывается z_0, алгоритм завершается по причине получения искомого решения (оно получается, если все свободные неизвестные приравнять нулю, а прочие определить из системы уравнений $w = q + Mz$). Если, наконец, покинувшая базис неизвестная отлична от z_0, алгоритм переходит к следующей итерации.

Пример. Рассмотрим задачу $LCP(q, M)$, где

$$q = \begin{pmatrix} -3 \\ 6 \\ -1 \end{pmatrix}, \ M = \begin{pmatrix} 0 & -1 & 2 \\ 2 & 0 & -2 \\ -1 & 1 & 0 \end{pmatrix}.$$

Ее уравнения (вместе с включенной в них искусственной переменной z_0) имеют вид

$$\begin{aligned} w_1 &= -3 + z_0 \phantom{{}+2z_1} - z_2 + 2z_3, \\ w_2 &= 6 + z_0 + 2z_1 \phantom{{}-z_2} - 2z_3, \\ w_3 &= -1 + z_0 - z_1 + z_2. \end{aligned} \tag{5.2}$$

Далее для краткости будем использовать только их табличное представление

$$\begin{array}{c|r|c|rrr|l} & & z_0 & z_1 & z_2 & z_3 & \\ \hline w_1 & -3 & 1 & 0 & -1 & 2 & 0 \\ w_2 & 6 & 1 & 2 & 0 & -2 & 9 \\ w_3 & -1 & 1 & -1 & 1 & 0 & 2 \\ \hline & & 3 & 0 & 0 & 0 & \end{array} \tag{5.3}$$

Пусть $z_1 = z_2 = z_3 = 0$. Минимальное значение искусственной переменной z_0, при котором базисные переменные w_1, w_2, w_3 оказываются неотрицательными, равно 3. При этом $w_1 = 0$.

Вводим искусственную переменную z_0 в базис на место переменной w_1. Таблица, отвечающая системе (5.2), разрешенной относительно нового набора базисных переменных z_0, w_2, w_3, служит начальной для метода Лемке

		w_1	z_1	z_2	z_3	
z_0	3	1	0	1	-2	3
w_2	9	1	2	1	-4	9
w_3	2	1	-1	2	-2	2
		0	0	0	0	

Следующей в базис будет вводиться переменная z_1, которая является дополнительной к покинувшей базис переменной w_1. Переменная w_3, чье место займет z_1, определяется по единственному отрицательному элементу столбца при вводимой в базис переменной. Преобразованная таблица примет вид

		w_1	w_3	z_2	z_3	
z_0	3	1	0	1	-2	3
w_2	13	3	-2	5	-8	13
z_1	2	1	-1	2	-2	2
		0	0	0	0	

Теперь в базис следует ввести переменную z_3. Блокирующая ее рост переменная z_1 покидает базис. В результате приходим к таблице

		w_1	w_3	z_2	z_1	
z_0	1	0	1	-1	1	1
w_2	5	-1	2	-3	4	5
z_3	1	$\frac{1}{2}$	$-\frac{1}{2}$	1	$-\frac{1}{2}$	1
		0	0	0	0	

Продолжая следовать правилам алгоритма, вводим в базис переменную w_1, которая заменяет там переменную w_2. Разрешая уравнения (5.2) относительно z_0, w_1, z_0, получаем четвертую таблицу

		w_2	w_3	z_2	z_1	
z_0	1	0	1	-1	1	1
w_1	5	-1	2	-3	4	5
z_3	$\frac{7}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{3}{2}$	$\frac{7}{2}$
		0	0	0	0	

Завершает работу смена ролей переменных z_0 и z_2 :

		w_2	w_3	z_0	z_1	
z_2	1	0	1	-1	1	1
w_1	2	-1	-1	3	1	2
z_3	3	$-\frac{1}{2}$	0	$\frac{1}{2}$	1	3
		0	0	0	0	

Поскольку искусственная переменная z_0 покинула базис, вектор $(w, z) = (2, 0, 0; 0, 1, 3)$ дает решение исходной задачи.

§ 6. Другие условия сходимости

Рассмотрим теперь вопрос, какой должна быть матрица M, чтобы описанный выше процесс заканчивался на альтернативном луче только в случае неразрешимости исходной задачи.

Теорема 6.1. *Завершение метода Лемке на луче влечет существование такого ненулевого неотрицательного вектора* $\bar{z}$, *что*

$$\bar{z}_i\,(M\bar{z})_i \le 0, \quad i = 1, \ldots, n. \tag{6.1}$$

Д о к а з а т е л ь с т в о. Ясно, что завершение алгоритма на альтернативном луче возможно только, если мы получаем допустимое базисное решение (w, z_0, z) системы (5.1), которое соответствует крайней точке множества $\mathcal{P}_0$, и для которой существует определяющее этот луч направление $(\bar{w}, \bar{z}_0, \bar{z})$ такое, что

$$\bar{w} = e_n \bar{z}_0 + M\bar{z}, \quad (\bar{w}, \bar{z}_0, \bar{z}) \ge 0, \tag{6.2}$$

и при всех $\lambda \ge 0$ выполняются условия

$$w + \lambda\bar{w} = q + e_n\,(z_0 + \lambda\bar{z}_0) + M\,(z + \lambda\bar{z}), \tag{6.3}$$

$$(w_i + \lambda\bar{w}_i)\,(z_i + \lambda\bar{z}_i) = 0, \quad i = 1, \ldots, n. \tag{6.4}$$

Случай $\bar{z} = 0$ невозможен, так как иначе $\bar{z}_0$ и затем $\bar{w} > 0$ (поскольку $(\bar{w}, \bar{z}_0, \bar{z}) \ne 0$). Последнее, в силу (6.4), значило бы, что $z + \lambda\bar{z} = z = 0$, т. е. конечный луч совпадает с начальным.

Далее, так как все точки луча почти удовлетворяют условиям дополнительности, то

$$z_i\,w_i = z_i\,\bar{w}_i = \bar{z}_i\,w_i = \bar{z}_i\,\bar{w}_i = 0, \quad i = 1, \ldots, n. \tag{6.5}$$

Выпишем уравнения системы (6.2) отдельно в виде

$$\bar{w}_i = \bar{z}_0 + (M\,\bar{z})_i, \quad i = 1, \ldots, n. \tag{6.6}$$

Умножение (6.6) почленно на $\bar{z}_i$ приводит, в силу (6.5), к равенствам

$$0 = \bar{z}_i\,\bar{z}_0 + \bar{z}_i\,(M\bar{z})_i, \quad i = 1, \ldots, n. \tag{6.7}$$

Отсюда вытекает искомое неравенство.

Опираясь на теорему 6.1, можно выделить классы матриц, для которых из условий (6.1) вытекала бы неразрешимость исходной задачи (например, несовместность системы ее ограничений). Одним из таких классов является класс *коположительных-плюс* матриц, другим — уже упоминавшийся класс строчно-достаточных матриц.

Матрица M называется коположительной-плюс, если она удовлетворяет условиям

$$x^\top M x \ge 0 \quad \text{при всех } x \ge 0, \tag{6.8}$$

$$(M + M^\top)x = 0 \quad \text{при всех } x \ge 0,\ x^\top M x = 0. \tag{6.9}$$

Класс таких матриц достаточно широк и, в частности, включает в себя

1) класс строго коположительных матриц, т. е. таких матриц M, для которых $x^\top Mx > 0$ при всех $0 \neq x \geq 0$;

2) класс положительно полуопределенных матриц, т. е. таких M, что $x^\top Mx \geq 0$ при всех x.

Положительные матрицы очевидно строго коположительны. Однако, как показывают примеры, матрицы вида

$$M = \begin{pmatrix} 0 & A \\ B^\top & 0 \end{pmatrix}, \quad \text{где } A > 0,\ B > 0,$$

не обязательно удовлетворяют условиям (6.9). Из матриц M_1 и M_2, удовлетворяющим условиям (6.8),(6.9), можно составить блочные матрицы вида

$$M = \begin{pmatrix} M_1 & 0 \\ 0 & M_2 \end{pmatrix},$$

которые также оказываются коположительными-плюс. Более того, если S — кососимметричная матрица, то $M+S$ также будет коположительной-плюс. Следовательно, блочная матрица вида

$$M = \begin{pmatrix} M_1 & -A^\top \\ A & M_2 \end{pmatrix}$$

будет коположительной-плюс тогда и только тогда, когда такими будут M_1 и M_2.

При доказательстве следующей теоремы будет использоваться факт, что система

$$w = q + Mz,\ z \geq 0,\ w \geq 0,$$

не имеет решений, если найдется вектор v такой, что

$$v^\top M \leq 0, \quad v^\top q < 0,\ v \geq 0 \tag{6.10}$$

(иначе получаем противоречие $0 \leq v^\top w = v^\top q + v^\top Mz < 0$).

Теорема 6.2. *Пусть матрица M — коположительная-плюс и метод Лемке завершается на луче. Тогда исходная задача не имеет допустимых решений.*

Доказательство. Завершение на луче означает, что достигнуто некоторое допустимое базисное решение (w, z_0, z), удовлетворяющее условиям (6.2)–(6.4), где

$$0 = \bar{z}^\top \bar{w} = \bar{z}^\top e_n \bar{z}_0 + \bar{z}^\top M\bar{z}. \tag{6.11}$$

Поскольку M — коположительная-плюс и $\bar{z} \geq 0$, оба слагаемых в правой части (6.11) неотрицательны и потому равны нулю. Скаляр $\bar{z}_0 = 0$, потому что $\bar{z}^\top e_n > 0$.

Равенство нулю квадратичной формы $\bar{z}^\top M\bar{z}$ по предположению означает, что

$$M\bar{z} + M^\top \bar{z} = 0.$$

В силу (6.2), равенство $\bar{z}_0 = 0$ влечет $\bar{w} = M\bar{z} \geq 0$, т. е. по предыдущему $M^\top \bar{z} \leq 0$ или, что то же, $\bar{z}^\top M \leq 0$. Далее, уже в силу (6.5),

$$0 = z^\top \bar{w} = z^\top M\bar{z} = z^\top (-M^\top \bar{z}) = \bar{z}^\top Mz$$

и снова, в силу (6.5),

$$0 = \bar{z}^\top w = \bar{z}^\top q + \bar{z}^\top e_n z_0 + \bar{z}^\top Mz = \bar{z}^\top q + \bar{z}^\top e_n z_0.$$

Поэтому $\bar{z}^\top q < 0$, так как $\bar{z}^\top e_n z_0 > 0$. Следовательно, имеет место (6.10) при $v = \bar{z}$, и исходная задача о дополнительности не имеет допустимых точек.

Анализ проведенного доказательства позволяет усилить последний результат для строго коположительных матриц, поскольку для них равенство нулю квадратичной формы $\bar{z}^\top M \bar{z}$ на неотрицательном ортанте возможно лишь при $\bar{z} = 0$.

Следствие 6.1. *Если M строго коположительна, метод Лемке завершается получением допустимого базисного решения системы* (5.1), *которое удовлетворяет условиям дополнительности.*

Это следствие обобщает, очевидно, результат теоремы 4.1.

В заключение заметим, что в методе Лемке искусственный столбец не обязан состоять из единиц. Достаточно, чтобы он просто был положителен. Более того, удачный выбор искусственного столбца гарантирует сходимость метода Лемке не более чем за $n+1$ итерацию.

Теорема 6.3. *Пусть M — невырожденная $(n \times n)$-матрица. Предположим, что существует положительный вектор p, удовлетворяющий условиям*

$$(M_{LL})^{-1} p_L \geq 0 \quad \forall L \subseteq \{1, \ldots, n\}; \tag{6.12}$$

здесь M_{LL} — главная подматрица, получаемая из M удалением строк и столбцов с номерами, не входящими в L; то же касается p_L. Тогда метод Лемке с p, взятым в качестве столбца при искусственной неизвестной, завершится получением решения $LCP(q, M)$ не более чем через $n+1$ итерацию.

Д о к а з а т е л ь с т в о. Рассмотрим очередную (после первой) итерацию метода Лемке. Пусть индексное множество L состоит из номеров базисных z-переменных и пусть индекс $t \notin L$ обозначает текущую небазисную пару (w_t, z_t), а K — дополнение $\{t\} \cup L$ до $\{1, \ldots, n\}$. Предположим, что неизвестная w_t только что покинула базис и потому следующая вводимая в базисе неизвестная — z_t. Преобразованная система уравнений имеет следующую структуру:

$$\begin{pmatrix} z_L \\ z_0 \\ w_K \end{pmatrix} = \begin{pmatrix} \square \\ \square \\ \square \end{pmatrix} + \begin{pmatrix} \square & \square & F & \square \\ \square & \square & f_0 & \square \\ \square & \square & \square & \square \end{pmatrix} \begin{pmatrix} w_L \\ w_t \\ z_t \\ z_K \end{pmatrix}.$$

В этом представлении нас интересует (и показан) только $\begin{pmatrix} z_L \\ z_0 \end{pmatrix}$-блок ведущего столбца

$$\begin{pmatrix} F \\ f_0 \end{pmatrix} = -\begin{pmatrix} M_{LL} & p_L \\ M_{tL} & p_t \end{pmatrix}^{-1} \begin{pmatrix} M_{Lt} \\ M_{tt} \end{pmatrix}.$$

Более аккуратные вычисления дают нам

$$F = \left(\frac{M_{tt} - M_{tL} M_{LL}^{-1} M_{Lt}}{p_t - M_{tL} M_{LL}^{-1} p_L} \right) M_{LL}^{-1} p_L - M_{LL}^{-1} M_{Lt}. \tag{6.13}$$

Здесь знаменатель отличен от нуля, в силу существования M_{LL}^{-1} и невырожденности базисной матрицы

$$B = \begin{pmatrix} M_{LL} & p_L \\ M_{tL} & p_t \end{pmatrix}.$$

По условию (6.12)

$$\begin{pmatrix} \bar{p}_L \\ \bar{p}_t \end{pmatrix} = \begin{pmatrix} M_{LL} & M_{Lt} \\ M_{tL} & M_{tt} \end{pmatrix}^{-1} \begin{pmatrix} p_L \\ p_t \end{pmatrix} \geq 0.$$

Более точно

$$\bar{p}_L = M_{LL}^{-1} \, p_L - \left(\frac{p_t - M_{tL} \, M_{LL}^{-1} \, p_L}{M_{tt} - M_{tL} \, M_{LL}^{-1} \, M_{Lt}} \right) M_{LL}^{-1} \, M_{Lt},$$

$$\bar{p}_L = (p_t - M_{tL} \, M_{LL}^{-1} \, p_L) \, / \, (M_{tt} - M_{tL} \, M_{LL}^{-1} \, M_{Lt}).$$

Снова невырожденность M гарантирует здесь неравенство нулю знаменателя.

Таким образом, (6.13) показывает, что интересующие нас z_L-элементы ведущего столбца равны

$$F = \bar{p}_L / \bar{p}_t \geq 0.$$

Следовательно, рост вводимой в базис неизвестной z_t не повлечет убывания z_L-неизвестных, уже находящихся в базисе, и очередное исключение может иметь место или в z_0-строке (в этом случае мы получаем решение исходной задачи), или в одной из w_K-строк (в этом случае множество L пополняется новым номером и все рассуждения можно повторить сначала). Завершение на луче невозможно, так как по крайней мере один элемент ведущего столбца отрицателен, а именно

$$f_0 = -1/\bar{p}_t < 0.$$

Подытожим сказанное. Метод начинает работу, когда неизвестные z и z_0 лежат вне базиса. Первая итерация состоит в вводе в базис неизвестной z_0 (в итоге получается первое допустимое базисное решение, почти удовлетворяющее условиям дополнительности). На каждой последующей итерации множество L, первоначально пустое, пополняется номером вводимой в базис z-неизвестной. Поскольку таких неизвестных n, решение будет получено не более чем через $n+1$ итерацию.

Разумеется, поиск подходящего вектора при искусственной переменной в методе Лемке сам может потребовать значительных вычислительных затрат.

§ 7. Матрицы с положительными главными минорами

Обратимся к другому важному матричному классу — классу матриц, все главные миноры которых положительны.

Лемма 7.1. *Пусть M — P-матрица. Тогда система неравенств*

$$Mz \leq 0, \quad z \geq 0, \tag{7.1}$$

имеет единственное решение $z = 0$.

Д о к а з а т е л ь с т в о. Утверждение очевидно при $n = 1$. Предполагая, что оно верно при размерности M, меньшей некоторого n, покажем, что оно верно и для n. Пусть $z = (z_1, \ldots, z_n)^\top$ удовлетворяет системе (7.1). Так как M — P-матрица, то M не вырождена и диагональные элементы обратной к ней матрицы $M^{-1} = (\beta_{ij})$ положительны. Поэтому каждый столбец матрицы M^{-1}, в том числе первый (обозначим его через b), имеет хотя бы одну положительную компоненту. Пусть

$$\theta = \min \left\{ \frac{z_i}{\beta_{i1}} \quad \text{по} \quad \text{всем } \beta_{i1} > 0 \right\}$$

и $i = k$ — номер, на котором этот минимум достигается. Тогда $\theta \geq 0$, $w = z - \theta b = (w_1, \ldots, w_n)^\top \geq 0$ и $w_k = 0$.

Заметим, что
$$Mw = Mz - \theta Mb = Mz - \theta e_1 \leq 0$$
(здесь $e_1 = (1, 0, \ldots, 0)^\top$ — орт пространства $\mathbb{R}^n$). Обозначим через $\bar{M}$ главную подматрицу, получаемую из M путем удаления k-й строки и столбца, а через $\bar{w}$ — $(n-1)$-вектор, получаемый из w удалением его k-й компоненты. Имеем
$$\bar{M}\,\bar{w} \leq 0, \quad \bar{w} \geq 0.$$
Поскольку $\bar{M}$ есть $((n-1)\times(n-1))$-матрица и все ее главные миноры положительны (совпадают с таковым у M), то по предположению индукции $\bar{w} = 0$. Это вместе с равенством $w_k = 0$ дает $w = 0$, так что $Mz = \theta e_1 \geq 0$. Следовательно, $Mz = 0$, и, ввиду невырожденности M, имеем $z = 0$, что и требовалось.

Следствие 7.1. *Если M — P-матрица, то система*
$$Mz > 0, \quad z > 0, \tag{7.2}$$
совместна.

Д о к а з а т е л ь с т в о. В силу однородности системы (7.2) ее совместность эквивалентна совместности системы
$$Mz \geq e, \quad z \geq e, \text{ где } e = (1, 1, \ldots, 1)^\top.$$
Последняя же совместна тогда и только тогда, когда система
$$M^\top w \leq 0, \quad w \geq 0, \tag{7.3}$$
не имеет отличных от нуля решений (следует из двойственности в линейном программировании). Но по только что доказанной лемме система (7.3) действительно не имеет ненулевых решений, так как матрица $M^\top$ принадлежит классу P вместе с матрицей M.

Доказанная лемма и ее следствие позволяют установить конечность метода Лемке для P-матриц.

Теорема 7.1. *Если все главные миноры матрицы M положительны, то метод Лемке при любом q завершается на допустимом базисном решении* (5.1), *удовлетворяющем условиям дополнительности, т. е. дает решение $LCP(q, M)$.*

Д о к а з а т е л ь с т в о. Мы уже видели, что завершение на луче подразумевает существование ненулевого вектора $\bar{z} \geq 0$, удовлетворяющего неравенствам (6.1). Однако по только что доказанной лемме в случае P-матриц это невозможно. Значит, для P-матриц невозможно и завершение метода Лемке на луче.

Приведем еще одну важную характеризацию P-матриц, для чего нам потребуется два вспомогательных утверждения.

Пусть векторы z, w связаны соотношением $w = Mz$. Будем говорить, что M *меняет знак* z, если $z_i w_i \leq 0$ для всех $i = 1, \ldots, u$.

Лемма 7.2. *Матрица M принадлежит классу P в том и только в том случае, когда она не меняет знак ни у одного вектора, кроме нулевого.*

Д о к а з а т е л ь с т в о. Прежде всего, доказывая необходимость, заметим, что можно ограничиться случаем $z \geq 0$. В самом деле, если $L = \{i: \ z_i < 0\} \neq \emptyset$, определим D — диагональную матрицу, получаемую из единичной матрицы сменой знака у тех ее строк, номера которых входят в L. Матрица $M' = DMD$ снова оказывается P-матрицей, поскольку мы просто сменили знаки у строк и столбцов M, номера которых входят в L. Кроме того, матрица M' меняет знак у вектора $Dz \geq 0$.

Итак, пусть $0 \neq z \geq 0$ и M меняет знак z. Определим множество $K = \{i: \ z_i > 0\}$. Предполагая, что $K \neq \emptyset$, выделим $\bar{M}$ — главную подматрицу M, получаемую из нее удалением тех строк и столбцов, номера которых не входят в K. Пусть также $\bar{z}$ — вектор, получаемый из z аналогичным образом. Матрица $\bar{M}$ — также P-матрица и меняет знак $\bar{z}$. Поскольку все $\bar{z}_i > 0$, то должно выполняться неравенство $\bar{M}\,\bar{z} \leq 0$. В силу леммы 7.1 это отвергает гипотезу о принадлежности M классу P. Необходимость доказана.

Перейдем к обоснованию достаточности. Пусть $\bar{M} = (a_{ij})$ — произвольная главная подматрица матрицы M, K — множество номеров входящих в нее строк и столбцов. Если ее определитель не положителен, $\bar{M}$ должна иметь хотя бы одно неположительное вещественное собственное значение и соответствующий ему ненулевой вещественный собственный вектор $\bar{z}$ (это следует из того, что определитель матрицы есть произведение всех ее собственных значений, причем комплексные значения образуют сопряженные пары). Если теперь дополнить вектор $\bar{z}$ до размерности n нулями в позициях вне K, мы получим вектор z, знак которого меняется над действием матрицы M. Полученное противоречие доказывает достаточность.

Напомним, что *элементарным преобразованием* системы линейных уравнений называют разрешение ее относительно одной из своих неизвестных (т. е. операцию исключения этой неизвестной из всех уравнений, кроме некоторого). Поскольку элементарное преобразование не меняет множества решений системы, простым следствием доказанной леммы является следующий факт.

Следствие 7.2. *Пусть имеется система уравнений*

$$w = q + Mz,$$

где M — P-матрица. Тогда элементарное преобразование этой системы, меняющее ролями переменные любой дополнительной пары (z_i, w_i), снова приводит к P-матрице.

Такое преобразование возможно, так как все $m_{ii} > 0 (\neq 0)$.

Теорема 7.2. *Линейная задача о дополнительности $LCP(q, M)$ имеет единственное решение при всех q тогда и только тогда, когда M — P-матрица.*

Д о к а з а т е л ь с т в о. Пусть M — P-матрица. По теореме 7.1 задача $LCP(q, M)$ разрешима при всех q, причем ее решение является некоторым допустимым базисным решением системы ее уравнений. Переходя в случае необходимости к преобразованной задаче, матрица которой также P-матрица, можно, не ограничивая общности, считать, что $q \geq 0$ и решение, получаемое методом Лемке, имеет вид $(q; 0)$ (т. е. $z = 0$, $w = q \geq 0$). Пусть оно не единственно и $(\bar{w}, \bar{z})$ — другое решение. Тогда не трудно видеть, что M меняет знак $\bar{z}$, т. е. по лемме 7.2 не может принадлежать P.

Обратно, пусть $LCP(q, M)$ имеет единственное решение при всех q, но M — не P-матрица. Тогда по лемме 7.2 существует ненулевой вектор $\bar{z}$ такой, что все $\bar{z}_i \cdot \bar{w}_i \leq 0$, где $\bar{w} = M\bar{z}$.

Определим $\bar{q} = (\bar{w})^+ - M(\bar{z})^+$, где плюс над вектором означает его положительную срезку, т. е. замену отрицательных координат нулями. Поскольку $\bar{z}_i^+ \cdot \bar{w}_i^+ = 0$, пара $(\bar{w}^+, \bar{z}^+) \geq 0$ является, очевидно, решением $LCP(\bar{q}, M)$. Другим очевидным решением $LCP(\bar{q}, M)$ будет пара $(\bar{w}^-, \bar{z}^-) \geq 0$, где минус над вектором означает положительную срезку противоположного ему вектора. При проверке всех условий опираемся на равенства $\bar{w} = M\bar{z}$, $\bar{w} = \bar{w}^+ - \bar{w}^-$, $\bar{z} = \bar{z}^+ - \bar{z}^-$, $\bar{z}_i^- \cdot \bar{w}_i^- = 0$.

В общем случае линейная задача о дополнительности не всегда разрешима, даже если ее ограничения совместны.

Упражнения

1. Задача $LCP(q, M)$ имеет единственное решение для всех $q > 0$ тогда и только тогда, когда матрица M полумонотонна. Докажите.

2. Убедитесь, что для задачи $LCP(q, M)$, где

$$q = \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \; M = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix},$$

метод Лемке завершается на луче, хотя задача имеет решение (какое?).

3. Проверьте, что при использовании стандартного выбора ведущей строки метод Лемке на задаче $LCP(q, M)$, где

$$q = \begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix}, \; M = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 2 \\ 2 & 0 & 1 \end{pmatrix},$$

зацикливается.

4. Докажите, что для задач с полумонотонными матрицами значение искусственной переменной в методе Лемке никогда не превышает своего начального значения, а для задач с P_0-матрицами оно или остается неизменным, или убывает.

5. Разберите ход вычислений в методе Лемке с искусственной переменной применительно к паре двойственных задач линейного программирования.

6. Дайте содержательную интерпретацию метода Лемке в применении к постановкам, возникающим в квадратичном программировании.

7. Матрица $M = (m_{ij})_{n \times n}$ называется *Z-матрицей,* если все ее внедиагональные элементы неположительны. Покажите, что всякая допустимая линейная задача о дополнительности с Z-матрицей разрешима.

8. Используя Бэйсик, Паскаль или другой язык программирования, опишите алгоритм пересчета таблиц в методе Лемке.

Глава 2

НЕЛИНЕЙНЫЕ ЗАДАЧИ О ДОПОЛНИТЕЛЬНОСТИ И ВАРИАЦИОННЫЕ НЕРАВЕНСТВА

Нелинейные задачи о дополнительности и вариационные неравенства являются обобщением для многих оптимизационных постановок, таких, как задачи математического (нелинейного) программирования, минимаксные задачи и задачи о седловой точке выпукло-вогнутых функций, задачи поиска равновесия в играх n лиц и др. Многие развиваемые для их решения итерационные методы могут быть с успехом применены и к линейным задачам о дополнительности.

§ 1. Постановка задач и их взаимосвязь

Пусть X — непустое подмножество пространства $\mathbb{R}^n$ и F — некоторое отображение X в $\mathbb{R}^n$. Задача решения *вариационного неравенства*, обозначаемая в дальнейшем $VI(X,F)$, состоит в отыскании вектора $\hat{x} \in X$ такого, что

$$F(\hat{x})^\top (x - \hat{x}) \geq 0 \ \text{ при всех } x \in X. \tag{1.1}$$

Обычно полагают, что множество X замкнуто и выпукло; в приложениях оно часто оказывается полиэдральным.

Геометрически условие (1.1) требует, чтобы вектор $F(\hat{x})$ составлял острый угол со всеми допустимыми векторами – направлениями, исходящими из $\hat{x}$. Более строго $\hat{x} \in X$ есть решение $VI(X,F)$ в том и только в том случае, когда $-F(\hat{x})$ лежит в нормальном (к множеству X в точке x) конусе $N_X(\hat{x})$, где

$$N_X(x) = \begin{cases} \{z \in R^n : z^\top (y - x) \leq 0 \ \ \forall y \in X\}, & \text{если } x \in X; \\ \emptyset \text{ — } & \text{в противном случае.} \end{cases}$$

Поэтому любое решение $\hat{x}$ задачи $VI(X,F)$ будет одновременно и решением задачи

$$\min\{\, F(\hat{x})^\top x : x \in X \,\}, \tag{1.2}$$

и наоборот, если $\hat{x} \in \operatorname{Arg}\,(1.2)$, то $\hat{x} \in \operatorname{Arg}\, VI(X,F)$.

Простейший пример задачи решения вариационного неравенства дает обычная система нелинейных уравнений

$$F(x) = 0,$$

где $F(x)$ — отображение $\mathbb{R}^n$ в себя. В самом деле, при $X = \mathbb{R}^n$ вектор $\hat{x}$ будет решением $VI(X, F)$, только если $F(\hat{x}) = 0$, так как нуль-вектор — единственный вектор, составляющий острый угол сразу со всеми векторами из $\mathbb{R}^n$.

Другой наглядный пример задачи решения вариационного неравенства дает дифференцируемая оптимизация. Если $F(x) = \nabla f(x)$ — градиент вещественной дифференцируемой функции $f : \mathbb{R}^n \to \mathbb{R}$ и множество X выпукло, то $VI(X, F)$ представляет собой просто перезапись необходимых условий оптимальности первого порядка для задачи

$$\min\{ f(x) : x \in X \}. \tag{1.3}$$

В самом деле, если $\hat{x}$ — решение задачи $VI(X, \nabla f)$, то в задаче (1.3) не существует допустимого направления спуска из $\hat{x}$. Если же $f(x)$ — псевдовыпукла (т. е. неравенство $\nabla f(x)^\top (y - x) \geq 0$ влечет за собой неравенство $f(y) \geq f(x)$ при всех $x, y \in X$), то любое решение $VI(X, F)$ будет также точкой глобального минимума в (1.3).

