А. В. НАЗИН
А. С. ПОЗНЯК

АДАПТИВНЫЙ ВЫБОР ВАРИАНТОВ

АДАПТИВНЫЙ ВЫБОР ВАРИАНТОВ

ТЕОРЕТИЧЕСКИЕ ОСНОВЫ
ТЕХНИЧЕСКОЙ КИБЕРНЕТИКИ

А. В. НАЗИН, А. С. ПОЗНЯК

АДАПТИВНЫЙ ВЫБОР ВАРИАНТОВ

РЕКУРРЕНТНЫЕ АЛГОРИТМЫ

МОСКВА «НАУКА»
ГЛАВНАЯ РЕДАКЦИЯ
ФИЗИКО-МАТЕМАТИЧЕСКОЙ ЛИТЕРАТУРЫ
1986

ББК 32.81
Н19
УДК 62 50

Назин А. В., Позняк А. С. **Адаптивный выбор вариантов: Рекуррентные алгоритмы.**— М.: Наука. Главная редакция физико-математической литературы, 1986.— 288 с.— (Теорет. основы техн. кибернетики).

Книга посвящена адаптивному управлению стохастическими системами с конечным множеством управляющих воздействий, или, иначе, проблеме адаптивного выбора вариантов. С единых позиций рассматриваются задачи безусловного и условного выбора, игровые задачи и задача адаптивного управления конечными однородными марковскими цепями. В каждой из них формулируется оптимизационная цель типа минимизации с вероятностью 1 предельных значений текущих средних потерь, устанавливается связь с соответствующей задачей стохастического программирования и изучаются свойства (сходимость и скорость сходимости) рекуррентных алгоритмов, обеспечивающих решение исходной задачи. Книга предназначена для научных работников и инженеров, работающих в области теории и практики адаптивных систем управления.

Табл. 5. Ил. 11. Библиогр. 150 назв.

Рецензент
доктор физико-математических наук *В. Г. Срагович*

Александр Викторович Назин, Александр Семенович Позняк

АДАПТИВНЫЙ ВЫБОР ВАРИАНТОВ
РЕКУРРЕНТНЫЕ АЛГОРИТМЫ

Редактор *М. К. Ермолова*

Художественный редактор *Т. Н. Кольченко*

Технический редактор *И. Ш. Аксельрод*

Корректоры *Л. И. Назарова, И. Я. Кришталь*

ИБ № 12898

Сдано в набор 06.05.85. Подписано к печати 25.02.86. Формат $84 \times 108^1/_{32}$. Бумага тип № 3. Гарнитура обыкновенная. Печать высокая. Усл. печ. л. 15,12. Усл. кр.-отт. 15,12. Уч.-изд. л. 16,68. Тираж 3600 экз. Заказ № 720. Цена 2 р. 80 к.

Ордена Трудового Красного Знамени издательство «Наука»
Главная редакция физико-математической литературы
117071 Москва В-71, Ленинский проспект, 15

4-я типография издательства «Наука»
630077 г. Новосибирск-77, Станиславского, 25

$Н\frac{1502000000-062}{053(02)-86}148\text{-}86$

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ

Одним из наиболее интенсивно развивающихся разделов современной теории управления является *теория адаптивных систем*. Значительный прогресс этой теории обусловлен, главным образом, тремя причинами:

— требованием практических разработчиков создавать алгоритмы, способные хорошо функционировать в ситуациях с ограниченным объемом априорной информации об управляемом объекте и внешней среде;

— интенсивным развитием вычислительной техники, позволяющей просто и быстро реализовать адаптивные алгоритмы;

— новыми результатами в математической теории случайных процессов, которые позволили конструктивно подойти к проблемам анализа и синтеза эффективных адаптивных алгоритмов.

Обширная литература, посвященная проблемам адаптивного управления, может быть разделена на две группы: работы по системам *с непрерывным управлением* (с континуальным множеством значений управляющих воздействий) и *с дискретным управлением* (с конечным или счетным множеством значений управляющих воздействий). Среди публикаций *по теории систем с непрерывным управлением* отметим монографии В. Н. Фомина, А. Л. Фрадкова, В. А. Якубовича [126], Д. П. Деревицкого и А. Л. Фрадкова [25], Г. Гудвина и К. Сина [22]. Основу математического аппарата, применяемого в этих работах, составили теория стохастических дифференциальных уравнений и теория мартингалов.

Исследование *адаптивных систем с дискретным управлением* было начато Г. Роббинсом [102], Р. Бушем и Ф. Мостеллером [12]. Новый импульс развитию этой области дали работы М. Л. Цетлина [132] и его последователей, проводившиеся долгое время независимо от работ зарубежных авторов. Эти исследования основывались на использовании другого математического аппарата — теории марковских цепей. При этом подход в задачах с дискретным управлением принципиально отличался от принятого в задачах с непрерывным управлением оптимизационного подхода: изучался поведенческий аспект того или иного конечного автомата, имитирующего целесообразное поведение живых существ в определенных ситуациях. В работе В. И. Варшавского и И. П. Воронцовой [15], касающейся поведения стохастических автоматов с формируемой структурой, фактически используется возможность подхода к вопросам синтеза и анализа таких автоматов с тех же позиций, которые сложились в теории адаптивных систем с непрерывным управлением (хотя авторы явно и не указывали на эту возможность). По-видимому, впервые эта возможность была реализована в работах В. Г. Сраговича и Ю. А. Флерова [119, 120] и независимо К. Нарендры и И. Шапиро [143]. В последней статье предложено для отыскания минимума функции в условиях помех применить обучающийся стохастический автомат, представляющий собой, по сути, алгоритм перестройки вероятностей выбора очередного значения аргумента минимизируемой функции. Я. З. Цыпкиным и А. С. Позняком [137] была установлена возможность использования метода стохастической аппроксимации для формирования рекуррентных рандомизированных процедур адаптации систем с дискретным управлением, и тем самым подход, развитый в монографии Я. З. Цыпкина [134], был распространен и на этот класс систем. Это направление, позволяющее исследовать уже известные и синтезировать новые адаптивные алгоритмы дискретного управления с единых оптимизационных позиций, разрабатывалось также и авторами настоящей книги. Аналогичный подход использовался и зарубежными авторами (К. Нарендра, М. Тхатхачар [78] и др.), хотя сами алгоритмы, в основном, носили эвристический характер. Эти исследования нашли свое отражение в недавно опубликованной монографии С. Лакшмиварахана [53]. Этапы развития теории адаптивных систем с дискретным управлением подробно отражены в монографиях Д. А. Поспелова [95, 96], В. И. Варшавского [14] и В. Г. Сраговича [117, 118].

Цель настоящей книги состоит в изложении как известных, так и новых результатов, полученных в теории адаптивных систем с дискретным управлением (с конечным числом значений управляющих воздействий). Задачи, изучаемые в этой теории, можно интерпретировать как задачи адаптивного выбора вариантов, состоящие в нахождении стратегии управления (или, иными словами, алгоритма формирования такой последовательности вариантов (управляющих воздействий)), которая в определенном вероятностном смысле обеспечивает минимум предельных средних потерь, характеризующих качество осуществляемого процесса выбора (управления). Использование термина «вариант» как бы подчеркивает дискретность множества значений управляющих воздействий. Единство принятого здесь подхода к решению различных задач адаптивного выбора вариантов проявляется не только в общности используемого математического аппарата и общей для всех разделов книги идеи рандомизации, но и в последовательности изложения материала по каждой рассматриваемой задаче: сначала приводится постановка задачи (в рамках сделанных предположений), обосновывается ее корректность, устанавливается связь с некоторой задачей стохастического программирования, а затем формируются и исследуются рекуррентные рандомизированные алгоритмы, обеспечивающие решение поставленной задачи. Структура книги такова, что изложенный в ней материал представлен в порядке возрастания сложности задач: от задач безусловного и условного выбора к игровым и динамическим задачам.

Первая глава посвящена общей проблематике адаптивного выбора и носит вводный характер. Отмечаются особенности рассматриваемого круга задач. Приводятся примеры задач выбора вариантов в условиях априорной неопределенности, встречающихся на практике. Затем излагается обзор известных методов и подходов к решению различных задач адаптивного выбора вариантов, который, однако, не претендует на исчерпывающую полноту и отражает лишь основные направления и результаты в этой области. Глава заканчивается обсуждением особенностей рандомизационного подхода.

Во второй главе детально исследуется задача безусловной минимизации предельных средних потерь, возникающих в процессе выбора. Особо рассматривается случай бинарных потерь. Исследуются свойства часто встречающихся в литературе алгоритмов Нарендры — Шапиро, Варшавского — Воронцовой, Льюса, Буша — Мостеллера. Устанавливаются условия, при которых они (или их модификации) решают поставленную задачу. Для задач с небинарными потерями предлагается и исследуется ряд алгоритмов стохастической аппроксимации, а также алгоритм метода случайного поиска.

Третья глава посвящена задаче условной минимизации средних потерь, когда выбор ограничивается некоторыми априори неизвестными ограничениями. Подробно обсуждаются методы штрафных функций и множителей Лагранжа, на основе которых строятся адаптивные алгоритмы. Устанавливаются условия их работоспособности и оценивается скорость сходимости, которая определяет и скорость решения исходной задачи.

В четвертой главе исследуется одна из возможных постановок игровой задачи адаптивного выбора вариантов, которая может быть интерпретирована как задача о поиске равновесия по Нэшу (в смешанных стратегиях) в матричной игре многих лиц в условиях априорной неопределенности. Предлагаются и исследуются алгоритмы ее решения.

Пятая глава посвящена проблеме адаптивного выбора в конечных динамических системах, а именно задаче адаптивного управления однородной конечной марковской цепью. Устанавливается корректность приводимой постановки задачи и обосновывается работоспособность предложенного здесь же алгоритма адаптивного управления, реализующего метод стохастической аппроксимации. Отдельно рассматриваются случаи регулярных и связных управляемых цепей, а также цепей более общего вида.

Авторы выражают глубокую благодарность своему учителю члену-корреспонденту АН СССР Я. З. Цыпкину за то, что он инициировал и поддерживал их работы в этом направлении. Краткие, но весьма полезные контакты с Б. Т. Поляком позволили осознать и преодолеть ряд трудностей, возникших при исследовании алгоритмов адаптации. Большую работу по рецензированию рукописи выполнил В. Г. Срагович, чьи многочисленные советы и критические замечания оказали существенное влияние на общий стиль книги. Своевременную помощь в оформлении рукописи оказывали Т. Г. Краузе и Н. В. Епихова. Авторы глубоко признательны всем им.

Все разделы книги писались авторами совместно, и поэтому претензии и замечания читателей авторы в равной степени принимают на свой счет и заранее благодарят всех за проявленный интерес к книге.

А. В. Назин, А. С. Позняк

СПИСОК ОБОЗНАЧЕНИЙ

Нумерация утверждений и формул двойная, самостоятельная в каждой главе, а номера утверждений и формул из приложения снабжены буквой П.

▲ — конец доказательства,

Ω — множество элементарных событий (исходов) ω,

ω — элементарный исход,

$\mathcal{F}$ — σ-алгебра событий на Ω,

P — вероятностная мера на $(\Omega, \mathcal{F})$,

$(\Omega, \mathcal{F}, \mathsf{P})$ — вероятностное пространство, на котором считаются заданными все встречающиеся случайные величины,

$\triangleq$ — равно по определению,

$\operatorname{int} D$ — множество, состоящее из внутренних точек множества D,

$\in$ — знак принадлежности элемента множеству,

$\notin$ — знак отсутствия элемента в множестве,

п. н. — почти наверное,

$\Leftrightarrow$ — тогда и только тогда, когда...,

R^k — k-мерное евклидово пространство,

R_+^k — множество векторов из R^k, имеющих неотрицательные компоненты,

$i = \overline{1, N}$ — $i = 1, 2, \ldots, N$,

$p = (p(1), \ldots, p(N))$ — вектор-столбец с компонентами $p(1), \ldots, p(N)$,

$$S^N = \{p \in R^N \mid p(i) \geqslant 0 \ \ (i = \overline{1, N}),\ \sum_{i=1}^{N} p(i) = 1\}$$ — $(N-1)$-мерный единичный симплекс,

$S_\varepsilon^N = \{p \mid p \in S^N,\ p(i) \geqslant \varepsilon\ (i = \overline{1, N})\}$ — $(N-1)$-мерный единичный ε-симплекс,

$\|p\|$ — евклидова норма вектора p,

$\langle \cdot, \cdot \rangle$ — скалярное произведение,

$\sigma(\cdot)$ — минимальная σ-алгебра, порожденная перечисленными в скобках случайными величинами,

M — символ математического ожидания,

$\mathsf{M}\{\cdot \mid \mathcal{F}_n\}$ — условное математическое ожидание по σ-алгебре $\mathcal{F}_n$,

$\mathsf{P}\{\cdot \mid \mathcal{F}_n\}$ — условная вероятность по σ-алгебре $\mathcal{F}_n$,

$\mathsf{M}\{\cdot \mid \mathcal{F}_n, x_n\}$ — условное математическое ожидание по минимальной σ-алгебре, содержащей $\mathcal{F}_n$ и $\sigma(x_n)$,

$\mathsf{M}\{\cdot \mid \mathscr{F}_n, x_n = x(i)\} \triangleq \mathsf{M}\{\cdot \mid \mathscr{F}_n, x_n\}|_{x_n = x(i)}$,

$\xi_n = \xi_n(x_n, \omega)$ — случайные потери за выбор варианта x_n в момент времени t_n,

$\Phi_n = \frac{1}{n}\sum_{t=1}^{n} \xi_t$ — текущие средние потери,

$\Phi = \overline{\lim_{n\to\infty}} \Phi_n$ — наибольшие асимптотические средние потери,

$\{\varkappa_n\}$ — бесконечная последовательность элементов $\varkappa_1, \ldots, \varkappa_n, \ldots$,

$O(\varkappa_n)$, $o(\varkappa_n)$, $O^*(\varkappa_n)$ — числовые последовательности, удовлетворяющие условиям $\overline{\lim_{n\to\infty}} |O(\varkappa_n)/\varkappa_n| < \infty$, $\lim_{n\to\infty} o(\varkappa_n)/\varkappa_n = 0$, $0 < \underline{\lim_{n\to\infty}} |O^*(\varkappa_n)/\varkappa_n| \leqslant \overline{\lim_{n\to\infty}} |O^*(\varkappa_n)/\varkappa_n| < \infty$,

$O(1)$, $o(1)$, $O^*(1)$ — $O(\varkappa_n)$, $o(\varkappa_n)$, $O^*(\varkappa_n)$ при $\varkappa_n \equiv 1$ соответственно,

$\varkappa_n \xrightarrow[n\to\infty]{} \varkappa \Leftrightarrow \lim_{n\to\infty} \varkappa_n = \varkappa$,

$\chi(\cdot)$ — индикаторная функция,

$\mathrm{Spur}(A)$ — след матрицы A,

т — верхний индекс означает транспонирование,

I — единичная матрица,

I_u — единичная матрица размером, соответствующим вектору u,

E — квадратная матрица, все элементы которой равны 1,

$e_i^N = (\underbrace{0, \ldots, 0, 1}_{i}, \underbrace{0, \ldots, 0}_{N-i})$,

$e^N = (1, \ldots, 1) \in R^N$,

$e(x) \triangleq \sum_{i=1}^{N} e_i^N \chi(x = x(i))$, если $X = \{x(1), \ldots, x(N)\}$,

δ_{ij} — символ Кронекера,

$\|\cdot\|_{\substack{i=\overline{1,m}\\ j=\overline{1,l}}}$ — матрица, имеющая m строк и l столбцов,

$\|\cdot\|_{i,j=\overline{1,K}}$ — квадратная $K \times K$-матрица, i — номер строки, j — номер столбца,

∇ — градиент,

∇_p — градиент по вектору p,

$\nabla_p^2 V$ — матрица вторых производных функции V по компонентам вектора p,

$\partial \bar{V}(c)/\partial c^{ij}$ — частная производная функции $\bar{V}(c)$ по переменной c^{ij},

$\partial \bar{V}(c)/\partial c$ — градиент функции $\bar{V}(c)$ по матрице c,

$X_1 \times X_2$ — прямое произведение множеств X_1, X_2,

$\varnothing$ — пустое множество,

π_ε^N, $\pi_\varepsilon^{\mathbf{C}}$, $\pi_\varepsilon^{\widehat{\mathbf{C}}}$ — операторы проектирования на множества S_ε^N, $\mathbf{C}_\varepsilon$ (5.55) и $\widehat{\mathbf{C}}_\varepsilon$ (5.64) соответственно,

$\tau \wedge n \triangleq \min\{\tau, n\}$.

ГЛАВА 1

ЗАДАЧИ АДАПТИВНОГО ВЫБОРА ВАРИАНТОВ

Обсуждается проблема выбора вариантов в условиях априорной неопределенности, возникающая в системах с дискретным управлением. Приводятся конкретные примеры таких задач. Дается обзор известных методов адаптивного выбора вариантов и обсуждаются достоинства и недостатки рандомизированных стратегий.