Третий пример связан с поиском седловых точек выпукло-вогнутых функций (и теми задачами из теории игр и математического программирования, которые сводятся к такому поиску). Пусть X и Y — непустые замкнутые выпуклые подмножества $\mathbb{R}^n$ и $\mathbb{R}^m$ соответственно, $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ — дифференцируемая числовая функция двух векторных аргументов. Седловой точкой функции $f(x, y)$ относительно области $X \times Y$ называется пара $(\hat{x}, \hat{y}) \in X \times Y$, удовлетворяющая неравенствам

$$f(\hat{x}, y) \leq f(\hat{x}, \hat{y}) \leq f(x, \hat{y}) \quad \forall x \in X, \ \forall y \in Y.$$

Как известно, необходимые условия того, что точка $(\hat{x}, \hat{y})$ является седловой, имеют вид

$$\begin{aligned} &\hat{x} \in X, \ \nabla_x f\,(\hat{x}, \hat{y})^\top (x - \hat{x}) \geq 0 \ \forall x \in X, \\ &\hat{y} \in Y, \ \ \nabla_y f\,(\hat{x}, \hat{y})^\top (y - \hat{y}) \leq 0 \ \ \forall y \in Y. \end{aligned} \tag{1.4}$$

Если функция $f(x, y)$ выпукла по x при всех фиксированных $y \in Y$ и вогнута по y при всех фиксированных $x \in X$, условия (1.4) носят также достаточный характер. Их можно переписать в виде вариационного неравенства $VI(X \times Y, F)$, где $F = (\nabla_x f, -\nabla_y f)$.

Известно, что непрерывно дифференцируемое отображение F является градиентом некоторой функции $f : \mathbb{R}^n \to R$ тогда и только тогда, когда его якобиан ∇F симметричен при всех $x \in \mathbb{R}^n$. В этом случае

$$f(x) = \int_0^1 F(\bar{x} + t(x - \bar{x}))^\top (x - \bar{x}) dt,$$

где $\bar{x}$ — произвольный фиксированный вектор из $\mathbb{R}^n$. Таким образом, симметричность ∇F является тем ключом, который позволяет переформулировать $VI(X, F)$ как обычную оптимизационную задачу.

В общем случае указанная выше переформулировка является редким исключением, пример которого дают задачи математического программирования и связанные с ними задачи поиска седловой точки ассоциированных с ними функций Лагранжа с необязательно симметричными матрицами вторых производных.

Важным частным случаем $VI(X, F)$ является *нелинейная задача о дополнительности*.

Пусть $F(x)$ — отображение $\mathbb{R}^n_+$ в $\mathbb{R}^n$. Нелинейная задача о дополнительности, обозначаемая $NCP(F)$, состоит в отыскании вектора $\hat{x} \in \mathbb{R}^n$ такого, что

$$\hat{x} \geq 0, \quad F(\hat{x}) \geq 0, \quad F(\hat{x})^\top \hat{x} = 0. \tag{1.5}$$

Если отображение F аффинно, т. е. $F(x) = q + Mx$ для некоторого $q \in \mathbb{R}^n$ и матрицы $M \in \mathbb{R}^{n \times n}$, задача (1.5) превращается в *линейную* задачу о дополнительности $LCP(q, M)$.

Геометрически нелинейная задача о дополнительности состоит в поиске неотрицательного вектора $\hat{x}$ такого, что образ $F(\hat{x})$ также неотрицателен и ортогонален $\hat{x}$.

Заменяя $\mathbb{R}^n_+$ на произвольный выпуклый конус X, приходим к *обобщенной задаче о дополнительности*. Последняя заключается в поиске $\hat{x} \in \mathbb{R}^n$ такого, что

$$\hat{x} \in X, \quad F(\hat{x}) \in X^*, \quad F(\hat{x})^\top \hat{x} = 0; \tag{1.6}$$

здесь X^* обозначает конус, двойственный к X, т. е.

$$X^* = \{y \in \mathbb{R}^n : y^\top x \geq 0 \quad \forall x \in X\}.$$

Эту задачу будем обозначать $GCP(X, F)$.

Между обобщенной (а значит, и нелинейной) задачей о дополнительности и задачей решения вариационного неравенства существует следующая простая связь.

Теорема 1.1. *Пусть X — выпуклый конус. Тогда множества решений задач $VI(X, F)$ и $GCP(X, F)$ совпадают*

Д о к а з а т е л ь с т в о. 1) Пусть $\hat{x}$ — решение $VI(X, F)$, т. е. $\hat{x} \in X$ и

$$F(\hat{x})^\top (x - \hat{x}) \geq 0 \quad \forall x \in X. \tag{1.7}$$

Покажем, что $F(\hat{x}) \in X^*$ и $F(\hat{x})^\top \hat{x} = 0$. Для этого заметим, что конус X вместе с любой своей точкой x содержит и все точки вида $tx, t \geq 0$. Подставляя в (1.7) вместо x вектор tx, получаем

$$F(\hat{x})^\top x \geq t^{-1} F(\hat{x})^\top \hat{x} \ \ \forall x \in X, \ t > 0.$$

Переходя здесь к пределу по $t \to \infty$, имеем $F(\hat{x})^\top x \geq 0 \ \ \forall x \in X$, т. е. $F(\hat{x}) \in X^*$. Чтобы убедиться в равенстве $F(\hat{x})^\top \hat{x} = 0$, достаточно в последнее неравенство подставить $x = \hat{x}$, а в неравенство (1.7) — точку $x = 0$.

2) Обратно, пусть $\hat{x}$ — решение $GCP\,(X, F)$, т. е. $\hat{x} \in X$, $F(\hat{x}) \in X^*$ и $F(\hat{x})^\top \hat{x} = 0$. Тогда

$$F(\hat{x})^\top (x - \hat{x}) = F(\hat{x})^\top x - F(\hat{x})^\top \hat{x} \geq 0 \quad \forall x \in X,$$

т. е. $\hat{x}$ — решение и $VI(X, F)$.

Из теоремы 1.1 следует, что любая обобщенная (и, в частности, нелинейная) задача о дополнительности есть задача решения некоторого вариационного неравенства.

Обратное в общем случае не верно. Тем не менее в случае, когда X определяется как множество решений некоторой системы неравенств, т. е.

$$X = \{x \in \mathbb{R}^n : g_i(x) \leq 0, \quad i = 1, \ldots, m\}, \tag{1.8}$$

задача $VI(X, F)$ может быть конвертирована в некоторую обобщенную задачу о дополнительности.

Теорема 1.2. *Пусть* X *определено соотношением* (1.8), *где все функции* $g_i(x)$ *непрерывно дифференцируемы. Тогда при выполнении некоторых стандартных условий регулярности любое решение* $\hat{x}$ *задачи* $VI(X,F)$ *можно дополнить вектором* $\hat{\pi} \in \mathbb{R}^m$ *до решения* $(\hat{x};\hat{\pi})$ *задачи* $GCP\,(\mathbb{R}^n \times \mathbb{R}^m_+; H)$, *где отображение* H : $\mathbb{R}^{n+m} \to \mathbb{R}^{n+m}$ *задается формулой*

$$H(x,\pi) = \begin{pmatrix} F(x) + \sum\limits_{i=1}^{m} \pi_i \nabla g_i(x) \\ -g(x) \end{pmatrix}.$$

Обратно, если все $g_i(x)$ *выпуклы и пара* $(\hat{x};\hat{\pi})$ *есть решение задачи* $GCP\,(\mathbb{R}^n \times \mathbb{R}^m_+; H)$, *то* $\hat{x}$ — *решение* $VI(X,F)$.

Д о к а з а т е л ь с т в о опирается на уже отмечавшийся простой факт: любое решение $\hat{x}$ задачи $VI(X,F)$ будет одновременно и решением задачи

$$\min\{\, F(\hat{x})^\top x : \; g_i(x) \le 0, \quad i = 1, \ldots, m\}, \tag{1.9}$$

и наоборот, если $\hat{x}$ — решение задачи (1.9), то $\hat{x} \in \operatorname{Arg}\, VI(X,F)$, где X определено соотношением (1.8). Поэтому утверждение теоремы есть просто другая формулировка необходимых условий Куна—Таккера оптимальности точки $\hat{x}$ в задаче (1.9):

$$\begin{aligned} &-F(\hat{x}) = \sum_{i=1}^{m} \hat{\pi}_i \nabla g_i(\hat{x}), \\ &\hat{\pi} = (\hat{\pi}_1, \ldots, \hat{\pi}_m) \ge 0, \; -g(\hat{x}) = (-g_1(\hat{x}), \ldots, -g_m(\hat{x})) \ge 0, \\ &g(\hat{x})^\top \hat{\pi} = 0. \end{aligned}$$

В случае выпуклости функций $g_i(x)$, $i = 1, \ldots, m$, эти условия приобретают также и достаточный характер.

Перечислим несколько простых условий регулярности, обеспечивающих справедливость теоремы Куна—Таккера:

1) все функции $g_i(x)$ аффинны;

2) градиенты активных ограничений $\nabla g_i(x)$ в точке $\hat{x}$ линейно независимы;

3) все функции $g_i(x)$ выпуклы, и существует точка $\bar{x}$ такая, что все $g_i(\bar{x}) < 0$ (точка Слейтера).

Обобщенная задача о дополнительности из теоремы 1.2 весьма специфична: только часть ее переменных неотрицательна, а компоненты H-отображения, соответствующие остальным переменным, должны равняться нулю. Такие задачи называют *смешанными* нелинейными задачами о дополнительности.

Заметим, что все предположения теоремы 1.2 относятся только к функциям $g_i(x)$; никаких условий на отображение F не было наложено. Недостатком рассмотренного приведения $VI(X,F)$ к задаче о дополнительности является рост числа неизвестных с n до $n+m$. С вычислительной точки зрения это может привести к росту затрат на нахождение искомого решения. Тем не менее такое приведение дает полезный инструмент анализа $VI(X,F)$.

Для вывода ряда теорем существования решения $VI(X,F)$ и развития многих итерационных методов их отыскания полезно переформулировать задачу решения вариационного неравенства как классическую задачу о неподвижной точке. Для этого напомним понятие проекции точки на непустое выпуклое замкнутое множество.

Пусть X — непустое выпуклое замкнутое подмножество $\mathbb{R}^n$. Под *проекцией* точки $y \in \mathbb{R}^n$ на множество X, обозначаемой в дальнейшем $P_X(y)$, понимают решение (оптимальный вектор) задачи

$$\min_{x \in X} (x-y)^\top (x-y).$$

Нетрудно убедиться, что неравенство

$$(P_X(x)-x)^\top (y-P_X(x)) \geq 0 \quad \forall y \in X \tag{1.10}$$

является характеристическим свойством введенной таким образом операции проектирования. Из него, в частности, вытекает, что отображение $P_X(\cdot)$ — не расширяющее, т. е.

$$\|P_X(x)-P_X(y)\| \leq \|x-y\| \qquad \forall x, y \in \mathbb{R}^n.$$

Кроме того, верно включение

$$x - P_X(x) \in N_X(P_X(x)).$$

Опираясь на понятие проекции, легко установить следующий результат.

Теорема 1.3. *Пусть X — непустое выпуклое замкнутое подмножество $\mathbb{R}^n$, $\alpha > 0$. Тогда любое решение $\hat{x}$ задачи $VI(X,F)$ удовлетворяет условию*

$$\hat{x} = P_X(\hat{x} - \alpha F(\hat{x})), \tag{1.11}$$

и обратно, любая точка $\hat{x}$, удовлетворяющая условию (1.11), *есть решение $VI(X,F)$.*

Несмотря на свою простоту, это утверждение носит важный характер.

Следствие 1.1. *Множество решений $VI(X,F)$ совпадает с множеством неподвижных точек отображения $H : \mathbb{R}^n \to \mathbb{R}^n$, определяемого формулой*

$$H(x) = P_X(x - \alpha F(x)), \quad \alpha > 0. \tag{1.12}$$

В случае нелинейной задачи о дополнительности $NCP\,(F)$, т. е. при $X = \mathbb{R}^n_+$, отображение (1.12) принимает вид

$$H(x) = (x - \alpha F(x))^+,$$

где + над вектором означает замену его отрицательных координат нулями (операция положительной срезки).

Вернемся к обсуждению вопроса о сведении задач о дополнительности и решении вариационного неравенства к обычным задачам оптимизации. Как уже отмечалось для последних, предположение о симметричности ∇F играет критическую роль. Однако для нелинейных задач о дополнительности имеет место следующий простой факт.

Теорема 1.4. *Пусть F — некоторое отображение $\mathbb{R}^n$ в себя. Вектор $\hat{x} \in \mathbb{R}^n$ решает $NCP(F)$ тогда и только тогда, когда он оптимален для задачи*

$$\min\{\, F(x)^\top x : \; F(x) \geq 0, \; x \geq 0 \,\} \tag{1.13}$$

и оптимальное значение последней равно нулю.

Вообще говоря, допустимая область задачи (1.13) необязательно выпукла. Однако в случае аффинности F (т. е. в случае линейной задачи о дополнительности) постановка (1.13) есть просто задача квадратичного программирования (необязательно выпуклого).

Для представления $VI(X,F)$ в виде оптимизационной задачи большую роль играет так называемая *функция скачка*.

Функция скачка (или разрыва) $g(x)$, ассоциированная с $VI(X,F)$, определяется формулой

$$g(x) = \max_{y \in X} F(x)^\top (x - y). \tag{1.14}$$

Заметим, что $g(x) \geq 0$ при всех $x \in X$, причем равенство имеет место только на решениях $VI(X,F)$. Таким образом, верна

Теорема 1.5. *Вектор $\hat{x} \in X$ есть решение $VI(X,F)$, если и только если он оптимален для задачи*

$$\min_{x \in X} g(x) = \min_{x \in X} \max_{y \in X} F(x)^\top (x - y) \tag{1.15}$$

и $g(\hat{x}) = 0$.

Нетрудно видеть, что в случае нелинейной задачи о дополнительности задача (1.15) сводится к (1.13).

Итак, функция скачка позволяет сформулировать $VI(X,F)$ как некоторую минимаксную задачу. К сожалению, подобно большинству задач такого рода, последняя, вообще говоря, не является гладкой, что затрудняет применение к ней численных методов решения.

Другие, более привлекательные в вычислительном отношении функции скачка будут рассмотрены в конце данной главы.

Упражнение

Введем в рассмотрение *двойственную* задачу решения вариационного неравенства $DVI(X,F)$, состоящую в поиске $\hat{y} \in X$ такого, что

$$F(x)^\top (x - \hat{y}) \geq 0 \quad \forall x \in X. \tag{1.16}$$

Для этой задачи можно определить свою функцию скачка $f(y)$ по формуле

$$f(y) = \min_{x \in X} F(x)^\top (x - y).$$

Покажите, что вектор $\hat{y} \in X$ есть решение $DVI(X,F)$, если и только если он оптимален для задачи

$$\max_{y \in X} f(y) = \max_{y \in X} \min_{x \in X} F(x)^\top (x - y) \tag{1.17}$$

и $f(\hat{y}) = 0$.

Покажите также, что вектор $\hat{x} \in X$ есть решение задачи $VI(X,F)$, а вектор $\hat{y} \in X$ — задачи $DVI(X,F)$ в том и только в том случае, когда

$$F(\hat{x})^\top (\hat{x} - \hat{y}) = \min_{x \in X} \max_{y \in X} F(x)^\top (x - y) = \max_{y \in X} \min_{x \in X} F(x)^\top (x - y),$$

т. е. когда пара $(\hat{x}, \hat{y})$ образует седловую точку функции

$$\psi(x, y) = F(x)^\top (x - y)$$

относительно декартова произведения $X \times X$. Воспользуйтесь тем, что значения минимакса и максимина в задачах (1.15), (1.17) в общем случае различаются знаком.

§ 2. Теоремы существования и единственности решения

Основной результат о существовании решений вариационного неравенства $VI(X, F)$ требует компактности и выпуклости множества X и непрерывности отображения F. Ряд следствий может быть получен из него путем замены требования компактности X на дополнительные условия относительно F. Сам результат получается применением к непрерывному H-отображению из (1.12) теоремы Брауэра о неподвижной точке. Он формулируется следующим образом.

Теорема 2.1. *Пусть X — непустое компактное и выпуклое подмножество $\mathbb{R}^n$, F — непрерывное отображение X в $\mathbb{R}^n$. Тогда задача $VI(X, F)$ разрешима.*

В случае неограниченности X (как это имеет место, например, в нелинейной задаче о дополнительности) можно опираться на условия, заведомо обеспечивающие ограниченность самого множества потенциальных решений для того, чтобы гарантировать наличие хотя бы одного из них. Это достигается различными путями. Например, можно пытаться указать внутри X такое ограниченное множество D, вне которого ни одна точка не может быть искомым решением. Строгий результат выглядит так.

Теорема 2.2. *Пусть X — непустое замкнутое и выпуклое подмножество $\mathbb{R}^n$, F — непрерывное отображение X в $\mathbb{R}^n$. Если существует такое непустое ограниченное подмножество D множества X, что*

$$\forall x \in X \backslash D \quad \exists y \in D: \; F(x)^\top (x - y) > 0, \tag{2.1}$$

то задача $VI(X, F)$ разрешима.

Доказательство. Пусть $B = \{x \in \mathbb{R}^n : \; \|x\| \leq 1\}$ — единичный шар в $\mathbb{R}^n$. По предыдущей теореме разрешима задача $VI(\bar{X}, F)$, где $\bar{X}$ — замыкание множества $(D{+}B){\cap}X$ — компакт. В силу (2.1), ее решение $\hat{x}$ лежит в D. Покажем, что оно является и решением $VI(X, F)$.

В самом деле, по построению

$$F(\hat{x})^\top (x - \hat{x}) \geq 0 \quad \forall x \in \bar{X}.$$

Пусть теперь $x \in X \backslash \bar{X}$. В этом случае $\alpha = \|x - \hat{x}\| > 1$ и вектор $y = \hat{x} + \alpha^{-1}(x - \hat{x}) \in \bar{X}$. Поэтому также

$$F(\hat{x})^\top (x - \hat{x}) = \alpha F(\hat{x})^\top (y - \hat{x}) \geq 0,$$

что и требовалось.

В общем случае указать заранее нужное множество D трудно. Однако, если отображение F обладает некоторыми дополнительными свойствами, то такое указание делается возможным. Введем ряд определений.

Отображение $F: \; \mathbb{R}^n \to \mathbb{R}^n$ называется

а) *монотонным* на X, если

$$(F(x) - F(y))^\top (x - y) \geq 0 \quad \forall x, \, y \in X;$$

б) *псевдомонотонным* на X, если

$$F(y)^\top (x - y) \geq 0 \quad \text{влечет} \quad F(x)^\top (x - y) \geq 0 \;\; \forall x, \, y \in X;$$

в) *строго монотонным* на X, если

$$(F(x) - F(y))^\top (x - y) > 0 \quad \forall x,\, y \in X, \quad x \neq y;$$

г) *сильно монотонным* на X, если найдется такая константа $\alpha > 0$, что

$$(F(x) - F(y))^\top (x - y) \geq \alpha \|x - y\|^2 \quad \forall x,\, y \in X;$$

д) *коэрцитивным* по отношению к X, если найдется вектор $\bar{x} \in X$ такой, что

$$\lim_{x \in X,\, \|x\| \to \infty} F(x)^\top (x - \bar{x}) / \|x\| = \infty.$$

Заметим, что если отображение F сильно монотонно на X, то оно коэрцитивно относительно X. Если $F(x) = q + Mx$ и $X = \mathbb{R}^n$, то F коэрцитивно только в случае своей сильной монотонности, что, в свою очередь, эквивалентно положительной определенности матрицы M. В общем случае, если отображение F непрерывно дифференцируемо, его различные свойства монотонности связаны с положительной полуопределенностью и положительной определенностью матриц ∇F. Среди перечисленных свойств наиболее слабым является, очевидно, свойство псевдомонотонности, затем следуют в порядке усиления свойства монотонности, строгой монотонности и сильной монотонности.

Хорошо известно, что оптимальное множество задачи выпуклого программирования выпукло. Следующее утверждение распространяет этот результат на вариационные неравенства псевдомонотонного типа.

Теорема 2.3. *Пусть X — непустое замкнутое выпуклое подмножество $\mathbb{R}^n$ и F — непрерывное псевдомонотонное отображение X в $\mathbb{R}^n$. Точка $\hat{x}$ решает задачу $VI(X, F)$ в том и только в том случае, когда $\hat{x} \in X$ и*

$$F(y)^\top (y - \hat{x}) \geq 0 \quad \forall y \in X. \tag{2.2}$$

В частности, множество решений задачи $VI(X, F)$ выпукло (если не пусто).

Д о к а з а т е л ь с т в о. 1) Пусть $\hat{x}$ — решение $VI(X, F)$, т. е. $\hat{x} \in X$ и

$$F(\hat{x})^\top (y - \hat{x}) \geq 0 \quad \forall y \in X.$$

Эти соотношения, в силу псевдомонотонности отображения F, влекут искомое

$$F(y)^\top (y - \hat{x}) \geq 0 \quad \forall y \in X.$$

2) Обратно, пусть имеет место (2.2). Зафиксируем произвольно вектор $y \in X$ и рассмотрим точки вида $y_\alpha = \alpha \hat{x} + (1 - \alpha) y$, $\alpha \in [0,\, 1]$. Эти точки лежат в X, в силу выпуклости последнего. Отсюда и из (2.2) следует, что

$$F(y_\alpha)^\top (y - \hat{x}) = \frac{1}{1 - \alpha} F(y_\alpha)^\top (y_\alpha - \hat{x}) \geq 0$$

при всех $\alpha \in [0,\, 1]$. Переходя в левой части к пределу по $\alpha \to 1 - 0$, в силу непрерывности F получаем искомое

$$F(\hat{x})^\top (y - \hat{x}) \geq 0 \quad \forall y \in X.$$

Обе части теоремы доказаны.

В общем случае задача $VI(X, F)$ может иметь более одного решения. Если, однако, отображение F строго монотонно, такое решение единственно (если вообще существует).

Теорема 2.4. *Если отображение F строго монотонно на X, то задача $VI(X,F)$ не может иметь более одного решения.*

Д о к а з а т е л ь с т в о. Пусть x и y — два различных решения задачи $VI(X,\, F)$. Тогда

$$F(x)^\top(y-x) \geq 0 \quad \text{и} \quad F(y)^\top(x-y) \geq 0.$$

Складывая эти два неравенства, получаем

$$(F(x)-F(y))^\top(y-x) \geq 0,$$

что противоречит предположению о строгой монотонности F.

Ни одно из приведенных утверждений 2.3, 2.4 не гарантирует еще существования решения задачи $VI(X,F)$. Общим приемом при выводе такого существования является опора на свойство коэрцитивности отображения F. Это свойство позволяет выбрать в качестве компакта D, требуемого в теореме 2.2, просто шар достаточно большого диаметра, пересекающийся с X. Таким образом, существование решения будет следовать из этой теоремы, и, более того, легко показать, что множество решений будет компактным.

Теорема 2.5. *Пусть X — непустое замкнутое выпуклое подмножество $\mathbb{R}^n$ и F — непрерывное отображение X в $\mathbb{R}^n$. Если F коэрцитивно относительно X, то задача $VI(X,F)$ имеет непустое компактное множество решений.*

Следующий результат показывает, что условие коэрцитивности можно несколько ослабить.

Теорема 2.6. *Пусть X — непустое замкнутое выпуклое подмножество $\mathbb{R}^n$ и F — непрерывное отображение X в $\mathbb{R}^n$. Если существует вектор $\bar{x} \in X$ такой, что множество*

$$D^0 = \{x \in X: \ F(x)^\top(x-\bar{x}) \leq 0\}$$

ограничено, то множество решений $VI(X,F)$ не пусто и компактно.

Д о к а з а т е л ь с т в о. Достаточно сослаться на теорему 2.2, взяв в качестве множества D множество D^0, а в качестве y — вектор $\bar{x}$.

Поскольку сильная монотонность отображения влечет как коэрцитивность, так и строгую его монотонность, то, комбинируя теоремы 2.4 и 2.5, можно получить следующий результат.

Следствие 2.1. *Пусть X — непустое замкнутое выпуклое подмножество $\mathbb{R}^n$ и F — непрерывное отображение X в $\mathbb{R}^n$. Если F сильно монотонно на X, то решение задачи $VI(X,F)$ существует и единственно.*

В случае, когда отображение F монотонно или псевдомонотонно, задача $VI(X,F)$ может не иметь решения. Однако псевдомонотонности F хватает, если дополнительно для X выполняются условия регулярности (типа Слейтера). Их точная формулировка дана ниже и опирается на понятие двойственного конуса к произвольному множеству X, который определяется точно так же, как и для выпуклого конуса.

Теорема 2.7. *Пусть X — непустое выпуклое замкнутое подмножество $\mathbb{R}^n$ и F — непрерывное отображение X в $\mathbb{R}^n$. Предположим, что F псевдомонотонно на X и существует вектор $\bar{x} \in X$ такой, что $F(\bar{x}) \in \operatorname{int}\,(X^*)$, где $\operatorname{int}\,(\cdot)$ означает внутренность множества. Тогда задача $VI(X,F)$ имеет непустое компактное выпуклое множество решений.*

Д о к а з а т е л ь с т в о. Снова воспользуемся теоремой 2.2. В качестве D возьмем множество

$$D^1 = \{x \in X : \; F(\bar{x})^\top (x - \bar{x}) \le 0\}. \tag{2.3}$$

Это множество не пусто, поскольку содержит $\bar{x}$, и ограничено, поскольку $F(\bar{x}) \in \operatorname{int}\,(X^*)$, т. е. $\exists \delta > 0$ такое, что

$$(F(\bar{x}) - \delta \frac{x}{\|x\|})^\top x \ge 0 \quad \forall x \in X,$$

что дает после преобразования $F(\bar{x})^\top x \ge \delta \|x\| \quad \forall x \in X$. Выполнимость остальных требований теоремы 2.2 к множеству D^1 следует из (2.3) и псевдомонотонности F.

В случае нелинейной задачи о дополнительности приведенные выше результаты можно усилить, благодаря специфике постановки последней.

Теорема 2.8. *Пусть отображение* $F : \mathbb{R}^n_+ \to \mathbb{R}^n_+$ *непрерывно и*

$$F(x)^\top x \ge 0 \quad \forall x \in C_R, \tag{2.4}$$

где $C_R = \{\, x \in \mathbb{R}^n_+ : \; \sum\limits_{i=1}^{n} x_i = \; R \,\}$, $R > 0$. *Тогда задача* $NCP(F)$ *разрешима.*

Д о к а з а т е л ь с т в о. Обозначим $C'_R = \{\, x \in \mathbb{R}^n_+ : \; \sum\limits_{i=1}^{n} x_i \le R \,\}$. В силу ограниченности, замкнутости и выпуклости этого множества существует $\bar{x} \in C'_R$ такое, что

$$F(\bar{x})^\top (x - \bar{x}) \ge 0 \quad \forall x \in C'_R. \tag{2.5}$$

Поскольку C'_R содержит ноль, то

$$F(\bar{x})^\top \bar{x} \le 0. \tag{2.6}$$

Далее, существует $\alpha > 0$ такое, что $x = \alpha e_i \in C'_R$ при всех $i = \; = 1, \ldots, n$ (здесь e_i — i-й единичный орт). Подставляя эти векторы в (2.5), получаем

$$F_i(\bar{x}) \ge \frac{1}{\alpha} \; \bar{x}^\top F(\bar{x}) \quad \text{ при всех } i = 1, \ldots, n. \tag{2.7}$$

Рассмотрим два случая.

1) Пусть $\bar{x} \in C_R$. Из условия (2.4) и из (2.6), (2.7) вытекает $F(\bar{x})^\top \bar{x} = 0$, $F(\bar{x}) \ge 0$, т. е. $\bar{x}$ — решение $NCP(F)$.

2) Пусть $\bar{x} \notin C_R$. Для каждого $i = 1, \ldots, n$ существует $t > 0$ такое, что $x = \bar{x} + te_i \in C'_R$. Подставляя эти точки в (2.5), получаем неравенства $F_i(\bar{x}) \ge 0 \quad \forall i$. Снова из (2.6) следует, что $\bar{x}$ — решение задачи $NCP(F)$.

Заметим, что в качестве множества C_R в только что доказанной теореме можно использовать пересечение $\mathbb{R}^n_+$ с шаром радиуса R. Доказательство не изменится.

Перейдем к рассмотрению обобщенной задачи о дополнительности.

Конус X называется *острым*, если $X \cap (-X) = \{0\}$, и *телесным*, если $\operatorname{int}\,(X) \ne \emptyset$.

Допустимым множеством обобщенной задачи о дополнительности $GCP(X, F)$ называется множество

$$\Omega(X, F) = \{x \in X : \; F(x) \in X^*\}.$$

Векторы из $\Omega(X, F)$ называются *допустимыми*; *допустимой* называется и задача $GCP(X, F)$, если ее множество $\Omega(X, F)$ не пусто.

Лемма 2.1. *Пусть X — острый замкнутый выпуклый телесный конус в $\mathbb{R}^n$ и $d \in \mathbb{R}^n$. Тогда включение $d \in \text{int }(X^*)$ эквивалентно любому из следующих условий:*

1) $x^\top d > 0$ *при всех* $0 \neq x \in X$*;*
2) *множества* $S(d;\alpha) = \{x \in X,\ 0 \le d^\top x \le \alpha\}$ *ограничены при всех* $\alpha > 0$.

Д о к а з а т е л ь с т в о. 1) Пусть от противного $d \in \text{int }(X^*)$, но существует $0 \neq \bar{x} \in X$ такой, что $d^\top \bar{x} = 0$. Поскольку d — внутренняя точка X^*, то при малом $\delta > 0$ точка $d - \delta\bar{x}$ также лежит в X^*, что приводит к противоречивому неравенству

$$0 \le (d - \delta\bar{x})^\top \bar{x} = d^\top \bar{x} - \delta\|\bar{x}\|^2 < 0.$$

Обратно, если $x^\top d > 0$ при всех $0 \neq x \in X$, но d — граничная точка X^*, то найдется последовательность точек $\{d^k\} \to d$ такая, что выполняются неравенства $d^k \notin X^*$, т. е. при некоторых $x^k \in X$

$$(d^k)^\top x^k < 0. \tag{2.8}$$

Поскольку X — замкнутый конус, то, не теряя общности, можно считать, что $\|x^k\| = 1\ \forall k$, причем $\{x^k\} \to \bar{x}$, $\|\bar{x}\| = 1$, $\bar{x} \in X$. Поэтому при переходе в (2.8) к пределу по $k \to \infty$, имеем $d^\top \bar{x} \le 0$, что противоречит постулированному в условии свойству вектора d. Следовательно, $d \in \text{int }(X^*)$.

2) Пусть от противного $d \in \text{int }(X^*)$, но множество $S(d;\,\alpha)$ не ограничено, т. е. существует последовательность его точек $\{x^k\}$ такая, что $\{x^k\} \to \infty$. Определим

$$y^k = x^k\|x^k\|^{-1} \in X.$$

Для точек этой последовательности имеем

$$d^\top y^k = \|x^k\|^{-1} d^\top x^k \le \alpha\|x^k\|^{-1}. \tag{2.9}$$

Поскольку $\|y^k\| = 1\ \ \forall k$, то существует подпоследовательность $\{y^{k_S}\}$, сходящаяся к некоторому $\bar{y} \in X$, и для этого $\bar{y}$ предельным переходом в (2.9) получаем неравенство $d^\top \bar{y} \le 0$. По первой части доказываемой теоремы это и означает, что d — граничная точка X^*.

Обратно, пусть все множества $S(d;\alpha)$ — компактны. Предположим от противного, что $d \notin \text{int }(X^*)$. По первой части теоремы существует $0 \neq \bar{x} \in X$ такой, что $d^\top \bar{x} = 0$. Но тогда, как легко видеть, $\lambda\bar{x} \in S(d;\alpha)$ при всех $\lambda > 0$, что противоречит условию ограниченности множеств $S(d;\alpha)$.

Монотонное отображение $F : X \to \mathbb{R}^n$ назовем *правильным* в точке $\bar{x} \in X$, если множество

$$B(\bar{x}) = \{x :\ x - \bar{x} \in X,\ F(x) - F(\bar{x}) \in X^*,\ (x - \bar{x})^\top (F(x) - F(\bar{x})) = 0\}$$

ограничено.

Теорема 2.9. *Пусть X — острый выпуклый замкнутый и телесный конус в $\mathbb{R}^n$, отображение F непрерывно и монотонно на X. Если задача $GCP(X, F)$ имеет допустимую точку $\bar{x}$, в которой F правильно, то она имеет и решение.*

Д о к а з а т е л ь с т в о. Возьмем любую точку $d \in \text{int }(X^*)$. Определим множества

$$S^k = \{x :\ x - \bar{x} \in X,\ \ (x - \bar{x})^\top d = k\},\quad k = 1, 2, \dots\ .$$

Так как все эти множества не пусты, выпуклы и компактны, а отображение $\overline{F}(x) = F(x) - F(\bar{x})$ непрерывно на них, то по теореме 1.1 разрешимы задачи $VI(S^k, \overline{F})$, т. е. существуют $x^k \in S^k$ такие, что

$$(F(x^k) - F(\bar{x}))^\top x \geq (F(x^k) - F(\bar{x}))^\top x^k \quad \forall x \in S^k.$$

Вычитая из обеих частей этого неравенства значение $(F(x^k) - F(\bar{x}))^\top \bar{x}$, в силу монотонности F получаем

$$(F(x^k) - F(\bar{x}))^\top (x - \bar{x}) \geq (F(x^k) - F(\bar{x}))^\top (x^k - \bar{x}) \geq 0 \quad \forall x \in S^k.$$

Последнее означает, что $F(X^k) - F(\bar{x}) \in X^*$. Но поскольку F правильно в точке $\bar{x}$, а последовательность $\{x^k\}$ не ограничена по построению, то найдется номер m такой, что

$$(F(x^m) - F(\bar{x}))^\top (x - \bar{x}) \geq (F(x^m) - F(\bar{x}))^\top (x^m - \bar{x}) > 0 \quad \forall x \in S^k.$$

Поэтому по предыдущей теореме $F(x^m) - F(\bar{x}) \in \text{int } (X^*)$, что, ввиду выпуклости конуса X^*, означает $F(x^m) \in \text{int } (X^*)$. А так как всякое монотонное отображение псевдомонотонно, то осталось сослаться на теорему 2.7.

Следствие 2.2. *Пусть X — острый выпуклый замкнутый телесный конус в $\mathbb{R}^n$, отображение F непрерывно и строго монотонно на X. Если задача $GCP(X, F)$ допустима, то она имеет решение, и оно единственно.*

Аналогично обобщенной задаче о дополнительности *допустимым множеством* нелинейной задачи о дополнительности $NCP(F)$ называется множество

$$\Omega(F) = \{x \in \mathbb{R}^n_+ : \ F(x) \in \mathbb{R}^n_+\}.$$

Векторы из $\Omega(F)$ называются *допустимыми*; *допустимой* называется и задача $NCP(F)$, если ее множество $\Omega(F)$ не пусто.

Следующее утверждение является следствием теоремы 2.9, но и его прямое доказательство не лишено интереса.

Теорема 2.10. *Если отображение F непрерывно и строго монотонно на $\mathbb{R}^n_+$, а $\Omega(X, F)$ не пусто, то задача $NCP(F)$ разрешима, и ее решение единственно.*

Доказательство. По предположению найдется $\bar{x} \geq 0$ такое, что $F(\bar{x}) \geq 0$. Если $F(\bar{x}) > 0$, то можно сразу воспользоваться теоремой 2.7. В противном случае покажем, что можно от $\bar{x}$ перейти к точке x^1 с меньшим числом нулевых компонент в образе $F(x^1)$. Действительно, пусть $F_k(\bar{x}) = 0$. Рассмотрим точки $x^\delta = \bar{x} + \delta e_k$, где $\delta > 0$, e_k — k-й единичный орт. В силу строгой монотонности F имеем

$$(F(x^\delta) - F(\bar{x}))^\top (x^\delta - \bar{x}) = \delta(F_k(x^\delta) - F_k(\bar{x})) > 0,$$

откуда $F_k(x^\delta) > 0$ при всех $\delta > 0$. Остается выбрать $\delta > 0$ настолько малым, чтобы те $F_i(x^\delta)$, что были положительны в точке $\bar{x}$, остались бы таковыми. Найденную точку обозначим через x^1. Если $F(x^1) > 0$, обращаемся к теореме 2.7, в противном случае повторяем описанную процедуру снова.

Приведем аналоги свойств монотонности, используемые при анализе нелинейных задач о дополнительности.

Отображение $F: \; R_+^N \to \mathbb{R}^n$ называется

а) *коположительным*, если

$$(F(x) - F(0))^\top x \geq 0 \quad \forall x \in \mathbb{R}^n_+;$$

б) *строго коположительным*, если

$$(F(x) - F(0))^\top x > 0 \quad \forall x \in \mathbb{R}^n_+, \; x \neq 0;$$

в) *сильно коположительным*, если существует число $\alpha > 0$ такое, что

$$(F(x) - F(0))^\top x \geq \alpha \|x\|^2 \quad \forall x \in \mathbb{R}^n_+.$$

Поскольку сильно коположительное отображение коэрцитивно относительно $\mathbb{R}^n_+$, то по теореме 2.5 задача $NCP(F)$ с сильно коположительным отображением всегда имеет непустое компактное множество решений. Если F лишь строго коположительно, то верен следующий результат.

Теорема 2.11. *Пусть отображение $F: \; \mathbb{R}^n_+ \to \mathbb{R}^n$ непрерывно и строго коположительно. Если существует числовая функция $c: \; \mathbb{R}_+ \to \mathbb{R}$ такая, что $c(\lambda) \to \infty$ при $\lambda \to \infty$ и*

$$(F(\lambda x) - F(0))^\top x \geq c(\lambda) \, (F(x) - F(0))^\top x$$

при всех $\lambda \geq 1, \; x \geq 0$, то $NCP(F)$ имеет непустое компактное множество решений.

Д о к а з а т е л ь с т в о. Обозначим $\gamma_1 = \min\limits_{\|x\|=1, \; x \geq 0} (F(x) - F(0))^\top x$ и $\gamma_2 = \max\limits_{\|x\|=1, \; x \geq 0} F(0)^\top x$. В силу непрерывности и строгой коположительности отображения F имеем $\gamma_1 > 0$. Пусть $x \in \mathbb{R}^n_+, \; x \neq 0$. По предположению

$$(F(x) - F(0))^\top x = \left(F\left(\|x\| \, \frac{x}{\|x\|}\right) - F(0)\right)^\top \|x\| \, \frac{x}{\|x\|} \geq$$

$$\geq c(\|x\|) \left(F\left(\frac{x}{\|x\|}\right) - F(0)\right)^\top \|x\| \, \frac{x}{\|x\|} \geq c\left(\|x\|\right) \|x\| \gamma_1.$$

Отсюда

$$F(x)^\top x \geq c(\|x\|) \, \|x\| \gamma_2 + F(0)^\top x \frac{\|x\|}{\|x\|} \geq \left(c(\|x\|)\gamma_1 - \gamma_2\right) \|x\|.$$

Правая часть этого неравенства неограниченно растет при $\|x\| \to \infty$. Следовательно, множество

$$\{x \in \mathbb{R}^n_+ : F(x)^\top (x - 0) \leq 0\}$$

ограничено, и можно сослаться на теорему 2.6, полагая в ней $\bar{x} = 0$.

Закончим раздел результатами о локальной единственности решений задач о дополнительности и вариационных неравенств, получаемых без предположений типа монотонности.

Решение $\hat{x}$ задачи $VI(X, F)$ называется *локально единственным* (или *изолированным*), если существует окрестность N точки $\hat{x}$ такая, что $\hat{x}$ — единственное решение $VI(X, F)$ в N.

Следующий результат дает достаточные условия локальной единственности решения вариационного неравенства.

Теорема 2.12. *Пусть F — непрерывно дифференцируемое отображение $\mathbb{R}^n$ в себя и $\hat{x}$ — решение $VI(X,F)$. Если матрица Якоби $\nabla F(\hat{x})$ положительно определена, это решение будет локально единственным.*

Доказательство. Выделим окрестность N точки $\hat{x}$, в которой матрицы $\nabla F(x)$ положительно определены (такая окрестность существует, поскольку матрица $\nabla F(x)$ непрерывна и положительно определена в точке $\hat{x}$). В этой окрестности $\hat{x}$ — единственное решение $VI(X,F)$.

В самом деле, пусть от противного $\hat{y}$ — другое такое решение из N. По определению решения

$$F(\hat{x})^\top(\hat{y}-\hat{x}) \geq 0 \text{ и } F(\hat{y})^\top(\hat{x}-\hat{y}) \geq 0,$$

откуда

$$(F(\hat{x})-F(\hat{y}))^\top(\hat{x}-\hat{y}) \leq 0.$$

Но, опираясь на функцию

$$\varphi(t) = (\hat{x}-\hat{y})^\top F(x(t)), \quad \text{где } x(t) = t\hat{x}+(1-t)\hat{y}, \quad t \in [0,\,1],$$

и на ее производную, равную

$$\varphi'(t) = (\hat{x}-\hat{y})^\top \nabla F(x(t))\,(\hat{x}-\hat{y}) > 0,$$

легко получить противоположное неравенство

$$(F(\hat{x})-F(\hat{y}))^\top(\hat{x}-\hat{y}) =$$

$$= \varphi(1)-\varphi(0) = \int_0^1 (\hat{x}-\hat{y})^\top \nabla F(x(t))\,(\hat{x}-\hat{y})dt > 0.$$

Это противоречие и завершает доказательство.

Теорема 2.12 не оговаривает специальных свойств множества X. Если, однако, это множество задано системой неравенств типа (1.8), условия положительной определенности $\nabla F(\hat{x})$ можно ослабить. Следующий результат обобщает достаточные условия оптимальности второго порядка, известные для оптимизационных постановок.

Теорема 2.13. *Пусть отображение $F:\ \mathbb{R}^n \to \mathbb{R}^n$ непрерывно дифференцируемо, множество X определено как в (1.8), где все функции $g_i(x)$ дважды непрерывно дифференцируемы, и $\hat{x}$ — решение задачи $VI(X,F)$. Предположим также, что существует вектор $\hat{\pi} \in \mathbb{R}^n$ такой, что*

$$\begin{aligned} &F(\hat{x}) + \sum_{i=1}^m \hat{\pi}_i \nabla g_i(\hat{x}) = 0, \\ &\hat{\pi}_i \leq 0,\ \hat{\pi}_i\, g_i(\hat{x}) = 0 \ \ \forall i = 1,\,\ldots,\,m, \end{aligned} \tag{2.10}$$

и, кроме того,

$$z^\top\Big(\nabla F(\hat{x}) + \sum_{i=1}^m \hat{\pi}_i\ \nabla^2 g_i(\hat{x})\Big)z > 0 \tag{2.11}$$

при всех $z \neq 0$ таких, что

$$\begin{aligned} &z^\top \nabla g_i(\hat{x}) = 0 \qquad \forall i \in I_1 = \{i:\ \hat{\pi}_i > 0\}, \\ &z^\top \nabla g_i(\hat{x}) \geq 0 \qquad \forall i \in I_2 = \{i:\ g_i(\hat{x}) = 0,\ \hat{\pi}_i = 0\}. \end{aligned}$$

Тогда $\hat{x}$ —локально единственное решение $VI(X,F)$.

Д о к а з а т е л ь с т в о. Пусть от противного существует сходящаяся к $\hat{x}$ последовательность $\{x^k\}$, где $x^k \in X$, $x^k \neq \hat{x}$ и

$$F(x^k)^\top (x - x^k) \geq 0 \quad \forall x \in X. \tag{2.12}$$

Определим последовательность векторов $\{y^k\}$ и чисел $\{\delta_k\}$ таких, что $x^k = \hat{x} + \delta_k y^k$, где $\|y^k\| = 1$, $\delta_k > 0$. Поскольку $x^k \in X$, то

$$g_i(x^k) - g_i(\hat{x}) \leq 0 \quad \forall i: \ g_i(\hat{x}) = 0; \tag{2.13}$$

$$F(\hat{x})^\top (x^k - \hat{x}) \geq 0. \tag{2.14}$$

Не ограничивая общности, можно считать, что $\delta_k \to 0$, $y^k \to \bar{y}$, причем $\|\bar{y}\| = 1$. Поделив обе части в (2.13) на δ_k и перейдя затем к пределу по $k \to \infty$, получим

$$\nabla g_i(\hat{x}) \bar{y} \leq 0 \quad \forall i: \ g_i(\hat{x}) = 0.$$

Из (2.14) следует, что

$$F(\hat{x})^\top y^k \geq 0 \quad \forall k, \tag{2.15}$$

а из (2.12) — что

$$F(\hat{x})^\top y^k \leq 0 \quad \forall k. \tag{2.16}$$

Поэтому

$$F(\hat{x})^\top y^k = 0 \quad \forall k. \tag{2.17}$$

Пусть найдется такой номер m, что $\hat{\pi}_m > 0$ и $\nabla g_m(\hat{x}) \bar{y} > 0$. Из (2.10) следует

$$F(\hat{x})^\top \bar{y} = \hat{\pi}^\top \nabla g(\hat{x}) \, \bar{y} = \sum_{i=1}^{m} \hat{\pi}_i \, \nabla g_i(\hat{x}) \, \bar{y} > 0,$$

что противоречит (2.17). Значит,

$$\nabla g_i(\hat{x}) \bar{y} = 0, \quad \forall i: \ \hat{\pi}_i > 0. \tag{2.18}$$

Далее по теореме о среднем

$$F(x^k) = F(\hat{x}) + \delta_k \begin{pmatrix} \nabla F_1(\eta_1^k) \\ \dots \\ \nabla F_n(\eta_n^k) \end{pmatrix} y^k. \tag{2.19}$$

В силу (2.16), из (2.19) вытекает

$$F(\hat{x})^\top y^k + \delta_k {y^k}^\top \begin{pmatrix} \nabla F_1(\eta_1^k) \\ \dots \\ \nabla F_n(\eta_n^k) \end{pmatrix} y^k \leq 0,$$

откуда, ввиду (2.15), уже следует

$$\delta_k {y^k}^\top \begin{pmatrix} \nabla F_1(\eta_1^k) \\ \dots \\ \nabla F_n(\eta_n^k) \end{pmatrix} y^k \leq 0.$$

Поделив обе части полученного неравенства на $\delta_k > 0$ и перейдя к пределу по $k \to \infty$, получим

$$\bar{y}^\top \nabla F(\hat{x}) \bar{y} \leq 0,$$

что противоречит условию (2.11).

Два последних утверждения можно отнести и к нелинейной задаче о дополнительности. Но, благодаря специальной структуре последней, справедливо более сильное утверждение (приводится без доказательства).

Теорема 2.14. *Пусть отображение* $F:\ \mathbb{R}^n \to \mathbb{R}^n$ *непрерывно дифференцируемо и* $\hat{x}$ *— решение* $NCP(F)$*. Определим индексные наборы* $I_+ = \{i:\ \hat{x}_i > 0\}$ и $I_0 = \{i:\ \hat{x}_i = 0,\ F_i(\hat{x}) = 0\}$ *и отвечающие им подвекторы*

$$x_{I_+} = \{x_i:\ i \in I_+\},\ x_{I_0} = \{x_i:\ i \in I_0\},$$

$$F_{I_+}(x) = \{F_i(x):\ i \in I_+\} \ \text{ и } \ F_{I_0}(x) = \{F_i(x):\ i \in I_0\}.$$

Если нет ненулевых векторов (x_{I_+}, x_{I_0})*, удовлетворяющих условиям*

$$\nabla_{I_+} F_{I_+}(\hat{x})\, x_{I_+} + \nabla_{I_0} F_{I_+}(\hat{x})\, x_{I_0} = 0,$$

$$\nabla_{I_+} F_{I_0}(\hat{x})\, x_{I_+} + \nabla_{I_0} F_{I_0}(\hat{x})\, x_{I_0} \geq 0,$$

$$x_{I_0} \geq 0 \ \text{ и } \ x_{I_0}^\top(\nabla_{I_0} F_{I_0}(\hat{x}) x_{I_+} + \nabla_{I_0} F_{I_0}(\hat{x}) x_{I_0}) = 0,$$

то $\hat{x}$ *— локально единственное решение* $NCP(F)$*. Здесь* $\nabla_N F_I(x)$ *обозначает* $I \times N$ *–блок матрицы* ∇F.

Важный частный случай описывается равенством $I_0 = \emptyset$.

Говорят, что решение $\hat{x}$ задачи $NCP(F)$ удовлетворяет условию *строгой дополнительности*, если $\hat{x} + F(\hat{x}) > 0$. В этом случае решение $\hat{x}$ называют *невырожденным*.

Теорема 2.15. *Пусть отображение* $F: \mathbb{R}^n \to \mathbb{R}^n$ *непрерывно дифференцируемо и* $\hat{x}$ *— невырожденное решение* $NCP(F)$*. Пусть* $I_+ = \{i:\ \hat{x}_i > 0\}$*. Если матрица* $\nabla_{I_+} F_{I_+}(\hat{x})$ *не вырождена, то* $\hat{x}$ *— локально единственное решение* $NCP(F)$.

Доказательство следует непосредственно из классической теоремы о неявной функции, так как в условиях невырожденности нелинейная задача о дополнительности сводится к системе нелинейных уравнений в некоторой окрестности решения.

Упражнения

1. Покажите, что функция расстояния $d_X(x)$ от точки x до выпуклого замкнутого множества X выпукла и $x - P_X(x) \in \partial d_X(x)$.

2. Покажите, что дифференцируемая функция $f: \mathbb{R}^n \to \mathbb{R}$ выпукла, псевдовыпукла, строго выпукла и сильно выпукла тогда и только тогда, когда отображение $F = \nabla f$ монотонно, псевдомонотонно, строго монотонно и сильно монотонно соответственно.

3. Покажите, что отображение $F = (\nabla_1 f, -\nabla_2 f)$, порождаемое дифференцируемой функцией двух аргументов $f: \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, будет монотонным тогда и только тогда, когда функция $f(x,y)$ при фиксированном значении второго аргумента выпукла по первому аргументу и при фиксированном значении первого аргумента вогнута по второму. Установите также связь между строгой и сильной монотонностью отображения F и свойствами строгой и сильной выпуклости-вогнутости функции f по ее аргументам.

4. Покажите, что аффинное отображение $F(z) = q + Mz$ всюду строго монотонно (просто монотонно) тогда и только тогда, когда матрица M положительно определена (полуопределена).

5. Покажите, что для аффинных отображений понятия строгой и сильной монотонности совпадают.

6. Покажите, что дважды непрерывно дифференцируемая функция $f : \mathbb{R}^n \to \mathbb{R}$ выпукла (сильно выпукла) тогда и только тогда, когда матрица ее вторых производных $H(x) = (\partial^2 f(x)/\partial x_i \partial x_j)_{n\times n}$ всюду положительно полуопределена (равномерно положительно определена, т. е. удовлетворяет условию $z^\top H(x) z \geq \delta > 0 \ \forall z, \ \|z\| = 1$, где δ не зависит от x).

7. Пусть отображение $F : S \to \mathbb{R}^n$ непрерывно на

$$S = \{x \in \mathbb{R}^n : \ x \geq 0, \ \sum\nolimits_{i=1}^n x_i = 1\}.$$

Покажите, что существует точка $\bar{x} \in S$, удовлетворяющая условиям:

a) $F(\bar{x})^\top (x - \bar{x}) \geq 0$ при всех $x \in S$;

b) $\bar{x}_k > 0 \Rightarrow F_k(\bar{x}) = \min_i F_i(\bar{x}) = \mu_0$;

c) $F_k(\bar{x}) > \mu_0 \Rightarrow \bar{x}_k = 0$.

§ 3. Проекционные методы решения

Обратимся к методам решения вариационных неравенств и нелинейных задач о дополнительности. Начнем с итерационных методов первого порядка, т. е. методов, не предполагающих дифференцируемости входящих в них отображений. Изложение будем вести для более общей задачи — задачи решения вариационного неравенства $VI(X, F)$.

Первым рассмотрим итерационный процесс, порождаемый следующими рекуррентными соотношениями:

$$x^{k+1} = P_X(x^k - \alpha F(x^k)), \quad k = 0, 1, \dots, \tag{3.1}$$

где $x^0 \in X$ — произвольное начальное приближение; $P_X(\cdot)$ — операция проектирования на множество X; $\alpha > 0$ — параметр.

Операция проектирования на множество X относится к обычным оптимизационным задачам и подробно здесь не обсуждается. Ее реализация может оказаться трудной, но часто оказывается и простой. В частности, в задачах о дополнительности $NCP(F)$, где $X = \mathbb{R}^n_+$, она решается взятием положительной срезки координат проектируемого вектора. На практике в сложных случаях проектирование осуществляется приближенно, а его точность согласуется с другими параметрами алгоритма.

Процесс (3.1) можно рассматривать как метод простой итерации для поиска неподвижной точки проекционного отображения H из (1.12). Как показывает следующее утверждение, он сходится глобально, но лишь при достаточно сильных предположениях относительно монотонности и непрерывности отображения F.

Теорема 3.1. *Пусть X — непустое выпуклое замкнутое подмножество $\mathbb{R}^n$ и пусть отображение $F : X \to \mathbb{R}^n$ сильно монотонно с константой $\gamma > 0$, т. е.*

$$(F(x) - F(y))^\top (x - y) \geq \gamma \|x - y\|^2 \quad \forall x, y \in X,$$

и удовлетворяет условию Липшица с константой $L > 0$, т. е.

$$\|F(x) - F(y)\| \leq L \, \|x - y\| \quad \forall x, y \in X.$$

Тогда вариационное неравенство $VI(X,F)$ разрешимо, и при любом $0 < \alpha < 2\gamma L^{-2}$ последовательность, порожденная процессом (3.1), *сходится к $\hat{x}$ — его единственному решению, причем скорость этой сходимости линейна: найдется $\varrho < 1$ такое, что*

$$\|x^{k+1} - \hat{x}\| \le \varrho\ \|x^k - \hat{x}\|$$

при всех k.

Д о к а з а т е л ь с т в о. По теореме 1.3 из данной главы точка $\hat{x}$ удовлетворяет равенству

$$\hat{x} = P_X(\hat{x} - \alpha F(\hat{x})) \quad \forall \alpha > 0.$$

Поэтому, в силу свойств нерасширяемости оператора проектирования и соотношений (3.1),

$$\|x^{k+1} - \hat{x}\|^2 = \|P_X(x^k - \alpha F(x^k)) - P_X(\hat{x} - \alpha F(\hat{x}))\|^2 \le$$
$$\le \|x^k - \alpha F(x^k) - \hat{x} + \alpha F(\hat{x})\|^2 =$$
$$= \|x^k - \hat{x}\|^2 + \alpha^2 \|F(x^k) - F(\hat{x})\|^2 - 2\alpha (F(x^k) - F(\hat{x}))^\top (x^k - \hat{x}).$$

Здесь, в силу липшицевости и сильной монотонности F, имеем

$$\|F(x^k) - F(\hat{x})\| \le L \|x^k - \hat{x}\|,$$

$$-(F(x^k) - F(\hat{x}))^\top (x^k - \hat{x}) \le -\gamma \|x^k - \hat{x}\|^2.$$

Поэтому окончательно

$$\|x^{k+1} - \hat{x}\|^2 \le (1 + \alpha^2 L^2 - 2\alpha\gamma) \|x^k - \hat{x}\|^2,$$

где $0 < 1 + \alpha^2 L^2 - 2\alpha\gamma < 1$ при $0 < \alpha < 2\gamma L^{-2}$. Это доказывает искомые свойства сходимости с $\varrho = (1 + \alpha^2 L^2 - 2\alpha\gamma)^{1/2}$.

Метод проекции можно применять не только для решения вариационных неравенств и нелинейных задач о дополнительности, но и для решения линейных задач о дополнительности, матрицы которых положительно определены. Действительно, аффинное отображение $F(x) = q + Mx$ всегда является липшицевым, а в случае положительно определенной матрицы будет и сильно монотонным с константой $\gamma = \max\limits_{\|x\|=1} x^\top M x$.

Заметим, что на практике константы γ и L редко известны заранее, и потому шаговый параметр α подбирают экспериментально, например, постепенным дроблением некоторого начального $\alpha_0 > 0$.

Если отображение F липшицево и монотонно, но не сильно монотонно (например, матрица в линейной задаче о дополнительности только положительно полуопределена, но не положительно определена), метод простой итерации не гарантирует получение искомого решения. В этом случае исходную задачу $VI(X,F)$ можно подвергнуть так называемой регуляризации, т. е. заменить на некоторую близкую задачу, обладающую предпочтительными вычислительными свойствами, например, на задачу $VI(X,F_\gamma)$, где $F_\gamma(x) = F(x) + \gamma x$, $\gamma > 0$ — малый параметр. Очевидно, отображение F_γ не только наследует свойства липшицевости и монотонности отображения F, но и приобретает свойство сильной монотонности, в силу чего сходимость процесса (3.1) для регуляризованной задачи имеет место. Разумеется, предельная точка процесса будет решать совсем другую задачу, а именно задачу $VI(X,F_\gamma)$. Обоснование методу регуляризации дает

Теорема 3.2. *Пусть X — непустое выпуклое замкнутое подмножество $\mathbb{R}^n$, отображение $F : X \to \mathbb{R}^n$ непрерывно и монотонно на X. Тогда для любого $\gamma > 0$ существует решение регуляризованной постановки, т. е. точка $x(\gamma) \in X$ такая, что*

$$(F(x(\gamma)) + \gamma x(\gamma))^\top (x - x(\gamma)) \geq 0 \quad \forall x \in X, \tag{3.2}$$

и если задача $VI(X, F)$ разрешима, то

$$\lim_{\gamma \to +0} x(\gamma) = \hat{x}, \tag{3.3}$$

где $\hat{x} = \arg \min\limits_{x \in \operatorname{Arg} VI(X,F)} \|x\|$ — нормальное решение $VI(X, F)$.

Д о к а з а т е л ь с т в о. Разрешимость задачи (3.2) вытекает из уже отмечавшегося свойства сильной монотонности отображения F_γ. Сосредоточимся на обосновании соотношения (3.3).