§ 1.1. Выбор вариантов в условиях неопределенности

При определении оптимальных режимов функционирования многих технических, биологических и организационных систем, способов управления различными технологическими процессами, методов организации вычислений в современных ЭВМ и т. д. приходится сталкиваться с одной важной и достаточно общей проблемой: *для достижения заданной цели необходимо многократно осуществлять выбор какого-либо варианта среди конечного числа возможных.*

В качестве *возможных вариантов*, среди которых производится выбор, могут выступать, например, положения переключателей, различные методики проведения технологических процессов, очередность обслуживания или выполнения работ, значения дискретных величин — параметров систем.

Цели, к достижению которых необходимо стремиться при осуществлении выбора, также могут быть различными: минимизация функционала потерь с учетом или без учета дополнительных ограничений, обеспечение выполнения некоторых целевых неравенств, достижение компромисса при оптимизации нескольких противоречивых критериев и т. д.

Во многих реальных ситуациях выбор вариантов приходится осуществлять в условиях *априорной неопределенности*, когда, в принципе, по имеющимся данным нельзя заранее указать, какие из возможных вариантов следует выбирать, чтобы обеспечить достижение заданной цели. Более того, наилучшую стратегию выбора невозможно точно определить в результате конечного числа реализаций, и поэтому цели адаптации должны быть асимптотическими, т. е. формулироваться в терминах предельных значений функционала потерь каждого из возможных вариантов. Априорная неопределенность может быть двух видов [134—136]:

1) неопределенность целей и требований, состоящая в отсутствии точной или корректной формулировки цели и требований к выбору вариантов;

2) неопределенность обстановки, заключающаяся в неполноте информации о ситуациях, в которых осуществляется выбор.

Далее рассматривается главным образом неопределенность второго вида. Отметим, что в том случае, когда при осуществлении выбора априорная информация достаточно полна, имеется принципиальная возможность заранее для каждой ситуации точно определить оптимальный вариант. В противном случае, когда выбор вариантов необходимо осуществлять в условиях априорной неопределенности (т. е. когда мы имеем дело с классом управляемых систем), достижение заданной цели возможно лишь на основе применения *адаптивного подхода* [117, 134]. Суть его состоит в надлежащем использовании текущей информации, получаемой в результате осуществления отдельных актов выбора, что позволяет компенсировать с течением времени недостаток априорной информации и тем самым реализовать оптимальную на классе систем стратегию управления. Задачу определения наилучшей стратегии выбора вариантов (управляющих воздействий) в условиях априорной неопределенности будем называть *задачей адаптивного выбора вариантов*.

В соответствии с адаптивным подходом эта задача решается непосредственно в процессе многократного выбора путем обработки текущей информации.

Рассмотрим ряд примеров, иллюстрирующих разнообразие возникающих на практике задач адаптивного выбора вариантов.

§ 1.2. Примеры задач адаптивного выбора вариантов

1. *Кодирование и декодирование в системах передачи дискретной информации.* В теории передачи информации [87, 88, 127] хорошо известен тот факт, что при наличии помех в канале связи за счет выбора надлежащего способа кодирования вероятность ошибочного приема может быть сделана сколь угодно малой, если скорость передачи информации не превышает пропускную способность канала связи [145, 146]. Разработанные способы кодирования и декодирования позволяют обнаруживать и исправлять различные виды ошибок, происходящих в канале связи. С точки зрения простоты технической реализации предпочтительными являются линейные циклические коды, для которых кодирующие и декодирующие устройства реализуются на основе сдвиговых регистров с линейной обратной связью.

Знаниями о корректирующих способностях кодов можно непосредственно воспользоваться лишь тогда, когда имеется достаточно полная информация о статистических характеристиках помех, действующих в канале связи. В случае априорной неопределенности относительно статистических свойств помех повышение достоверности передачи информации возможно на основе применения адаптивного подхода [99, 121, 144]. Этот подход может быть реализован следующим образом: чередуя передаваемую информацию с тестовыми сигналами, можно получать текущую информацию о свойствах помех в канале связи; обработка этой информации соответствующим алгоритмом адаптации позволяет так изменять способ кодирования и декодирования, что с течением времени достоверность передачи информации повышается до требуемого уровня.

Для построения адаптивной системы передачи выберем r кодов K_i $(i = \overline{1, r})$ с различными корректирующими способностями. При их выборе должна учитываться априорная информация о наиболее вероятных вариантах ошибок, происходящих в канале связи, а также ограничения на скорость передачи и сложность технической реализации. Каждый код полностью определяет параметры кодирующего устройства в рамках выбранной схемы кодирования. Параметры декодирующего устройства при фиксированном способе декодирования полностью определяются кодом и теми ошибками, для обнаружения и исправления которых он предназначается. Поскольку один и тот же код, в общем случае, может быть использован для исправления различных видов ошибок [87, 88], для реализации этой возможности каждому коду K_i поставим в соответствие s_i наборов значений параметров декодирующего устройства D_{ij} $(j=\overline{1, s_i})$, при каждом из которых i-м кодом исправляется j-й вид ошибок. Таким образом, задано множество пар (K_i, D_{ij}), которое

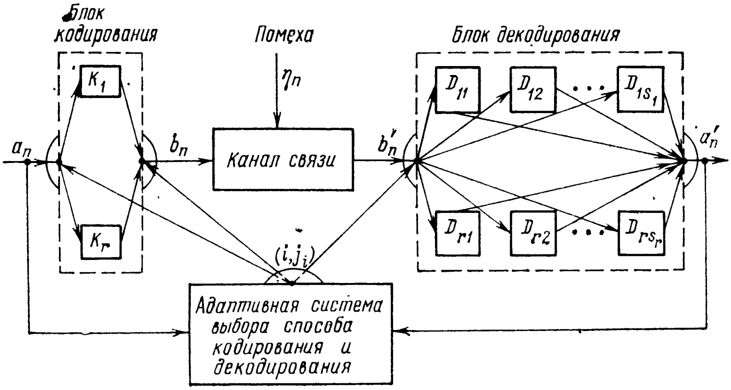

Рис. 1

является множеством возможных вариантов в рассматриваемой задаче. Каждый вариант полностью определяет параметры кодирующего и декодирующего устройств, а следовательно, и степень достоверности передачи информации.

Рассмотрим систему передачи, изображенную на рис. 1. Пусть тестовые сигналы поступают на вход системы в моменты времени $t_1, \ldots, t_n, \ldots$ В те же моменты времени может производиться смена параметров кодирующего и декодирующего устройств. В результате кодирования кодом K_i тестового сигнала a_n, поступившего в момент времени t_n, по каналу связи передается сообщение b_n. Вследствие действия случайной помехи в канале связи $\eta_n(\omega)$*) на блок декодирования с параметрами D_{ij} поступает искаженное сообщение b_n'. Сравнение тестового сигнала a_n с сигналом a_n', полученным в результате

*) Здесь и далее ω — элементарный исход.

декодирования сообщения b_n', позволяет определить величину

$$\xi_n = \xi_n(K_i, D_{ij}, \omega) \triangleq \begin{cases} 1, & \text{если} \quad a_n \neq a_n', \\ 0, & \text{если} \quad a_n = a_n', \end{cases}$$

которую будем трактовать как потери системы передачи на интервале времени (t_n, t_{n+1}), соответствующие варианту (K_i, D_{ij}). Тогда величина текущих средних потерь

$$\Phi_n \triangleq \frac{1}{n} \sum_{t=1}^{n} \xi_t, \quad n = 1, 2, \ldots \tag{1.1}$$

будет характеризовать качество последовательности вариантов (K_i, D_{ij}), выбранных в моменты времени $t_1, \ldots$ $\ldots, t_n$, а предельные значения последовательности $\{\Phi_n\}$ — качество всей последовательности вариантов. Очевидно, величина Φ_n представляет собой текущую частоту ошибочного приема тестовых сигналов.

Потери $\xi_n(K_i, D_{ij}, \omega)$ представляют собой текущую информацию о достоверности передачи в соответствии с вариантом (K_i, D_{ij}). Эту информацию можно использовать в целях формирования последовательности вариантов кодирования и декодирования, минимизирующей в асимптотике текущие средние потери Φ_n (1.1). Отметим, что для реализации этой возможности в системе передачи необходимо наличие надежной обратной связи, по которой должны передаваться указания на изменение используемого кода.

Таким образом, целью адаптации здесь является *безусловная минимизация* наибольшего предельного значения текущих средних потерь Φ_n (1.1), что соответствует минимизации вероятности ошибочного приема.

2. *Водоохлаждение химических реакторов.* Многие современные способы производства ряда химических продуктов (например, полипропилена, изопрена и др.) связаны с проведением экзотермических реакций, для поддержания нормальных условий протекания которых требуется вовремя и в нужном количестве отводить от реакторов выделяемое тепло. Для обеспечения необходимого теплового режима используются системы водоохлаждения, в которых отбор тепла осуществляется водой, циркулирующей по замкнутому контуру. Нагретая вода

проходит через специальные холодильники (градирни) и затем вновь омывает реактор (рис. 2).

Количество тепла, отводимого из реактора, прямо пропорционально произведению интенсивности водяного потока и разности температур в выходной и входной магистралях. Регулировка величины потока осуществляется

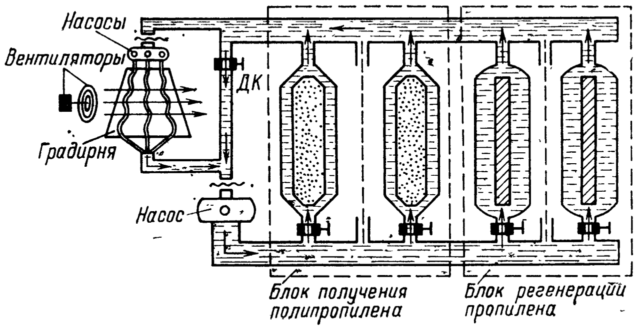

Рис. 2

с помощью специального дренажного клапана (ДК), имеющего конечное число N степеней перекрытия потока в обводной магистрали. Такой способ регулирования связан с двумя противоположными по результату явлениями: с одной стороны, увеличение степени открытия клапана приводит к увеличению водяного потока, с другой стороны — к уменьшению перепада температур в магистралях. Таким образом, при фиксированных внешних условиях существует некоторое наилучшее положение клапана, при котором обеспечивается максимальный теплосъем.

Температура окружающей среды, влияющая на процесс теплосъема, не является, очевидно, стационарным случайным процессом. Однако для каждого сезона (зима, лето, весна — осень) и каждого суточного периода (день, ночь, утро — вечер) этот процесс можно считать стационарным. Рассмотрим последовательность суточных периодов, соответствующих одному и тому же режиму стационарности атмосферной температуры. Пусть моменты времени t_n $(n = 1, 2, \ldots)$ переключения дренажного клапана в одно из возможных положений $x(i)$ $(i = \overline{1, N})$ отстоят друг от друга на величину τ, равную времени затухания переходного процесса теплосъема ($\tau \approx 10$ мин). Измерив с помощью расходомера и термопары интенсивность водяного потока и разность температур в магистралях, можно рассчитать количество отводимого из реактора тепла в единицу времени $Q_n \triangleq Q_n(x(i), \omega)$. Случайный характер величины Q_n обусловлен различными факторами: колебаниями атмосферной температуры и интенсивности водяного потока, нестабильностью подачи сырья в реактор, неоднородностью состава сырья, погрешностями измерений и т. д.

Определим вличину потерь системы водоохлаждения на интервале времени (t_n, t_{n+1}), соответствующих положению дренажного клапана $x_n = x(i)$, как $\xi_n \triangleq -Q_n(x(i), \omega)$. Задача формирования последовательности $\{x_n\}$ положений дренажного клапана, обеспечивающей в асимптотике минимум текущих средних потерь

$$\Phi_n \triangleq \frac{1}{n}\sum_{t=1}^{n}\xi_t = -\frac{1}{n}\sum_{t=1}^{n}Q_t(x_t, \omega), \tag{1.2}$$

является задачей адаптивного выбора режимов (вариантов) водоохлаждения химического реактора. Величина $-\Phi_n$ представляет собой среднее на интервале (t_1, t_{n+1}) количество тепла, отведенного из реактора. Таким образом, сформулированная задача также относится к классу задач *безусловной минимизации* наибольшего предельного значения текущих средних потерь (1.2), что соответствует максимизации среднего, отводимого из реактора количества тепла при существующих внешних условиях.

3. *Испытание машинных агрегатов.* При создании новых машин и механизмов, как правило, необходимо проводить многочасовые испытания опытных образцов с целью выявления их свойств и возможностей. При этом в процессе испытаний необходимо определять наиболее рациональные и производительные режимы работы и возможно большую часть времени проводить испытания именно в этих режимах. С этой целью испытатель должен по нескольку раз опробовать каждый режим, выбирая среди них наиболее эффективный. Таким образом, здесь мы также сталкиваемся с задачей адаптивного выбора вариантов, поскольку процесс испытания проходит в условиях неопределенности: недостаточно точно известны свойства испытываемой машины, свойства окружающей среды, законы взаимодействия с ней машины и т. д.

В качестве конкретного примера рассмотрим задачу адаптивного выбора вариантов, возникающую при испытании бульдозерных агрегатов в процессе рытья траншей. В соответствии с принятыми нормами эксплуатационные испытания проводятся в объеме 5000—6000 моточасов.

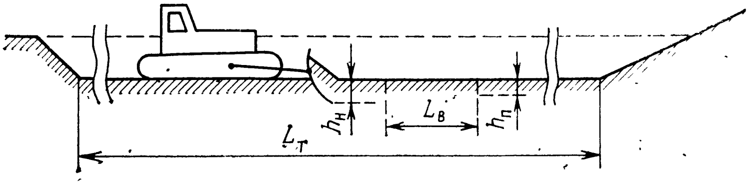

Рис. 3

При этом одной из важнейших задач является определение максимальной средней производительности бульдозера (на грунте заданной категории) при допустимом среднем расходе топлива. Испытания проводятся на траншеях определенных размеров (15—25 м, 30—50 м и 60—100 м), при этом производительность вычисляется как отношение объема извлеченного грунта ко времени, затраченному на всю работу [19].

Основными этапами в работе бульдозерного агрегата при рытье траншеи являются (рис. 3):

1) набор грунта (срез пласта с накоплением максимально возможного объема),

2) транспортировка, или волочение набранного грунта (при этом часть набранного грунта рассыпается),

3) подбор грунта (дополнительный срез грунта по ходу движения с целью компенсации рассыпавшейся части),

4) работа на кавальер (транспортировка и отсыпка грунта вне траншеи),

5) холостой ход (передвижение бульдозера в начало очередного участка набора). Режим работы бульдозерного агрегата определяется следующими тремя параметрами: глубиной набора грунта $h_н$ (при наборе на постоянной глубине), длиной участка волочения $L_в$ и глубиной подбора грунта $h_п$.

Параметры $h_н$, $L_в$ и $h_п$ являются непрерывными, т. е., в принципе, могут принимать любые значения в допустимых или разумных пределах. Так, толщина слоя $h_н$ и глубина подбора $h_п$ могут равняться любой величине из отрезка $[0, H]$, где H — максимально возможная величина заглубления отвала. Однако на практике водитель бульдозера может реализовать с достаточной точностью лишь четыре значения, равные $0{,}25\,H$, $0{,}5\,H$, $0{,}75\,H$ и H. Аналогично, лишь конечное число значений длины участка волочения $L_{\text{в}}$, равных целым количествам метров, имеет практический смысл. Кроме того, поскольку, как оказывается, производительность бульдозера значительно менее чувствительна к вариациям длины участка волочения $L_{\text{в}}$, чем глубины подбора $h_{\text{п}}$, можно ограничиться только частью этих значений. При этом множество значений $L_{\text{в}}$ зависит от длины траншеи $L_{\text{т}}$. Так, например, при $L_{\text{т}} = 50$ м в качестве возможных значений $L_{\text{в}}$ достаточно взять следующие величины: 10, 15, 20, 25 и 30 метров.

Пусть множества возможных значений величин $h_{\text{п}}$, $L_{\text{в}}$ и $h_{\text{н}}$ равны $\{h(i)\,|\,i = \overline{1,\ N_1}\}$, $\{L(j)\,|\,j = \overline{1,\ N_2}\}$ и $\{h(k)\,|\,k = \overline{1,\ N_3}\}$ соответственно; $n = 1,\ 2,\ \ldots$ — номер очередной траншеи (некоторой фиксированной длины $L_{\text{т}}$), разрабатываемой в процессе испытаний бульдозера. Пусть, кроме того, в момент времени t_n $(n = 1,\ 2,\ \ldots)$ начала разработки n-й траншеи в качестве значений величин $h_{\text{н}}$, $L_{\text{в}}$, $h_{\text{п}}$ выбраны соответственно значения $h_{\text{н},\,n}$, $L_{\text{в},\,n}$ и $h_{\text{п},\,n}$, определяющие один из возможных вариантов разработки траншеи. По окончании разработки траншеи вычисляется производительность $\Pi_n = \Pi_n(h_{\text{н},\,n},\ L_{\text{в},\,n},\ h_{\text{п},\,n},\ \omega)$, показанная бульдозером, а также расход топлива $R_n = R_n(h_{\text{н},\,n},\ L_{\text{в},\,n},\ h_{\text{п},\,n},\ \omega)$ в единицу времени. Тогда в качестве величин ξ_n^0 и ξ_n^1, характеризующих потери на интервале времени $(t_n,\ t_{n+1})$, можно принять $\xi_n^0 \triangleq -\Pi_n$ и $\xi_n^1 \triangleq R_n - \overline{R}$, где $\overline{R}$ — максимальный допустимый средний расход топлива. Случайность величин ξ_n^0, ξ_n^1 обусловлена несколькими факторами: неровностью поверхности земли, случайностью коэффициентов трения грунта о грунт и сопротивления грунта резанью, случайностью плотности грунта, отклонениями от заданного варианта разработки и т. д. Средние потери

$$\Phi_n^j \triangleq \frac{1}{n}\sum_{t=1}^{n} \xi_t^j, \quad j = 0, 1; \quad n = 1, 2, \ldots \tag{1.3}$$

характеризуют качество последовательности реализованных к моменту времени t_n вариантов разработки траншей: $-\Phi_n^0$ представляет собой текущую среднюю производительность, а Φ_n^1 — текущий средний перерасход топлива против максимально допустимого.

Таким образом, задачу адаптивного выбора вариантов разработки траншей можно сформулировать как задачу минимизации наибольшего предельного значения текущих средних потерь Φ_n^0 при условии, что все предельные значения последовательности Φ_n^1 неположительны. Другими словами, здесь требуется обеспечить максимальную среднюю производительность агрегата при допустимом среднем расходе топлива. Эта задача относится к классу задач *условной минимизации.*

4. *Диспетчеризация рабочих программ ЭУМ для узла коммутации телефонной сети связи.* Управление процессом приема, обработки и выдачи информации при обслуживании абонентов современных узлов коммутации (УК) осуществляется специализированной электронной управляющей машиной (ЭУМ) [4, 51]. Для обеспечения достаточно высокого уровня надежности обслуживания вызовов в состав рабочих программ ЭУМ включают программы контроля правильности установления соединений между коммутационной системой и различными типами передатчиков и приемников, а также соединений внутри самой коммутационной системы [52]. Выполнение этих программ ЭУМ в процессе обслуживания вызовов требует определенного времени и, следовательно, снижает пропускную способность узла коммутации. При малых нагрузках на УК это не имеет значения. Однако в часы наибольшей нагрузки низкая величина пропускной способности УК приводит к потерям вызовов, причем эти потери могут превосходить возможные потери из-за ненадежного обслуживания в отсутствие рабочих программ контроля [115]. Целесообразно поэтому изменять частоту использования программ контроля правильности установления соединений при изменении нагрузки на УК [52].

Обозначим через l число рабочих программ контроля, входящих в состав математического обеспечения ЭУМ, а сами эти программы — через $r_1, \ldots, r_l$. Время, необходимое для выполнения той или иной рабочей программы, в общем случае зависит от состояния коммутационной системы и, таким образом, может рассматриваться как случайная величина. Время обслуживания вызовов в ЭУМ определяется с учетом многоэтапности процесса обслуживания вызовов на УК. Без нарушения общности можно считать, что каждому этапу обслуживания соответствует рабочая программа.

Пусть на УК поступают вызовы m различных видов. Для каждого вида образуется очередь, в которой вызовы ожидают момента начала обслуживания. Решение о выборе на обслуживание вызова того или иного вида принимается программой-диспетчером. В каждый момент времени на обслуживании в ЭУМ может находиться лишь один вызов. При выборе на обслуживание вызова k-го вида $(k=\overline{1,\ m})$ программа-диспетчер указывает, какие из рабочих программ и в какой последовательности должны быть выполнены.

Процесс диспетчеризации рабочих программ можно организовать следующим образом. Разобьем мысленно время на последовательные интервалы некоторой длительности T и перенумеруем их числами 1, 2, ... в порядке наступления. Поставим в соответствие интервалу с номером n вектор $x_n = \left(x_n^1, \ldots, x_n^l\right)$, каждая компонента которого принимает два значения: 0 или 1. При этом в течение n-го интервала времени программа r_k включается в процесс обслуживания вызовов, если $x_n^k = 1$ и не включается, если $x_n^k = 0$ $(k=\overline{1,\ l})$. Отметим, что конкретная величина T не имеет принципиального значения. Она должна лишь удовлетворять следующему условию: быть в несколько раз больше среднего времени обслуживания одного вызова в ЭУМ и одновременно быть значительно меньше продолжительности интервала стационарности нагрузки на УК. Как правило, достаточно взять $T = 1$ с. Отметим, что здесь моментом времени t_n является момент начала n-го интервала времени T, а вариантом, выбранным в этот момент времени,— вектор x_n, причем множеством значений k-й его компоненты является $X_k = \{0, 1\}$.

При формировании последовательности вариантов необходимо обеспечить выполнение следующих двух требований: 1) среднее число вызовов, поступивших на обслуживание за время T не должно превышать среднего числа вызовов, обслуженных за это же время, и 2) уровень надежности обслуживания вызовов должен быть возможно более высоким. Для формализации этих требований поступим следующим образом. Поставим в соответствие каждой рабочей программе контроля r_k $(k=\overline{1,\ l})$

последовательность случайных величин

$$\xi_n^k = \eta_n^{\text{вх}} - \eta_n^{\text{вых}} + \alpha_k (T - \tau_n), \quad n = 1, 2, \ldots,$$

где n — номер текущего интервала времени длительности T, $\eta_n^{\text{вх}}$ и $\eta_n^{\text{вых}}$ — количество поступивших на УК и обслуженных за этот интервал времени вызовов, α_k — некоторые положительные весовые константы, τ_n — оценка времени обслуживания вызовов в ЭУМ, вычисляемая на основе априорных сведений о средних временах работы выполняемых программ:

$$\tau_n = \sum_{i=1}^{m} \zeta_n^i \left(\tau_i^R + \sum_{j=1}^{l} x_n^j \tau_j^r \right);$$

здесь ζ_n^i — количество вызовов i-го вида, обслуженных в ЭУМ в течение n-го интервала времени T, τ_i^R — среднее суммарное время выполнения рабочих программ (обязательных для каждого вызова) при обслуживании одного вызова i-го вида, τ_j^r — среднее время выполнения программы контроля r_j. Разность $\eta_n^{\text{вх}} - \eta_n^{\text{вых}}$ показывает, как ЭУМ справляется с обслуживанием поступающего на УК потока вызовов. Величина $T - \tau_n$ представляет собой ту часть n-го интервала времени T, в течение которой ЭУМ простаивала. Это время (если оно достаточно велико) могло быть использовано для работы программ контроля, поэтому разность $T - \tau_n$ характеризует возможность увеличения надежности обслуживания вызовов в ЭУМ. То, какая из двух разностей, $\eta_n^{\text{вх}} - \eta_n^{\text{вых}}$ или $T - \tau_n$, вносит больший вклад в величину ξ_n^k, зависит от весового коэффициента α_k. Будем трактовать ξ_n^k как потери при выборе того или иного значения x_n^k $(k = \overline{1, l})$. Именно, при формировании последовательности $\{x_n^k\}$ будем преследовать такую цель: обеспечить (с ростом n) возможно меньшее значение средних потерь

$$\Phi_n^k \triangleq \frac{1}{n} \sum_{t=1}^{n} \xi_t^k, \quad k = \overline{1, l}. \tag{1.4}$$

Другими словами, требуется определить асимптотически оптимальное (в смысле критериев (1.4)) децентрализованное управление [65].

Процесс диспетчеризации рабочих программ в ЭУМ полностью определяется последовательностью вариантов $\{x_n\}$. Поскольку характеристики нагрузки на УК априори неизвестны, определение оптимальной в каком-либо смысле последовательности $\{x_n\}$ возможно лишь в процессе нормальной работы УК на основе адаптивного подхода [77]. Это означает, что последовательность $\{x_n\}$ должна формироваться непосредственно в процессе работы УК в соответствии с тем или иным алгоритмом

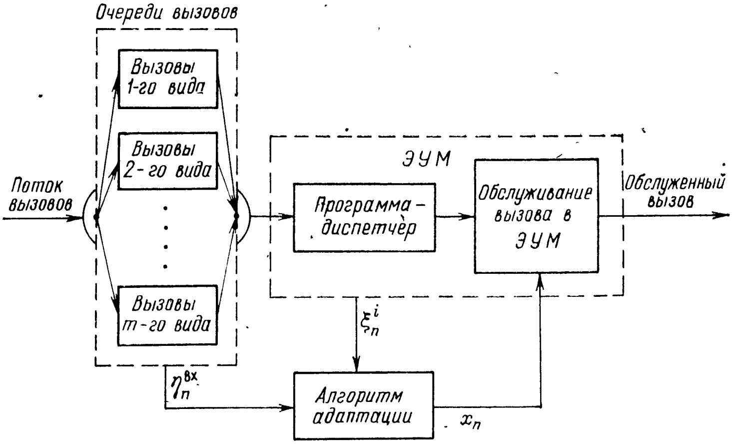

Рис. 4

адаптации (рис. 4). Итак, можно сформулировать следующую задачу: обеспечить формирование такой последовательности вариантов $\{x_n\}$, для которой с вероятностью 1 при каждом $k = \overline{1, l}$ предельные значения последовательности $\{\Phi_n^k\}$ (1.4) минимальны по всем возможным последовательностям $\{x_n^k\}$ при фиксированных остальных $\{x_n^j\}$ $(j \neq k)$. Эта задача является *игровой задачей* адаптивного выбора вариантов.

5. *Двухуровневая система снабжения.* Современные системы снабжения представляют собой сложные разветвленные системы складов, организованные в многоуровневые иерархические структуры. Как правило, пополнение складов нижнего уровня осуществляется из крупного центрального склада (ЦС) более высокого уровня, который обслуживает несколько «низовых» складов (НС), связанных непосредственно с потребителем [109, 128].

Задача организации (управления) разветвленной системы снабжения состоит в нахождении таких стратегий пополнения запасов на НС, при которых обеспечивается требуемое качество функционирования каждого склада.

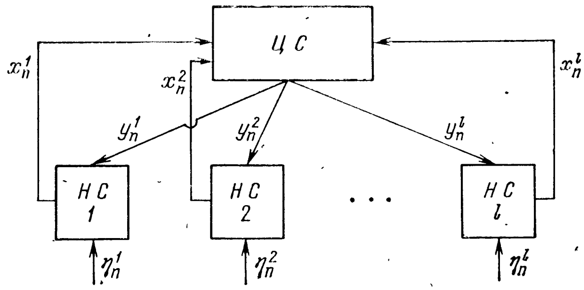

Рис. 5

Рассмотрим подробнее двухуровневую систему снабжения (рис. 5), в которой НС, действуя независимо друг от друга, запрашивают у ЦС необходимое количество товара и получают некоторое его количество (не обязательно равное затребованному) для реализации потребителю в течение времени T (например, $T = 1$ сутки). По истечении этого времени нереализованный товар теряет свою товарную стоимость и удаляется из НС.

Пусть в момент времени t_n ($n = 1, 2, \ldots$), соответствующий окончанию очередного интервала времени T, k-й НС запрашивает у ЦС количество товара в размере x_n^k $(k = \overline{1, l})$ единиц, не имея информации об аналогичных запросах от других НС. ЦС, получив все заявки от НС, отпускает k-му НС $y_n^k \triangleq y^k(x_n^1, \ldots, x_n^l, \omega)$ единиц, где $y^k(x^1, \ldots, x^l, \omega)$ — функция, определяющая стратегию ЦС по распределению товара среди НС. Будем считать, что величина x_n^k принимает значения из конечного множества $X_k \triangleq \{x^k(1), \ldots, x^k(N_k)\}$. Если каждый конкретный потребитель формирует величину своего спроса у НС на основании имеющейся информации о векторе заявок $x_n \triangleq (x_n^1, \ldots, x_n^l)$, то общий спрос η_n^k на товар с k-го НС будет также зависеть от x_n, т. е. $\eta_n^k \triangleq \eta^k(x_n^1, \ldots, x_n^l, \omega)$.

Потери ξ_n^k k-го НС на n-м интервале времени T зависят от разности $y_n^k - \eta_n^k$ между количеством имеющегося товара и величиной спроса на него, т. е. $\xi_n^k \triangleq \xi_n^k(y_n^k - \eta_n^k)$. В условиях дефицита, когда $y_n^k - \eta_n^k < 0$, значение ξ_n^k характеризует потери k-го НС за неудовлетворение спроса потребителей. Если $y_n^k - \eta_n^k > 0$, то ξ_n^k представляет собой потери за нереализованный товар.

Поскольку потери ξ_n^k являются случайными функциями вектора заявок x_n, причем статистические характеристики ξ_n^k априори предполагаются неизвестными, а НС действуют независимо друг от друга, стремясь минимизировать свои средние потери за счет выбора значения x_n^k, то можно сформулировать следующую *игровую задачу* адаптивного выбора вариантов: каждому k-му НС ($k = \overline{1, l}$) указать алгоритм формирования последовательности заявок $\{x_n^k\}$, для которой с вероятностью 1 минимизируются предельные значения соответствующих средних потерь

$$\Phi_n^k \triangleq \frac{1}{n} \sum_{t=1}^{n} \xi_t^k$$

по всем возможным последовательностям $\{x_n^k\}$ при фиксированных остальных $\{x_n^j\}$ $(j \neq k)$.

6. *Управление несколькими динамическими объектами в режиме разделения времени.* Рассмотрим систему, содержащую N динамических объектов и устройство управления, работающее в режиме разделения времени. Подключение к тому или иному объекту устройства управления, содержащего ЦВМ с фиксированным набором управляющих программ, осуществляется коммутатором (рис. 6) в заданные моменты времени $t_1, \ldots, t_n, \ldots$ Состояние z_n^k объекта O_k $(k = \overline{1, N})$ в момент времени t_n зависит от состояния z_{n-1}^k, стационарных случайных воздействий w_t^k, а также от того, осуществлялось ли управление этим объектом на интервале времени (t_{n-1}, t_n). Введем в рассмотрение состояния системы объектов

$$z_n \triangleq (z_n^1, \ldots, z_n^N)$$

и будем предполагать, что последовательность состояний $\{z_n\}$ описывается *управляемой цепью Маркова.* Управлением x_n для этой марковской цепи в момент времени t_n является номер объекта, к которому подключено управляющее устройство на интервале времени (t_n, t_{n+1}). Оно, таким образом, принимает конечное число значений.

Пусть потери $\xi^k(t)$ в момент времени t объекта O_k $(k = \overline{1, N})$ зависят от текущего состояния объекта, причем на интервалах управления (когда к объекту O_k

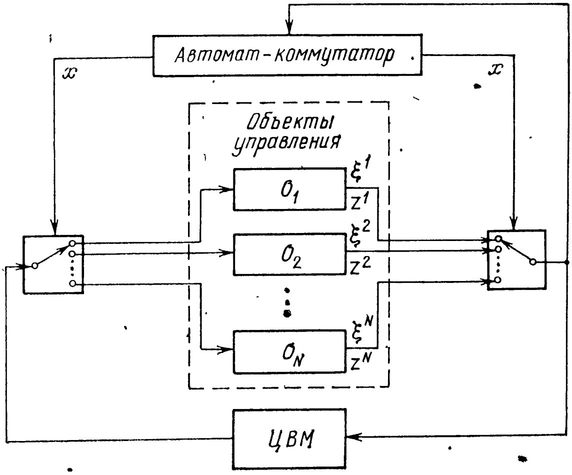

Рис. 6

подключено управляющее устройство) эти потери уменьшаются, а на интервалах отсутствия управления они увеличиваются (рис. 7). Определим потери ξ_n всей системы на интервале времени (t_n, t_{n+1}) как

$$\xi_n \triangleq \sum_{k=1}^{N} \lambda_k \frac{1}{t_{n+1} - t_n} \int_{t_n}^{t_{n+1}} \xi^k(\tau)\, d\tau,$$

где λ_k — заданные весовые коэффициенты. Средние потери системы за n тактов переключения управляющего устройства

$$\Phi_n \triangleq \frac{1}{n} \sum_{t=1}^{n} \xi_t, \quad n = 1, 2, \ldots \tag{1.5}$$

характеризуют качество работы объектов O_k $(k = \overline{1, N})$ до момента времени t_{n+1} и, следовательно, зависят от управлений $x_1, \ldots, x_n$. Теперь можно сформулировать задачу адаптивного управления системой динамических объектов в режиме разделения времени следующим образом: по наблюдениям состояний z_n системы объектов и потерь $\xi^k(t)$, $t_n < t < t_{n+1}$ $(k = \overline{1, N},\ n = 1, 2, \ldots)$ сформировать такую последовательность управлений $\{x_n\}$,

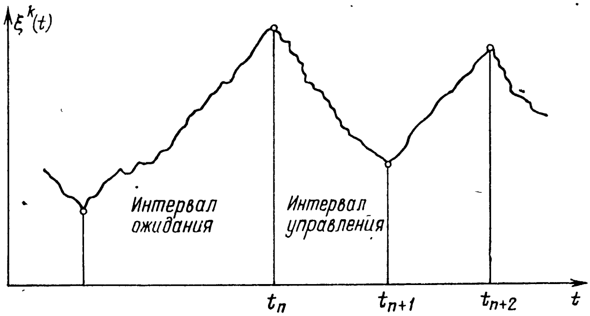

Рис. 7

которая обеспечила бы минимальное предельное значение средних потерь Φ_n (1.5) по всем возможным последовательностям $\{x_n\}$. Эта задача относится к классу задач *управления марковской цепью* в условиях априорной неопределенности.