Отметим вначале, что, в силу непустоты, выпуклости и замкнутости множества $\operatorname{Arg} VI(X, F)$, его минимальный по норме элемент $\hat{x}$ (нормальное решение) существует и является единственным. Он удовлетворяет условию

$$F(\hat{x})^\top (x - \hat{x}) \geq 0 \quad \forall x \in X.$$

Подставим сюда $x = x(\gamma)$ и $x = \hat{x}$ — в соотношение (3.2). Складывая полученное, приходим к неравенству

$$(F(x(\gamma)) + \gamma x(\gamma) - F(\hat{x}))^\top (\hat{x} - x(\gamma)) \geq 0.$$

После перегруппировки слагаемых имеем

$$\gamma x(\gamma)^\top (\hat{x} - x(\gamma)) \geq (F(\hat{x}) - F(x(\gamma)))^\top (\hat{x} - x(\gamma)).$$

Далее, в силу монотонности F правая часть этого неравенства неотрицательна, следовательно, неотрицательна и левая и потому

$$x(\gamma)^\top \hat{x} \geq \|x(\gamma)\|^2.$$

Отсюда $\|x(\gamma)\|^2 \leq \|\hat{x}\| \, \|x(\gamma)\|$, т. е.

$$\|x(\gamma)\| \leq \|\hat{x}\|. \tag{3.4}$$

Таким образом, последовательность $\{x(\gamma)\}$, где $\gamma \to +0$, является ограниченной и имеет предельные точки. Последние являются решением $VI(X, F)$, в чем легко убедиться путем перехода в соотношениях (3.2) к пределу по сходящимся подпоследовательностям. Но, в силу (3.4) и единственности минимального элемента $\hat{x}$, все такие предельные точки совпадают и равны этому элементу.

Процесс постепенного уменьшения параметра регуляризации можно и целесообразно совместить с процессом решения регуляризованного неравенства. Это соображение реализуется в методе итеративной регуляризации А.Б.Бакушинского, соотношения которого имеют вид

$$x^{k+1} = P_X(x^k - \alpha_k(F(x^k) + \gamma_k x^k)), \quad k = 0, 1, 2, \ldots, \tag{3.5}$$

здесь $x^0 \in X$ — произвольное начальное приближение к решению, $P_X(\cdot)$ — операция проектирования на множество X, положительные параметры α_k, γ_k согласованно меняются в ходе вычислений.

Прежде чем сформулировать условия сходимости метода итеративной регуляризации, докажем ряд вспомогательных утверждений.

Лемма 3.1. *Пусть* X *— непустое выпуклое замкнутое подмножество* $\mathbb{R}^n$, *отображение* $F : X \to \mathbb{R}^n$ *непрерывно и монотонно на* X *и задача* $VI(X,F)$ *разрешима. Пусть* $x(\gamma)$ *— решение* $VI(X,F_\gamma)$, *где* $F_\gamma(x) = F(x) + \gamma x$, $\gamma > 0$. *Тогда*

$$\|x(\gamma') - x(\gamma'')\| \leq \|\hat{x}\| \, (\gamma' - \gamma'')/\gamma',$$

где $\hat{x}$ *— нормальное решение* $VI(X,F)$, $\gamma' > \gamma'' > 0$ *— любые.*

Д о к а з а т е л ь с т в о. Точка $x(\gamma')$ удовлетворяет неравенствам

$$(F(x(\gamma')) + \gamma' x(\gamma'))^\top (x - x(\gamma')) \geq 0 \qquad \forall x \in X,$$

а точка $x(\gamma'')$ — неравенствам

$$(F(x(\gamma'')) + \gamma'' x(\gamma''))^\top (x - x(\gamma'')) \geq 0 \qquad \forall x \in X.$$

Подставляя $x = x(\gamma'')$ в первое из них и $x = x(\gamma')$ — во второе и складывая полученное, имеем

$$(F(x(\gamma'')) + \gamma'' x(\gamma'') - F(x(\gamma')) - \gamma' x(\gamma'))^\top (x(\gamma') - x(\gamma'')) \geq 0.$$

Отсюда

$$(\gamma'' x(\gamma'') - \gamma' x(\gamma'))^\top (x(\gamma') - x(\gamma'')) \geq$$
$$\geq (F(x(\gamma')) - F(x(\gamma'')))^\top (x(\gamma') - x(\gamma'')).$$

Правая часть этого неравенства неотрицательна ввиду монотонности F. Поэтому левая часть также неотрицательна. Используем это в следующей оценке:

$$\gamma' \|x(\gamma') - x(\gamma'')\|^2 =$$
$$= (\gamma' x(\gamma') - \gamma'' x(\gamma''))^\top (x(\gamma') - x(\gamma'')) + (\gamma'' - \gamma') x(\gamma'')^\top (x(\gamma') - x(\gamma'')) \leq$$
$$\leq (\gamma' - \gamma'') \cdot \|x(\gamma'')\| \cdot \|x(\gamma') - x(\gamma'')\|.$$

Остается применить соотношение (3.4).

Лемма 3.2. *Пусть отображение* F *монотонно и удовлетворяет условию Липшица с константой* $L > 0$, *т. е.*

$$\|F(x) - F(y)\| \leq L \|x - y\| \quad \forall x, y \in X.$$

Тогда $\|x' - \alpha F_\gamma(x') - x'' + \alpha F_\gamma(x'')\| \leq (1 - r(\alpha,\gamma)) \|x' - x''\|$, *где* $F_\gamma = F(x) + \gamma x$, $\gamma > 0$, $r(\alpha,\gamma) = \alpha\gamma - \alpha^2 (L+\gamma)^2/2$.

Д о к а з а т е л ь с т в о. В силу монотонности и липшицевости отображения F имеем

$$\|x' - \alpha F_\gamma(x') - x'' + \alpha F_\gamma(x''\|^2 =$$
$$= \|x' - x''\|^2 + \alpha^2 \|F_\gamma(x') - F_\gamma(x'')\|^2 - 2\alpha (F_\gamma(x') - F_\gamma(x''))^\top (x' - x'') \leq$$
$$\leq \|x' - x''\|^2 + \alpha^2 (\|F(x') - F(x'')\| + \gamma \|x' - x''\|)^2 - 2\alpha\gamma \|x' - x''\|^2 \leq$$
$$\leq (1 + \alpha^2 (L+\gamma)^2 - 2\alpha\gamma) \|x' - x''\|^2.$$

Отсюда, ввиду неравенства $\sqrt{1+a} \leq 1 + a/2$, следует требуемое утверждение.

Следующее утверждение хорошо известно из теории оптимизации ([9], с.96) и потому приводится без доказательства.

Лемма 3.3. *Пусть числовая последовательность $\{\Delta_k\}$ удовлетворяет соотношениям*

$$0 \le \Delta_{k+1} \le (1-\delta_k)\Delta_k + \sigma_k \quad (k = 0, 1, 2, \ldots),$$

где $\delta_k \in (0,1], \sigma_k > 0, \sum\limits_{k=1}^{\infty} \delta_k = \infty, \lim\limits_{(k)} \sigma_k \delta_k^{-1} = 0.$ *Тогда* $\lim\limits_{(k)} \Delta_k = 0.$

Сформулируем теперь основной результат сходимости для метода итеративной регуляризации.

Теорема 3.3. *Пусть X — непустое выпуклое замкнутое подмножество $\mathbb{R}^n$, отображение $F : X \to R^n$ монотонно и удовлетворяет на X условию Липшица с константой $L > 0$, т. е.*

$$\|F(x) - F(y)\| \le L\|x - y\| \quad \forall x, y \in X.$$

Если задача $VI(X, F)$ разрешима, а параметры метода удовлетворяют условиям

$$\alpha_k \to +0,\ \gamma_k > \gamma_{k+1} > 0,\ \gamma_k \to +0,\ \sum\nolimits_{k=1}^{\infty} \alpha_k\gamma_k = \infty, \tag{3.6}$$

$$\gamma_k = (\gamma_k - \gamma_{k+1})/(\alpha_k\gamma_k^2) \to 0,\ 0 < \alpha_k\gamma_k^{-1} < \varepsilon_0 = 1/L^2, \tag{3.7}$$

то последовательность (3.5) *сходится к одному из ее решений.*

Д о к а з а т е л ь с т в о. Обозначим через $\check{x}^k$ решения задач $VI(X, F_k)$, где $F_k(x) = F(x) + \gamma_k x$. Напомним, что по теореме 1.3 из данной главы

$$\check{x}^k = P_X(\check{x}^k - \alpha(F(\check{x}^k) + \gamma_k\check{x}^k)) \quad \forall \alpha > 0.$$

Покажем, что $\|x^k - \check{x}^k\| \to 0$ при $k \to \infty$. В самом деле, по лемме 3.1

$$\|x^{k+1} - \check{x}^{k+1}\| \le \|x^{k+1} - \check{x}^k\| + \|\check{x}^k - \check{x}^{k+1}\| \le$$
$$\le \|x^{k+1} - \check{x}^k\| + \|\hat{x}\|\,(\gamma_k - \gamma_{k+1})/\gamma_k,$$

где $\hat{x}$ — нормальное решение $VI(X, F)$. По лемме 3.2 и свойствам операции проектирования

$$\|x^{k+1} - \check{x}^k\| =$$
$$= \|P_X(x_k - \alpha_k(F(x^k) + \gamma_k x^k)) - P_X(\check{x}^k - \alpha_k(F(\check{x}^k -) + \gamma_k\check{x}^k))\| \le$$
$$\le \|x_k - \alpha_k(F(x^k) + \gamma_k x^k) - \check{x}^k + \alpha_k(F(\check{x}^k) + \gamma_k\check{x}^k)\| \le$$
$$\le (1 - r(\alpha_k, \gamma_k))\|x^k - \check{x}^k\|.$$

В итоге

$$\|x^{k+1} - \check{x}^{k+1}\| \le (1 - r(\alpha_k, \gamma_k))\|x^k - \check{x}^k\| + \frac{\gamma_k - \gamma_{k+1}}{\gamma_k}\|\hat{x}\|.$$

Остается воспользоваться леммой 3.3 и предположениями (3.6), (3.7) относительно параметров шага и регуляризации.

Класс последовательностей, для которых условия приведенной теоремы верны, достаточно широк. В частности, в него входят последовательности вида $\alpha_k = (1 + k)^{-1/2}$, $\gamma_k = (1 + k)^{-1/3}$.

Альтернативой методу регуляризации с его сложными правилами согласования шаговых параметров может служить экстраполяционный метод проекции, в котором в процесс вычислений вовлечены сразу две последовательности. Одна из них является основной, а вторая играет роль своеобразного поводыря. Рекуррентные соотношения экстраполяционного метода имеют вид

$$x^{k+1} = P_X(x^k - \alpha F(\bar{x}^k)), \ \bar{x}^k = P_X(x^k - \alpha F(x^k)), \quad \text{где } k = 0, 1, 2, \ldots, \tag{3.8}$$

здесь x^k — точки основной последовательности; $\bar{x}^k$ — точки вспомогательной последовательности; $x^0 \in X$ — произвольное начальное приближение; $P_X(\cdot)$ — операция проектирования на множество X; $\alpha > 0$ — параметр.

Условия сходимости экстраполяционного метода дает

Теорема 3.4. *Пусть X — непустое выпуклое замкнутое подмножество $\mathbb{R}^n$, отображение F монотонно и удовлетворяет на X условию Липшица с константой L, т. е.*

$$\|F(x) - F(y)\| \le L\|x - y\| \quad \forall x, y \in X.$$

Если задача $VI(X, F)$ разрешима и $0 < \alpha < 1/L$, обе последовательности из (3.8) сходятся к одному из ее решений.

Доказательство. Пусть $\hat{x}$ — произвольное решение $VI(X, F)$. Воспользуемся свойством операции проектирования:

$$\|u - v\|^2 \ge \|u - P_X(u)\|^2 + \|v - P_X(u)\|^2 \qquad \forall u \in \mathbb{R}^n, \quad \forall v \in X.$$

Подставляя сюда $v = \hat{x}$, $u = x^k - \alpha F(\bar{x}^k)$ и учитывая (3.8), получаем неравенство

$$\|x^{k+1} - \hat{x}\|^2 \le \|x^k - \alpha F(\bar{x}^k) - \hat{x}\|^2 - \|x^k - \alpha F(\bar{x}^k) - x^{k+1}\|^2 =$$

$$= \|x^k - \hat{x}\|^2 - \|x^k - x^{k+1}\|^2 + 2\alpha F(\bar{x}^k)^\top(\hat{x} - x^{k+1}).$$

Поскольку $\hat{x}$ — решение $VI(X, F)$, то последнее слагаемое можно оценить так:

$$F(\bar{x}^k)^\top(\hat{x} - x^{k+1}) = F(\bar{x}^k)^\top(\hat{x} - \bar{x}^k) + F(\bar{x}^k)^\top(\bar{x}^k - x^{k+1}) \le$$

$$\le F(\bar{x}^k)^\top(\bar{x}^k - x^{k+1}).$$

С учетом этого продолжим основную цепочку:

$$\|x^{k+1} - \hat{x}\|^2 \le \|x^k - \hat{x}\|^2 - \|x^k - x^{k+1}\|^2 + 2\alpha F(\bar{x}^k)^\top(\bar{x}^k - x^{k+1}) =$$

$$= \|x^k - \hat{x}\|^2 - \|x^k - \bar{x}^k\|^2 - \|\bar{x}^k - x^{k+1}\|^2 -$$

$$-2(x^k - \bar{x}^k)^\top(\bar{x}^k - x^{k+1}) + 2\alpha F(\bar{x}^k)^\top(\bar{x}^k - x^{k+1}) =$$

$$= \|x^k - \hat{x}\|^2 - \|x^k - \bar{x}^k\|^2 - \|\bar{x}^k - x^{k+1}\|^2 +$$

$$+2(x^k - \alpha F(\bar{x}^k) - \bar{x}^k)^\top(x^{k+1} - \bar{x}^k).$$

Вновь обратимся к последнему слагаемому. Разложим его на сумму двух других слагаемых, одно из которых по свойствам операции проектирования неположительно, а другое оценивается сверху неравенством Коши—Буняковского:

$$(x^k - \alpha F(\bar{x}^k) - \bar{x}^k)^\top(x^{k+1} - \bar{x}^k) =$$

$$= (x^k - \alpha F(x^k) - \bar{x}^k)^\top(x^{k+1} - \bar{x}^k) + \alpha(F(x^k) - F(\bar{x}^k))^\top(x^{k+1} - \bar{x}^k) \le$$

$$\le \alpha\|F(x^k) - F(\bar{x}^k)\| \cdot \|x^{k+1} - \bar{x}^k\|.$$

Кроме того,
$$\|x^k-\bar{x}^k\|^2+\|\bar{x}^k-x^{k+1}\|^2 \geq 2\|x^k-\bar{x}^k\|\cdot\|\bar{x}^k-x^{k+1}\|.$$
Возвращаясь к основной цепочке неравенств, ввиду липшицевости F, имеем
$$\|x^{k+1}-\hat{x}\|^2 \leq \|x^k-\hat{x}\|^2-\|x^k-\bar{x}^k\|^2-\|\bar{x}^k-x^{k+1}\|^2+$$
$$+2\alpha L\|x^k-\bar{x}^k\|\cdot\|x^{k+1}-\bar{x}^k\| \leq \|x^k-\hat{x}\|^2-\|x^k-\bar{x}^k\|^2-\|\bar{x}^k-x^{k+1}\|^2+$$
$$+\alpha^2L^2\|x^k-\bar{x}^k\|^2+\|x^{k+1}-\bar{x}^k\|^2.$$
В итоге
$$\|x^{k+1}-\hat{x}\|^2 \leq \|x^k-\hat{x}\|^2-(1-\alpha^2L^2)\|x^k-\bar{x}^k\|^2. \tag{3.9}$$
А так как по предположению $1-\alpha^2L^2>0$, то из (3.9) следует, что $\|x^k-\bar{x}^k\|\to 0$ при $k\to\infty$ и что последовательность $\|x^k-\hat{x}\|$ — невозрастающая, а значит, ограничена и имеет предельные точки. Пусть $\check{x}$ — одна из таких предельных точек. Поскольку
$$\bar{x}^k = P_X(x^k-\alpha F(x^k)) \quad \forall k$$
и $\|x^k-\bar{x}^k\|\to 0$, то, переходя в последнем равенстве к пределу по сходящейся подпоследовательности, получаем
$$\check{x}=P_X(\check{x}-\alpha F(\check{x})),$$
т. е. $\check{x}$ — решение задачи $VI(X,F)$. Отождествляя $\check{x}$ с $\hat{x}$ из оценки (3.9), получаем сходимость к $\hat{x}$ всей последовательности $\{x^k\}$, а значит, и $\{\bar{x}^k\}$.

Рассмотренная экстраполяционная вычислительная схема не является единственно возможной. В частности, на каждой итерации можно использовать одно и то же направление спуска, что уменьшает вычислительные затраты метода. Модифицированная схема вычислений имеет вид
$$\begin{gathered} x^{k+1}=P_X(x^k-\alpha F(\bar{x}^k)),\ \bar{x}^{k+1}=P_X(x^{k+1}-\alpha F(\bar{x}^k)),\\ \text{где } k=0,1,2,\dots, \end{gathered} \tag{3.10}$$
здесь x^k — точки основной последовательности; $\bar{x}^k$ — точки вспомогательной последовательности; $x^0,\ \bar{x}^0\in X$ — произвольные начальные приближения; $P_X(\cdot)$ — операция проектирования на множество X; $\alpha>0$ — параметр.

Приведем без доказательства условия сходимости модифицированной схемы.

Теорема 3.5. *Пусть X — непустое выпуклое замкнутое подмножество $\mathbb{R}^n$, отображение F монотонно и удовлетворяет на X условию Липшица с константой L. Если задача $VI(X,F)$ разрешима и $0<\alpha<1/3L$, обе последовательности из* (3.10) *сходятся к одному из ее решений.*

Заметим, что технику ведущей точки (поводыря) можно использовать и при работе с сильно монотонными отображениями. Сходимость метода от этого только выигрывает.

§ 4. Методы решения второго порядка

Рассматриваемые ниже методы второго порядка решения вариационных неравенств и нелинейных задач о дополнительности привлекают своей квадратичной скоростью сходимости, но требуют повышенной гладкости отображений, более сложны в реализации и сходятся, вообще говоря, лишь локально (начальное приближение должно быть близко к искомому решению задачи).

Общая схема методов второго порядка имеет вид

$$x^{k+1} \in \text{Arg } VI(X, F^k), \quad k = 0, 1, 2, \dots, \tag{4.1}$$

где $F^k(x) = F(x^k) + A(x^k)\,(x - x^k)$ — линейная аппроксимация $F(x)$ в точке x^k; $A(x^k)$ — $(n \times n)$-матрица. Существует множество конкретных реализаций общей схемы, различающихся выбором $A(x^k)$:

$A(x^k) = \nabla F(x^k)$ — метод Ньютона,

$A(x^k) \approx \nabla F(x^k)$ — квазиньютоновские методы,

$A(x^k) = D(x^k)$ — линеаризованный метод Якоби

и др. Здесь $D(x^k)$ — диагональная часть якобиана $\nabla F(x^k)$.

Приведем несколько теорем, обосновывающих сходимость методов второго порядка.

Теорема 4.1. *Пусть X — непустое выпуклое замкнутое подмножество $\mathbb{R}^n$ и отображения F и A непрерывны. Пусть $\hat{x}$ — решение $VI(X, F)$. Если существуют такие $\alpha > \beta > 0$, что матрица $A(\hat{x}) - \alpha E$ положительно полуопределена и для всех x, y из некоторой окрестности N_1 точки $\hat{x}$ выполнены неравенства*

$$\|F(x) - F(y) - A(y)\,(x - y)\| \le b\|x - y\|, \tag{4.2}$$

то при условии близости начального приближения x^0 к $\hat{x}$ процесс (4.1) определен и сходится к $\hat{x}$.

Доказательство. Как следует из условия, матрица $A(\hat{x})$ положительно определена, и, ввиду непрерывности A, существует окрестность N_2 точки $\hat{x}$ такая, что $A(y)$ положительно определена при всех $y \in N_2$. Отсюда следует, что задача $VI(X, F^y)$, где $F^y(x) = F(y) + A(y)(x - y)$, разрешима и имеет единственное решение при всех $y \in N_2$.

Выберем $\varepsilon > 0$ такое, что $0 < r = b/(\alpha - \varepsilon) < 1$. Пусть N_3 — окрестность точки $\hat{x}$, в которой

$$\|A(\hat{x}) - A(y)\| \le \varepsilon \qquad \forall y \in N_3.$$

Все окрестности определены относительно $\|\cdot\|$. Возьмем $N = \bigcap\limits_{i=1}^{3} N_i$. Пусть $x^0 \in N$. Тогда x^1 — решение $VI(X, F^0)$ — существует и единственно. Оно удовлетворяет неравенству

$$(\hat{x} - x^1)^\top (F(x^0) + A(x^0)\,(x^1 - x^0)) \ge 0,$$

так как $\hat{x} \in X$. Аналогично, так как $\hat{x}$ — решение исходной задачи и $x^1 \in X$,

$$(x^1 - \hat{x})^\top F(\hat{x}) \ge 0.$$

Складывая эти два неравенства и перегруппировывая слагаемые, получаем

$$(x^1 - \hat{x})^\top A(x^0)(x^1 - \hat{x}) \le (x^1 - \hat{x})^\top (F(\hat{x}) - F(x^0) - A(x^0)\,(\hat{x} - x^0)).$$

Оценим левую часть этого неравенства:

$$(x^1 - \hat{x})^\top A(x^0)\,(x^1 - \hat{x}) = (x^1 - \hat{x})^\top A(\hat{x})\,(x^1 - \hat{x}) +$$

$$+(x^1-\hat{x})^\top (A(x^0)-A(\hat{x}))\,(x^1-\hat{x}) \geq \alpha\|x^1-\hat{x}\|^2 - \varepsilon\|x^1-\hat{x}\|^2 =$$
$$= (\alpha-\varepsilon)\|x^1-\hat{x}\|^2.$$

Теперь оценим правую часть:

$$(x^1-\hat{x})^\top (F(\hat{x})-F(x^0)-A(x^0)\,(\hat{x}-x^0)) \leq$$
$$\leq \|x^1-\hat{x}\|\cdot\|F(\hat{x})-F(x^0)-A(x^0)\,(\hat{x}-x^0)\| \leq b\|\hat{x}-x^0\|\cdot\|x^1-\hat{x}\|.$$

В итоге получаем, что

$$\|x^1-\hat{x}\| \leq r\|\hat{x}-x^0\|.$$

По определению r приближение x^1 остается в окрестности N. Поэтому мы можем повторить все рассуждения и установить, что

$$\|x^{k+1}-\hat{x}\| \leq r\|x^k-\hat{x}\| \quad \forall k.$$

Это и означает сходимость $\{x^k\}$ к $\hat{x}$, так как $r<1$.

Теорема 4.1 сформулирована в весьма общих предположениях. Приведем более конкретные условия, обеспечивающие неравенство (4.2).

Следствие 4.1. *Пусть X, F, A и $\hat{x}$ — те же, что и в теореме 4.1, причем отображение F непрерывно дифференцируемо в окрестности $\hat{x}$. Если существует $\alpha>0$ такое, что матрица $A(\hat{x})-\alpha E$ положительно полуопределена и в некоторой окрестности N точки $\hat{x}$ выполнены неравенства*

$$\|A(x)-\nabla F(x)\| \leq \delta < \alpha \quad \forall x \in N, \tag{4.3}$$

то заключение теоремы 4.1 верно.

Д о к а з а т е л ь с т в о. По теореме о среднем

$$F(x)-F(y)-A(y)\,(x-y) = (B(x,y)-A(y))^\top (x-y) =$$
$$= (B(x,y)-\nabla F(y))^\top (x-y) + (\nabla F(y)-A(y))^\top (x-y),$$

где $B(x,y)$ — матрица, строки которой есть $\nabla F_i(x+t_i(y-x))$ при некоторых $t_i \in (0,1)$. В силу (4.3) и непрерывной дифференцируемости F, отсюда вытекает условие (4.2) теоремы 4.1.

Преимуществом методов второго порядка является то, что при определенных предположениях они демонстрируют квадратичную скорость сходимости. Пример таких предположений дает следующая теорема.

Теорема 4.2. *Пусть X, F, $\hat{x}$ — те же, что и в теореме 4.1, причем отображение F непрерывно дифференцируемо, а матрица $\nabla F(\hat{x})$ положительно определена. Тогда существует окрестность точки $\hat{x}$ такая, что если точка x^0 выбрана из N, то последовательность (4.1) при $A(x)=\nabla F(x)$ определена и сходится к $\hat{x}$. Более того, если $\nabla F(x)$ локально удовлетворяет условию Липшица, т. е. существует $L>0$ такое, что*

$$\|\nabla F(x)-\nabla F(y)\| \leq L\|x-y\| \quad \forall x,y \in N,$$

то при некотром $\varrho>0$ имеем

$$\|x^{k+1}-\hat{x}\| \leq \varrho\,\|x^k-\hat{x}\|^2 \quad \forall k,$$

что и означает квадратичную скорость сходимости процесса.

Д о к а з а т е л ь с т в о. При доказательстве теоремы 4.1 нами было получено неравенство

$$\alpha(1-\varepsilon)\|x^{k+1}-\hat{x}\|^2 \leq (x^{k+1}-\hat{x})^\top (F(\hat{x})-F(x^k)-\nabla F(x^k)(\hat{x}-x^k)),$$

где $\alpha = \min\limits_{\|y\|=1} y^\top \nabla F(\hat{x})y > 0$, ε — некоторая величина, меньшая 1. Отсюда, по теореме о среднем, вытекает оценка

$$\alpha(1-\varepsilon)\|x^{k+1}-\hat{x}\| \leq$$

$$\leq \sup_{0\leq t\leq 1} \|\nabla F(x^k+t(\hat{x}-x^k))-\nabla F(x^k)\|\cdot\|x^k-\hat{x}\| \leq L\|x^k-\hat{x}\|^2.$$

Из нее мы и заключаем, что скорость сходимости процесса (4.1) квадратичная с $\varrho = L[\alpha(1-\varepsilon)]^{-1}$.

Заметим, что каждый шаг методов второго порядка требует решения вспомогательного вариационного неравенства. Если текущее приближение близко к решению и якобиан $\nabla F(\hat{x})$ положительно определен, отображение F^k вспомогательной задачи сильно монотонно. Таким образом, для ее решения можно опираться на рассмотренные ранее проекционные методы.

§ 5. Проксимальное отображение и его свойства

Пусть X — непустое замкнутое выпуклое подмножество $\mathbb{R}^n$, F — непрерывное и монотонное отображение X в $\mathbb{R}^n$. Отображение, любому $y \in \mathbb{R}^n$ ставящее в соответствие вектор $x(y) \in X$ такой, что

$$(F(x(y))+x(y)-y)^\top (x-x(y)) \geq 0 \quad \forall x \in X, \tag{5.1}$$

называется *проксимальным* (относительно F) и обозначается через $\mathcal{R}_F(\cdot) : \mathbb{R}^n \to X$.

Теорема 5.1. *Пусть X — непустое замкнутое выпуклое подмножество $\mathbb{R}^n$, F — непрерывное и монотонное отображение X в $\mathbb{R}^n$. Тогда проксимальное отображение $\mathcal{R}_F$ определено и однозначно на всем $\mathbb{R}^n$.*

Д о к а з а т е л ь с т в о вытекает из сильной монотонности отображения в задаче (5.1).

Изучим свойства проксимального отображения.

Теорема 5.2. *Проксимальное отображение является монотонным и нерасширяющим, т. е.*

$$\|\mathcal{R}_F(x)-\mathcal{R}_F(y)\| \leq \|x-y\| \quad \forall x,y \in \mathbb{R}^n.$$

Д о к а з а т е л ь с т в о. Пусть $x, y \in \mathbb{R}^n$. По определению проксимального отображения

$$(F(\mathcal{R}_F(x))+\mathcal{R}_F(x)-x)^\top (u-\mathcal{R}_F(x)) \geq 0 \quad \forall u \in X,$$

$$(F(\mathcal{R}_F(y))+\mathcal{R}_F(y)-y)^\top (v-\mathcal{R}_F(y)) \geq 0 \quad \forall v \in X.$$

Подставим сюда $u = \mathcal{R}_F(y)$, $v = \mathcal{R}_F(x)$ и сложим полученное. После перегруппировки слагаемых, ввиду монотонности F, получаем

$$(\mathcal{R}_F(x)-\mathcal{R}_F(y))^\top (x-y) \geq \|\mathcal{R}_F(x)-\mathcal{R}_F(y)\|^2 +$$

$$+(F(\mathcal{R}_F(x)) - F(\mathcal{R}_F(y)))^\top(\mathcal{R}_F(x) - \mathcal{R}_F(y)) \geq$$
$$\geq \|\mathcal{R}_F(x) - \mathcal{R}_F(y)\|^2.$$

Отсюда сразу следует монотонность проксимального отображения, а также его нерасширяемость, поскольку левая часть последнего неравенства оценивается как

$$(\mathcal{R}_F(x) - \mathcal{R}_F(y))^\top(x-y) \leq \|\mathcal{R}_F(x) - \mathcal{R}_F(y)\| \cdot \|x-y\|.$$

Утверждение доказано.

Заметим, что из свойства нерасширяемости проксимального отображения следует его непрерывность.

Теорема 5.3. *Пусть X — непустое замкнутое выпуклое подмножество $\mathbb{R}^n$, F — непрерывное и монотонное отображение X в $\mathbb{R}^n$. Если задача $VI(X, F)$ разрешима и $\hat{x}$ — ее решение, то $\hat{x}$ — неподвижная точка проксимального отображения $\mathcal{R}_F$.*

Д о к а з а т е л ь с т в о. Достаточно подставить $\hat{x}$ в неравенство (5.1) вместо y и убедиться, что оно выполняется при $x(y) = \hat{x}$.

Рассмотрим итерационный процесс, порождаемый проксимальным отображением, а именно пусть

$$x^{k+1} = \mathcal{R}_F(x^k), \quad k = 0, 1, 2, \ldots, \quad x^0 \in \mathbb{R}^n; \tag{5.2}$$

здесь $x^0 \in X$ — произвольная начальная точка. Процесс представляет собой метод простой итерации для нахождения неподвижной точки проксимального отображения. Хотя последнее не обязательно сжимающее, верна

Теорема 5.4. *Пусть X — непустое выпуклое замкнутое множество, отображение $F : X \to R^n$ непрерывно и монотонно. Тогда процесс (5.2) сходится к $\hat{x}$ — некоторому решению $VI(X, F)$, если только последние существуют.*

Д о к а з а т е л ь с т в о. По построению [см.(5.1), (5.2)]

$$(\,F(x^{k+1}) + x^{k+1} - x^k\,)^\top(x - x^{k+1}) \geq 0 \qquad \forall x \in X.$$

Это значит, что $h^{k+1} = -F(x^{k+1}) + x^k - x^{k+1} \in N_X(x^{k+1})$ при всех k. Отсюда $x^k = x^{k+1} + F(x^{k+1}) + h^{k+1}$, и если $\hat{x}$ — произвольное решение $VI(X, F)$, то

$$\begin{gathered}\|x^k - \hat{x}\|^2 = \|x^{k+1} - \hat{x}\|^2 + \|F(x^{k+1}) + h^{k+1}\|^2 + \\ +2(F(x^{k+1}) + h^{k+1})^\top(x^{k+1} - \hat{x}) \geq \\ \geq \|x^{k+1} - \hat{x}\|^2 + \|F(x^{k+1}) + h^{k+1}\|^2 + 2(F(\hat{x}),\, x^{k+1} - \hat{x}) \geq \\ \geq \|x^{k+1} - \hat{x}\|^2 + \|F(x^{k+1}) + h^{k+1}\|^2;\end{gathered} \tag{5.3}$$

здесь использовались монотонность F, включение $h^{k+1} \in N_X(x^{k+1})$ и то, что $\hat{x}$ — решение $VI(X, F)$.

Неравенство (5.3) устанавливает монотонное убывание $\|x^k - \hat{x}\|$ по k и сходимость к нулю $\|F(x^k) + h^k\|$, т. е. $\|x^k - x^{k+1}\| \to 0$. Ограниченность последовательности $\{x^k\}$ влечет существование хотя бы одной ее предельной точки $\check{x} \in X$. Предельным переходом по сходящейся подпоследовательности в (5.1) получаем, что

$$F(\check{x})^\top(x - \check{x}) \geq 0 \quad \forall x \in X,$$

т. е. $\check{x}$ — решение $VI(X,F)$. Отождествляя $\check{x}$ с $\hat{x}$ из (5.3), получаем монотонную сходимость к $\hat{x}$ всей последовательности $\{x^k\}$.

Заметим, что итерационный процесс (5.2) относится к разряду двухуровневых: на каждой его итерации нужно решить вспомогательное вариационное неравенство, требующее своего, вообще говоря, бесконечного вычислительного процесса. Как и в методах второго порядка, мы опускаем здесь анализ влияния погрешности решения вспомогательной подзадачи на сходимость основного процесса.

§ 6. Метод оценочных функций

Пусть дано множество $C \in \mathbb{R}^n$. *Оценочной функцией,* или *функцией оценки* задачи $VI(X,F)$ (соответственно $NCP(F)$), называется неотрицательная функция $M: \ C \to R$, удовлетворяющая условию: точка $\hat{x}$ будет решением $VI(X,F)$ (соответственно $NCP(F)$) в том и только в том случае, когда $\hat{x} \in C$ и $M(\hat{x}) = 0$, т. е. когда точки глобального оптимума задачи

$$\min\{\, M(x): \ x \in C\,\} \tag{6.1}$$

совпадают с решениями $VI(X,F)$ $(NCP(F))$ в предположении разрешимости последней.

Построить функцию оценки не так сложно. Например, для задачи $NCP(F)$ такой функцией является функция $M(x) = F(x)^\top x$, неотрицательная на $C = \{x : F(x) \geq 0, \ x \geq 0\}$, а для задачи $VI(X,F)$ — обычная функция скачка $g(x) = \max\limits_{y \in X} F(x)^\top (y-x)$, заданная на $C = X$. Соответствующие теоремы эквивалентности были приведены в главе 2 §1, 2.

Гораздо сложнее построить оценочную функцию, которая обладала бы рядом свойств, полезных в вычислительном отношении. В частности, только что приведенные в качестве примера функции такими свойствами не обладают, — уже отмечалось, что $g(x)$, вообще говоря, не является гладкой и даже не имеет конструктивно очерченной области определения; вторая из приведенных функций, хотя и может считаться гладкой, порождает задачу, найти точки глобального оптимума которой достаточно сложно.

Свойства, которые выглядят привлекательно с вычислительной точки зрения, таковы:

— M является гладкой функцией;

— любая ее стационарная относительно C точка является точкой глобального минимума в (6.1);

— множества уровня $L(\alpha) = \{x \in C: \ M(x) \leq \alpha\}$ ограничены

и т.п.

Рассмотрим некоторые из известных оценочных функций, обладающих перечисленными свойствами.

Пусть имеется задача $VI(X,F)$, где X — непустое выпуклое замкнутое подмножество $\mathbb{R}^n$, отображение $F: X \to \mathbb{R}^n$ непрерывно. При этих предположениях множество решений $VI(X,F)$ совпадает, как уже отмечалось выше, с множеством неподвижных точек отображения

$$H(x) = P_X(x - \alpha F(x)), \quad \alpha > 0.$$

Нетрудно видеть, что точка $\hat{y} = H(x)$ будет единственным решением задачи

$$\min_{y\in X} \varphi(y), \text{ где } \varphi(y) = F(x)^\top (y - x) + \frac{1}{2\alpha}\|y - x\|^2. \tag{6.2}$$

В самом деле, функция $\varphi(y)$ выпукла и дифференцируема, а ее градиент $\nabla\varphi(y) = F(x) + \frac{1}{\alpha}(y - x)$, вычисленный в точке $\hat{y} = H(x)$, удовлетворяет условиям

$$\nabla\varphi(\hat{y})^\top (y - \hat{y}) \geq 0 \quad \forall y \in X.$$

Очевидно, что оптимальное значение задачи (6.2) не положительно.

Теперь мы подготовлены к описанию свойств функции оценки, называемой функцией Фукушимы. Она определена формулой

$$M_0(x) = -F(x)^\top (F(x) - x) - \frac{1}{2\alpha}\|H(x) - x\|^2. \tag{6.3}$$

Эта функция задана при всех $x \in X$ и принимает на X неотрицательные значения, так как отличается от оптимального значения задачи (6.2) только знаком.

Рассмотрим задачу

$$\min\{\, M_0(x) :\ x \in X \,\}. \tag{6.4}$$

Следующая теорема устанавливает эквивалентность этой задачи задаче решения вариационного неравенства $VI(X, F)$.

Теорема 6.1. *Пусть функция $M_0(x)$ определена в соответствии с* (6.3). *Тогда $M_0(x) \geq 0$ при всех $x \in X$ и $M_0(x) = 0$ тогда и только тогда, когда x — решение задачи $VI(X, F)$.*

Д о к а з а т е л ь с т в о. Функцию $M_0(x)$ можно переписать в виде

$$M_0(x) = \frac{1}{2\alpha}\Big\{\|\alpha F(x)\|^2 - \|H(x) - x + \alpha F(x)\|^2\Big\}.$$

Так как $\|\alpha F(x)\|$ совпадает с расстоянием между x и $x - \alpha F(x)$, а $\|H(x) - x + \alpha F(x)\|$ — с расстоянием между $x - \alpha F(x)$ и проекцией этой точки на X, то $M_0(x) \geq 0$ при всех $x \in X$. Более того, определение $H(x)$ подразумевает, что эти расстояния равны, только если $x = H(x)$. Последнее имеет место в случае, когда x — решение задачи $VI(X,\, F)$.

Заметим, что в случае $X = \mathbb{R}^n_+$, т. е. когда задача $VI(X,\, F)$ превращается в задачу о дополнительности $NCP(F)$, функция Фукушимы приобретает форму

$$M_0(x) = F(x)^\top x + \frac{1}{2\alpha}\,(\|(x - \alpha F(x))^+\|^2 - \|x\|^2).$$

В самом деле, здесь $H(x) = (x - \alpha F(x))^+$ и $(x - \alpha F(x))^\top H(x) = \|H(x)\|^2$. Поэтому

$$-F(x)^\top H(x) - \frac{1}{2\alpha}\,\|H(x) - x\|^2 =$$

$$= -F(x)^\top (x - \alpha F(x))^+ - \frac{1}{2\alpha}\|(x - \alpha F(x))^+ - x\|^2 =$$

$$= \frac{1}{\alpha}\,(x - \alpha F(x))^\top (x - \alpha F(x))^+ - \frac{1}{2\alpha}\,\|(x - \alpha F(x))^+\|^2 - \frac{1}{2\alpha}\,\|x\|^2 =$$

$$= \frac{1}{2\alpha}\,\|(x-\alpha F(x))^{+}\|^2 - \frac{1}{2\alpha}\,\|x\|^2.$$

Теорема 6.1 не накладывает особых условий на то, какими должны быть отображение F и функция M_0. Однако практические соображения требуют, чтобы последняя обладала рядом полезных вычислительных свойств. К счастью, оказывается, что если X — непустое выпуклое замкнутое множество, то функция M_0 непрерывно дифференцируема, если дифференцируемо отображение F.

Теорема 6.2. *Если отображение F непрерывно, а множество X непусто, выпукло и замкнуто, то функция M_0, определяемая соотношениями* (6.3), *непрерывна также. Далее, если отображение F непрерывно дифференцируемо, то функция M_0 также непрерывно дифференцируема и ее градиент равен*

$$\nabla M_0(x) = F(x) - \Big(\nabla F(x) - \frac{1}{\alpha}E\Big)\,(H(x)-x). \tag{6.5}$$

Д о к а з а т е л ь с т в о. Если отображение F непрерывно, функция M_0 непрерывна как суперпозиция непрерывных функций. Чтобы доказать вторую часть утверждения, рассмотрим функцию $h:\ \mathbb{R}^n \times X \to R$, определенную формулой

$$h(x,y) = F(x)^{\top}(y-x) + \frac{1}{2\alpha}\,\|y-x\|^2.$$

Если отображение F — непрерывно дифференцируемо, функция h также будет непрерывно дифференцируемой по x и ее градиент равен

$$\nabla_x g(x,y) = F(x) - (\nabla F(x) - \frac{1}{\alpha}E)(y-x). \tag{6.6}$$

По определению

$$M_0(x) = -\min_{y\in X} h(x,y),$$

причем минимум достигается в единственной точке $\hat{y} = H(x)$. В соответствии с классическими теоремами о производных функции оптимума параметрических задач оптимизации это обеспечивает дифференцируемость M_0 и равенство

$$\nabla M_0(x) = -\nabla_x h(x,\,H(x)).$$

Формула (6.5) следует отсюда с учетом (6.6).

Взглянем на связь задачи (6.4) с задачей решения вариационного неравенства $VI(X,\,F)$ более пристально. Теорема 6.1 говорит, что решение $VI(X,\,F)$ сводится к отысканию точек глобального оптимума в (6.4). Но поскольку функция M_0, вообще говоря, не выпукла, задача (6.4) может иметь точки локального минимума и стационарные точки, которые не являются точками глобального минимума. К счастью, перечисленные трудности исчезают в случае, когда отображение F непрерывно дифференцируемо и матрица $\nabla F(x)$ положительно определена при всех $x\in X$, т. е. $y^{\top}\nabla F(x)y > 0 \quad \forall y\in\mathbb{R}^n,\ \forall x\in X$.

Теорема 6.3. *Пусть X — непустое выпуклое замкнутое подмножество $\mathbb{R}^n$, отображение $F:\ \mathbb{R}^n \to \mathbb{R}^n$ непрерывно дифференцируемо и матрица Якоби $\nabla F(x)$ положительно определена при всех $x\in X$. Если $\hat{x}$ — стационарная точка в задаче* (6.4), *т. е.*

$$\nabla M_0(\hat{x})^{\top}(y-\hat{x}) \ge 0 \quad \forall y\in X, \tag{6.7}$$

то она является также точкой ее глобального оптимума и, следовательно, решением задачи $VI(X,\,F)$.

Д о к а з а т е л ь с т в о. Пусть $\hat{x}$ удовлетворяет условиям (6.7). Подставляя в эти условия формулу (6.6) и полагая $y = H(\hat{x})$, получаем после перегруппировки слагаемых, что

$$\Big(F(\hat{x}) + \frac{1}{\alpha}(H(\hat{x}) - \hat{x})\Big)^\top (H(\hat{x}) - \hat{x}) \geq (H(\hat{x}) - \hat{x})^\top \nabla F(\hat{x})\,(H(\hat{x}) - \hat{x}).$$

Левая часть этого неравенства неположительна. Это следует из соотношения

$$(\hat{x} - \alpha F(\hat{x}) - H(\hat{x}))^\top (y - H(\hat{x})) \leq 0 \qquad \forall y \in X,$$

которому $H(\hat{x})$ удовлетворяет как проекция точки $\hat{x} - \alpha F(\hat{x})$ на выпуклое множество X. Полагая здесь $y = \hat{x}$, получаем искомое неравенство

$$\Big(F(\hat{x}) + \tfrac{1}{\alpha}(H(\hat{x}) - \hat{x})\Big)^\top (H(\hat{x}) - \hat{x}) = $$
$$= \tfrac{1}{\alpha}\Big(\hat{x} - \alpha F(\hat{x}) - H(\hat{x})\Big)^\top (\hat{x} - H(\hat{x})) \leq 0.$$

Таким образом, $(H(\hat{x}) - \hat{x})^\top \nabla F(\hat{x})\,(H(\hat{x}) - \hat{x}) \leq 0$, что, в силу положительной определенности матрицы Якоби, влечет равенство $\hat{x} - H(\hat{x}) = 0$, т. е. $\hat{x}$ действительно оказывается решением $VI(X,\,F)$, а значит, по теореме 6.1 и точкой глобального оптимума в задаче (6.4).

Пусть X — непустое выпуклое замкнутое подмножество $\mathbb{R}^n$, отображение $F : \mathbb{R}^n \to \mathbb{R}^n$ непрерывно дифференцируемо. По теореме 6.2 целевая функция M_0 задачи (6.4) также непрерывно дифференцируема и ее градиент может быть найден из (6.5). Поэтому для решения задачи (6.4) можно применять любой из существующих алгоритмов градиентного типа. Ниже, однако, мы остановимся на возможности использования в качестве направления спуска вектора

$$d = H(x) - x = P_X(x - \alpha F(x)) - x, \quad \alpha > 0. \tag{6.8}$$

Поскольку соотношение (6.8) не содержит $\nabla F(x)$, вектор d может быть определен с гораздо меньшими вычислительными затратами по сравнению с градиентным направлением ∇M_0, особенно в случаях, когда вычисление матрицы ∇F затруднено. Следующее утверждение показывает, что при условии монотонности F вектор d, задаваемый соотношениями (6.8), действительно является направлением спуска функции M_0 в точке x.

Теорема 6.4. *Пусть X — непустое выпуклое замкнутое подмножество $\mathbb{R}^n$, отображение $F : \mathbb{R}^n \to \mathbb{R}^n$ непрерывно дифференцируемо и матрица Якоби $\nabla F(x)$ положительно определена при всех $x \in X$. Тогда при любом $x \in X$ вектор $d = H(x) - x$ удовлетворяет условию*

$$\nabla M_0(x)^\top d < 0, \tag{6.9}$$

если только $d \neq 0$. Более того, если отображение F сильно монотонно на X с константой $\gamma > 0$, т. е.

$$(F(x) - F(y))^\top (x - y) \geq \gamma \|x - y\|^2$$

при всех $x, y \in X$, то при всех $x \in X$ также

$$\nabla M_0(x)^\top d < -\gamma \|d\|^2. \tag{6.10}$$

Д о к а з а т е л ь с т в о. Из (6.6) и (6.8) вытекает, что

$$\nabla M_0(x)^\top d = \Big(F(x) + \frac{1}{\alpha}\,(H(x)-x)\Big)^\top (H(x)-x) -$$
$$- (H(x)-x)^\top \nabla F(x)\,(H(x)-x). \tag{6.11}$$

Но поскольку $H(x) = P_X(x - \alpha F(x))$, то

$$(x - \alpha F(x) - H(x))^\top (y - H(x)) \le 0 \qquad \forall y \in X.$$

Отсюда

$$(F(x) + \frac{1}{\alpha}\,(H(x)-x))^\top (H(x)-x) \le 0. \tag{6.12}$$

Комбинируя (6.12) и (6.11), получаем

$$\nabla M_0(x)^\top d \le -(H(x)-x)^\top \nabla F(x)\,(H(x)-x) = -d^\top \nabla F(x) d.$$

Таким образом, если матрица Якоби $\nabla F(x)$ положительно определена, то $\nabla M_0(x)^\top d < 0$ при $d \ne 0$. Если же отображение F сильно монотонно с константой $\gamma > 0$, то условие (6.10) просто следует из того, что $d^\top \nabla F d \ge \gamma \|d\|^2$.

Следующая теорема устанавливает глобальную сходимость алгоритма, использующего направление спуска $d = H(x) - x$ вкупе с точным правилом линейного поиска.

Теорема 6.5. *Пусть $\{x^k\}$ — последовательность, порождаемая рекуррентными соотношениями*

$$x^{k+1} = x^k + t_k d^k, \qquad k = 0, 1, 2, \dots\,,$$

где $d^k = H(x^k) - x^k$, а шаговые параметры $t_k \in [0,1]$ определены из условия

$$M_0(x^k + t_k d^k) = \min\{M_0(x^k + t d^k) :\ 0 \le t \le 1\}. \tag{6.13}$$

Предположим также, что X — выпуклый компакт и отображение $F:\ \mathbb{R}^n \to \mathbb{R}^n$ непрерывно дифференцируемо, причем матрица $\nabla F(x)$ положительно определена при всех $x \in X$. Тогда независимо от начального приближения $x^0 \in X$ последовательность $\{x^k\}$ лежит в X и сходится к $\hat{x}$ — единственному решению задачи $VI(X,F)$.

Д о к а з а т е л ь с т в о. Заметим вначале, что непрерывность отображения F вместе с условием компактности X гарантирует разрешимость задачи $VI(X,F)$, а условие положительной определенности матрицы $\nabla F(x)$ — единственность ее решения.

Далее, так как $t_k \in [0,1]$ и множество X — выпукло, то из включений $x^k \in X$, $H(x^k) = x^k + d^k \in X$ следует, что $x^{k+1} \in X$. Но $x^0 \in X$, и потому вся последовательность $\{x^k\}$ лежит в X. Следовательно, любая предельная точка $\bar{x}$ этой последовательности также лежит в X.

Предположим теперь от противного, что $\bar{d} = H(\bar{x}) - \bar{x} \ne 0$. Тогда $\bar{d}$ — направление спуска функции M_0 в точке $\bar{x}$. Поэтому

$$\bar{\delta} = M_0(\bar{x} + \bar{t}\bar{d}) - M_0(\bar{x}) < 0,$$

где величина $\bar{t}$ определена из условия

$$M_0(\bar{x} + \bar{t}\,\bar{d}) = \min\{M_0(\bar{x} + t\,\bar{d}) :\ 0 \le t \le 1\}.$$

Воспользуемся теперь свойством непрерывности отображений F и H, а также функции M_0 и отображения, ставящего каждой паре (x, d) величину шага t по правилу (6.13). Для сходящейся к $\bar{x}$ подпоследовательности $\{x^{k_S}\}$ имеем

$$\lim_{(s)} \Big(M_0(x^{k_S} + t_{k_S}\, d^{k_S}) - M_0(x^{k_S}) \Big) = \bar{\bar{\delta}} < 0.$$

Полученное строгое неравенство и очевидная монотонность последовательности значений $M_0(x^k)$ противоречит ограниченности последней снизу.

Рассмотренный метод тесно связан с методом простой итерации из §3 данной главы, поскольку и там и здесь участвует одно и то же направление $d = H(x) - x$. Однако в методе простой итерации параметр $\alpha > 0$ постоянен и не связан с минимизацией какой бы то ни было функции.

В заключение раздела приведем (без подробного анализа вычислительных свойств) две другие функции оценки, предназначенные для сведения нелинейной задачи о дополнительности $NCP(F)$ к обычной задаче безусловной оптимизации типа $\min\limits_x M(x)$. Первая известна как функция Мангасарьяна—Солодова. Она имеет вид

$$M_1(x) = F(x)^\top x + \tfrac{1}{4}\, (\|(x - 2F(x))^+\|^2 - \\ - \|x\|^2 + \|(F(x) - 2x)^+\,\|^2 - \|F(x)\|^2)\,.$$

Эта функция непрерывно дифференцируема и, если якобиан $\nabla F(x)$ отображения F положительно определен при всех $x \in \mathbb{R}^n$, любая стационарная точка функции M_1 относительно всего пространства является ее точкой глобального минимума и решением задачи $NCP(F)$, если последняя разрешима.

Вторая функция, функция Фишера, имеет вид

$$M_2(x) = \sum_{i=1}^{n} \varphi(x_i, F_i(x))^2,$$

где $\varphi(a,b) = \sqrt{a^2 + b^2} - (a + b)$. Нетрудно заметить, что

$$\varphi(a,b) = 0 \iff a \geq 0, b \geq 0, ab = 0.$$

Функция M_2 также непрерывно дифференцируема и в случае, когда отображение F представляет собой P_0 — функцию, т. е. когда для любых x и $y \in \mathbb{R}^n$, таких, что $x \neq y$, существует индекс i такой, что

$$x_i \neq y_i,\ (x_i - y_i)\,(F_i(x) - F_i(y)) \geq 0,$$

любая стационарная точка функции M_2 относительно всего пространства дает точку ее глобального минимума и соответственно решение задачи $NCP(F)$.

Глава 3

ПОИСК ЭКОНОМИЧЕСКОГО РАВНОВЕСИЯ

Перейдем к обсуждению важнейшей области приложения теории и методов решения конечномерных вариационных неравенств и задач о дополнительности — задачам поиска равновесия в бескоалиционных играх n лиц (сторон). В математической экономике этому соответствуют задачи поиска общего экономического равновесия, прогнозирование транспортных потоков и тарифов на грузоперевозки и многие другие.

§ 1. Равновесие по Нэшу в играх n лиц

Рассмотрим бескоалиционную игру n лиц, в которой каждый из участников с номером $i \in \{1, \ldots, n\}$ характеризуется своим множеством допустимых стратегий $X_i \subseteq \mathbb{R}^{m_i}$, $m_i \geq 1$, и своей функцией полезности $u_i(x_1, \ldots, x_n) : \prod\limits_{i=1}^{n} X_i \to \mathbb{R}$, значение которой он стремится максимизировать.

Обозначим

$$N = \{1, \ldots, n\}, \; x_{N\setminus\{i\}} = (x_j : j \in N, \; j \neq i), \; X = \prod_{i=1}^{n} X_i.$$

Точка $\hat{x} = (\hat{x}_1, \ldots, \hat{x}_n) \in X$ называется *точкой равновесия по Нэшу* в бескоалиционной игре n лиц $NE(X, u)$, если для всех ее участников

$$u_i(\hat{x}_i, \hat{x}_{N\setminus\{i\}}) \geq u_i(x_i, \hat{x}_{N\setminus\{i\}}) \quad \forall x_i \in X_i,$$

т. е. ни один из них не может в одиночку улучшить значение своей функции полезности.

Следующий результат о сведении задачи поиска равновесия по Нэшу к некоторому вариационному неравенству может быть легко установлен суммированием необходимых условий оптимальности первого порядка для задач максимизации функции полезности каждого из участников игры. При этом полное допустимое множество в задаче решения вариационного неравенства представляет собой декартово произведение множеств допустимых стратегий участников игры.

Теорема 1.1. *Если все X_i — непустые замкнутые выпуклые подмножества $\mathbb{R}^{m_i}$, где $m_i \geq 1$, и функции полезности всех игроков $u_i : X \to \mathbb{R}$ непрерыв-*

но дифференцируемы и псевдовогнуты по x_i, *то точка* $\hat{x} \in X$ *будет точкой равновесия по Нэшу в том и только в том случае, когда*

$$\sum_{i=1}^{n} F_i(\hat{x})^{\top}(x_i - \hat{x}_i) \geq 0 \quad \forall x \in X,$$

т. е. когда эта точка является решением вариационного неравенства $VI(X, F)$, *определяемого множеством* X *и отображением*

$$F = (F_i : i \in N), \; F : X \to \prod_{i=1}^{n} \mathbb{R}^{m},$$

где $F_i : X \to \mathbb{R}^{m_i}$, $F_i\,(x) = -\nabla_{x_i} u_i(x)$.

Таким образом, задача поиска равновесия по Нэшу весьма естественно формулируется в терминах теории вариационных неравенств. Существование и единственность равновесия по Нэшу может быть установлена при помощи результатов раздела 2. В частности, наиболее известный результат для бескоалиционных игр является простым следствием теоремы 2.1.

Теорема 1.2. *Пусть* X_i *— непустые компактные выпуклые подмножества* $\mathbb{R}^{m_i}$, $m_i \geq 1$, *и функции полезности* $u_i : X \to \mathbb{R}$ *непрерывно дифференцируемы и псевдовогнуты по* x_i, $i = 1, \ldots, n$. *Тогда решение задачи* $NE(X, u)$ *существует.*

Необходимость численного решения задач о равновесии по Нэшу была и остается главным побудительным мотивом развития алгоритмов для задач о дополнительности и вариационных неравенств.

Перейдем к рассмотрению содержательных игровых постановок экономического характера.

§ 2. Прогнозирование потоков в транспортных сетях

Задача прогнозирования грузовых потоков в транспортной сети дает сильный побудительный мотив к развитию алгоритмов решения вариационных неравенств *большой размерности*. Модель, лежащая в основе этой задачи, использует понятие равновесия по Вардропу. Последнее есть просто поведенческий принцип, утверждающий, что пользователи сети (водители автомашин) ведут некооперативное соревнование за сетевые ресурсы с целью минимизации индивидуальных дорожных расходов. В итоге задача прогнозирования грузовых потоков может рассматриваться как частный случай поиска равновесия по Нэшу в некоторой игре n лиц.

Дадим обозначения, необходимые для постановки рассматриваемой задачи:

$G(V, A)$ — транспортная сеть, описываемая множествами своих вершин V и дуг A;

W — множество выделенных пар вершин «поставщик—потребитель», т. е. пар $w = (i, j) \in W$;

P_w — множество путей, соединяющих пару $w \in W$;

$P = \bigcup\limits_{w \in W} P_w$ — совокупность всех путей P_w в сети;

$f = (f_a : a \in A)$, где f_a — грузовой поток вдоль дуги $a \in A$;

$h = (h_p : p \in P)$, где h_p — грузовой поток вдоль пути $p \in P$;

$$\delta_{ap} = \begin{cases} 1, & \text{если путь } p \in P \text{ содержит дугу } a \in A; \\ 0 & \text{— в противном случае;} \end{cases}$$

$\mathcal{M} = (\,\delta_{ap}\,)$ — матрица инцидентности дуг и путей;

$c(f) = (c_a(f) : a \in A)$, где $c_a(f)$ — функция средних транспортных расходов для дуги $a \in A$;

$C(h) = (C_p(h) : p \in P)$, где $C_p(h)$ — функция средних транспортных расходов для пути $p \in P$;

$u = (u_w : w \in W)$, где u_w —минимальные транспортные расходы на перевозку груза для пары вершин «поставщик—потребитель» $w \in W$, т. е. $u_w = \min\limits_{p \in P_w} \{C_p(h)\}$;

$T(u) = \{T_w(u) : w \in W\}$, где $T_w(u)$ — функция спроса на транспортировку груза для пары вершин «поставщик—потребитель» $w \in W$.

Очевидно, что потоки и затраты вдоль дуг связаны с потоками и затратами вдоль путей транспортной сети линейно:

$$f = \mathcal{M}h, \; C(h) = \mathcal{M}^{\top} c\,(f).$$

Принцип равновесия Вардропа постулирует, что водители автомашин, перевозящие груз от источника к потребителю, выбирают пути, вдоль которых их индивидуальные затраты на перевозку груза минимальны. Таких путей может быть несколько, но пути с затратами, превосходящими минимальные, не используются.

Пара «поток–затраты» $(\hat{f}, \hat{u})$ образует *точку равновесия по Вардропу*, если выполнены следующие условия:

$$h_p\,(C_p(h) - u_w) = 0, \; C_p\,(h) - u_w \geq 0, \; h_p \geq 0 \quad \forall w \in W, \; p \in P_w;$$

$$\sum_{p \in P_w} h_p - T_w(u) = 0, \quad u_w \geq 0 \quad \forall w \in W.$$

Задачу поиска равновесия по Вардропу обозначим $UE(c, T)$.

Если бы не условия неотрицательности для u_w, задача $UE(c, T)$ являлась бы просто смешанной нелинейной задачей о дополнительности.