§ 1.3. Методы адаптивного выбора вариантов (обзор)

Задача адаптивного выбора вариантов является одной из важнейших задач теории адаптивных систем, предмет изучения которой составляют разнообразные *адаптивные алгоритмы*, позволяющие оптимизировать функционирование систем в условиях априорной неопределенности [134, 135]. Суть этих алгоритмов заключается в том, что они указывают, как следует распоряжаться текущей информацией и в результате ее обработки воздействовать на работу системы, изменяя режим, или вариант ее функционирования, с тем, чтобы обеспечить достижение заданной цели. В задачах адаптивного выбора вариантов такой текущей информацией являются реализации потерь, получаемых в результате выбора конкретных вариантов.

Приведенные в § 1.2 примеры показывают, что можно выделить четыре группы задач, различающиеся между собой целями, к достижению которых требуется стремиться в процессе выбора вариантов, а также возможной зависимостью потерь от последовательности реализованных к текущему моменту времени вариантов:

1) задачи безусловной минимизации средних потерь,
2) задачи условной минимизации средних потерь,
3) игровые задачи,
4) задачи управления конечными марковскими цепями. На рис. 8 изображена соответствующая классификация задач адаптивного выбора вариантов.

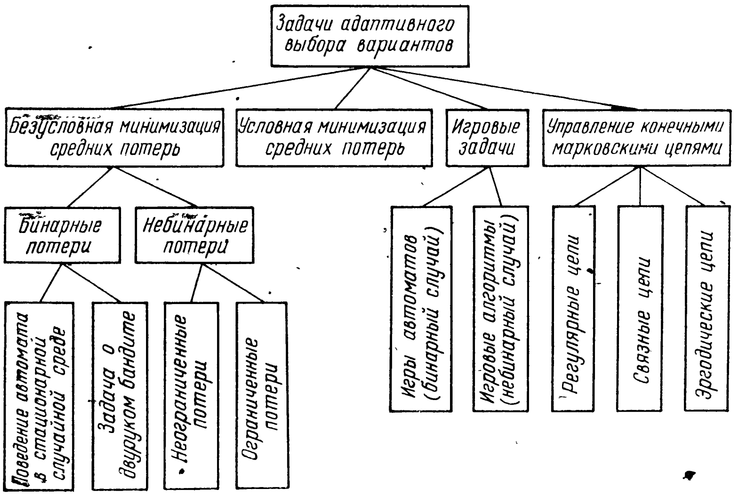

Рис. 8

Приведем краткий обзор существующих методов решения перечисленных задач.

1. *Безусловная минимизация средних потерь.* Задачи этой группы отличаются тем, что качество выбранного варианта характеризуется в них скалярной случайной величиной, а целью является обеспечение минимально возможного среднего ее значения. Здесь различают задачи с *бинарными потерями*, в которых величина, характеризующая качество отдельного акта выбора, принимает только два значения (1, что соответствует «штрафу», и 0, что трактуется как «нештраф»), и задачи с *небинарными потерями*.

Большинство работ в этом направлении посвящено задаче с бинарными потерями, составляющей предмет исследования *теории поведения автоматов* в стационарных случайных средах [14, 117, 133]. Алгоритмы адаптивного выбора вариантов рассматриваются в этой теории как конечные автоматы (вообще говоря, стохастические), имеющие двоичный вход и множество действий, совпадающее с множеством возможных вариантов. Совершая то или иное действие, автомат указывает на выбор соответствующего варианта, определяющего вероятность того, что автомат будет оштрафован средой. В ответ на реакцию среды (0 или 1), автомат производит смену состояния (в соответствии с матрицей переходов), совершая затем новое действие и вызывая очередную реакцию среды и т. д.

Это направление исследований начато Роббинсом [101] и развито Цетлиным [132], предложившим конструкцию так называемого автомата с линейной тактикой и изучившим его поведение в стационарной случайной среде. Другие конструкции автоматов (как детерминированных, так и стохастических) были предложены и исследованы Кринским [46], Пономаревым [94] и Крыловым [47]. Каждый из этих автоматов представляет собой *автомат с постоянной структурой* (т. е. матрица переходов и отображение «состояние — действие» не меняются во времени). Исследование поведения такого автомата в стационарной случайной среде сводится к вычислению и анализу финальных вероятностей состояний соответствующей дискретной цепи Маркова. Перечисленные выше конструкции автоматов обладают свойством ε-оптимальности: за счет выбора достаточно большого числа состояний автомата (глубины памяти), приходящихся на одно действие, можно обеспечить такое предельное значение среднего штрафа, которое будет сколь угодно мало отличаться от минимально возможной величины.

Следующий этап исследования поведения автоматов в случайных средах был начат работами, в которых предлагалось расширить класс изучаемых автоматов, а именно исследовать поведение *стохастических автоматов с переменной структурой*. Такие автоматы при наблюдении реакции среды на совершенное действие первоначально изменяют матрицу переходов, а затем производят смену состояния в соответствии с этой измененной матрицей. Алгоритм изменения матрицы переходов может быть различным.

Предложенный Варшавским и Воронцовой [15, 16] алгоритм был изучен ими путем моделирования на ЦВМ. Полученные при этом результаты позволили сделать следующий вывод: изменением структуры автомата, у которого число состояний равно числу действий, можно добиться оптимального его поведения в стационарной случайной среде.

Сраговичем и Флеровым [119] была предложена конструкция серии стохастических автоматов Мура, которые обеспечивали ε-оптимальный средний выигрыш в стационарной среде за конечное число тактов путем соответствующей перестройки условных вероятностей выбора действий на основе текущих оценок средних выигрышей в каждом из состояний. Эти же автоматы исследовались теми же авторами [120] в игровой ситуации с партнером, применяющим стационарные смешанные стратегии.

В дальнейшем в основном изучались такие автоматы, у которых матрица переходов имеет одинаковые строки, а число состояний равно числу действий (каждому состоянию соответствует «свое» действие). Такие автоматы получили название «автомат-строка». Они полностью определяются стохастическим вектором, компоненты которого суть вероятности совершения автоматом соответствующего действия. При этом вероятность совершения какого-либо действия не зависит от того действия, которое автомат совершил на предыдущем шаге. Алгоритмы изменения этого стохастического вектора в процессе взаимодействия «автомата-строки» со средой, представляющие собой алгоритмы изменения его структуры, называют *автоматными алгоритмами*. Различные автоматные алгоритмы были предложены Варшавским и Воронцовой [15], Шапиро и Нарендрой [143], Лакшмиварахановым и Тхатхачаром [54] и другими [78]. Они были получены из эвристических соображений, а основным инструментом исследования их свойств являлось моделирование на ЦВМ. Цыпкин и Позняк в [137] установили, что все эти алгоритмы можно изучать, исходя из единых соображений, основанных на методе стохастической аппроксимации. Такой подход позволил провести теоретическое исследование сходимости указанных алгоритмов [5, 110], которое показало, что ни один из них не гарантирует с вероятностью 1 оптимального поведения в произвольной среде.

Теория поведения автоматов в случайных средах возникла и долгое время развивалась независимо от более ранних исследований зарубежных авторов, изучавших так называемую *задачу о двуруком бандите* [102, 111, 114], основная разновидность которой по существу представляет собой частный случай задачи адаптивного выбора вариантов с бинарными потерями, когда имеется всего два возможных варианта. Предложенные методы решения этой задачи во многом аналогичны тем, которые были получены при изучении поведения в случайных средах автоматов с постоянной структурой. Так, конструкция автомата Роббинса [101] соответствует автомату Кринского [46], автоматы Исбелла [34] и Смита и Пайка [114] — автомату Пономарева [94], а автомат Самуэлса [111] — автомату Крылова [47]. В последние годы, начиная с работы Кавера и Хеллмана [36], основная проблема, связанная с задачей о двуруком бандите, которая обсуждается в литературе, посвящена определению точной нижней границы предельной вероятности выбора неоптимального варианта, если выбор осуществлять с помощью какого-либо конечного стохастического автомата с фиксированным числом состояний, а также построению таких автоматов, которые бы реализовали эту нижнюю границу или были бы сколь угодно близки к ней [122, 140]. Не останавливаясь более на других работах, посвященных задаче о двуруком бандите, укажем на интересную с нашей точки зрения работу Херкенрата и Теодореску [130], в которой исходная задача сводится к задаче нахождения максимума функции регрессии и для ее решения применяется процедура Кифера — Вольфовица [42].

Перейдем к рассмотрению методов решения задач адаптивного выбора вариантов с *небинарными потерями.* Здесь будем различать два случая: *задачи с ограниченными потерями*, когда величина ξ_n ($n = 1, 2, \ldots$), характеризующая качество выбранного варианта, с вероятностью 1 принимает значения из некоторого известного конечного отрезка $[a, b]$, и *задачи с неограниченными потерями*, в которых либо отрезок $[a, b]$ неизвестен, либо ξ_n с положительной вероятностью может быть сколь угодно большой по абсолютной величине.

Задачи адаптивного выбора вариантов с ограниченными потерями могут решаться методами, разработанными для решения задач с бинарными потерями. Без ограничения общности можно считать, что характеристика качества ξ_n принимает значения из отрезка $[0, 1]$. В [14, 117] предложено преобразовывать величины ξ_n в последовательность нулей и единиц с помощью датчика случайных чисел: с вероятностью ξ_n выбирать 1, а с вероятностью $(1-\xi_n)$ выбирать 0. Если выход такого преобразователя рассматривать вместо потерь ξ_n как штраф (1) или нештраф (0) за осуществленный выбор варианта и стремиться к минимизации среднего штрафа, то мы приходим к задаче с бинарными потерями, которая эквивалентна исходной.

Коноваловым [44] предложена конструкция семейства автоматов, позволяющих (при ограниченных с вероятностью 1 потерях) добиваться ε-оптимальности не только в стационарных, но и в периодически меняющихся средах.

Другой подход к решению задач с ограниченными потерями предложили Лакшмиварахан и Тхатхачар [55]. Они распространили автоматные алгоритмы на случай, когда потери принимают конечное число значений и континуум значений, и показали, что для этих алгоритмов справедливы те же выводы об их сходимости, что и в бинарном случае: с положительной вероятностью задача может быть решена ошибочно. Отметим, что ограниченность потерь существенно используется в этих алгоритмах для обеспечения «стохастичности» вектора вероятностей выбора вариантов при его изменениях.

Для решения задач адаптивного выбора вариантов с неограниченными потерями можно воспользоваться одним из алгоритмов, предложенных Сраговичем [116] и Флеровым [124]. Все эти алгоритмы основаны на вычислении (в процессе перебора возможных вариантов) оценок средних значений потерь, соответствующих разным вариантам, и изменении стратегии выбора в зависимости от этих оценок. Например, в одном из алгоритмов Флерова на каждом шаге (кроме заранее заданных, частота появления которых стремится к нулю) осуществляется выбор того варианта, которому соответствует наименьшая текущая оценка средних потерь (специальным приемом обеспечивается, чтобы ни один из вариантов не мог быть выбран конечное число раз в бесконечной серии актов выбора, что необходимо для обеспечения сходимости всех оценок средних потерь к своим истинным значениям). Для этого алгоритма частота выбора оптимального варианта стремится к единице почти наверное. Остальные из упомянутых алгоритмов Сраговича и Флерова являются автоматными, в которых вероятности выбора вариантов изменяются в зависимости от текущих оценок средних потерь. Сходимость этих оценок к оптимальной или ε-оптимальной стратегии выбора вариантов обеспечивается за счет сходимости оценок средних потерь для каждого варианта к их истинным значениям.

Автоматный алгоритм, предназначенный для решения задач с неограниченными потерями, при работе которого нет необходимости производить оценивание средних потерь, был получен Назиным и Позняком [72] на основе методов стохастической аппроксимации, проекции градиента и регуляризации. Изменение вероятностей выбора вариантов в этом алгоритме производится на каждом шаге в зависимости от потерь, полученных в результате последнего акта выбора. Доказана сходимость с вероятностью 1 этого алгоритма к оптимальной стратегии выбора. Там же получен и исследован двухуровневый иерархический автоматный алгоритм, предназначенный для решения той же задачи. Отличие этого алгоритма от предыдущего состоит в том, что в нем на каждом шаге выбор варианта осуществляется в два этапа: сначала с некоторыми вероятностями производится случайный выбор одной из групп вариантов, на которые разбито множество всех возможных вариантов, а затем среди вариантов, образующих выбранную группу, с некоторыми вероятностями случайно выбирается один вариант. После получения величины потерь, соответствующей выбранному варианту, производится изменение вероятностей выбора группы и вариантов внутри группы. Такой алгоритм позволяет эффективно решать задачи адаптивного выбора вариантов повышенной размерности.

Отметим, что использование регулярных методов, какими являются методы проекции градиента и стохастической аппроксимации, для построения автоматных алгоритмов решения задач адаптивного выбора вариантов позволяет единообразно проводить теоретическое исследование свойств получаемых алгоритмов (нахождение условий сходимости, скорости сходимости, оптимальных параметров алгоритмов и т. д.).

2. *Условная минимизация средних потерь.* В задачах этой группы качество выбранного варианта характеризуется несколькими случайными величинами (показателями), а целью является обеспечение минимально возможного среднего значения одной из них при условии, что средние значения остальных величин не превышают за-

данных уровней. Такая задача адаптивного выбора вариантов впервые рассматривалась в [35]. Позняком [89] был предложен и исследован автоматный алгоритм, основанный на методе штрафных функций, при работе которого необходимо производить оценку средних значений потерь по всем показателям для каждого из возможных вариантов. Там же описан аналогичный двухуровневый иерархический автоматный алгоритм.

Алгоритм (неавтоматного типа), решающий задачу условной минимизации с использованием оценок средних значений потерь, был предложен Сраговичем в работе [118].

Автоматный алгоритм, при работе которого не требуется оценивать средние значения потерь, получен Цыпкиным и Позняком [137]. Он также основан на методе штрафных функций. Каждый шаг этого алгоритма делается по результатам двух актов выбора, производимых при одном и том же значении вектора вероятностей выбора вариантов. Получены условия сходимости алгоритма к оптимальной стратегии выбора и показано, что скорость сходимости имеет порядок $n^{-1/3}$.

Назиным [66] предложено два автоматных алгоритма, основанных на методе множителей Лагранжа (параллельно с изменением вероятностей выбора вариантов производится коррекция множителей Лагранжа). В первом алгоритме на каждом шаге необходимо наблюдать по одной реализации потерь для каждого показателя, в то время как во втором — реализацию потерь лишь по одному из показателей, причем то, по какому именно, определяется случайным образом с вероятностями, зависящими от текущих значений множителей Лагранжа. Это позволяет второму алгоритму «обучаться» существенности ограничений и с течением времени чаще получать текущую информацию о выполнении более существенных ограничений. Таким образом, второй алгоритм является более эффективным в задачах с большой долей несущественных или малосущественных ограничений. Получены условия сходимости с вероятностью 1 обоих алгоритмов к оптимальной стратегии выбора.

3. *Игровые задачи* в наиболее часто встречающейся в литературе постановке рассматриваются как бескоалиционные матричные игры многих лиц, в которых игроки априори не знают матриц выигрышей и вынуждены в процессе многократного разыгрывания партий по наблю-

дениям реализаций выигрышей обучаться равновесию по Нэшу.

Игровые задачи адаптивного выбора вариантов впервые рассматривались в теории коллективного поведения автоматов в случайных средах [14, 133] (задачи с бинарными потерями). Изучались игры автоматов как с постоянной, так и с переменной структурой, причем основное внимание уделялось играм двух автоматов с нулевой суммой. Исследования Цетлина [133], Варшавского [14] и других (основанные на теории цепей Маркова) показали, что в таких играх автоматы с постоянной структурой далеко не всегда приходят к цене игры по фон Нейману. С другой стороны, с помощью моделирования на ЦВМ Варшавский показал, что в играх некоторых автоматов с переменной структурой средний штраф с течением времени приближается к цене игры по фон Найману, хотя стратегии автоматов могут и не сходиться к оптимальным смешанным стратегиям, соответствующим седловой точке игры [14]. Исследование игр других автоматов с переменной структурой, которые при взаимодействии со стационарной случайной средой обладают ε-оптимальным поведением, проводили Чандрасекаран и Шен [141] и Висванатан и Нарендра [17]. Последние, в частности, показали, что для игр с нулевой суммой двух таких автоматов средний штраф в пределе может быть сколь угодно близок к цене игры по фон Нейману, если матрица игры имеет единственную седловую точку.

В [32] Ершов и Позняк исследовали модификацию известного алгоритма Брауна — Робинсон определения седловой точки матричной игры двух лиц с нулевой суммой, использующую текущие оценки элементов матрицы игры.

Теоретические результаты, изложенные в [14], были основаны на так называемой «гипотезе медленности» Волконского, которая, как позднее показал Гурвич [23], неверна. Предложенный [23] метод асимптотического исследования игр автоматов позволил установить, что в игре одинаковых автоматов некоторого описанного класса предельный средний доход наиболее проигравшего автомата достигает максимума.

В [73] Назин и Позняк рассмотрели игру двух стохастических автоматов с нулевой суммой, в которой потери могут принимать любые значения (небинарный случай). Этот автоматный алгоритм основан на методах стохастической аппроксимации, проекции градиента и регуляризации. Получены условия сходимости алгоритма с вероятностью 1 и в среднеквадратическом смысле к седловой точке (в смешанных стратегиях) соответствующей матричной игры двух лиц с нулевой суммой, а также оценка скорости сходимости. В [74] эти результаты обобщаются на игры N автоматов, удовлетворяющие условию диагональной выпуклости. Другие исследования по бескоалиционным играм N автоматов в основном сводятся к моделированию тестовых задач [14, 78, 133].

4. *Управление конечными марковскими цепями.* В условиях полной априорной информации задача управления цепями Маркова впервые была сформулирована в монографии Ховарда [131], где также приведено ее решение на основе метода динамического программирования. Методы решения этой задачи были развиты в работах Дермана [26], Майна и Осаки [62], Бертсекаса [8], Денардо [24] и Росса [108].

При отсутствии априорной информации эта задача рассматривалась в работе Риордана [100] и была названа задачей адаптивного управления марковской цепью. В этой работе было предложено оценивать вероятности переходов марковской цепи и величины средних потерь и использовать их в алгоритмах оптимального управления при полной априорной информации. Мандлом [64] было установлено так называемое «условие идентифицируемости» марковской цепи, выполнение которого гарантирует работоспособность соответствующей адаптивной управляемой марковской цепи. Позднее Боркар, Варайя [9] и Доши и Шрив [27] показали, что эти «условия идентифицируемости» в ряде случаев могут быть ослаблены. Это направление, связанное с оцениванием параметров марковской цепи, получило развитие в работах Любчика и Позняка [61], Эль-Фаттаха [148, 149], Кумара и Беккера [48], Кумара и Лина [49], Коновалова [43] и Сраговича [118]. Гёссель и Срагович [20] рассматривали задачу адаптивного управления марковской цепью в ситуации, когда состояния цепи не наблюдаются, и для ее решения предложили алгоритм, реализующий ε-оптимальную стратегию в классе программных стратегий.

Приведенный здесь обзор охватывает наиболее важные и интересные, по мнению авторов, направления исследований в области адаптивного выбора вариантов, хотя и не претендует на исчерпывающую полноту. Более подробная библиография по этому вопросу содержится в монографиях [14, 53, 96, 97, 117, 118, 133] и обзорах [11, 78, 125, 138].

§ 1.4. Стратегии выбора

В приведенных в § 1.2 примерах практических задач адаптивного выбора вариантов имела место следующая общая ситуация, возникающая при оптимизации систем с дискретным управлением. В каждый из последовательных моментов времени t_n $(n=1, 2, \ldots)$ необходимо выбирать вариант x_n из конечного множества возможных вариантов (управлений) X.

Возникающие в результате произведенного выбора потери системы ξ_n представляют собой случайную величину (функцию элементарного исхода ω) и зависят от x_n и, возможно, состояний системы (случай управляемой марковской цепи, см. пример 6 § 1.2). Реализуемая при этом последовательность вариантов (управлений) $\{x_n\}$ должна быть такой, чтобы достигалась заданная цель, формулируемая в терминах предельных значений текущих средних потерь.

Наличие априорной неопределенности, состоящей в отсутствии точной информации о потерях системы и ее характеристиках, приводит к тому, что формирование последовательности вариантов $\{x_n\}$, обеспечивающей достижение целевого условия решаемой задачи, следует, как отмечалось в § 1.1, осуществлять в соответствии с адаптивным подходом. При этом выбор очередного варианта x_{n+1} производится на основе полученной к данному моменту времени последовательности потерь $\xi_1, \ldots, \xi_n$, соответствующей реализованной последовательности вариантов $x_1, \ldots, x_n$. Это значит, что x_{n+1} является функцией от $x_1, \ldots, x_n, \xi_1, \ldots, \xi_n$ и, возможно, от момента времени n и элементарного исхода ω, т. е.

$$x_{n+1}=T_n(x_1, \ldots, x_n;\ \xi_1, \ldots, \xi_n;\ \omega), \quad n=1, 2, \ldots,$$

где ξ_n, в зависимости от задачи, либо скаляр, либо вектор. Функцию T_n будем называть *правилом выбора* варианта x_{n+1}. Эта функция может быть как детерминированной, так и случайной (рандомизированной). Последовательность $\{T_n\}$ правил выбора определяет *стратегию выбора вариантов* (стратегию управления).

Детерминированные стратегии выбора составляют предмет изучения теории поведения детерминированных автоматов в случайных средах [14, 117, 133].

Эти стратегии допускают возможность простой реализации с помощью детерминированных конечных автоматов. Они в основном ориентированы на задачи с бинарными потерями, хотя, как отмечалось выше, могут применяться и в других случаях. Кроме того, для них характерно обеспечение приближенно оптимального поведения, близость которого к оптимальному возрастает с увеличением глубины памяти автомата, что влечет за собой, однако, уменьшение скорости достижения решения и увеличивает сложность (число состояний) соответствующего автомата. Это же свойственно и стохастическим автоматам с постоянной структурой [14, 47, 133], которые реализуют рандомизированные стратегии выбора.

Отметим также, что в условиях полной информации оптимальная стратегия всегда принадлежит классу детерминированных стратегий и, более того, программных (когда $x_{n+1} = T_n(\omega)$, $n = 1, 2, \ldots$) [30].

Наличие априорной неопределенности приводит к необходимости использовать более сложные *рандомизированные стратегии*. В теории поведения автоматов таким стратегиям соответствуют стохастические автоматы с переменной структурой [14—16, 78, 138]. Большинство из них реализуют рандомизированные правила выбора следующего вида:

$$p_{n+1} = R_n(x_1, \ldots, x_n;\ p_1, \ldots, p_n;\ \xi_1, \ldots, \xi_n), \quad n = 1, 2, \ldots, \tag{1.6}$$

где R_n — вектор-функция со значениями в симплексе S^N, p_n — вектор условных вероятностей выбора вариантов $x(1), \ldots, x(N)$ в момент времени t_n. Выбору очередного варианта x_{n+1} предшествует вычисление в соответствии с (1.6) вектора p_{n+1}. Вариант x_{n+1} представляет собой дискретную случайную величину, принимающую значения $x(1), \ldots, x(N)$ с условными вероятностями $p_{n+1}(1), \ldots \ldots, p_{n+1}(N)$ при фиксированной предыстории $(x_1, \ldots, x_n;$ $\xi_1, \ldots, \xi_n)$. Такую случайную величину можно реализовать с помощью датчика случайных чисел, равномерно распределенных на отрезке $[0, 1]$, используя *метод деле-*

ния отрезка: если η_n — выход датчика в момент времени t_n, то $x_n \triangleq x(i)$, где номер i определяется условием

$$\sum_{j=1}^{i-1} p_n(j) \leqslant \eta_n < \sum_{j=1}^{i} p_n(j), \quad n = 1, 2, \ldots$$

Такой способ формирования последовательности вариантов и составляет существо рандомизационного подхода.

Можно отметить следующие особенности рандомизированных стратегий:

1) исходные задачи адаптивного выбора вариантов, как будет показано далее, эквивалентны оптимизационным задачам с непрерывными переменными — компонентами $p_n(1), \ldots, p_n(N)$ рандомизированных правил типа (1.6), что дает возможность применить хорошо разработанный аппарат стохастической аппроксимации для построения и исследования алгоритмов адаптивного выбора вариантов;

2) применение таких стратегий позволяет единообразно формировать алгоритм адаптивного выбора вариантов при решении всех рассматриваемых задач;

3) использование рандомизированных стратегий позволяет решать широкий класс задач адаптивного выбора вариантов, включая задачи с небинарными и даже неограниченными потерями ξ_n;

4) эти стратегии позволяют в асимптотике (по времени) реализовать оптимальный режим адаптивного выбора.

Отметим, что среди рандомизированных правил выбора (1.6) содержатся так называемые *марковские правила*:

$$p_{n+1} = Q_n(x_n, p_n, \xi_n), \quad n = 1, 2, \ldots \tag{1.7}$$

Рандомизированные стратегии, определяемые последовательностью правил вида (1.7), будем далее называть *рекуррентными алгоритмами адаптивного выбора вариантов*. Эти алгоритмы достаточно просто реализуются, поскольку они на каждом шаге n используют минимальную информацию о предыстории процесса. Исследованию алгоритмов вида (1.7), решающих различные задачи адаптивного выбора вариантов, и посвящены последующие главы.

АЛГОРИТМЫ БЕЗУСЛОВНОЙ МИНИМИЗАЦИИ

Дается постановка задачи безусловной минимизации средних потерь. Проводится исследование сходимости и скорости сходимости алгоритмов Нарендры — Шапиро, Льюса, Варшавского — Воронцовой и Буша — Мостеллера, предназначенных для решения задач с бинарными потерями. Для задач с небинарными потерями предлагается и исследуется ряд алгоритмов, реализующих методы стохастической аппроксимации и случайного поиска.

§ 2.1. Задача безусловной минимизации средних потерь

Пусть в дискретные моменты времени t_n $(n=1, 2, \ldots;$ $t_{n+1} > t_n)$ в соответствии с некоторой стратегией производится выбор одного из N возможных вариантов $x(1), \ldots$ $\ldots, x(N)$ работы некоторой системы на очередном временном интервале (t_n, t_{n+1}). Множество $X=\{x(1), \ldots$ $\ldots, x(N)\}$ — *множество возможных вариантов*. Если в момент времени t_n выбран вариант $x_n \in X$, то до наступления момента времени t_{n+1} наблюдается реализация случайной величины $\xi_n \triangleq \xi^n(x_n, \omega)$, имеющая смысл потерь системы на интервале времени (t_n, t_{n+1}). Здесь $\xi_n(x, \omega)$ при любых фиксированных $n=1, 2, \ldots$ и $x \in X$ есть случайная величина *), принимающая значения из множества $\Xi \subseteq R^1$. Величина текущих средних потерь

$$\Phi_n \triangleq \frac{1}{n} \sum_{t=1}^{n} \xi_t, \quad n = 1, 2, \ldots \tag{2.1}$$

*) Все рассматриваемые случайные процессы и величины будем считать заданными на некотором полном вероятностном пространстве $(\Omega, \mathcal{F}, \mathsf{P})$, где $(\Omega, \mathcal{F})$ — измеримое пространство, состоящее из множества Ω элементарных событий ω и определенной на нем σ-алгебры $\mathcal{F}$, а P — вероятностная мера на множествах из $\mathcal{F}$. Зависимость случайных величин от ω часто для удобства опускаем.

характеризует качество последовательности вариантов $\{x_1, \ldots, x_n\}$, выбранных на первых n тактах. Сделаем предположения относительно свойств последовательностей $\{x_n\}$ и $\{\xi_n\}$, необходимых для формулировки цели адаптации:

П1. Для всех $n = 1, 2, \ldots$ совокупности $\{\xi_n(x, \omega) \mid x \in \in X\}$ и $\{\xi_t(x, \omega), x_k \mid x \in X, t = \overline{1, n-1}, k = \overline{1, n}\}$ независимы.

П2. При любых $x(i)$ $(i = \overline{1, N})$

$$\mathsf{M}\{\xi_n(x(i), \omega)\} \triangleq v_i, \quad \sup_n \mathsf{M}\{(\xi_n(x(i), \omega) - v_i)^2\} \triangleq \sigma_i^2 < \infty$$

для всех $n = 1, 2, \ldots$, где $\mathsf{M}\{\cdot\}$ — символ математического ожидания.

Обозначим текущую частоту выбора варианта $x(i)$ через

$$f_n(i) \triangleq \frac{1}{n} \sum_{t=1}^{n} \chi(x_t = x(i)) \qquad (i = \overline{1, N}), \tag{2.2}$$

где $\chi(\cdot)$ — характеристическая функция, т. е.

$$\chi(A) = \begin{cases} 1, & \text{если событие } A \text{ выполняется,} \\ 0, & \text{если событие } A \text{ не выполняется.} \end{cases}$$

Лемма 2.1. *Пусть выполнены предположения* П1, П2. *Тогда*

1) *для того чтобы случайная величина* Φ *была частичным пределом последовательности* $\{\Phi_n\}$ *для почти всех* ω, *необходимо и достаточно, чтобы среди предельных точек последовательности векторов* $f_n \triangleq (f_n(1), \ldots, f_n(N))$ *нашелся, вообще говоря, случайный вектор* $p \triangleq (p(1), \ldots$

$$\ldots, p(N)) \in S^N \triangleq \left\{ p \in R^N \mid p(i) \geqslant 0 \ (i = \overline{1, N}), \ \sum_{i=1}^{N} p(i) = 1 \right\},$$

для которого

$$\Phi = \sum_{i=1}^{N} p(i) v_i \triangleq V(p); \tag{2.3}$$

2) *все предельные точки последовательности* $\{\Phi_n\}$ *заключены в отрезке* $[v_-, v_+]$ *с вероятностью* 1, *где*

$$v_- \triangleq \min_{i=\overline{1,N}} v_i, \quad v_+ \triangleq \max_{i=\overline{1,N}} v_i.$$

Доказательство. Представим Φ_n (2.1) в виде

$$\Phi_n = \sum_{i=1}^{N} f_n(i)\,\bar{\xi}_n(x(i), \omega),$$

где

$$\bar{\xi}_n(x(i), \omega) \triangleq$$

$$\triangleq \begin{cases} \dfrac{\sum\limits_{t=1}^{n} \chi(x_t = x(i))\, \xi_t(x(i), \omega)}{\sum\limits_{t=1}^{n} \chi(x_t = x(i))}, & \text{если } \sum\limits_{t=1}^{n} \chi(x_t = x(i)) > 0, \\ 0, & \text{если } \sum\limits_{t=1}^{n} \chi(x_t = x(i)) = 0 \end{cases}$$

— текущие средние потери для варианта $x(i)$ $(i = \overline{1, N})$. В силу леммы П.12 для почти всех

$$\omega \in B_i \triangleq \left\{\omega \,\middle|\, \sum_{t=1}^{\infty} \chi(x_t = x(i)) = \infty\right\}$$

$$\lim_{n\to\infty} \bar{\xi}_n(x(i), \omega) = v_i.$$

Для почти всех $\omega \bar{\in} B_i$ значение $\sup\limits_n |\bar{\xi}_n(x(i), \omega)| < \infty$ и $\lim\limits_{n\to\infty} f_n(i) = 0$. Вектор f_n (2.2) — стохастический, т. е. $f_n \in S^N$, а значит, все предельные точки последовательности $\{f_n\}$ также лежат в симплексе S^N. Отсюда следует, что любой частичный предел Φ последовательности $\{\Phi_n\}$ с вероятностью 1 представим в виде

$$\Phi = \sum_{i=1}^{N} p(i)\, v_i, \quad p \in S^N,$$

где p — частичный предел последовательности $\{f_n\}$. Из того, что

$$v_- = \min_{p \in S^N} V(p), \quad v_+ = \max_{p \in S^N} V(p),$$

следует требуемое утверждение. Лемма доказана. ▲

Из леммы 2.1 вытекает, что для любой стратегии выбора вариантов x_n предельные значения текущих средних потерь Φ_n (любой частичный предел) не могут быть меньше величины v_- при сделанных предположениях.

Ясно, что существуют последовательности $\{x_n\}$, для которых $\lim_{n\to\infty} \Phi_n = v_-$ с вероятностью 1. Такими последовательностями являются, например, последовательности вида

$$x_n = x(i),\quad i \in I_{\min} \triangleq \{i \mid v_i = v_-\},\quad n \geqslant n_0,$$

где n_0 — произвольное натуральное число. Поскольку множество $I_{\min}$ априори неизвестно, а точные значения условных средних потерь v_i можно определить лишь путем осуществления бесконечного числа актов выбора каждого варианта, то естественно сформулировать цель адаптации в терминах предельных значений текущих средних потерь Φ_n: для почти всех ω минимизировать предельные значения $\{\Phi_n\}$, т. е. обеспечить выполнение условия

$$\overline{\lim_{n\to\infty}} \Phi_n \overset{\text{п.н.}}{=} v_-. \tag{2.4}$$

Отметим, что в силу леммы 2.1 цель (2.4) эквивалентна условию

$$\lim_{n\to\infty} \Phi_n \overset{\text{п.н.}}{=} v_-.$$

Учитывая сказанное, сформулируем *задачу безусловной минимизации средних потерь* следующим образом:

> найти последовательность правил (стратегию), позволяющую в каждый момент времени t_n на основе имеющейся совокупности данных (x_1, ξ_1), (x_2, ξ_2), ..., (x_{n-1}, ξ_{n-1}) осуществить выбор очередного варианта x_n так, чтобы сформированная таким образом последовательность $\{x_n\}$ обеспечивала достижение цели (2.4).