Предполагая, однако, транспортные расходы вдоль всех путей и дуг строго положительными, а функции спроса — неотрицательными, задачу $UE(c, T)$ можно переформулировать как «чистую» нелинейную задачу о дополнительности.

Теорема 2.1. *Пусть $c_a(f) > 0$ при всех $a \in A$ и $T_w(u) \geq 0$. Пара $(\hat{f}, \hat{u})$ будет решением задачи $UE(c, T)$ в том и только в том случае, когда $(\hat{f}, \hat{u})$ решает следующую задачу $NCP(F)$:*

$$\sum_{p \in P} F_p(h, u) h_p + \sum_{w \in W} F_w(h, u) u_w = 0,$$

$$F_p(h, u) \geq 0, \; h_p \geq 0, \; F_w(h, u) \geq 0, \; u_w \geq 0,$$

где

$$F_p(h, u) = C_p(h) - u_w, \; F_w(h, u) = \sum_{p \in P_w} h_p - T_w(u),$$

$$f = \mathcal{M}h, \quad C(h) = \mathcal{M}^{\top} c(f).$$

Таким образом, нелинейная задача о дополнительности есть весьма естественная формулировка равновесия по Вардропу, а введенные выше необходимые условия этого достаточно слабые. Однако задача имеет очень большую размерность и, в частности, требует перенумерации всех путей в транспортной сети.

Последнюю проблему можно решить в случае, когда функция $T(u)$ имеет обратную, которую обозначим $u = \Phi(T)$. Теперь задачу $UE(c,T)$ можно записать как задачу решения вариационного неравенства на полиэдральном множестве X, в котором переменные, характеризующие пути, трактуются как скрытые.

Теорема 2.2. *Пусть функция $T(u)$ обратима. Пара $(\hat{f},\hat{T})$ есть решение задачи $UE(c,T)$ в том и только в том случае, когда $(\hat{f},\hat{T})$ решает следующую задачу $VI(X, F = (c, -\Phi))$:*

$$c\,(\hat{f})^\top(f-\hat{f}) - \Phi\,(\hat{T})^\top(T-\hat{T}) \geq 0 \quad \forall (f,T) \in X,$$

где

$$X = \Big\{(f,T) : f = \mathcal{M}h,\ \sum_{p\in P_w} h_p = T_w \quad \forall w \in W,\ h \geq 0,\ T \geq 0\Big\}.$$

Определенное здесь множество X является полиэдральным конусом. В случае фиксированного спроса (т. е. при $T(u) = T$) обратная функция Φ исчезает из постановки вообще, а множество X обращается просто в многогранник (т. е. ограниченное полиэдральное множество). Сама задача превращается при этом просто в серию задач поиска кратчайшего пути в сети.

Существование и единственность равновесия в $UE(c,T)$ можно установить, применяя различные теоремы из главы 2 §2. В частности, справедливо следующее утверждение.

Теорема 2.3. *Пусть все $c_a(f)$, $a \in A$, — строго положительные и непрерывные функции, отделенные от нуля, и пусть $T_w(u)$ — строго положительные, непрерывные функции, ограниченные сверху при всех $w \in W$. Тогда решение $NCP(F)$ из теоремы 2.1 существует и единственно.*

Заметим, что единственной является пара $(\hat{f},\hat{u})$; что касается потоков h, то они редко являются единственными в задаче определения грузовых потоков.

Важным расширением рассмотренной выше задачи является задача развития транспортной сети. В последней оптимизируется некоторая мера производительности транспортной сети, пользователи которой подчиняются поведенческому принципу Вардропа. Важность задачи развития можно проиллюстрировать известным парадоксом Браеса: при неудачном добавлении в сеть новой дуги ее производительность может снизиться(!). Поэтому следует аккуратно подходить к вопросу о включении в транспортную сеть новых и удалению из нее старых дуг.

§ 3. Пространственное равновесие цен

Другое обобщение задачи определения грузовых потоков в сети можно получить, если заменить функции спроса $T_w(u)$ на более базовые понятия спроса—предложения товара в каждой из ее вершин. Введем необходимые обозначения:

$\pi = (\pi_i : i \in V)$, где π_i — цена некоторого однородного товара в вершине $i \in V$;

$D(\pi) = (D_i(\pi) :\ i \in V)$, где $D_i(\pi)$ — спрос на этот товар в вершине $i \in V$;

$S(\pi) = (S_i(\pi) : \ i \in V)$, где $S_i(\pi)$ — предложение этого товара в вершине $i \in V$.

Принцип пространственного равновесия цен утверждает, что конкуренция при перевозке товаров между регионами приводит к отсутствию положительной прибыли при перевозках; если же эта прибыль отрицательна, движение товаров отсутствует. Приведем точное математическое определение.

Пара «поток—цена» $(\hat{f}, \hat{\pi})$ описывает *пространственное равновесие цен*, если

$$S_i(\pi) - D_i(\pi) + \sum_{w=(k,i)\in W} \sum_{p\in P_w} h_p \ - \sum_{w=(i,j)\in W} \sum_{p\in P_w} h_p = 0, \ \pi_i \geq 0, \ \forall_i \in V,$$

$$(\pi_i + C_p(h) - \pi_j)\, h_p = 0, \quad \pi_i + C_p(h) - \pi_j \geq 0, \quad h_p \geq 0,$$
$$\forall w = (i,j) \in W, \quad p \in P_w.$$

Задачу поиска пространственного равновесия цен будем обозначать через $SPE(c, S, D)$.

Предполагая, что функции затрат строго положительны, а функции спроса и предложения удовлетворяют условию

$$\pi_i = 0 \Rightarrow D_i(\pi) \geq S_i(\pi), \quad \forall_i \in V, \tag{3.1}$$

задачу $SPE(c, SD)$ можно переформулировать как нелинейную задачу о дополнительности.

Теорема 3.1. *Пусть функции* $c_a(f)$, $a \in A$, *строго положительны и выполнены условия* (3.1). *Пара* $(\hat{f}, \hat{\pi})$ *будет решением* $SPE(c, S, D)$ *в том и только в том случае, когда* $(\hat{f}, \hat{\pi})$ *решает следующую задачу* $NCP(F)$:

$$\sum_{i\in V} F_i(h,\pi)\pi_i + \sum_{w\in W} \sum_{p\in P_w} F_p(h,\pi)h_p = 0,$$

$$F_i(h,\pi) \geq 0, \quad \pi_i \geq 0, \quad F_p(h,\pi) \geq 0, \quad h_p \geq 0,$$

где

$$f = \mathcal{M}h, \quad C(h) = \mathcal{M}^\top c(f),$$
$$F_i(h,\pi) = S_i(\pi) - D_i(\pi) + \sum_{w=(k,i)\in w} \sum_{p\in P_w} h_p \ - \sum_{w=(i,j)\in w} \sum_{p\in P_w} h_p,$$
$$F_p(h,\pi) = \pi_i + C_p\,(h) - \pi_j \quad \text{при всех} \ \ p \in P_{(i,j)}.$$

Как и в случае задачи определения грузовых потоков, при предположении об обратимости функций $S(\pi)$ и $D(\pi)$ задачу $SPE(c, S, D)$ можно представить как задачу решения некоторого вариационного неравенства относительно полиэдрального множества.

Теорема 3.2. *Пусть функции* $S\,(\pi)$ *и* $D(\pi)$ *обратимы, причем обратные к ним функции* $\Psi(S)$ *и* $\Theta(D)$ *соответственно существуют и неотрицательны. Тройка* $(\hat{f}, \hat{S}, \hat{D})$ *порождает равновесие в задаче* $SPE(c, S, D)$ *тогда и только тогда, когда она решает следующую задачу* $VI(X, F = (\Psi, c, -\Theta))$:

$$\Psi(\hat{S})^\top (S - \hat{S}) + c\,(\hat{f})^\top (f - \hat{f}) - \Theta(\hat{D})^\top (D - \hat{D}) \geq 0 \quad \forall (f, S, D) \in X,$$

где

$$X = \Big\{ (f, S, D) : S_i - D_i + \sum_{w=(k,i)\in W} \sum_{p\in P_w} h_p \ - \sum_{w=(i,j)\in W} \sum_{p\in P_w} h_p = 0$$
$$\forall i \in V, \ f = \mathcal{M}h, \ S \geq 0, \ D \geq 0, \ h \geq 0 \Big\}.$$

Приведем один из результатов о существовании и единственности решения $SPE(c, S, D)$. Он получается как следствие теоремы 2.6 из главы 2.

Теорема 3.3. *Пусть функции* $c_a(f)$, $a \in A$, *строго положительны, непрерывны и отделены от нуля, а функции* $S(\pi)$ *и* $D(\pi)$ *непрерывны, удовлетворяют условию* (3.1) *и следующему условию: существует вектор цен* $\hat{\pi}$ *такой, что*

$$\pi_i \geq \hat{\pi}_i \Rightarrow S_i(\pi) \geq D_i(\pi).$$

Тогда решение задачи $NCP(F)$ *из теоремы* 3.1 *существует и единственно.*

Отметим, что рассмотренные модели часто входят в качестве составных частей в более общие математические постановки и имеют большое значение при планировании и управлении развитием сложных городских транспортных сетей.

§ 4. Общее экономическое равновесие по Вальрасу

Модель, к рассмотрению которой мы переходим, предназначена для прогнозирования экономической активности в замкнутой экономике, т. е. для вычисления равновесных цен и объемов производства различного вида товаров и услуг, которые ее составляют. Эта хорошо известная модель является основой большинства современных экономико-математических исследований и доказала свою пригодность для изучения вопросов ценообразования, для анализа международной торговли, энергетической политики и др. Она имеет весьма богатую историю теоретических исследований и приложений.

Из различных математических формализаций модели общего экономического равновесия остановимся на следующей, предложенной Масиесеном. Пусть

M — множество различных продуктов (товаров), производимых в рамках некоторой экономики;

N — множество технологических способов (отраслей, предприятий и т.п.), ее составляющих;

x_j — интенсивность использования i-го технологического способа (отрасли, предприятия), $x = (x_j : j \in N)$;

π_i — цена i-го продукта, $\pi = (\pi_i : i \in M)$;

b_i — начальный запас i-го продукта, $b = (b_i : i \in M)$;

c_j — затраты на реализацию единичной интенсивности i-го технологического способа (оплата труда и т.д.), $c = (c_j : j \in N)$;

$d_i(\pi)$ — функция спроса на i-й продукт, $d(\pi) = (d_i(\pi); i \in M)$;

$A(\pi) = (a_{ij}(\pi))$ — технологическая матрица Леонтьева, ее положительные элементы $a_{ij}(\pi)$ равны объему выпуска i-го продукта при единичной интенсивности реализации j-го технологического способа, а отрицательные — объему его затрат.

При помощи этих обозначений модель общего экономического равновесия определяется следующим образом.

Пара «цена—интенсивности» $(\hat{\pi}, \hat{x})$ образует *точку общего экономического равновесия* $GE(b, c, d, A)$, если она удовлетворяет следующим соотношениям:

1) $c - A(\pi)^\top \pi \geq 0$ — бесприбыльность всех технологических способов;

2) $b + A(\pi)x - d(\pi) \geq 0$ — обеспечение спроса на все виды продуктов;

3) $\pi \geq 0, x \geq 0$ — неотрицательность всех цен и интенсивностей реализации всех технологических способов;

4) $(c - A(\pi)^\top \pi)^\top x = 0$ — технологические способы с отрицательной прибылью не используются (используются с нулевой интенсивностью);

5) $(b + A(\pi)x - d(\pi))^\top \pi = 0$ — закон Вальраса — товары, произведенные в избытке, имеют нулевую цену.

Нетрудно видеть, что условия 1 – 5 прямо переводятся в нелинейную задачу о дополнительности.

Теорема 4.1. *Пара $(\hat{\pi}, \hat{x})$ является точкой общего равновесия в задаче $GE(b, c, d, A)$ в том и только в том случае, когда $(\hat{\pi}, \hat{x})$ решает следующую задачу $NCP(F)$:*

$$F_x(\pi, x)^\top x + F_\pi(\pi, x)^\top \pi = 0,$$

$$F_x(\pi, x) \geq 0, \quad \pi \geq 0, \quad F_\pi(\pi, x) \geq 0, \quad x \geq 0,$$

где

$$F_x(\pi, x) = c - A(\pi)^\top \pi, \quad F_\pi(\pi, x) = b + A(\pi)x - d(\pi).$$

Приведенная формулировка включает в себя как цены, так и интенсивности применения технологических способов. В случае когда технологическая матрица $A(\pi)$ не зависит от цен, т. е. $A(\pi) \equiv A$, можно дать сокращенную формулировку задачи поиска точки общего экономического равновесия в виде некоторого вариационного неравенства над полиэдральным множеством относительно только вектора цен.

Теорема 4.2. *Вектор цен $\hat{\pi}$ есть решение задачи $VI(X, F)$, где*

$$X = \{\pi : c - A^\top \pi \geq 0, \quad \pi \geq 0\}, \quad F(\pi) = b - d(\pi),$$

в том и только в том случае, когда существует вектор интенсивностей $\hat{x}$ такой, что пара $(\hat{\pi}, \hat{x})$ образует точку общего экономического равновесия $GE(b, c, d, A)$; при этом $\hat{x}$ есть вектор множителей Лагранжа ограничений общего вида, определяющих множество X.

Литература по модели общего экономического равновесия содержит огромное число результатов по существованию и единственности ее решения. Один простой результат может быть получен как непосредственное следствие теоремы 1.1. Поскольку функции спроса однородны степени ноль, то без потери общности можно нормализовать все цены. Определим

$$\bar{X} = X \cap \Big\{\pi : \sum_{i \in M} \pi_i = 1\Big\}.$$

Теорема 4.3. *Пусть $d(\pi)$ — непрерывная функция такая, что $d(\alpha\pi) = d(\pi) \quad \forall \alpha \in \mathbb{R}_+$. Тогда существует решение $\hat{\pi}$ задачи $VI(\bar{X}, F)$, определенной в теореме 4.2.*

Среди численных методов решения общей задачи о равновесии проекционные методы выделяются своей интересной экономической интерпретацией. Переопределение итерационной точки в них ведется отдельно для вектора интенсивностей использования технологических способов и отдельно для вектора цен на продукцию. Нетрудно проверить, что направление смещения для первого из них дается вектором $A(\pi)^\top \pi - c$, а для второго — вектором $d(\pi) - b - A(\pi)x$. Таким образом, на каждом шаге вычислительного процесса несколько возрастает интенсивность использования технологических способов, характеризующихся положительной прибылью (в текущих ценах на продукцию), и снижается интенсивность использования технологических способов, прибыль которых отрицательна. Одновременно растут цены на продукты, пользующиеся повышенным спросом, и падают цены на продукты, остающиеся в избытке. Жесткие условия на свойства отображений и шаговые параметры, при которых имеет место сходимость итерационных последовательностей, являются адекватным отражением известной и признаваемой в экономической науке неустойчивости классического рыночного механизма регулирования спроса и предложения и необходимости дополнительных мер по его стабилизации (например, со стороны государства).

Упражнение

Обсудите условия сходимости и экономическую интерпретацию регуляризованных и экстраполяционных методов первого порядка решения вариационных неравенств применительно к задаче общего экономического равновесия.

§ 5. Многоуровневое принятие решений

Пусть заданы X — непустое подмножество пространства $\mathbb{R}^n$ и $\{F_1, F_2, \ldots, F_m\}$ — конечный упорядоченный набор отображений X в $\mathbb{R}^n$. Рассмотрим *лексикографическую* задачу решения вариационного неравенства $LVI(X, F_1, \ldots, F_m)$, которая заключается в поиске элементов множества X_m, последнего в следующем рекурсивном ряду множеств

$$X_{i+1} = \operatorname{Arg} VI(X_i, F_{i+1}), \quad i = 0, 1, \ldots, m-1; \tag{5.1}$$

здесь, как и ранее, $\operatorname{Arg} VI$ — множество решений соответствующего вариационного неравенства (обычного), $X_0 = X$.

Лексикографические вариационные неравенства возникают как естественное обобщение классических задач лексикографической (последовательной) оптимизации, а также как обобщение аппарата, развитого разными авторами при анализе обычных оптимизационных задач с ограничениями в форме вариационных неравенств, в задачах поиска аппроксимационных корней монотонных отображений, задачах оптимальной коррекции неразрешимых (несобственных) минимаксных задач и ряде других.

Рассмотрим, например, проблематику двухуровневой технологии принятия решений. Двухуровневое программирование изучает задачи принятия решений несколькими неравноценными партнерами, один из которых является лидером, а все прочие — ведомыми. Лидер оптимизирует свою целевую функцию (функцию *верхнего уровня*) при ограничениях, включающих в себя реакцию ведомой стороны (сторон) на его решения. Реакция ведомой стороны описывается задачей математического программирования *нижнего уровня*, целевая функция и ограничения которой параметрически зависят от переменных верхнего уровня. В более строгой записи задача двухуровневого программирования выглядит следующим образом:

$$\min_x \{F(x,y): \ (x,y) \in X, \ y \in \operatorname{Arg} \min_{z \in Y(x)} f(x,z)\},$$

здесь $F(x,y)$ — целевая функция лидера; x — вектор его переменных; y — вектор переменных ведомой стороны; $f(x,y)$ — ее целевая функция; X и $Y(x)$ описывают общие ограничения и ограничения на поведение ведомой стороны в зависимости от поведения лидера.

Допустимую область и условия оптимальности в этой задаче можно описать вариационными неравенствами. Сама задача возникает, например, при государственном регулировании цен на энергетические носители. В этом случае переменные x интерпретируются как цены на нефть, газ, электроэнергию и т.п., а переменные y — как реакция промышленного сектора на их изменение. Целевая функция, оптимизируемая на верхнем уровне, формируется не только на основе финансовых соображений, но также и на основе общеполитических, социальных, экологических и других аспектов. На нижнем уровне промышленные предприятия, ориентируясь на цены, предложенные государством, стремятся распорядиться своими ресурсами и технологиями так, чтобы обеспечить себе максимальную прибыль.

Вернемся к лексикографической задаче $LVI(X, F_1, \ldots, F_m)$. В прямой постановке эта задача, являясь обобщением обычной задачи последовательного (лексикографического) программирования, сохраняет свойства неустойчивости последней. Одним из приемов преодоления указанной неустойчивости в последовательном программировании служит переход от векторного критерия к скалярному, представляющему собой взвешенную сумму частных критериев. При этом упорядоченность частных критериев находит свое отражение в выборе весовых множителей. Перенесем этот прием на лексикографическую задачу решения вариационного неравенства.

Пусть $\lambda = (\lambda_1,\ \lambda_2,\ \ldots,\ \lambda_m)$ — набор положительных весовых коэффициентов. Определим взвешенное отображение $F^\lambda : X \to \mathbb{R}^n$ по правилу

$$F^\lambda(x) = \lambda_1 F_1(x) + \lambda_2 F_2(x) + \ldots + \lambda_m F_m(x). \tag{5.2}$$

Введем условия:

1. Множество X не пусто, выпукло и замкнуто.
2. Отображения $F_1,\ F_2,\ \ldots,\ F_m$ непрерывны и монотонны на X.

Следующий результат характеризует принципиальную связь лексикографической задачи (5.1) с решениями вариационного неравенства, порождаемого отображением (5.2) и множеством X при различном выборе весовых множителей λ_i.

Теорема 5.1. *Пусть выполнены условия* 1, 2 *и при всех* $\lambda > 0$ *задача* $VI(X,\, F^\lambda)$ *имеет единственное решение* $x\ (\lambda)$. *Если существует повторный предел*

$$\lim_{\lambda_{m-1}\to\infty} \lim_{\lambda_{m-2}\to\infty} \ldots \lim_{\lambda_1\to\infty} x\ (\lambda) = x^*,$$

то x^* *— одно из решений лексикографической задачи* (5.1).

Д о к а з а т е л ь с т в о. Заметим, что рассматриваемый повторный предел не зависит от значения λ_m, которое можно считать равным 1. Введем обозначения:

$$\begin{aligned}
x^1\ (\lambda) &= \lim_{\lambda_1\to\infty} x\ (\lambda),\\
x^2\ (\lambda) &= \lim_{\lambda_2\to\infty} x^1\ (\lambda) = \lim_{\lambda_2\to\infty}\lim_{\lambda_1\to\infty} x\ (\lambda),\\
&\ldots\\
x^{m-1}(\lambda) &= \lim_{\lambda_{m-1}\to\infty} x^{m-2}(\lambda) = \lim_{\lambda_{m-1}\to\infty}\lim_{\lambda_{m-2}\to\infty}\ldots\lim_{\lambda_1\to\infty} x\ (\lambda)\ =\ x^*.
\end{aligned}$$

По условию отображения $F_1, F_2, \ldots, F_m$ монотонны и все $\lambda_i > 0$. Поэтому отображение (5.2) также мотонно, и по теореме 2.3 из главы 2

$$(\lambda_1 F_1(x) + \lambda_2 F_2(x) + \ldots + \lambda_m F_m(x),\ x - x(\lambda)) \geq 0 \quad \forall x \in X. \tag{5.3}$$

Деля обе части этого неравенства на $\lambda_1 > 0$ и переходя к пределу по $\lambda_1 \to \infty$, получаем

$$(F_1(x),\ x - x^1(\lambda)) \geq 0 \quad \forall x \in X = X_0,$$

что, в силу той же теоремы, означает $x^1(\lambda) \in X_1 \neq \emptyset$.

Если $m > 1$, перепишем неравенство (5.3) в виде

$$(\lambda_2\, F_2(x) + \lambda_3\, F_3(x) + \ldots + \lambda_m\, F_m(x),\ x - x(\lambda)) \geq$$
$$\geq (\lambda_1 F_1(x),\ x(\lambda) - x) \quad \forall x \in X_0.$$

Отсюда, с учетом неравенств $\lambda_1 > 0$ и $(F_1(x),\ x(\lambda) - x) \geq 0 \quad \forall x \in X_1$ и включения $X_1 \subseteq X_0$, вытекает

$$(\lambda_2\, F_2(x) + \lambda_3\, F_3(x) + \ldots + \lambda_m\, F_m(x),\ x - x(\lambda)) \geq 0 \quad \forall x \in X_1,$$

или, после перехода к пределу по $\lambda_1 \to \infty$,

$$(\lambda_2\, F_2(x) + \lambda_3\, F_3(x) + \ldots + \lambda_m\, F_m(x),\ x - x^1(\lambda)) \geq 0 \quad \forall x \in X_1. \tag{5.4}$$

Деля обе части этого неравенства на $\lambda_2 > 0$ и переходя к пределу по $\lambda_2 \to \infty$, получаем

$$(F_2(x),\ x - x^2(\lambda)) \geq 0 \quad \forall x \in X_1,$$

что, в силу все той же теоремы 2.3, означает $x^2(\lambda) \in X_2 \neq \emptyset$.

Если $m > 2$, повторим все приведенные выше рассуждения применительно к неравенству (5.4). Перепишем его в виде

$$(\lambda_3\, F_3(x) + \ldots + \lambda_m\, F_m(x),\ x - x(\lambda)) \geq (\lambda_2 F_2(x),\ x^1(\lambda) - x) \quad \forall x \in X_1.$$

С учетом того что $\lambda_2 > 0$ и $(F_2(x),\ x^1(\lambda) - x) \geq 0 \quad \forall x \in X_2$, а также включения $X_2 \subseteq X_1$, имеем

$$(\lambda_3\, F_3(x) + \ldots + \lambda_m\, F_m(x),\ x - x^1(\lambda)) \geq 0 \quad \forall x \in X_2,$$

или, после перехода к пределу по $\lambda_2 \to \infty$,

$$(\lambda_3\, F_3(x) + \ldots + \lambda_m\, F_m(x),\ x - x^2(\lambda)) \geq 0 \quad \forall x \in X_2. \tag{5.5}$$

Деля обе части этого неравенства на $\lambda_3 > 0$ и переходя к пределу по $\lambda_3 \to \infty$, получаем

$$(F_3(x),\ x - x^3(\lambda)) \geq 0 \quad \forall x \in X_2,$$

что и означает $x^3(\lambda) \in X_3 \neq \emptyset$.

Повторяя приведенные выше рассуждения снова и снова, будем получать аналоги неравенств (5.3)–(5.5) вида

$$(\lambda_k\, F_k(x) + \ldots + \lambda_m\, F_m(x),\ x - x^{k-1}(\lambda)) \geq 0 \quad \forall x \in X_{k-1}, \quad k = 4, \ldots, m-1,$$

пока, наконец, не дойдем до неравенства

$$(\lambda_m\, F_m(x),\ x - x^{m-1}(\lambda)) \geq 0 \quad \forall x \in X_{m-1}.$$

Деля обе части этого неравенства на $\lambda_m > 0$ и замечая, что $x^{m-1}(\lambda) = = x^*$, в силу все той же теоремы 2.3, приходим к заключительному выводу: $x^* \in X_m \neq \emptyset$.

Ввиду сложности вычисления повторного предела доказанный результат носит скорее принципиальный, чем практический характер. Поэтому перейдем к выяснению условий, при которых можно ограничиться обычным пределом или даже конечными значениями весовых множителей. Хотя по прежнему веса более важных критериев будут больше весов менее важных критериев, в оставшейся части параграфа нам удобнее будет считать эти веса сходящимися не к бесконечности, а к нулю.

Итак, пусть значения весовых коэффициентов в (5.2) зависят от некоторого натурального параметра k. Выпишем дополнительные условия:

3. $\lambda_1^k = 1, \quad \lim\limits_{k\to\infty} \lambda_i^k = 0 \quad (i = 2, \ \ldots, \ m).$

4. $\lim\limits_{k\to\infty} \lambda_i^k / \lambda_{i-1}^k = 0 \quad (i = 3, \ \ldots, \ m).$

Введем также условие регулярности для промежуточных подзадач:

5. $\rho(x, X_i) \le C_i(F_i(x), \ x - P_i(x))$ при всех $i < m$ и $x \in X_{i-1}$.

Здесь $P_i(x)$ — проекция точки x на множество X_i; $C_i > 0$ — некоторые константы.

Теорема 5.2. *Пусть выполнены условия* $1-5$ *и последовательность, составленная из векторов* $x^k = x(\lambda_k) \in \operatorname{Arg} VI(X, \ F_{\lambda^k})$, *ограничена. Тогда ее элементы при всех достаточно больших* k *будут решением лексикографической задачи* (5.1).

Д о к а з а т е л ь с т в о. Вначале покажем, что $x^k \in X_1$ при всех достаточно больших k. В самом деле, по самому способу определения точек x^k имеем

$$(F_1(x^k) + \frac{1}{\lambda_1^k} \sum_{i=2}^{m} \lambda_i^k \, F_i(x^k), \ x^k - x) \le 0 \quad \forall k \ \ \forall x \in X = X_0. \tag{5.6}$$

Отсюда, в силу условия 5, следует

$$\rho(x^k, \ X_1) \le C_1(F_1(x^k), \ x^k - P_1(x^k)) \le \frac{C_1}{\lambda_1^k} \sum_{i=2}^{m} \lambda_i^k \, (F_i(x^k), \ P_1(x^k) - x^k) \le$$

$$\le C_1 \sum_{i=2}^{m} \max_i \{\lambda_i^k / \lambda_1^k\} \, \|F_i(x^k)\| \, \rho(x^k, \ X_1) \qquad \forall k.$$

Поэтому

$$(1 - \mu_k^{(1)}) \ \rho(x^k, \ X_1) \le 0,$$

где

$$\mu_k^{(1)} = C_1 \max_i \{\lambda_i^k / \lambda_1^k\} \sum_{i=2}^{m} \|F_i(x^k)\| \to 0 \ \text{ при } \ k \to \infty.$$

При достаточно больших k имеем $\mu_k^{(1)} < 1$, т. е. $\rho(x^k, \ X_1) = 0$.

Далее, по индукции, предполагая уже установленным включение $x^k \in X_i$ при всех достаточно больших k для $i = 1, \ldots, n-1 < m$, покажем, что при больших k это включение верно также и для $i = n$.

Рассмотрим два случая.

1) Пусть $n < m$. В силу формул (5.6) и условия 4 и ввиду монотонности рассматриваемых отображений, имеем

$$\lambda_n^k \rho(x^k,\ X_n) \le \lambda_n^k C_n (F_n(x^k),\ x^k - P_n(x^k)) \le$$

$$\le C_n \sum_{i=1}^{n-1} \lambda_i^k \left(F_i(x^k),\ P_n(x^k) - x^k\right) + C_n \sum_{i=n+1}^{m} \lambda_i^k \left(F_i(x^k),\ P_n(x^k) - x^k\right) \le$$

$$\le C_n \sum_{i=1}^{n-1} \lambda_i^k \left(F_i(P_n(x^k)),\ P_n(x^k) - x^k\right) + C_n \sum_{i=n+1}^{m} \lambda_i^k \left(F_i(x^k),\ P_n(x^k) - x^k\right) \le$$

$$\le C_n \sum_{i=n+1}^{m} \max_{i>n}\{\lambda_i^k\} \, \|F_i(x^k)\| \, \rho(x^k,\ X_n),$$

здесь использовались неравенства $(F_i(P_n(x^k)),\ P_n(x^k) - x^k) \le 0$, которые верны благодаря включениям $P_n(x^k) \in \operatorname{Arg} VI(X_{i-1},\ F_i)$ при всех $i = 1, \ldots, n$.