Перейдем теперь к вопросу о разрешимости этой задачи. Каждой стратегии выбора вариантов можно поставить в соответствие последовательности условных вероятностей

$$p_n(i) \triangleq \mathsf{P}\{x_n = x(i) \mid x_1, \ldots, x_{n-1};\ \xi_1, \ldots, \xi_{n-1}\},\quad i = \overline{1, N}, \tag{2.5}$$

представляющих собой измеримые функции наблюдений $x_1, \ldots, x_{n-1}, \xi_1, \ldots, \xi_{n-1}$. Наоборот, совокупность последовательностей таких функций определяет некоторую стратегию. Поэтому далее, говоря о стратегии выбора вариантов, или о последовательности правил выбора вариантов, мы будем подразумевать последовательность $\{p_n\}$ векторов $(p_n(1), \ldots, p_n(N)) \triangleq p_n$ условных вероятностей (2.5). Возникает вопрос: какой должна быть последовательность $\{p_n\}$ для того, чтобы выполнялось целевое условие (2.4)? На него отвечает следующая лемма.

Лемма 2.2. *Пусть выполнены условия* П1, П2 *и существует* $\tau \in (0, 1)$, *для которого*

$$\overline{\lim_{n\to\infty}} n^{\tau} \mathsf{M}\{[V(p_n) - v_-]^2\} \triangleq C < \infty, \tag{2.6}$$

где

$$V(p) \triangleq \sum_{i=1}^{N} v_i p_i,$$

$$p \in S^N \triangleq \left\{p = (p_1, \ldots, p_N) \,\middle|\, \sum_{i=1}^{N} p_i = 1,\ p_i \geqslant 0\ (i = \overline{1, N})\right\}.$$

Тогда с вероятностью 1 *выполняется целевое условие* (2.4), *причем*

$$\overline{\lim_{n\to\infty}} n^{\tau} \mathsf{M}\{(\Phi_n - v_-)^2\} \leqslant C\left(1 - \frac{\tau}{2}\right)^{-2}, \tag{2.7}$$

т. е. последовательность текущих средних потерь $\{\Phi_n\}$ *сходится к* v_- *также и в среднеквадратическом смысле со скоростью, не меньшей* $n^{-\tau}$.

Доказательство. Имеем из (2.1)

$$\Phi_n = \left(1 - \frac{1}{n}\right)\Phi_{n-1} + \frac{1}{n}\xi_n, \quad n = 1, 2, \ldots$$

Введем σ-алгебру $\mathscr{F}_n \triangleq \sigma(x_1, \ldots, x_n; \xi_1, \ldots, \xi_n)$ и $d_n \triangleq (\Phi_n - v_-)^2$. Тогда

$$d_n = \left(1 - \frac{1}{n}\right)^2 d_{n-1} + \frac{2}{n}\left(1 - \frac{1}{n}\right)(\Phi_{n-1} - v_-)(\xi_n - v_-) + \frac{1}{n^2}(\xi_n - v_-)^2$$

и

$$\mathsf{M}\{d_n|\mathscr{F}_{n-1}\} \leqslant \left(1 - \frac{1}{n}\right)^2 d_{n-1} + \frac{2}{n}\left(1 - \frac{1}{n}\right)(\Phi_{n-1} - v_-)(\mathsf{M}\{\xi_n|\mathscr{F}_{n-1}\} - v_-) + \frac{2}{n^2}(\mathsf{M}\{\xi_n^2|\mathscr{F}_{n-1}\} + v_-^2).$$

В силу свойств условного математического ожидания и условия П1 имеем

$$\mathsf{M}\{\xi_n \mid \mathcal{F}_{n-1}\} = \sum_{i=1}^{N} \mathsf{M}\{\chi(x_n = x(i))\, \xi_n(x(i), \omega) \mid \mathcal{F}_{n-1}\} =$$

$$= \sum_{i=1}^{N} \mathsf{M}\{\chi(x_n = x(i))\, \mathsf{M}\{\xi_n(x(i), \omega) \mid \mathcal{F}_{n-1}, x_n\} \mid \mathcal{F}_{n-1}\} =$$

$$= \sum_{i=1}^{N} v_i \mathsf{M}\{\chi(x_n = x(i)) \mid \mathcal{F}_{n-1}\} = V(p_n)$$

и, аналогично, из условия П2 $\mathsf{M}\{\xi_n^2 \mid \mathcal{F}_{n-1}\} \leqslant K < \infty$. Учитывая это и используя первый раз неравенство $2ab \leqslant a^2 + b^2$ и второй раз неравенство Коши — Буняковского, получим, соответственно, соотношения

$$\mathsf{M}\{d_n \mid \mathcal{F}_{n-1}\} \leqslant \left(1 - \frac{1}{n}\right)^2 d_{n-1} +$$

$$+ \frac{1}{n}\left(1 - \frac{1}{n}\right)(d_{n-1} + (V(p_n) - v_-)^2) + \frac{2}{n^2}(K + v_-^2) \leqslant$$

$$\leqslant \left(1 - \frac{1}{n}\right) d_{n-1} + \frac{1}{n}(V(p_n) - v_-)^2 + \frac{2}{n^2}(K + v_-^2)$$

и

$$\mathsf{M}\{d_n\} \leqslant \left(1 - \frac{1}{n}\right)^2 \mathsf{M}\{d_{n-1}\} + \frac{1}{n^2}\mathsf{M}\{(\xi_n - v_-)^2\} +$$

$$+ \frac{2}{n}\left(1 - \frac{1}{n}\right)[\mathsf{M}\{d_{n-1}\}\, \mathsf{M}\{(V(p_n) - v_-)^2\}]^{1/2}.$$

Из первого неравенства в силу (2.6) и леммы П.11 вытекает $d_n \to 0$ с вероятностью 1 при $n \to \infty$. Из второго неравенства условий (2.6), П2 и леммы П.2 Приложения 1 (случай а)) получаем (2.7). Лемма доказана. ▲

Эта лемма показывает, что к анализу и синтезу алгоритмов адаптивного выбора вариантов, предназначенных для решения сформулированной задачи, можно подходить с точки зрения минимизации функции средних потерь $V(p)$ на симплексе S^N (поскольку $v_- = \min\limits_{p \in S^N} V(p)$).

Далее, в основном будем считать, что

$$v_- \triangleq v_\alpha < v^- \triangleq \min_{i \neq \alpha} v_i, \tag{2.8}$$

т. е. *оптимальный вариант* $x(\alpha)$ с минимальными условными средними потерями $v_- = v_\alpha$ единствен. Это предположение позволяет упростить исследование рассматриваемых ниже алгоритмов. В § 2.6 для алгоритма типа стохастической аппроксимации будут получены результаты как в предположении (2.8), так и без него.

§ 2.2. Рекуррентные алгоритмы адаптивного выбора вариантов

В табл. 1 приведен ряд алгоритмов адаптивного выбора вариантов вида (1.7), предназначенных для решения задачи безусловной минимизации. Как правило, они ориентированы на задачи с бинарными потерями $\xi_n \in \{0, 1\}$ (алгоритмы № 1—5). Бинарность потерь используется в них для обеспечения выполнения условия нормировки ($p_n \in S^N$) при каждом $n = 1, 2, \ldots$ Отметим, что нормировка в этих алгоритмах сохраняется и для потерь вида $\xi_n \in [0, 1]$.

В алгоритмах № 6—9 потери ξ_n могут принимать любые значения. Нормировка в этих алгоритмах обеспечивается за счет введения оператора проектирования π_ε^N на ε-симплекс:

$$S_\varepsilon^N \triangleq \{p \mid p \in S^N, p(i) \geqslant \varepsilon\,(i = \overline{1, N})\}, \quad \varepsilon \in [0, 1/N). \tag{2.9}$$

Этот оператор определяется следующими условиями:

$$\| p - \pi_\varepsilon^N \{q\} \| \leqslant \| p - q \|, \quad \pi_\varepsilon^N \{q\} \in S_\varepsilon^N \tag{2.10}$$

для любых $p \in S_\varepsilon^N$ и $q \in R^N$. В частности, условиям (2.10) удовлетворяет оператор π_ε^N, переводящий заданную точку $q \in R^N$ в ближайшую к ней точку ε-симплекса S_ε^N [90]. Простой способ реализации такого оператора (требующий конечного числа вычислительных операций) и соответствующая ему подпрограмма для ЭВМ на языке фортран-4 приведены в Приложении 2. Известны и другие варианты оператора проектирования, удовлетворяющего условиям (2.10) [45].

Алгоритмы адаптивного выбора вариантов можно подразделить на *беспроекционные алгоритмы* вида

$$p_{n+1} = p_n - \gamma_n R(x_n, p_n, \xi_n) \tag{2.11}$$

и *проекционные алгоритмы*

$$p_{n+1} = \pi_{\varepsilon_{n+1}}^N \{p_n - \gamma_n R(x_n, p_n, \xi_n)\}, \tag{2.12}$$

Таблица 1

Номер	Алгоритм	Авторы
1	$p_{n+1} = p_n + \gamma_n (1 - \xi_n) [e(x_n) - p_n]$, $\xi_n \in \{0, 1\}, \quad \gamma_n \in (0, 1)$	И. Шапиро, К. Нарендра [143] (при $\gamma_n = \mathrm{const}$)
2	$p_{n+1} = p_n + \gamma p_n^{\mathrm{T}} e(x_n) (e(x_n) - p_n) \times$ $\times \left[\frac{1 - \xi_n}{a + (1-a) p_n^{\mathrm{T}} e(x_n)} - \frac{\xi_n}{1 - (1-a) p_n^{\mathrm{T}} e(x_n)} \right]$, $\xi_n \in \{0, 1\}, \quad \gamma \in (0, 1), \quad a \in (0, 1]$.	Р. Д. Льюс [59]
3	$p_{n+1} = p_n + \gamma p_n^{\mathrm{T}} e(x_n) (e(x_n) - p_n)(1 - 2\xi_n)$, $\xi_n \in \{0, 1\}, \quad \gamma \in (0, 1)$	В. И. Варшавский, И. П. Воронцова [15]
4	$p_{n+1} = p_n + \gamma (e(x_n) - p_n) [\lambda(p_n)(1 - \xi_n) - \mu(p_n) \xi_n]$, $\xi_n \in \{0, 1\}, \quad \gamma \in (0, 1)$, $\lambda(p) \triangleq c_\lambda \prod_{i=1}^{N} [p(i)]^{\lambda_i}, \quad c_\lambda \in [0, 1]$, $\mu(p) \triangleq c_\mu \prod_{i=1}^{N} [p(i)]^{\mu_i}, \quad c_\mu \in [0, 1]$, $\lambda_i, \mu_i \geqslant 0, \quad i = \overline{1, N}$	С. Лакшмиварахан, М. Тхатхачар [54]

Продолжение табл. 1

Номер	Алгоритм	Авторы
5	$p_{n+1}=p_n+\gamma_n\left(e(x_n)-p_n+\frac{e^N-Ne(x_n)}{N-1}\xi_n\right)$, $\xi_n\in\{0,1\}$, $\gamma_n\in(0,1)$	Р. Буш, Ф. Мостеллер [12] (при $\gamma_n=\gamma$)
6	$\tilde{p}_{n+1}=\pi^N_{\varepsilon_{n+1}}\{\tilde{p}_n-\gamma_n\xi_n q_n\}$, $p_{n+1}=\tilde{p}_n+\frac{\varepsilon_n}{2}(I-N^{-1}E)\,q_n$, $\gamma_n\geqslant 0$, $\varepsilon_n\in(0,N^{-1})$, $q_n\triangleq(q_{1n},\dots,q_{Nn})$, $q_{in}\in[-1,1]$, $\mathbf{M}\{q_{in}\}=0\ (i=\overline{1,N})$, $\mathbf{M}\{q_nq_n^{\mathrm{T}}\}=I\sigma_q^2>0$	А. В. Назин, А. С. Позняк
7	$p_{n+1}=\pi^N_{\varepsilon_{n+1}}\left\{p_n-\gamma_n\frac{\xi_n-\Delta}{e^{\mathrm{T}}(x_n)\,p_n}\,e(x_n)\right\}$, $\gamma_n\geqslant 0$, $\varepsilon_n\in(0,N^{-1})$	А. С. Позняк [89] (при $\Delta=0$), А. В. Назин [69]

Продолжение табл. 1

Номер	Алгоритм	Авторы
8	$p_{n+1} = \pi^{N_1}_{\varepsilon'_{n+1}} \left\{ p_n - \gamma'_n \frac{\xi_n - \Delta}{p_n^{\mathrm{T}} e^{N_1}(x_n^1)} e^{N_1}(x_n^1) \right\},$ $p_{n+1}^i = \pi^{N_2}_{\varepsilon''_{n+1}} \left\{ p_n^i - \gamma''_n \frac{\xi_n - \Delta_i}{p_n^{i\mathrm{T}} e^{N_2}(x_n^2)} \chi(x_n^1 = x^1(i))\, e^{N_2}(x_n^2) \right\},$ $\gamma'_n, \gamma''_n \geqslant 0, \ \varepsilon'_n \in (0, N_1^{-1}), \quad \varepsilon''_n \in (0, N_2^{-1}).$	А. В. Назин [70] (при $\Delta = 0$)
9	$p_{n+1} = \pi^{N_1}_{\varepsilon'_{n+1}} \left\{ p_n - \gamma'_n \frac{\xi_n - \Delta}{p_n^{\mathrm{T}} e^{N_1}(x_n^1)} e^{N_1}(x_n^1) \right\},$ $p_{n+1}^i = \pi^{N_2}_{\varepsilon''_{n+1}} \left\{ p_n^i - \gamma''_n \frac{(\xi_n - \Delta_i)\, \chi(x_n^1 = x^1(i))}{[p_n^{\mathrm{T}} e^{N_1}(x_n^1)][p_n^{i\mathrm{T}} e^{N_2}(x_n^2)]} e^{N_2}(x_n^2) \right\},$ $\gamma'_n, \gamma''_n \geqslant 0, \quad \varepsilon'_n \in (0, N_1^{-1}), \quad \varepsilon''_n \in (0, N_2^{-1})$	А. В. Назин, А. С. Позняк [72] (при $\Delta = \Delta_i = 0$)

где $R(x_n, p_n, \xi_n) \triangleq R_n$ — вектор движения алгоритма, γ_n — скалярный множитель, называемый далее *длиной шага алгоритма*, $n = 1, 2, \ldots$ — номер шага, $\varepsilon_n \in [0, 1/N]$ — параметр проектора π_ε^N на n-м шаге. Относительное усложнение проекционных алгоритмов (за счет наличия оператора проектирования π_ε^N) по сравнению с беспроекционными компенсируется возможностью их использования для решения более широкого класса задач (как с бинарными, так и с небинарными потерями ξ_n).

Для обеспечения работоспособности алгоритмов (2.11), (2.12) необходимо ограничить свойства векторов движения R_n. Будем рассматривать класс так называемых *псевдоградиентов* относительно *функции Ляпунова* $W(p)$, $p \in S^N$, обладающей свойствами:

1) $W(p) > 0$ для любых $p \in S^N$, $p \neq p^0$ и $W(p^0) = 0$, где $p^0 \in S^N$ — некоторый фиксированный вектор;

2) $W(p)$ дифференцируема.

Класс таких векторов движения R_n определяется *условием псевдоградиентности* [93]

$$\rho_n(p) \triangleq \langle \mathsf{M}\{R(x_n, p_n, \xi_n) \mid p_n = p\}, \nabla W(p)\rangle \geqslant 0 \quad (2.13)$$

для всех $p \in S^N$, которое означает, что на каждом шаге n вектор движения R_n составляет в среднем острый угол с градиентом функции Ляпунова (за исключением точек $p \in S^N$, в которых $\rho_n(p) = 0$). В (2.13) $\langle \cdot, \cdot \rangle$ — скалярное произведение, соответствующее евклидовой норме.

В качестве $W(p)$ для каждого конкретного алгоритма, как правило, приходится брать свою функцию Ляпунова, при этом точкой p^0 должна быть предполагаемая предельная точка последовательности $\{p_n\}$. Так, например, для беспроекционных алгоритмов № 1 — 5 табл. 1 удобно выбрать

$$W(p) = 1 - p(\alpha), \quad p^0 = e_\alpha^N \triangleq p^*,$$

где α — номер оптимального варианта

$$e_\alpha^N = (\underbrace{0, \ldots, 0, 1}_{\alpha}, \underbrace{0, \ldots, 0}_{N-\alpha}).$$

Для проекционного алгоритма № 7, учитывая наличие оператора π_ε^N, приходится брать функцию Ляпунова вида

$$W(p) = \|p - p^0\|^2, \; p^0 = p_\varepsilon^N = (1 - \varepsilon N)\, p^* + \varepsilon e^N,$$
$$e^N \triangleq (1, \ldots, 1).$$

В табл. 2 приводятся функции Ляпунова, соответствующие левой части выражения (2.13), а также ограничения на величины условных средних потерь v_i, обеспечивающие псевдоградиентность рассматриваемых алгоритмов. Из этой таблицы видно, что для алгоритмов № 1 — 5 функция $\rho_n(p)$ не зависит от n, т. е. $\rho_n(p) \triangleq \rho(p)$, $n = 1, 2, \ldots$ и $\rho(p) = 0$ в любой вершине симплекса S^N. Независимость $\rho_n(p)$ от n объясняется линейностью по ξ_n векторов движения $R(x_n, p_n, \xi_n)$ в этих алгоритмах и условиями П1, П2 из § 2.1.

Наличие общих свойств алгоритмов № 1—5 из табл. 1 (общая функция Ляпунова и указанные свойства $\rho(p)$) позволяет доказать для них следующее утверждение.

Теорема 2.1. *Пусть выполнены предположения* П1, П2 *из* § 2.1 *и условие* (2.8), *а p_n — вектор условных вероятностей выбора вариантов $x_n \in X$ на n-м шаге* (2.5), *формируемый по алгоритму* (2.11). *Пусть, кроме того,*

1) *вектор движения R_n удовлетворяет условию псевдоградиентности* (2.13) *относительно функции Ляпунова*

$$W(p) = 1 - p(\alpha);$$

2) *функция $\rho_n(p)$ не зависит от времени, т. е. $\rho_n(p) = \rho(p)$, непрерывна и обращается в нуль только на множестве S^* вершин симплекса S^N;*

3) $\sum\limits_{n=1}^{\infty} \gamma_n = \infty$, $\gamma_n \geqslant 0$ *для всех* $n = 1, 2, \ldots$

Тогда при любом начальном векторе $p_1 \in S^N$ с положительными компонентами $p_1(j) > 0$ $(j = \overline{1, N})$ последовательность $\{p_n(\alpha)\}$ сходится с вероятностью 1 *к случайной величине $\tilde{p}(\alpha)$, принимающей значения* 0 *или* 1. *Если, кроме того,*

$$\varliminf_{n\to\infty} \gamma_n > 0,$$

то последовательность векторов $\{p_n\}$ сходится к случайному вектору $\tilde{p}$, распределенному на множестве S^*.

Доказательство. Обозначим $\eta_n \triangleq W(p_n)$, $\mathcal{F}_n \triangleq \sigma(x_t, \xi_t \mid t = \overline{1, n-1})$. Тогда в силу алгоритма (2.11), учитывая (2.13), получим при любом $n = 1, 2, \ldots$

$$\mathsf{M}\{\eta_{n+1} \mid \mathcal{F}_n\} = \eta_n - \gamma_n \rho(p_n) \leqslant \eta_n.$$

Отсюда

$$\sum_{n=1}^{\infty} \gamma_n \mathsf{M}\{\rho(p_n)\} < \infty \tag{2.14}$$

Таблица 2

Номер алгоритма в табл. 1	$W(p)$	$\rho_n(p)$	Условия, обеспечивающие псевдоградиентность $\rho_n(p) \geqslant 0$
1	$1-p(\alpha)$	$p(\alpha)\left(\sum_{i=1}^{N} v_i p(i) - v_\alpha\right)$	$v_i \in [0, 1]$ $(i=\overline{1, N})$
2	$1-p(\alpha)$	$ap(\alpha)\left\{p(\alpha)\left[\frac{1-v_\alpha}{a+(1-a)p(\alpha)} - \frac{v_\alpha}{1-(1-a)p(\alpha)}\right] - \sum_{i=1}^{N} p^2(i) \times \left[\frac{1-v_i}{a+(1-a)p(i)} - \frac{v_i}{1-(1-a)p(i)}\right]\right\}$	$v_\alpha \leqslant \frac{a}{1+a}$, $v_j \geqslant \frac{1}{1+a}$, $j \neq \alpha$ $(j=\overline{1, N})$
3	$1-p(\alpha)$	$p(\alpha)\left(p(\alpha)\bar{v}_\alpha - \sum_{i=1}^{N} p_i^2 \bar{v}_i\right)$ $\bar{v}_i \triangleq 1-2v_i \quad (i=\overline{1, N})$	$v_\alpha \leqslant 1/2$, $v_j \geqslant 1/2$, $j \neq \alpha$ $(j=\overline{1, N})$
4	$1-p(\alpha)$	$(\lambda(p)+\mu(p))\,p(\alpha)\left[\sum_{i=1}^{N} v_i p(i) - v_\alpha\right]$	$v_i \in [0, 1]$ $(i=\overline{1, N})$
5	$1-p(\alpha)$	$(N-1)^{-1}\left[\sum_{i=1}^{N} v_i p(i) - N v_\alpha p(\alpha)\right]$	ни при каких v_i $(i=\overline{1, N})$
6	$\lVert\tilde{p}-p^0\rVert^2$	$\frac{\varepsilon_n}{2}\sigma_q^2\left[V(\tilde{p}) - V(p^0)\right]$, $V(p) \triangleq \sum_{i=1}^{N} v_i p(i)$	$v_i \in (-\infty, \infty)$ $(i=\overline{1, N})$
7	$\lVert p-p^0\rVert^2$	$V(p)-V(p^0)$, $V(p) \triangleq \sum_{i=1}^{N} v_i p(i)$	$v_i \in (-\infty, \infty)$ $(i=\overline{1, N})$

Продолжение табл. 2

Номер алгоритма в табл. 1	$W(p)$	$\rho_n(p)$	Условия, обеспечивающие псевдоградиентность $\rho_n(p) \geqslant 0$
8	$\|p-p^0\|^2$, $\|p^i-p^{i0}\|^2$ $(i=\overline{1, N_1})$	$\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} v_{ij}[p(i)-p^0(i)]\,p^i(j)$, $\sum_{j=1}^{N_2} v_{ij}p(i)[p^i(j)-p^{i0}(j)]$ $(i=\overline{1, N_1})$	$v_{ij} \in (-\infty, \infty)$ $(i=\overline{1, N_1}, j=\overline{1, N_2})$
9	$\|p-p^0\|^2$, $\|p^i-p^{i0}\|^2$ $(i=\overline{1, N_1})$	$\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} v_{ij}[p(i)-p^0(i)]\,p^i(j)$, $\sum_{j=1}^{N_2} v_{ij}[p^i(j)-p^{i0}(j)]$ $(i=\overline{1, N_1})$	$v_{ij} \in (-\infty, \infty)$ $(i=\overline{1, N_1}, j=\overline{1, N_2})$

и $p_n(\alpha) \xrightarrow{\text{п.н.}} \tilde{p}(\alpha)$ при $n \to \infty$ в силу леммы П.10. Из условия 3) теоремы и (2.14) следует существование подпоследовательности $\{n_k\}$, для которой

$$\mathsf{M}\{\rho(p_{n_k})\} \xrightarrow[k\to\infty]{\text{п.н.}} 0.$$

Следовательно, учитывая условие 2) и сходимость $p_n(\alpha)$ к $\tilde{p}(\alpha)$, получаем $\tilde{p}(\alpha) \in \{0, 1\}$.

Если $\varliminf_{n\to\infty} \gamma_n > 0$, то из (2.14) и леммы Фату следует

$$\sum_{n=1}^{\infty} \rho(p_n) \overset{\text{п.н.}}{<} \infty,$$

откуда вытекает, что $\rho(p_n) \xrightarrow{\text{п.н.}} 0$ при $n \to \infty$. Теорема доказана. ▲

Важность этой теоремы состоит в том, что она охватывает случай известных алгоритмов Нарендры — Шапиро [143], Льюса [59] и Варшавского — Воронцовой [15], в которых $\gamma_n = \gamma \in (0, 1)$, $n = 1, 2, \ldots$ Из теоремы видно также, что выполнения условия псевдоградиентности (2.13) может быть не достаточно для обеспечения сходимости алгоритмов к оптимальной чистой стратегии $p^* = e_\alpha^N$. Ниже показывается, что при $\gamma_n = \mathrm{const}$ алгоритмы № 1—3 могут порождать неоптимальные рандомизированные стратегии. Кроме того, устанавливаются достаточные условия сходимости алгоритмов табл. 1, а также изучается скорость их сходимости.

§ 2.3. Алгоритм Нарендры — Шапиро

Рассмотрим алгоритм Нарендры — Шапиро

$$p_{n+1} = p_n + \gamma_n (1 - \xi_n) [e(x_n) - p_n], \tag{2.15}$$

предназначенный для решения задачи адаптивного выбора вариантов с бинарными потерями $\xi_n \in \{0, 1\}$, описанной в § 2.1. В (2.15) вектор $e(x_n) = \sum_{i=1}^{N} e_i^N \chi(x_n = x(i))$ — вершина симплекса S^N, соответствующая варианту x_n, $e_i^N = (\underbrace{0, \ldots, 0, 1}_{i}, \underbrace{0, \ldots, 0}_{N-i})$, $\gamma_n \in (0, 1)$ — длина шага, $p_1 \in S^N$, $p_1(j) > 0$ $(j = \overline{1, N})$, $n = 1, 2, \ldots$

Алгоритм (2.15) при постоянной длине шага $\gamma_n = \gamma \in (0, 1)$ рассматривался впервые в работе [143]. Долгое время факт сходимости с вероятностью 1 этого алгоритма считался доказанным. Однако в работе [40] было показано, что даже в простейшем случае двух вариантов ($N = 2$) этот алгоритм обеспечивает сходимость $p_n \xrightarrow[n\to\infty]{} e_\alpha^N$ лишь с вероятностью, меньшей единицы. Этот факт содержится в следующем утверждении.

Теорема 2.2. *Пусть выполнены предположения* П1, П2 *из* § 2.1, *условие* (2.8), $\gamma_n \triangleq \gamma \in (0, 1)$ и $\xi_n \in \{0, 1\}$ $(n = 1, 2, \ldots)$. *Тогда для алгоритма* (2.15) *справедливы неравенства*

$$\mathrm{P}_{\text{н}} \triangleq p_1(\alpha) \exp\left\{B\left[1 - \frac{\gamma B}{2(2-\gamma)}\right]\right\} \leqslant \mathrm{P}\left\{\lim_{n\to\infty} p_n(\alpha) = 1\right\} \leqslant$$

$$\leqslant 1 - \sum_{s \neq \alpha} (1 - v_s)^{1 - \ln^{-1} v_s} (p_1(s))^{1 - \ln^{-1}\gamma + \ln^{-2}\gamma} \triangleq \mathrm{P}_{\text{в}},$$

где

$$B \triangleq \sum_{i=1}^{N} v_i p_1(i) - v_\alpha = V(p_1) - V(e_\alpha^N).$$

Доказательство. Установим сначала справедливость верхней оценки. Рассмотрим события

$$A_k(s) \triangleq \{\omega \mid x_n = x(s)\,(n = n_k, \ldots, n_{k+1} - 1)\},$$

$$B_k(s) \triangleq \{\omega \mid \exists n \in \{n_k, \ldots, n_{k+1} - 1\} : \xi_n = 0\},$$

где $s = \overline{1, N}$, $s \neq \alpha$, $\{n_k\}$ — последовательность чисел, определяемая рекуррентным соотношением

$$n_{k+1} = n_k + k \qquad (k = 1, 2, \ldots), \quad n_1 = 1.$$

В силу алгоритма (2.15) для почти всех $\omega \in \bigcap_{k=1}^{\infty} (A_k(s) \cap$ $\cap B_k(s))$ имеем

$$p_{n+1}(s) \geqslant p_n(s), \quad n = 1, 2, \ldots,$$

$$p_{n_k}(s) \geqslant \gamma p_{n_{k-1}} + 1 - \gamma \geqslant \ldots \geqslant \gamma^{k-1} p_1(s) + 1 - \gamma^{k-1},$$

$$k = 1, 2, \ldots, \qquad (2.16)$$

откуда следует, что

$$\lim_{n\to\infty} p_n(s) = 1 \qquad (s \neq \alpha).$$

Таким образом, учитывая теорему 2.1, получаем

$$\mathsf{P}\{\lim_{n\to\infty} p_n(\alpha) = 1\} = 1 - \sum_{s\neq\alpha} \mathsf{P}\{\lim_{n\to\infty} p_n(s) = 1\} \leqslant$$

$$\leqslant 1 - \sum_{s\neq\alpha} \mathsf{P}\left\{\bigcap_{k=1}^{\infty} (A_k(s) \cap B_k(s))\right\}. \qquad (2.17)$$

Используя формулу условных вероятностей, учитывая неравенства (2.16) и предположения П1, П2 из § 2.1, получаем

$$\mathsf{P}\left\{\bigcap_{k=1}^{K} (A_k(s) \cap B_k(s))\right\} =$$

$$= \prod_{t=1}^{K} \mathsf{P}\left\{A_t(s) \mid B_t(s) \cap \bigcap_{k=1}^{t-1} (A_k(s) \cap B_k(s))\right\} \times$$

$$\times \mathsf{P}\left\{B_t(s) \,\middle|\, \bigcap_{k=1}^{t-1} (A_k(s) \cap B_k(s))\right\} \geqslant$$

$$\geqslant \prod_{k=1}^{K} (1 - v_s^k) \left[1 - (1 - p_1(s))\,\gamma^{k-1}\right]^k.$$

Переходя к пределу при $K\to\infty$ и используя (2.17), получаем

$$\mathsf{P}\left\{\lim_{n\to\infty} p_n(\alpha)=1\right\}\leqslant 1-\sum_{s\neq\alpha}\prod_{k=1}^{\infty}(1-v_s^k)\left(1-(1-p_1(s))\gamma^{k-1}\right)^k.$$

Требуемая верхняя оценка следует из этого неравенства и леммы П.3.

Установим теперь справедливость нижней оценки [138]. Из алгоритма (2.15) для любого $n=1, 2, \ldots$ получаем

$$\mathsf{M}\{p_{n+1}(\alpha)\}=$$
$$=\mathsf{M}\{p_n(\alpha)\}+\gamma\mathsf{M}\left\{\sum_{i=1}^{N}p_n(i)(1-v_i)(\delta_{i\alpha}-p_n(\alpha))\right\}\geqslant$$
$$\geqslant\mathsf{M}\{p_n(\alpha)\}+\gamma\mathsf{M}\left\{p_n(\alpha)\sum_{i\neq\alpha}p_n(i)(v_i-v_\alpha)\right\}$$

и

$$p_n(i)\geqslant p_{n-1}(i)(1-\gamma)\geqslant\ldots\geqslant p_1(i)(1-\gamma)^{n-1},$$

откуда следует неравенство

$$\mathsf{M}\{p_{n+1}(\alpha)\}\geqslant\mathsf{M}\{p_n(\alpha)\}\left[1+\gamma(1-\gamma)^{n-1}\sum_{i\neq\alpha}p_1(i)(v_i-v_\alpha)\right]\geqslant\ldots$$
$$\ldots\geqslant p_1(\alpha)\prod_{t=1}^{n}\left(1+\gamma(1-\gamma)^{t-1}\sum_{i\neq\alpha}p_1(i)(v_i-v_\alpha)\right).$$

В силу теоремы 2.1 и теоремы Лебега о мажорируемой сходимости [147] имеем

$$\mathsf{P}\left\{\lim_{n\to\infty}p_n(\alpha)=1\right\}=\mathsf{M}\left\{\lim_{n\to\infty}p_n(\alpha)\right\}=\lim_{n\to\infty}\mathsf{M}\{p_n(\alpha)\},$$

поэтому, переходя к пределу при $n\to\infty$, получаем неравенство

$$\mathsf{P}\left\{\lim_{n\to\infty}p_n(\alpha)=1\right\}\geqslant p_1(\alpha)\prod_{k=1}^{\infty}\left[1+\gamma(1-\gamma)^{k-1}\sum_{i\neq\alpha}(v_i-v_\alpha)p_1(i)\right].$$

Отсюда и из леммы П.3 следует требуемая нижняя оценка. Теорема доказана. ▲

Нижняя $\mathsf{P}_{\text{н}}$ и верхняя $\mathsf{P}_{\text{в}}$ оценки, приведенные в теореме 2.2, зависят как от длины шага γ в алгоритме (2.15), так и от начальных вероятностей $p_1(i)$ $(i=\overline{1,N})$ выбора возможных вариантов: с увеличением начальной вероятности $p_1(\alpha)$ выбора оптимального варианта $x(\alpha)$ верхняя и нижняя оценки вероятностей $\mathsf{P}\left\{\lim\limits_{n\to\infty}p_n(\alpha)=1\right.$

также увеличиваются и стремятся к единице при $p_1(\alpha)\to 1$. Заметим, что нижняя оценка увеличивается с уменьшением γ и при $\gamma\to 0$ стремится к величине $p_1(\alpha)e^{B}$.

В табл. 3 приведены значения нижней $\mathsf{P}_{\text{н}}$ и верхней $\mathsf{P}_{\text{в}}$ оценок вероятности $\mathsf{P}\left\{\lim\limits_{n\to\infty} p_n(\alpha)=1\right\}$ для $N=2$ при равновероятном начальном выборе вариантов $(p_1(1)=p_1(2)=0{,}5)$ и ряда значений параметра γ и условных средних потерь v_1, v_2.

Итак, при постоянной длине шага $\gamma_n=\gamma\in(0,1)$ алгоритм Нарендры — Шапиро (2.15) не позволяет решить поставленную задачу адаптивного выбора вариантов (безусловной минимизации средних потерь) с вероятностью 1. Оказывается, что задача может быть решена с помощью этого алгоритма, если в процессе работы длину шага алгоритма уменьшать (с определенной скоростью) до нуля.