Из полученного неравенства следует

$$(1 - \mu_k^{(n)})\ \rho(x^k,\ X_n) \le 0,$$

где

$$\mu_k^{(n)} = C_n \max_{i>n}\{\lambda_i^k / \lambda_n^k\} \sum_{i=n+1}^{m} \|F_i(x^k)\| \to 0 \quad \text{при} \quad k \to \infty.$$

Поэтому при достаточно больших k имеем $\mu_k^{(n)} < 1$, что влечет $\rho(x^k,\ X_n) = 0$.

2) Пусть $n = m$. Как и выше, из формул (5.6) при достаточно больших k в силу монотонности всех отображений следует

$$\lambda_m^k (F_m(x^k),\ x^k - x) \le$$

$$\le \sum_{i=1}^{m-1} \lambda_i^k (F_i(x^k),\ x - x^k) \le \sum_{i=1}^{m-1} \lambda_i^k (F_i(x),\ x - x^k) \le 0 \qquad \forall x \in X_{m-1},$$

здесь использовались неравенства $(F_i(x),\ x - x^k) \le 0$, которые верны благодаря определению множеств X_i и их последовательной вложенности друг в друга.

В качестве примера условий, при которых решения задач $VI(X, F_\lambda)$ существуют и ограничены при малых $\lambda_2, \ldots, \lambda_m$ в совокупности, приведем следующее:

6. Существует точка $x^0 \in X$ и окружающий ее компакт D такие, что $(F_1(x),\ x - x^0) > 0$ при всех $x \in D \cap X$.

Напомним, что множество D окружает точку x^0, если для любого l существует $\gamma = \gamma(l) > 0$ такое, что $x^0 + \gamma l \in D$ (простейшим примером окружающего множества является сфера $\{x : \ \|x - x^0\| = r\}$ с центром в точке x^0 и радиусом $r > 0$).

Теорема 5.3. *Пусть выполнены условия* 1, 2 *и* 6. *Тогда при постоянном* $\lambda_1 > 0$ *и всех достаточно малых* $\lambda_2 > 0, \ldots, \lambda_m > 0$ *задачи* $VI(X, F_\lambda)$ *разрешимы и их решения ограничены в совокупности.*

Доказательство. Для обоснования разрешимости задачи $VI(X, F_\lambda)$ покажем, что для всех достаточно малых $\lambda_2 > 0, \ldots, \lambda_m > 0$ существует ограниченное множество $M(\lambda)$, обладающее свойством: для любой точки $x \in X \setminus M(\lambda)$ найдется точка $y \in M(\lambda)$ со свойством

$$(F^\lambda(x),\ x - y) > 0$$

(это достаточные условия существования решения монотонного вариационного неравенства с непустым выпуклым замкнутым исходным множеством X и с непрерывным оператором, каковым при сделанных предположениях является F^λ).

В самом деле, можно взять $M(\lambda) = M = \operatorname{conv} D$ и $y = x^0$. Выполнимость неравенства $(F^\lambda(x),\ x - x^0) > 0$ для всех $x \in X \setminus M$ следует из монотонности F^λ и того, что это неравенство при достаточно малых $\lambda_2 > 0, \ldots, \lambda_m > 0$ выполняется для любой точки из D, в том числе точки, принадлежащей отрезку, соединяющему x^0 и x. Действительно, пусть z — такая точка. Тогда при некоторых $\mu_1 > 0$, $\mu_2 > 0$ имеем

$$x - x^0 = \mu_1(x - z) = \mu_2(z - x^0),$$

и потому

$$(F^\lambda(x),\ x - x^0) = \mu_1(F^\lambda(x),\ x - z) \geq \mu_1(F^\lambda(z),\ x - z) =$$
$$= \mu_2(F^\lambda(z),\ z - x^0) > 0.$$

Ограниченность множеств решений вытекает из того, что все они лежат в M, а M в нашем случае не зависит от λ.

Легко привести и другие примеры необходимых условий.

ПРИЛОЖЕНИЕ

В этом разделе приведен дополнительный материал по строчно- и столбцово-достаточным матрицам и методу Данцига—Коттла главных исключений для решения линейных задач о дополнительности. Последний в ряде случаев оказывается более предпочтительным по сравнению с методом Лемке.

§ 1. Достаточные матрицы

Введем операцию умножения векторов по Адамару:

$$x * y = Xy = (x_1 y_1, \dots, x_n y_n)^\top.$$

Напомним, что матрица M называется строчно(столбцово)-достаточной, если $x * (M^\top x) \le 0 \Rightarrow x * (M^\top x) = 0$ (соответственно $x * Mx \le 0 \ \Rightarrow \ x * Mx = 0.$)

Рассмотрим ряд свойств операции умножения векторов по Адамару и строчно(столбцово)-достаточных матриц (часть этих свойств носит тривиальный характер и отмечалась ранее).

Свойство 1. Пусть $u, v \in \mathbb{R}^n$ и P — произвольная перестановочная $(n \times n)$-матрица. Тогда

$$P^\top (u * v) = (P^\top u) * (P^\top v).$$

Свойство 2. Пусть M — произвольная $(n \times n)$-матрица и P — произвольная перестановочная $(n \times n)$-матрица. Тогда

$$P^\top (x * (Mx)) = (P^\top x) * ((P^\top M P)\,(P^\top x)) \quad \forall x \in \mathbb{R}^n.$$

Следует из свойства 1 и того, что $PP^\top = E$.

Под *главным переупорядочением* матрицы M будем понимать любую матрицу вида $P^\top M P$, где P — перестановочная матрица.

Свойство 3. Любое главное переупорядочение строчно(столбцово)-достаточной матрицы также строчно(столбцово)-достаточно.

Следует из свойства 2 и определения классов RS, CS.

Свойство 4. Пусть M — произвольная $(n \times n)$-матрица и $D = = \operatorname{diag}\,(\delta_1, \dots, \delta_n)$. Тогда

$$x * ((DMD)x) = (Dx) * (M(Dx)) \quad \forall x \in \mathbb{R}^n.$$

В самом деле,

$$(x * ((DMD)x))_i = x_i (DMDx)_i =$$
$$= x_i \,(\delta_i \sum_{j=1}^n m_{ij}\, \delta_j\, x_j) = (\delta_i x_i) \sum_{j=1}^n m_{ij}\, (\delta_j\, x_j) =$$
$$= ((Dx) * (M(Dx)))_i.$$

Свойство 5. Если M — строчно(столбцово)-достаточная матрица, то такой же будет и матрица DMD, где матрица D — диагональная.

Свойство 6. Каждая главная подматрица строчно(столбцово)-достаточной матрицы сама строчно(столбцово)-достаточна.

Свойство 7. Все главные миноры строчно(столбцово)-достаточной матрицы не отрицательны.

Свойство 8. Пусть a и b — произвольные вещественные числа, произведение которых отрицательно. Тогда (2×2)-матрица

$$M = \begin{pmatrix} 0 & a \\ b & 0 \end{pmatrix}$$

является достаточной.

В самом деле,

$$X * (Mx) = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} * \begin{pmatrix} ax_2 \\ bx_1 \end{pmatrix} = x_1 \cdot x_2 \begin{pmatrix} a \\ 1 \end{pmatrix} \leq 0$$

только при $x_1 \cdot x_2 = 0$. Но тогда и $x * (Mx) = 0$.

Свойство 9. Пусть $a \geq 0$, $b \neq 0$ — вещественные числа. Тогда матрицы

$$M_1 = \begin{pmatrix} a & 0 \\ b & 0 \end{pmatrix}, \quad M_2 = \begin{pmatrix} 0 & b \\ 0 & a \end{pmatrix}$$

не могут быть строчно-достаточными. Соответственно $M_1^\top$ и $M_2^\top$ не могут быть столбцово-достаточными.

Действительно, для вектора $x = (1, -ab^{-1} - \delta)^\top$, где $\delta > 0$, верно $x * (M_1^\top x) \leq 0$, но $x_1(M_1^\top x)_1 < 0$. Аналогично для M_2.

Свойство 10. Пусть M — строчно-достаточная матрица. Если $m_{ii} = 0$ при некотором i, причем $m_{ji} \geq 0$ для всех $j = 1, \ldots, n$, то $m_{ij} \leq 0$ для всех $j = 1, \ldots, n$.

В самом деле, в силу свойства 6 достаточно доказать это утверждение при $n = 2$. По свойству 3 без ограничения общности можно считать $i = 1$. В силу свойства 7, $m_{22} \geq 0$. Поэтому случай, когда $m_{21} = 0$, вытекает из свойства 9. Если же $m_{21} > 0$, то по свойству 7 неравенство $m_{12} > 0$ невозможно. Тем самым всегда $m_{12} \leq 0$.

Выпишем линейную задачу о дополнительности $LCP(q, M)$: найти векторы $z, w \in \mathbb{R}^n$ такие, что

$$z \geq 0, w \geq 0, \tag{1.1}$$

$$w = q + Mz, \tag{1.2}$$

$$z^\top w = 0; \tag{1.3}$$

здесь матрица $M = (m_{ij})_{n \times n}$ и вектор $q \in \mathbb{R}^n$ заданы.

Метод главных исключений основан на идее перебора базисных решений системы (1.2), удовлетворяющих условиям дополнительности (1.3), но, возможно, не удовлетворяющих условиям (1.1).

Рассмотрим систему (1.2). Пусть для некоторого набора индексов $\alpha \subset \{1, \ldots, n\}$ главная подматрица $M_{\alpha\alpha}$ матрицы M не вырождена. Прибегая в случае необходимости к подходящему главному переупорядочению M, можно считать $M_{\alpha\alpha}$ ведущей главной подматрицей. Уравнения (1.2) примут вид

$$\begin{aligned} w_\alpha &= q_\alpha + M_{\alpha\alpha} z_\alpha + M_{\alpha\bar{\alpha}} z_{\bar{\alpha}}, \\ w_{\bar{\alpha}} &= q_{\bar{\alpha}} + M_{\bar{\alpha}\alpha} z_\alpha + M_{\bar{\alpha}\bar{\alpha}} z_{\bar{\alpha}}, \end{aligned} \tag{1.4}$$

здесь $\bar{\alpha} = \{1, \ldots, n\} \backslash \alpha$.

В представлении (1.4) будем говорить о переменных w как о базисных, а о переменных z — как о независимых (небазисных).

Поскольку по предположению $\det\,(M_{\alpha\alpha}) \neq 0$, то в (1.4) можно поменять ролями переменные z_α и w_α. Применяя стандартную процедуру исключения, получаем

$$\begin{aligned} z_\alpha &= q'_\alpha + M'_{\alpha\alpha} w_\alpha + M'_{\alpha\bar\alpha} z_{\bar\alpha}, \\ w_{\bar\alpha} &= q'_{\bar\alpha} + M'_{\bar\alpha\alpha} w_\alpha + M'_{\bar\alpha\bar\alpha} z_{\bar\alpha}, \end{aligned} \tag{1.5}$$

где

$$\begin{gathered} q'_\alpha = -M^{-1}_{\alpha\alpha}\, q_\alpha, \quad q'_{\bar\alpha} = q_{\bar\alpha} - M_{\bar\alpha\alpha}\, M^{-1}_{\alpha\alpha}\, q_\alpha, \\ M'_{\alpha\alpha} = M^{-1}_{\alpha\alpha}, \quad M'_{\alpha\bar\alpha} = -M^{-1}_{\alpha\alpha}\, M_{\alpha\bar\alpha}, \\ M'_{\bar\alpha\alpha} = M_{\bar\alpha\alpha}\, M^{-1}_{\alpha\alpha}, \; M'_{\bar\alpha\bar\alpha} = M_{\bar\alpha\bar\alpha} - M_{\bar\alpha\alpha}\, M^{-1}_{\alpha\alpha}\, M_{\alpha\bar\alpha}. \end{gathered}$$

Система (1.5) называется преобразованной системой для задачи $LCP(q, M)$. Матрица $M_{\alpha\alpha}$ называется ведущим блоком. Чтобы показать, что матрица M' получена из M при помощи главного преобразования с ведущим блоком $M_{\alpha\alpha}$, будем писать

$$M' = \rho_\alpha(M).$$

Сделаем еще замечание о матрице смены знаков. Под последней будем понимать диагональную матрицу $S_\alpha = \mathrm{diag}\ (\sigma_1, \ldots, \sigma_n)$, где

$$\sigma_i = \begin{cases} \ \ 1, & \text{если } i \in \alpha; \\ -1, & \text{если } i \in \bar\alpha. \end{cases}$$

Свойство 11. Пусть $M_{\alpha\alpha}$ — невырожденная подматрица матрицы M. Тогда

$$(\rho_\alpha\,(M))^\top = S_\alpha\,(\rho_\alpha\,(M^\top))\,S_\alpha.$$

Следует из того, что матрица S_α меняет знаки только у внедиагональных блоков, и того, что $(M^\top)_{\alpha\alpha} = (M_{\alpha\alpha})^\top$, $(M^{-1}_{\alpha\alpha})^\top = (M_{\alpha\alpha})^{-1}$ и $(M_{\alpha\beta})^\top = (M^\top)_{\beta\alpha}$.

Вообще говоря, нас интересуют только строчно-достаточные матрицы. Однако нам удобнее получить следующий результат вначале для столбцово-достаточных матриц.

Теорема 1.1. *Пусть $M_{\alpha\alpha}$ — невырожденная подматрица матрицы M. Если M столбцово-достаточна, то такой же будет и $M' = = \rho_\alpha\,(M)$.*

Д о к а з а т е л ь с т в о. Не ограничивая общности, можно считать матрицу $M_{\alpha\alpha}$ ведущей (т. е. $\alpha = \{1, \ldots, k\}$, $k < n$). Обозначим $w = = M'z$ и предложим, что $z * w \leq 0$. Имеем

$$\begin{pmatrix} w_\alpha \\ w_{\bar\alpha} \end{pmatrix} = \begin{pmatrix} M'_{\alpha\alpha} & M'_{\alpha\bar\alpha} \\ M'_{\bar\alpha\alpha} & M'_{\bar\alpha\bar\alpha} \end{pmatrix} \begin{pmatrix} z_\alpha \\ z_{\bar\alpha} \end{pmatrix}.$$

Условие $z * w \leq 0$ означает, что

$$\begin{pmatrix} z_\alpha \\ z_{\bar\alpha} \end{pmatrix} * \begin{pmatrix} w_\alpha \\ w_{\bar\alpha} \end{pmatrix} = \begin{pmatrix} z_\alpha * w_\alpha \\ z_{\bar\alpha} * w_{\bar\alpha} \end{pmatrix} = \begin{pmatrix} w_\alpha * z_\alpha \\ z_{\bar\alpha} * w_{\bar\alpha} \end{pmatrix} \leq 0.$$

Поскольку $M' = \rho_\alpha\,(M)$, то

$$\begin{pmatrix} z_\alpha \\ w_{\bar\alpha} \end{pmatrix} = \begin{pmatrix} M_{\alpha\alpha} & M_{\alpha\bar\alpha} \\ M_{\bar\alpha\alpha} & M_{\bar\alpha\bar\alpha} \end{pmatrix} \begin{pmatrix} w_\alpha \\ z_{\bar\alpha} \end{pmatrix}.$$

Но $M \in CS$, откуда
$$\begin{pmatrix} w_\alpha \\ z_{\bar\alpha} \end{pmatrix} * \begin{pmatrix} z_\alpha \\ w_{\bar\alpha} \end{pmatrix} = 0.$$
Следовательно, и $z * w = 0$, т. е. $M' \in CS$ также.

Перейдем теперь к тому результату, который нам действительно нужен.

Теорема 1.2. *Пусть $M_{\alpha\alpha}$ — невырожденная главная подматрица матрицы M. Если M строчно-достаточна, то такой же будет и $M' = \rho_\alpha(M)$.*

Доказательство. Мы уже отмечали, что $M \in RS$ тогда и только тогда, когда $M^\top \in CS$. Поэтому достаточно показать, что $(M')^\top$ — столбцово-достаточна. По предположению столбцово-достаточной является $M^\top$. Теорема 1.1 утверждает, что $\rho_\alpha(M^\top)$ столбцово-достаточна. Но по определению M' и свойству 11
$$(M')^\top = (\rho_\alpha(M))^\top = S_\alpha(\rho_\alpha(M^\top))\,S_\alpha.$$
Остается применить свойство 5.

Таким образом, классы строчно- и столбцово-достаточных матриц оказываются инвариантными относительно преобразования ρ_α.

§ 2. Метод Данцига—Коттла

Метод Данцига—Коттла, подобно другим конечным методам решения линейных задач о дополнительности, использует симплексные преобразования системы уравнений
$$w = q + Mz. \tag{2.1}$$
Ниже мы будем использовать верхний индекс k как счетчик итераций. Начальное значение этого счетчика $k = 0$, и система (2.1) может быть переписана как
$$w^0 = q^0 + M^0 z^0. \tag{2.2}$$
После k симплексных преобразований система будет выглядеть так:
$$w^k = q^k + M^k z^k. \tag{2.3}$$
Каждый из векторов w^k и z^k системы (2.3), первый из которых мы будем называть вектором базисных, а второй — небазисных переменных, вообще говоря, скомпонован как из переменных w_i, так и z_j.

Система (2.3) может быть также представлена в наглядной табличной форме:

$$\begin{array}{c|c|ccc|}
 & & z_1^k & \dots & z_n^k \\ \cline{2-5}
w_1^k & q_n^k & m_{11}^k & \dots & m_{1n}^k \\
\vdots & \vdots & \vdots & & \vdots \\
w_n^k & q_n^k & m_{n1}^k & \dots & m_{nn}^k \\ \cline{2-5}
\end{array}$$

Метод Данцига—Коттла опирается на главные симплексные преобразования (порядка 1 и 2) с целью достижения одного из двух возможных конечных состояний исходной таблицы. Первое из них характеризуется неотрицательностью столбца свободных членов $q^k \geq 0$. Второе — наличием в таблице строки со свойством

$$q_r^k < 0 \;\text{ и }\; m_{rj}^k \leq 0, \quad j = 1, \ldots, n.$$

Первое состояние говорит о достижении решения исходной задачи, второе устанавливает факт пустоты множества ее допустимых точек.

Строго говоря, второе состояние не проверяется явно. В случае, когда M является P-матрицей, оно вообще не наступает. Когда M является строчно-достаточной матрицей, второе состояние проявляет себя в наличии столбца со свойством

$$q_r^k < 0, \quad m_{rr}^k = 0 \;\text{ и }\; m_{ir}^k \geq 0 \quad \forall i \neq r.$$

Наличие такого столбца обнаруживается в алгоритме при проведении теста на минимальное отношение.

Метод состоит в серии главных итераций, каждая из которых начинается с выбора *отмеченной* переменной — переменной, чье текущее значение отрицательно. Отмеченная переменная остается таковой на протяжении всей главной итерации. Цель операций, реализуемых внутри главной итерации, — рост отмеченной переменной до нуля, если это вообще осуществимо. Каждая такая операция дает прирост одной из небазисных переменных, способных вывести отмеченную переменную на нулевой уровень. Наращиваемая небазисная переменная называется ведущей. В соответствии с соглашениями метода все переменные, текущие значения которых неотрицательны, должны оставаться таковыми. Начальная точка имеет вид $(w, z) = (q^0, 0)$. Это значит, что вначале по крайней мере n переменных неотрицательны. Для переменных с отрицательным начальным значением $q_i^0 < 0$ вводится произвольная нижняя граница

$$\lambda < \min_{1 \leq i \leq n} \{q_i^0\},$$

она также отрицательна и остается неизменной в течение всех вычислений. Эта искусственно вводимая в условия задачи граница используется всегда, за исключением случая, когда $M \in P$. По соглашениям метода переменные, текущие значения которых отрицательны, не должны принимать значения, меньше λ. Мы также расширим понятие небазисной переменной, разрешив ей принимать не только нулевое значение, но и значение λ. Решение системы (2.1) будем называть невырожденным, если не более n ее переменных принимает значения 0 или λ. В противном случае решение называется вырожденным.

Для большей ясности введем различие между *названием* переменной и ее *значением*. Последнее будем отмечать черточкой сверху над именем переменной. Так, в начале главной итерации мы всегда будем иметь $\bar{z}_i^k = 0$ или $\bar{z}_i^k = \lambda, \quad i = 1, \ldots, n$. Мы будем также пользоваться обозначением

$$W^k(z^k) = M^k z^k + q^k$$

(хотя численно это то же, что и w^k, но подчеркивает роль z^k как аргумента).

С учетом сказанного опишем правила выбора ведущей переменной.

Если в самом начале главной итерации отмеченная переменная выбрана среди базисных, первая ведущая переменная должна быть к ней дополнительной. Так, пусть w_r^k — отмеченная переменная, тогда z_r^k — ведущая переменная на первом шаге. Однако отмеченная переменная не обязана быть базисной. В соответствии с используемой здесь более широкой трактовкой базисного решения текущее решение

$(\bar{w}^k, \bar{z}^k)$ может иметь в начале главной итерации отрицательную небазисную компоненту $z_r^k = \lambda < 0$. В этом случае переменная z_r^k будет одновременно и отмеченной, и ведущей. Ее рост всегда ограничен сверху тем, что некоторая базисная переменная, убывая, достигнет своей нижней границы (0 или λ), или тем, что z_r^k в процессе роста достигнет нуля (в этом случае главная итерация завершается).

Опишем перечисленные соглашения формально в виде последовательности шагов алгоритма.

Шаг 0. Полагаем $k = 0$ *и* $(\bar{w}^0, \bar{z}^0) = (q^0, 0)$. *Определяем нижнюю отрицательную границу* $\lambda < \min\limits_{(i)} \{q_i^0\}$.

Шаг 1. Если $q^k \geq 0$ *или выполнено неравенство* $(\bar{w}^k, \bar{z}^k) \geq (0, 0)$, *то работа алгоритма завершается получением решения задачи — вектора* $(q^k, 0)$. *В противном случае определяем номер* r *такой, что* $\bar{z}_r^k = \lambda$, *или (если такой переменной нет) такой, что* $\bar{w}_r^k < 0$.

Шаг 2. Определяем ζ_r^k *— максимальное значение переменной* $z_r^k \geq \bar{z}_r^k$, *удовлетворяющее следующим условиям:*

(1) $z_r^k \leq 0$, *если* $\bar{z}_r^k = \lambda$.

(2) $W_r^k\,(\bar{z}_1^k, \ldots, \bar{z}_{r-1}^k, z_r^k, \bar{z}_{r+1}^k, \ldots, \bar{z}_n^k) \leq 0$, *если* $\bar{w}_r^k < 0$.

(3) $W_i^k\,(\bar{z}_1^k, \ldots, \bar{z}_{r-1}^k, z_r^k, \bar{z}_{r+1}^k, \ldots, \bar{z}_n^k) \geq 0$, *если* $\bar{w}_i^k > 0$.

(4) $W_r^k\,(\bar{z}_1^k, \ldots, \bar{z}_{r-1}^k, z_r^k, \bar{z}_{r+1}^k, \ldots, \bar{z}_n^k) \geq \lambda$, *если* $\bar{w}_i^k < 0$.

Шаг 3. Если $\zeta_r^k = +\infty$, *то задача не имеет допустимых точек и работа алгоритма завершается. Если* $\zeta_r^k = 0$, *полагаем* $\bar{z}_r^{k+1} = 0$, $\bar{z}_i^{k+1} = \bar{z}_i^k$ *при все* $i \neq r$ *и*

$$\bar{w}^{k+1} = W^{k+1}\,(\bar{z}^{k+1}) - W^k\,(\bar{z}^{k+1}).$$

Возвращаемся к шагу 1, меняя k *на* $k+1$. *Если* $0 < \zeta_r^k < +\infty$, *выделяем единственный номер* s, *определяемый условиями* $(2) - (4)$ *шага 2.*

Шаг 4. Если $m_{ss}^k > 0$, *проводим изменение базиса порядка 1: вводим в него переменную* z_s^k *и выводим переменную* w_s^k. *Полагаем*

$$\bar{z}_s^{k+1} = W_s^k(\bar{z}_1^k, \ldots, \bar{z}_{r-1}^k, \zeta_r^k, \bar{z}_{r+1}^k, \ldots, \bar{z}_n^k),$$

$$\bar{w}^{k+1} = W^{k+1}(\bar{z}^{k+1}).$$

Если $s = r$, *возвращаемся к шагу 1, меняя* k *на* $k+1$. *В противном случае возвращаемся к шагу 2, меняя* k *на* $k+1$.

Шаг 5. Если $m_{ss}^k = 0$, *проводим преобразование порядка 2: переменные* z_r^k *и* z_s^k *вводятся в базис, а переменные* w_r^k *и* w_s^k *его покидают. Полагаем*

$$\bar{w}_r^{k+1} = \bar{z}_s^k,\ \bar{w}_s^{k+1} = \zeta_r^k,\ \bar{z}_i^{k+1} = \bar{z}_i^k$$

при всех $i \notin \{r, s\}$. *После этого вычисляем* $\bar{w}_i^{k+1} = W_i^{k+1}\,(\bar{z}^{k+1})$ *для всех* $i \notin \{r, s\}$. *Возвращаемся к шагу 2, меняя* k *на* $k+1$ *и* r *на* s.

Перейдем к обсуждению того, что, собственно, делает алгоритм и почему он решает произвольную линейную задачу о дополнительности, матрица M которой строчно-достаточна.

Все главные итерации алгоритма начинаются с шага 1, где производится проверка, не получено ли уже решение исходной задачи. Этому соответствует случай $(\bar{w}^k, \bar{z}^k) \geq (0, 0)$, поскольку $(\bar{w}^k, \bar{z}^k)$ будет неотрицательным решением (2.1) при $\bar{z}^k = 0$. Как будет ниже проиллюстрировано на примере, может произойти так, что столбец свободных членов q^k станет неотрицательным раньше, чем z^k. И в этом случае, переопределяя z^k как $\bar{z}^k = 0$, получаем искомое решение. Если же ни одна из этих ситуаций не реализовалась, существует номер r такой, что $\bar{z}_r^k < 0$ или $\bar{w}_r^k < 0$. Такая переменная объявляется отмеченной.

Напомним, что исходная задача (2.1)–(2.3) имеет $2n$ переменных. Число отрицательных компонент в решении системы (2.1) назовем *индексом недопустимости*. Условия, включенные в шаг 2, не допускают роста этого индекса. Следовательно, с каждым возвратом к шагу 1 алгоритм продуцирует базисное решение с меньшим индексом недопустимости по сравнению с предыдущим. Поскольку число базисных решений конечно, конечно и число возвратов к шагу 1. Доказательство конечности алгоритма упирается, таким образом, в необходимость показать конечность всех операций внутри главной итерации.

Прекращение вычислений предусмотрено и в шаге 3. Оно происходит при $\zeta_r^k = +\infty$, что возможно только, если w_r^k — отмеченная переменная. Также должны выполняться условия

$$m_{rr}^k = 0 \quad \text{и} \quad m_{ir}^k \geq 0 \quad \forall i \neq r.$$

По свойству 10 отсюда вытекает, что $m_{r_j}^k \leq 0, \;\; j = 1, \ldots, n.$ Кроме того, небазисные переменные $\bar{z}_j^k \leq 0$ при всех j и

$$\bar{w}_r^k = q_r^k + \sum_{j=1}^{n} m_{rj}^k \, \bar{z}_j^k < 0.$$

Следовательно, $q_r^k < 0$, и уравнение

$$w_r^k = q_r^k + \sum_{j=1}^{n} m_{r_j}^k \, z_j^k$$

не имеет неотрицательных решений.

Другой исход шага 3 — $\zeta_r^k = 0$. В этом случае, в силу предположения о невырожденности системы (2.1), отмеченная и ведущая переменные совпадают: это некоторая небазисная переменная z_r^k, которая возрастает со значения λ до нуля. Главная итерация завершается без смены базиса.

Наконец, последняя возможность $0 < \zeta_r^k < +\infty$ означает, что некоторая базисная переменная блокирует рост ведущей переменной z_r^k. Возникающие здесь альтернативы разбираются в шаге 4.

А именно, если $m_{ss}^k > 0$, то переменные z_s^k и w_s^k меняются ролями: первая из них вводится в базис, а вторая — выводится. Пересчет таблицы требует конечного числа операций. Если к тому же $r = s$, то отмеченная переменная выходит на нулевой уровень, что приводит к возврату на шаг 1 и уменьшению индекса недопустимости по крайней мере на единицу. Если же $r \neq s$, после смены базиса рост ведущей переменной z_r^k продолжается (в соответствии с правилами шага 2).

Если $m_{ss}^k = 0$, то заведомо $r \neq s$. Факт блокировки z_r^k переменной w_s^k означает, что $m_{sr}^k < 0$. В этом случае проводят главное симплексное преобразование порядка 2 с ведущим блоком

$$\begin{pmatrix} m_{rr}^k & m_{rs}^k \\ m_{sr}^k & m_{ss}^k \end{pmatrix}.$$

Это можно сделать, так как, в силу строчной достаточности M, отрицательность m_{sr}^k влечет положительность $m_{rs}^k > 0$ (см. свойство 9). Значения всех переменных сразу после смены базиса будут такими, какими они были в момент блокировки. При возврате на шаг 2 переменная w_r^k занимает небазисную позицию z_s^{k+1}, а переменная w_s^k — позицию z_r^{k+1}. Естественный порядок восстанавливается их главным переупорядочением (одновременно меняются местами строки r и s и небазисные переменные z_r^{k+1} и z_s^{k+1}).

Как уже отмечалось, нужно показать, что число возвращений к шагу 2 будет конечным. Это следует из того, что число всех главных преобразований порядка 2 конечно и конечно число различных конфигураций значений небазисных переменных (вариантов их распределения между значениями 0 и λ). Что касается ведущей небазисной переменной z_r^k, ее значение, как и значение дополнительной к ней переменной w_r^k, монотонно растет, а их сумма растет строго монотонно на протяжении главной итерации. Отсюда определение ζ_r^k и ζ_r^{k+1} $(k > 0)$ делает невозможным ситуацию, при которой

$$z_i^k = z_i^{k+t}\,(i = 1, \ldots, n) \quad \text{и} \quad \bar{z}_i^k = \bar{z}_i^{k+t}\,(i \neq r)$$

и которая возникала бы, если число возвратов к шагу 2 было бесконечным.

Пример. Вновь рассмотрим задачу $LCP(q, M)$ из главы 1 §5, где

$$q = \begin{pmatrix} -3 \\ 6 \\ -1 \end{pmatrix}, \; M = \begin{pmatrix} 0 & -1 & 2 \\ 2 & 0 & -2 \\ -1 & 1 & 0 \end{pmatrix}.$$

Матрица M не является ни P-, ни PD-, ни даже PSD-матрицей. Но она строчно-достаточна, как нетрудно проверить. Прямой подстановкой легко убедиться, что вектор $(\bar{w}, \bar{z}) = (2, 0, 0; 0, 1, 3)$ является решением поставленной задачи. Применим к ней метод Данцига—Коттла.

Запишем данные задачи в табличной форме:

		z_1	z_2	z_3	
w_1	-3	0	-1	2	-3
w_2	6	2	0	-2	6
w_3	-1	-1	1	0	-1
		0	0	0	

Нижней границей отрицательных значений переменных w_i может служить число $\lambda = -4$. Выберем w_1 в качестве отмеченной переменной.

Дополнительно к ней переменная z_1 будет ведущей. Блокирующей переменной, мешающей неограниченному росту z_1, будет w_3. Убывая, эта переменная достигнет нижней границы -4 при $z_1 = 3$. Поскольку $m_{33} = 0$, необходимо поменять ролями две пары переменных — w_3 с z_1 и w_1 с z_3. Новая таблица примет вид

$$
\begin{array}{c|c|ccc|c}
 & & w_3 & z_2 & w_1 & \\
\cline{2-5}
z_3 & \frac{3}{2} & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\
w_2 & 1 & -2 & 1 & -1 & \frac{1}{2} \\
z_1 & -1 & -1 & 1 & 0 & 3 \\
\cline{2-5}
 & & -4 & 0 & -3 &
\end{array}
$$

Теперь отмеченной переменной становится небазисная переменная w_1, которая может наращиваться сама, выступая в роли ведущей. Она блокирует сама себя. В результате первая главная итерация завершается таблицей

$$
\begin{array}{c|c|ccc|c}
 & & w_3 & z_2 & w_1 & \\
\cline{2-5}
z_3 & \frac{3}{2} & 0 & \frac{1}{2} & \frac{1}{2} & \frac{3}{2} \\
w_2 & 1 & -2 & 1 & -1 & 9 \\
z_1 & -1 & -1 & 1 & 0 & 3 \\
\cline{2-5}
 & & -4 & 0 & 0 &
\end{array}
$$

Для следующей главной итерации имеется единственный кандидат на отмеченную переменную — w_3. Она становится и ведущей переменной, рост которой блокирует переменная z_1; она обращается в ноль при $w_3 = -1$. Снова необходимо поменять ролями две пары переменных — w_2 с z_3 и w_3 с z_2. Это приводит к таблице

$$
\begin{array}{c|c|ccc|c}
 & & z_1 & w_3 & w_2 & \\
\cline{2-5}
w_1 & 2 & 1 & -1 & -1 & 3 \\
z_3 & 3 & 1 & 0 & -\frac{1}{2} & 3 \\
z_2 & 1 & 1 & 1 & 0 & 0 \\
\cline{2-5}
 & & 0 & -1 & 0 &
\end{array}
$$

Отмеченная переменная все еще w_3, чье текущее значение равно -1. Будучи использована в качестве ведущей, w_3 блокирует сама себя. В результате получаем решение. Еще лучше сразу заметить, что вектор свободных членов неотрицателен. В любом случае решением служит вектор $(\bar{w}, \bar{z}) = (2, 0, 0; 0, 1, 3)$.

§ 3. Параметрический вариант метода Лемке

Вернемся к методу Лемке из главы 1 и дадим иную интерпретацию его искусственной переменной. Будем трактовать ее как параметр, начальное положительное значение которого в ходе вычислений постепенно доводится до нуля. Такой подход позволяет более тесно увязать вопросы сходимости метода с типом матрицы в линейной задаче о дополнительности.

Выпишем линейную задачу о дополнительности: найти $w,\ z \in \mathbb{R}^n$ такие, что

$$w = q + Mz \geq 0, \quad z \geq 0, \tag{3.1}$$

$$w^\top z = 0; \tag{3.2}$$

здесь $(n \times n)$-матрица M и n-вектор q заданы, w и z — векторы неизвестных.

Поместим задачу (3.1), (3.2) в параметрическое семейство задач того же типа : найти $w,\ z \in \mathbb{R}^n$ такие, что

$$w = q + fz_0 + Mz \geq 0, \quad z \geq 0, \tag{3.3}$$

$$w^{\top} z = 0. \tag{3.4}$$

Здесь $z_0 \geq 0$ — числовой параметр; $f = (f_1, \ldots, f_n) > 0$ — произвольный вектор (например, единичный). При $z_0 = 0$ задача (3.3), (3.4) совпадает с задачей (3.1), (3.2).

Как обычно, будем предполагать, что системы уравнений в (3.1) и (3.3) невырождены, т. е. любое их базисное решение содержит ровно n ненулевых компонент.

Организуем исходные данные задачи (3.3), (3.4) в таблицу

$$\begin{array}{c|c|c|c|c|} & & & z_0 & z_1 \ldots z_n \\ \hline w_1 & & & & \\ \vdots & \bar{q} & q & f & M \\ w_n & & & & \\ \hline \end{array} \tag{3.5}$$

Здесь $\bar{q} = \bar{q}(z_0) = q + f z_0$. Будем называть переменные $w_1, \ldots, w_n$ базисными, а переменные $z_1, \ldots, z_n$ — свободными (небазисными). Каждая базисная переменная соответствует своей строке, а свободная переменная — своему столбцу матрицы M. Еще раз подчеркнем, что z_0 играет роль параметра.

Как и раньше, метод будет состоять из конечной последовательности итераций, в ходе которых одни переменные будут вводиться в базис, другие — покидать его. Столбцы q и f будут переопределяться отдельно и, если $\bar{z}_0$ обозначает текущее значение параметра z_0, столбец $\bar{q}$ таблицы (3.5) определяется как $\bar{q} = q + f \bar{z}_0$.

Итерации делятся на большие и малые. Каждая большая итерация начинается с определения нового значения параметра z_0. Затем следует серия малых итераций, во время которых значение z_0 не меняется. Цель малых итераций — создать условия для возможного снижения этого значения. Ввод и вывод из базиса переменных производится так, чтобы поддерживать неотрицательность вектора $\bar{q}$. Эта неотрицательность поддерживается и при переопределении z_0.

Таблицу, получаемую на k-й итерации, будем записывать так:

$$\begin{array}{c|c|c|c|c|} & & & z_0 & z_1^k \ldots z_n^k \\ \hline w_1^k & & & & \\ \vdots & \bar{q}^k & q^k & f^k & M^k \\ w_n^k & & & & \\ \hline \end{array} \tag{3.5.k}$$

Здесь, как и раньше, компоненты вектора $(w_1^k, \ldots, w_n^k,\ z_1^k, \ldots, z_n^k)$ представляют собой некоторую перестановку компонент исходного вектора $(w_1, \ldots, w_n,\ z_1, \ldots, z_n)$, определяемую порядком смены базисных и свободных переменных задачи на предыдущих итерациях.

Пусть таблица (3.5.k) соответствует большой итерации; при $k = 0$ она совпадает с таблицей (3.5). Таблица большой итерации должна удовлетворять условиям:

1) из каждой пары переменных, связанных условиями дополнительности, ровно одна является базисной и ровно одна — свободной;

2) параметр z_0 принимает такое неотрицательное значение, при котором среди компонент вектора $\bar{q} = q + f z_0$ ровно одна равна нулю, а остальные — положительны.

Очевидно, что в самом начале условие 1 выполняется автоматически. Выполнения условия 2 добиваются следующим образом. Поскольку $f > 0$, то при достаточно большом z_0 сумма $\bar{q} = q + f z_0$ также будет положительной. Постепенно снижая значение параметра z_0, останавливаются в тот момент, когда одна из компонент вектора

$\bar{q}$, убывая, обратится в нуль [в силу предположения о невырожденности системы (3) она окажется единственной]. Если этого не случилось, т. е. в процессе снижения z_0 до нуля все компоненты вектора $\bar{q}$ остались положительными, то $q = \bar{q} - f z_0 = \bar{q} > 0$ и задача (3.1), (3.2) имеет очевидное решение $(w, z) = (q, 0)$. То же самое можно сказать и о большой итерации, на которой параметр z_0 обратился в нуль: вектор, свободные компоненты которого равны нулю, а базисные — соответствующим компонентам вектора $\bar{q} \geq 0$, является решением задачи (3.1), (3.2).

Пусть зафиксированное на большой итерации значение $\bar{z}_0$ параметра z_0 положительно. Обозначим через r номер единственной нулевой компоненты вектора $\bar{q} = q + f\bar{z}_0$ (все остальные положительны). Переменную w_r^k назовем отмеченной. Цель последующих малых итераций — заставить отмеченную переменную покинуть базис. Для этого на первой малой итерации в базис вводится переменная z_r^k, дополнительная к w_r^k. Если отмеченная переменная не покинула базис, применяется основное правило метода Лемке: на каждой следующей малой итерации в базис вводится переменная, дополнительная к той, что покинула базис на предыдущей итерации.

Переходя к описанию содержания малой итерации, будем считать, что таблица (3.5.k) соответствует ее началу и что s — номер вводимой в базис свободной переменной z_s^k.

Если $m_{rs}^k \neq 0$, производится смена ролей переменных w_r^k и z_s^k, т. е. первая из них покидает базис, а вторая — входит в него. Поскольку $\bar{q}_r^k = 0$, при проведении операции исключения Жордана—Гаусса с ведущим элементом m_{rs}^k вектор $\bar{q}^k$ не изменится. При этом новая таблица будет удовлетворять условию 1, т. е. из каждой пары переменных, связанных условием дополнительности, ровно одна окажется в базисе и ровно одна — свободной. Более того, появляется возможность переопределить значение параметра z_0, т. е. перейти к большой итерации. Действительно, в силу невырожденности системы уравнений (3.1), $q_r^k \neq 0$. Поэтому также $f_r^k = -q_r^k \, \bar{z}_0^{-1} \neq 0$, и при $f_r^k > 0$ параметр z_0 можно неограниченно наращивать, а при $f_r^k < 0$ — уменьшать без потери свойства неотрицательности $\bar{q}_r$.

Поскольку остальные компоненты вектора $\bar{q}$ были положительны, они останутся таковыми и при малом отклонении z_0 от предыдущего значения $\bar{z}_0^k$. С ростом этого отклонения (ростом z_0 при $f_r^k > 0$ и его уменьшением при $f_r^k < 0$) одна из компонент вектора $\bar{q} = q^k + f^k z_0$, уменьшаясь, может обратиться в нуль [в силу невырожденности системы (3.3) она окажется единственной]. Если это случится, соответствующее значение параметра z_0 фиксируется, и мы получаем таблицу, удовлетворяющую условиям 1, 2, т. е. можно переходить к следующей большой итерации. Если же параметр z_0 неограниченно растет, а все компоненты вектора $\bar{q} = q^k + f^k z_0$ остаются неотрицательными, метод завершается на так называемом z_0-луче (без получения решения исходной задачи). Наконец, если параметр z_0, убывая, достиг нуля, а $\bar{q} = q^k \geq 0$, мы получаем решение задачи (3.1), (3.2), а именно вектор, все свободные компоненты которого равны нулю, а базисные — соответствующим компонентам вектора q^k.

Рассмотрим теперь случай, когда на малой итерации $m_{rs}^k = 0$. В этой ситуации базис покидает базисная переменная w_t^k с номером t, на котором достигается $\theta = \min\{-\bar{q}_i^k / m_{is}^k$ по всем $i : m_{is}^k < 0\}$. В соответствии с этим преобразованный столбец свободных членов $\bar{q}$ остается неотрицательным, а условие $m_{rs}^k = 0$ сохранит равенство $\bar{q}_r^k = 0$ (переменная w_r^k остается отмеченной). Можно переходить к следующей малой итерации.

И наконец, если величину θ определить нельзя, т. е. все $m_{is}^k \geq 0$, $i = 1, \ldots, n$, метод завершает работу на альтернативном луче (также без получения решения исходной задачи).

Завершая параметрическое изложение метода Лемке, заметим, что последовательности таблиц, получаемые обычным методом и его параметрическим вариантом, совпадают в следующем смысле: если в таблице параметрического метода поменять ролями отмеченную и искусственную переменные, получим соответствующую таблицу обычного метода [это можно сделать, поскольку в (3.5.k) всегда $f_r^k \neq 0$].

Пример. Вернемся к задаче, решение которой с применением метода Лемке уже было разобрано в главе 1 §5 (она же была решена в предыдущем параграфе с применением метода Данцига—Коттла). Начальная таблица для нее в параметрическом изложении имеет вид

			z_0	z_1	z_2	z_3
w_1	0	-3	1	0	-1	2
w_2	9	6	1	2	0	-2
w_3	2	-1	1	-1	1	0
			3	0	0	0

Начальное значение параметра z_0 равно 3 (минимальное значение, обеспечивающее неотрицательность базисных переменных w_1, w_2, w_3). Поскольку базисная переменная $w_1 = 0$, она объявляется отмеченной. Вводим в базис дополнительную к отмеченной переменную z_1. Выводимая переменная определяется по обычному тесту минимального отношения. Это w_3. Выполняя операцию исключения, приходим к таблице

			z_0	w_3	z_2	z_3
w_1	0	-3	1	0	-1	2
w_2	13	4	3	-2	2	-2
z_1	2	-1	1	-1	1	0
			3	0	0	0

Далее вводим в базис переменную z_3, которая является дополнительной к только что покинувшей базис переменной w_3. Поскольку элемент, стоящий на пересечении z_3-столбца отмеченной строки, отличен от нуля, отмеченная переменная покидает базис. Получаем таблицу

			z_0	w_3	z_2	w_1
z_3	1	$\frac{3}{2}$	$-\frac{1}{2}$	0	$\frac{1}{2}$	$\frac{1}{2}$
w_2	5	1	4	-2	1	-1
z_1	0	-1	1	-1	1	0
			1	0	0	0

Новое значение параметра z_0 равно 1 (снизилось), а z_1 становится новой отмеченной переменной. Вводим в базис дополнительную к z_1 переменную w_1. Поскольку в w_1-столбце элемент из отмеченной строки равен нулю, применяем тест на минимальное отношение. В соответствии с ним базис покидает переменная w_2. Получаем очередную преобразованную таблицу

$$
\begin{array}{c|c|c|c|cccc|}
\multicolumn{1}{c}{} & \multicolumn{1}{c}{} & \multicolumn{1}{c}{} & \multicolumn{1}{c}{z_0} & w_3 & z_2 & \multicolumn{1}{c}{w_2} \\
\cline{2-7}
z_3 & \frac{7}{2} & 2 & \frac{3}{2} & -1 & 1 & -\frac{1}{2} \\
w_1 & 5 & 1 & 4 & -2 & 1 & -1 \\
z_1 & 0 & -1 & 1 & -1 & 1 & 0 \\
\cline{2-7}
\multicolumn{1}{c}{} & \multicolumn{1}{c}{} & \multicolumn{1}{c}{} & \multicolumn{1}{c}{1} & 0 & 0 & \multicolumn{1}{c}{0}
\end{array}
$$

Теперь следует ввести в базис переменную z_2. Так как элемент z_2-столбца, лежащий в отмеченной строке, отличен от нуля, выводим из базиса переменную z_1 и переопределяем значение параметра z_0. Получаем завершающую таблицу

$$
\begin{array}{c|c|c|c|cccc|}
\multicolumn{1}{c}{} & \multicolumn{1}{c}{} & \multicolumn{1}{c}{} & \multicolumn{1}{c}{z_0} & w_3 & z_1 & \multicolumn{1}{c}{w_2} \\
\cline{2-7}
z_3 & 3 & 3 & \frac{1}{2} & 0 & 1 & -\frac{1}{2} \\
w_1 & 2 & 2 & 3 & -1 & 1 & -1 \\
z_2 & 1 & 1 & -1 & 1 & 1 & 0 \\
\cline{2-7}
\multicolumn{1}{c}{} & \multicolumn{1}{c}{} & \multicolumn{1}{c}{} & \multicolumn{1}{c}{0} & 0 & 0 & \multicolumn{1}{c}{0}
\end{array}
$$

Значение z_0 снизилось до нуля, а вектор $(w, z) = (2, 0, 0; 0, 1, 3)$ является решением задачи.

ЗАКЛЮЧЕНИЕ

Мы познакомили читателя лишь с кратким введением в теорию, методы и экономические приложения конечномерных вариационных неравенств и задач о дополнительности. Естественно, что многие важные и интересные аспекты обсуждаемой проблематики остались неосвещенными или освещенными недостаточно.

Так, изложение ограничено однозначными отображениями, в то время как многие приложения приводят к задачам с отображениями точечно-множественного характера, анализ которых более сложен и требует большого числа новых понятий. Не рассмотрены также спорные и не до конца проработанные вопросы двойственности для перечисленных математических постановок, а также вопросы их устойчивости и параметрического анализа. Из всего разнообразия вычислительных методов решения нелинейных задач в основном отобраны те, что апеллируют к свойствам монотонности тех или иных отображений. Не описаны другие (кроме метода Лемке и Данцига—Коттла) конечные методы решения линейной задачи о дополнительности и матричные классы, связанные с ними. Не отмечены важные в прикладном отношении методы декомпозиции задач большой размерности.

Познакомиться более глубоко с теорией конечномерных вариационных неравенств и задач о дополнительности поможет предлагаемый ниже список литературы.

Список литературы

1. *Антипин А.С.* О сходимости и оценках скорости сходимости проксимальных методов к неподвижным точкам экстремальных отображений // Журн. вычисл. математики и мат. физики. 1995. Т.35, N5. С.688–704.

2. *Астафьев Н.Н.* Линейные неравенства и выпуклость: Учеб.пособие. Свердловск: УрГУ, 1980.

3. *Бакушинский А.Б.* Методы решения монотонных вариационных неравенств, основанные на принципе итеративной регуляризации // Журн. вычисл. математики и мат. физики. 1977. Т. 17, N 6. С. 1350–1362.

4. *Бакушинский А.Б., Гончарский А.В.* Итеративные методы решения некорректных задач. М.: Наука, 1989.

5. *Булавский А.А.* Методы релаксации для систем неравенств: Учеб. пособие. Новосибирск: НГУ, 1981.

6. *Булавский А.А.* Обобщенные решения и регуляризация систем неравенств // Вычисл. методы линейной алгебры. Новосибирск: Наука, 1985. С. 161–174.

7. *Вайнберг М.М.* Вариационный метод и метод монотонных операторов. М.: Наука, 1972.

8. *Васильев Ф.П.* Методы решения экстремальных задач. М.: Наука, 1981.

9. *Васильев Ф.П.* Численные методы решения экстремальных задач. М.: Наука, 1988.

10. Введение в нелинейное программирование / Эльстер К.-Х., Рейнгард Р., Шойбле М., Донат Г.; Пер. с нем. под ред. И.И.Еремина. М.: Наука, 1985.

11. *Гловински Р., Лионс Ж.-Л., Тремольер Р.* Численное исследование вариационных неравенств. М:. Мир. 1979.

12. *Гольштейн Е.Г., Третьяков Н.В.* Модифицированные функции Лагранжа. Теория и методы оптимизации. М.: Наука, 1989. (Сер. «Экономико-мат. б-ка»).

13. *Еремин И.И.* О задачах последовательного программирования // Сиб. мат. журн. 1973. Т. 14, N 1. С. 124–129.

14. *Еремин И.И.* Противоречивые модели оптимального планирования. М.: Наука, 1988.

15. *Кокурин М.Ю.* Операторная регуляризация и исследование нелинейных монотонных задач. Йошкар-Ола: Мар. гос. ун-т, 1998.

16. *Коннов И.В.* Методы решения конечномерных вариационных неравенств: Курс лекций. Казань: ДАС, 1998.

17. *Корпелевич Г.М.* Экстраградиентный метод для отыскания седловых точек и других задач // Эконом. и мат. методы. 1976. Т.12, N4. С.747–756.

18. *Панин В.М., Скопецкий В.В., Лаврина Т.В.* Модели и методы конечномерных вариационных неравенств // Кибернетика и системный анализ. 2000. N 2. С.47–64.

19. *Попов Л.Д.* Модификация метода Эрроу—Гурвица поиска седловых точек // Мат. заметки. 1980. Т. 28, вып.5. С.777–784.

20. *Попов Л.Д.* Метод обобщенного градиентного спуска для задачи последовательного программирования // Методы аппроксимации несобственных задач математического программирования. Свердловск: УНЦ АН СССР, 1984. С. 76–82.

21. *Попов Л.Д.* Аппроксимационные корни неразрешимых уравнений с монотонными отображениями в левой части // Изв. высш. учеб. заведений. Математика. 1993. N 12. С. 70–80.

22. *Попов Л.Д.* Применение метода проекции для нахождения аппроксимационных корней монотонных отображений // Там же. 1995. N 12. С. 74–80.

23. *Aganagic M., Cottle R.W.* A constructive characterization of Q_0-matrices with nonnegative principal minors // Math. Progr. 1987. Vol.37, N 2. P.223–231.

24. *Ahn B.H.* Computation of Market Equilibria for Policy Analysis: The Project Independed Evaluation Study (PIES) Approach. Garland, N.Y, 1979.

25. *Aubin J.P.* Mathematical methods of game and economic theory. North-Holland, Amsterdam, 1979.

26. *Bertsekas D.P., Gafni E.M.* Projection methods for variational inequalities with application to the traffic assignment problem // Math. Prog. Study. 1982. Vol. 17. P. 139–159.

27. *Cottle R.W.* The principal pivoting problem // Math. Progr. Ser. B. 1990. Vol.48, N3. P.369–386.

28. *Cottle R.W., Dantzig G.B.* Complementarity pivote theory of mathematical programming // Linear Algebra and its Applications. 1968. Vol.1, N1. P.103–125.

29. *Cottle R.W., Habetler G.J., Lemke C.E.* Quadratic forms semidefinite over convex cone // Kuhn H.W., ed. Proc. of the Princeton Symp. on Math. Progr. Princeton University Press, Princeton. N.J. 1970, P. 551–565.

30. *Dafermos S.* Traffic equilibria and variational inequalities // Transportation Science. 1980. 14. P. 42–54.

31. *Eaves B.C.* On the basic theorem of complementarity // Math. Progr. 1971. N 1. P.68–75.

32. *Eaves B.C.* The linear complementarity problem // Management Science. Theory ser. 1971. Vol.17, N 9. P.612–634.

33. *Eaves B.C.* Homotopies for computation of fixed points // Math. Progr. 1972. Vol.3. P. 1–22.

34. *Ferris M.C., Meeraus A., Rutherford T.F.* Computing Wardropian equilibria in a complementarity framework // Optimization: methods and software. 1999. Vol.10. N5. P.669–686.

35. *Friesz T.L, Tobin R.L., Smith T.E. and Harker P.T.* A nonlinear complementary formulation and solution procedure for the general derived demand network equilibrium problem // J. of Regional Science. 1983. 23. P. 337–359.

36. *Fukushima M.* Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems // Math. Progr. 1992. Vol.53, N 1. P.99–110.

37. *Gabay D. and Moulin H.* On the uniqueness and stability of Nashequilibria in noncooperative games // Bensoussan A., Kleindorfer P., Tapiero C.S., eds. Applied Stochastic Control in Econometrics and Managment Science. Amsterdam, 1980. P. 271–292.

38. *Garcia C.B., Zangwill W.I.* Pathways to Solutions, Fixed Points and Equilibria. Prentice-Hall, Englewood Cliffs, N.J, 1981.

39. *Harker P.T.* A variational inequality approach for the determination of oligopolistic market equilibrium // Math. Progr. 1984. Vol.30. P. 105–111.

40. *Harker P.T.* Predicting Intercity Freight Flows. VNU Science Press, Utrecht, The Netherlands, 1987.

41. *Harker P.T., Pang J.-S.* Finite-dimensional variational inequalities and nonlinear complementarity problems: a survey of theory, algorithms and applications // Math. Progr. Ser. B. 1990. Vol.48, N2. P.161–220.

42. *Karamardian S.* Generalized complementarity problem // J. Optimiz. Theor. & Appl. 1971. N 8. P. 161–167.

43. *Karamardian S.* The complementarity problem // Math. Progr. 1972. Vol.2, N 1. P.103–129.

44. *Kuhn H.W.* Simplicial approximation of fixed points // Proc. of the National Academy of Sciences U.S.A. 1968. 61. P. 1238–1242.

45. *Kuhn H.W., Tucker A.W.* Nonlinear programming. Proceedings of the Second Symposium on Mathematics Statistics and Probability. Berkeley, University of California Press, 1951. P. 481–492.

46. *Lawphongpanich S., Hearn D.W.* Benders decompozition for variational inequalities // Math. Progr. 1990. Vol.48. N 2. P.231–248.

47. *Lemke C.E.* Some pivote schemes for linear complementarity problem // Math. Progr. Study. 1978. Vol.27. P.15–35.

48. *Lemke C.E. and Howson J.T.* Equilibrium points of bimatrix games // SIAM Review. 1964. 12. P. 45–78.

49. *Lions J.L. and Stampacchia G.* Variational inequalities // Communications on Pure and Applied Mathematics. 1967. 20. P. 493–519.

50. *Mathiensen L.* Computation of economic equilibria by a sequence of linear complementarity problems // Math. Progr. Study. 1985. Vol.23. P. 144–162.

51. *Megiddo N.* A monotone complementarity problem with feasible solutions but no complementary solutions // Math. Progr. 1977. Vol. 12. P. 131–132.

52. *Megido N., Kojima M.* On the existence and uniqueness of solutions in nonlinear complementarity theory // Math. Progr. 1977. Vol. 12. P. 110–130.

53. *Minty G.J.* On the Maximal Domain of a «Monotone» Function // The Michigan Math. J. 1961. Vol.8. N2. P. 135–138.

54. *Moreau J.J.* Proximite et dualite dans un espace Hilberiten // Bull. of Soc. of Math. of France. 1965. 93. P. 273–299.

55. *Murty K.G.* Linear Complementarity, Linear and Nonlinear Programming. B., 1988.

56. *Nash J.F.* Equilibrium points in n-person games // Proc. of the Nat. Acad. of Sciences. 1950. Vol. 36. P. 48–49.

57. *Neumann J., Morgenstern O.* Theory of games and economic behavior. 3th ed., Princeton, 1953.

58. *Pang J.S.* Solution of the general multicommodity spatial equilibrium problem by variational and complemetarity methods // J. of Regional Science. 1984. 24. P.403–414.

59. *Pang J.S., Chan D.* Iterative methods for variational and complementarity problems // Math. Progr. 1982. Vol. 24. P. 284–313.

60. *Pang J.-S., Chandrasekaran R.* Linear complementarity problem solvable by a polynomial bounded pivoting algorithm // Math. Progr. Study. 1987. Vol.25. Pt.2. P.13–27.

61. *Peng J.-M.* Equivalence of variational inequality problems to unconstrained minimization // Math.

Progr. 1997. Vol.78, N3. P.347–355.

62. *Rockafellar R.T.* Monotone operators and the proximal point algorithm // SIAM J. on Control and Optimization. 1976. Vol.14. P. 877–898.

63. *Scarf H.E. and Hansen T.* Computation of Economic Equilibriua. Yale University Press, New Haven, CT, 1973.

64. *Stone J.C.* Sequential optimization and complementarity techniques for computing economics equilibria // Math. Progr. Study. 1985. Vol. 23. P. 173–191.

65. *Todd M.J.* The Computation of Fixed Points and Applications. Springer, B., 1976.

66. *Variational Inequalities and Network Equilibrium Problems.* Edited by F.Giannessi and A.Maugeri. Plenum Press, N.Y., 1995.

Учебное издание

Попов Леонид Денисович

ВВЕДЕНИЕ В ТЕОРИЮ, МЕТОДЫ И ЭКОНОМИЧЕСКИЕ ПРИЛОЖЕНИЯ ЗАДАЧ О ДОПОЛНИТЕЛЬНОСТИ

Учебное пособие

Редактор Т. А. Сасина
Компьютерная верстка — Л. Д. Попов

ЛР. N 020257 от 22.11.96 г. Подписано в печать 18.06.2001.
Формат 60×84 1/16. Бумага для множительных аппаратов.
Усл. печ. л. 5,2. Уч.-изд. л. 5,6. Тираж 100 экз. Заказ .

Издательство Уральского университета. Екатеринбург, пр. Ленина, 51.

Отпечатано на ризографе в Гос. целевом фонде высш. шк. Свердл. обл.
Екатеринбург, Тургенева, 4.