Теорема 2.3 [75]. *Пусть выполнены предположения* П1, П2 *из* § 2.1, *условия* (2.8), $\xi_n\in\{0,1\}$ $(n=1,2,\ldots)$ *и*

$$\sum_{n=1}^{\infty}\gamma_n=\infty,\quad \gamma_n\in(0,1),$$

$$\varlimsup_{n\to\infty}\gamma_n\prod_{k=1}^{n}(1-\gamma_k)^{-1}\triangleq C_0<\frac{v^{-}-v_{-}}{1-v^{-}}\,p_1(\alpha).$$

Тогда последовательность векторов $\{p_n\}$, *порождаемая алгоритмом* (2.15), *сходится с вероятностью* 1 *и в среднеквадратическом к* $p^*\triangleq e_\alpha^N$.

Доказательство. Из алгоритма (2.15) следует, что при любом $n=1,2,\ldots$ с вероятностью 1 $p_n(\alpha)>0$. Рассмотрим последовательность случайных величин

$$W_n\triangleq\frac{1-p_n(\alpha)}{p_n(\alpha)}$$

и σ-алгебр $\mathscr{F}_n\triangleq\sigma(x_1,\xi_1,\ldots,x_{n-1},\xi_{n-1})$. В силу алгоритма (2.15) имеем с вероятностью 1

$$\mathsf{M}\{W_{n+1}\mid\mathscr{F}_n\}\leqslant$$

$$\leqslant W_n\left[1-\frac{\gamma_n(v^{-}-v_{-})}{1-\gamma_n}+\frac{\gamma_n^2(1-v_{-})}{(1-\gamma_n)\left[\gamma_n+(1-\gamma_n)p_n(\alpha)\right]}\right],$$

$$p_n(\alpha)\geqslant p_{n-1}(\alpha)(1-\gamma_{n-1})\geqslant\ldots\geqslant p_1(\alpha)\prod_{k=1}^{n-1}(1-\gamma_k).$$

Таблица 3

$v_1 = v_-$	$v_2 = v^-$	γ	$P_н$	$P_в$
0,1	0,3	0,01	0,552	0,783
		0,1	0,547	0,831
		0,5	0,525	0,977
		0,9	0,505	1,0
0,1	0,5	0,01	0,609	0,923
		0,1	0,598	0,940
		0,5	0,552	0,992
		0,9	0,510	1,0
0,1	0,7	0,01	0,673	0,996
		0,1	0,654	0,997
		0,5	0,579	0,9996
		0,9	0,515	1,0
0,1	0,9	0,01	0,743	1,0
		0,1	0,714	1,0
		0,5	0,607	1,0
		0,9	0,520	1,0
0,3	0,5	0,01	0,552	0,923
		0,1	0,547	0,940
		0,5	0,525	0,992
		0,9	0,505	1,0
0,3	0,7	0,01	0,609	0,996
		0,1	0,598	0,997
		0,5	0,552	0,9996
		0,9	0,510	1,0
0,3	0,9	0,01	0,673	1,0
		0,1	0,654	1,0
		0,5	0,579	1,0
		0,9	0,515	1,0
0,5	0,7	0,01	0,552	0,996
		0,1	0,547	0,997
		0,5	0,525	0,9996
		0,9	0,505	1,0
0,5	0,9	0,01	0,609	1,0
		0,1	0,598	1,0
		0,5	0,552	1,0
		0,9	0,510	1,0
0,7	0,9	0,01	0,552	1,0
		0,1	0,547	1,0
		0,5	0,525	1,0
		0,9	0,505	1,0

Отсюда, учитывая условия теоремы, получим

$$\mathsf{M}\{W_{n+1}\,|\,\mathcal{F}_n\}\leqslant$$

$$\leqslant W_n\left[1-\frac{\gamma_n(v^- - v_-)}{1-\gamma_n}+\frac{\gamma_n^2(1-v_-)}{(1-\gamma_n)\left[\gamma_n+p_1(\alpha)\prod\limits_{k=1}^{n}(1-\gamma_k)\right]}\right]\leqslant$$

$$\leqslant W_n\left[1-\gamma_n\left(v^- - v_- - \frac{1-v_-}{1+p_1(\alpha)\left(C_0^{-1}+o(1)\right)}\right)\right], \qquad (2.18)$$

где $o(1)\to 0$ при $n\to\infty$. Используя теперь лемму П.11 и условия теоремы, получаем, что $\lim\limits_{n\to\infty} W_n \overset{\text{п.н.}}{=} 0$, а значит, $\lim\limits_{n\to\infty} p_n(\alpha) \overset{\text{п.н.}}{=} 1$. По теореме Лебега [147] отсюда следует и сходимость в среднеквадратическом. Теорема доказана. ▲

Следствие. *В классе последовательностей* $\{\gamma_n\}$ *вида*

$$\gamma_n=\frac{\gamma}{(n+a)^{\varkappa}}$$

условия теоремы 2.3 *выполняются при* $\varkappa=1$ *и*

а) *либо* $\gamma\in(0,\ 1)$, $a>\gamma-1$,

б) *либо* $\gamma=1$, $a>\dfrac{1}{p_1(\alpha)}\,\dfrac{1-v^-}{v^- - v_-}$.

Справедливость этого следствия вытекает из п. б) леммы П.4.

Следующее утверждение устанавливает скорость сходимости алгоритма (2.15).

Теорема 2.4. *Пусть выполнены условия теоремы* 2.3 *и*

$$\gamma_n=\frac{\gamma}{n+a},\quad n=1,2,\ldots, \qquad (2.19)$$

где параметры γ *и* a *удовлетворяют одному из условий* а) *или* б) *следствия из теоремы* 2.3. *Тогда для всех* $n=1,\ 2,\ \ldots$

1) *при* $\gamma<1$ *и* $a>\gamma$

$$\mathsf{M}\{p_n(\alpha)\}\geqslant 1-\frac{1-p_1(\alpha)}{p_1(\alpha)}\,A\left(\frac{n+a-\gamma+1}{a-\gamma+1}\right)^{-\gamma(v^- - v_-)}, \qquad (2.20)$$

где

$$A\triangleq\exp\left\{\frac{(1-v_-)\,\gamma^2}{p_1(\alpha)\,(1-\gamma)\,(a-\gamma)}\right\};$$

2) *при* $\gamma = 1$

$$\mathsf{M}\{p_n(\alpha)\} \geqslant 1 - \frac{1-p_1(\alpha)}{p_1(\alpha)}\left(\frac{a}{n+a-H}\right)^H, \qquad (2.21)$$

где

$$H \triangleq v^- - v_- - \frac{1-v_-}{1+p_1(\alpha)\,a} > 0.$$

Доказательство. Усредняя (2.18) и используя нижнюю оценку п. а) леммы П.4, получаем при $\gamma < 1$

$$\mathsf{M}\{W_n\} \leqslant \mathsf{M}\{W_1\}\prod_{k=1}^{n}\left[1 - \right.$$
$$\left. - \frac{\gamma}{k+a-\gamma}\left(v^- - v_- - \frac{1-v_-}{1+p_1(\alpha)\,\gamma^{-1}(a-\gamma)^{\gamma}(k+a)^{1-\gamma}}\right)\right].$$

Воспользуемся неравенством $1+z \geqslant \ln z$ $(z>0)$:

$$\mathsf{M}\{W_n\} \leqslant \mathsf{M}\{W_1\}\exp\left\{-\sum_{k=1}^{n}\frac{\gamma}{k+a-\gamma}\left(v^- - v_- - \right.\right.$$
$$\left.\left. - \frac{1-v_-}{1+p_1(\alpha)\,\gamma^{-1}(a-\gamma)^{\gamma}(k+a)^{1-\gamma}}\right)\right\}.$$

Оценивая суммы соответствующими определенными интегралами, получим

$$\mathsf{M}\{W_n\} \leqslant \mathsf{M}\{W_1\}\exp\left\{\frac{(1-v_-)\,\gamma^2}{p_1(\alpha)\,(1-\gamma)\,(a-\gamma)}\right\}\left(\frac{a-\gamma+1}{n+a-\gamma+1}\right)^{\gamma(v^- - v_-)}.$$

Требуемое соотношение для случая $\gamma < 1$ следует из этого неравенства с учетом оценки $W_n \geqslant 1 - p_n(\alpha)$.

Аналогично, используя п. б) леммы П.4, получаем при $\gamma = 1$

$$\mathsf{M}\{W_n\} \leqslant$$
$$\leqslant \mathsf{M}\{W_1\}\prod_{k=1}^{n}\left[1 - \frac{1}{k+a-1}\left(v^- - v_- - \frac{1-v_-}{1+p_1(\alpha)\,a}\right)\right].$$

Применяя теперь к этому неравенству верхнюю оценку пункта а) леммы П.4, получаем требуемое соотношение для случая $\gamma = 1$. Теорема доказана. ▲

Следствие. *В условиях теоремы* 2.4 *последовательность текущих средних потерь* $\{\Phi_n\}$ *с вероятностью* 1 *сходится* *к минимальным условным средним потерям* v_-, *при этом*

$$\overline{\lim_{n\to\infty}}\, n^{\tau}\mathrm{M}\{(\Phi_n - v_-)^2\} \leqslant C_1\left(1-\frac{\tau}{2}\right)^{-2},$$

где

а) *при* $\gamma < 1$

$$\tau = \gamma(v^- - v_-),\quad C_1 = (v_+ - v_-)^2\,\frac{1-p_1(\alpha)}{p_1(\alpha)}\,A(a-\gamma+1)^{\tau};$$

б) *при* $\gamma = 1$

$$\tau = H,\quad C_1 = (v_+ - v_-)^2\,\frac{1-p_1(\alpha)}{p_1(\alpha)}\,a^{\tau}.$$

Это следствие вытекает из леммы 2.2 и соотношений

$$(V(p) - v_-)^2 = \Big[\sum_{i\neq\alpha}(v_i - v_-)\,p(i)\Big]^2 \leqslant (v_+ - v_-)^2\,(1-p(\alpha))^2,\quad p\in S^N.$$

Оценки скорости сходимости алгоритма Нарендры — Шапиро (2.15), полученные в теореме 2.4, позволяют рассмотреть вопрос об оптимальных параметрах длины шага алгоритма γ и a в (2.19). Анализируя эти оценки, можно прийти к следующим выводам:

1) *асимптотическая скорость сходимости алгоритма* (2.15) *с* γ_n *вида* (2.19), *гарантируемая теоремой* 2.4, *не превышает по порядку*

$$n^{-(v^- - v_-)}$$

и достигает этого порядка либо при $\gamma \to 1-0$ в оценке (2.20), *либо при* $a \to +\infty$ *в* (2.21);

2) *при фиксированном* n *и* $\gamma \to 1-0$ *в* (2.20) $A \to +\infty$ *и эта оценка становится тривиальной, поэтому предпочтительнее оценка* (2.21) *и использование в* (2.19) $\gamma = 1$;

3) *при* $\gamma = 1$ *для каждого* n *существует оптимальное значение параметра* $a = a^*(n)$, *которое максимизирует правую часть в оценке* (2.21), *при этом* $a^*(n) \to +\infty$ *при* $n \to \infty$.

Итак, теорема 2.4 и следствие из нее устанавливают условия, при выполнении которых задача адаптивного выбора вариантов (типа безусловной минимизации) с бинарными потерями может быть решена с помощью алгоритма Нарендры — Шапиро, а также дают оценки скорости достижения минимальных средних потерь.

§ 2.4. Алгоритмы Льюса и Варшавского — Воронцовой

Рассмотрим теперь алгоритм Льюса [59]

$$p_{n+1} = p_n + \gamma p_n^{\mathrm{T}} e(x_n)(e(x_n) - p_n) \times$$
$$\times \left[\frac{1-\xi_n}{a + (1-a)\, p_n^{\mathrm{T}} e(x_n)} - \frac{\xi_n}{1-(1-a)\, p_n^{\mathrm{T}} e(x_n)} \right], \qquad (2.22)$$

который также предназначен для решения задачи адаптивного выбора вариантов с бинарными потерями $\xi_n \in \in \{0, 1\}$. В отличие от алгоритма Нарендры — Шапиро (2.15) алгоритм Льюса (2.22) решает эту задачу с вероятностью 1 при постоянной, но достаточно малой длине шага γ. Этот факт содержится в следующем утверждении.

Теорема 2.5. *Пусть выполнены предположения* П1, П2 *из* § 2.1, *условие* (2.8), $\xi_n \in \{0, 1\}$ $(n = 1, 2, \ldots)$ *и*

$$v_- < \frac{a}{1+a}, \quad v^- > \frac{1}{1+a}.$$

Тогда при $a \in (0, 1]$, $\gamma \in (0, \bar\gamma)$, *где*

$$\bar\gamma \triangleq \min\{a - (a+1)v_-, \ (a+1)v^- - 1\},$$

и произвольном начальном векторе p_1, $p_1(i) > 0$ $(i = \overline{1, N})$, $\sum_{i=1}^{N} p_1(i) = 1$ *последовательность* $\{p_n\}$, *порождаемая алгоритмом Льюса* (2.22), *сходится с вероятностью* 1 *и в среднеквадратическом к* e_α^N. *При этом имеет место экспоненциальная скорость сходимости*

$$\mathsf{M}\{(1 - p_n(\alpha))^2\} \leqslant \frac{1 - p_1(\alpha)}{p_1(\alpha)} \lambda^{n-1},$$

где $\lambda \triangleq \lambda(\gamma, a, v_-, v^-, N) \in (0, 1)$, $n = 1, 2, \ldots$

Доказательство. Введем последовательности величин

$$W_n \triangleq \frac{1 - p_n(\alpha)}{p_n(\alpha)} \qquad (2.23)$$

и σ-алгебр $\mathcal{F}_n \triangleq \sigma\{x_t, \xi_t \mid t = \overline{1, n-1}\}$. В силу (2.22) и предположений П1, П2 для любого $n = 1, 2, \ldots$ с вероятностью 1

$$\mathsf{M}\{W_{n+1}\,|\,\mathcal{F}_n\} =$$
$$= p_n(\alpha)\, W_n\left[v_- \frac{1+\gamma p_n(\alpha)\,[1-(1-a)\,p_n(\alpha)]^{-1}}{1-\gamma(1-p_n(\alpha))\,[1-(1-a)\,p_n(\alpha)]^{-1}} + \right.$$
$$\left. + (1-v_-)\frac{1-\gamma p_n(\alpha)\,[a+(1-a)\,p_n(\alpha)]^{-1}}{1+\gamma(1-p_n(\alpha))\,[a+(1-a)\,p_n(\alpha)]^{-1}}\right] +$$
$$+ \sum_{i\neq\alpha} p_n(i)\left[v_i\,\frac{1-p_n(\alpha)-\gamma p_n(\alpha)\,p_n(i)\,[1-(1-a)\,p_n(i)]^{-1}}{p_n(\alpha)+\gamma p_n(\alpha)\,p_n(i)\,[1-(1-a)\,p_n(i)]^{-1}} + \right.$$
$$\left. + (1-v_i)\,\frac{1-p_n(\alpha)+\gamma p_n(\alpha)\,p_n(i)\,[a+(1-a)\,p_n(i)]^{-1}}{p_n(\alpha)-\gamma p_n(\alpha)\,p_n(i)\,[a+(1-a)\,p_n(i)]^{-1}}\right].$$

Правая часть этого равенства не уменьшится, если, используя (2.8), заменить v_i, $i\neq\alpha$ на v^- и выражения вида $1-(1-a)p_n(i)$ и $a+(1-a)p_n(i)$ для каждого $i=\overline{1,N}$ заменить либо на максимум, либо на минимум (в зависимости от знака перед соответствующими множителями) по $p_n(i)\in[0,1]$. Проведя эти замены, после несложных преобразований получим неравенство

$$\mathsf{M}\{W_{n+1}\,|\,\mathcal{F}_n\} \leqslant W_n + \gamma W_n p_n(\alpha)\left(\frac{v_-}{a-\gamma} - \frac{1-v_-}{1+\gamma}\right) +$$
$$+ \frac{\gamma}{p_n(\alpha)}\left(\frac{1-v^-}{a-\gamma} - \frac{v^-}{1+\gamma}\right)\sum_{i\neq\alpha} p_n^2(i).$$

Полагая $\gamma\in(0,\bar\gamma)$, увеличиваем правую часть этого неравенства, заменив $\sum\limits_{i\neq\alpha} p_n^2(i)$ на максимум по $p_n(i)$, $i\neq\alpha$ при условии $\sum\limits_{i\neq\alpha} p_n(i) = 1-p_n(\alpha)$, равный, как легко убедиться, $(1-p_n)(\alpha))^2/(N-1)$. Получаем

$$\mathsf{M}\{W_{n+1}\,|\,\mathcal{F}_n\} \leqslant W_n\left\{1+\gamma\left[p_n(\alpha)\left(\frac{v_-}{a-\gamma} - \frac{1-v_-}{1+\gamma}\right) + \right.\right.$$
$$\left.\left. + \frac{1-p_n(\alpha)}{N-1}\left(\frac{1-v^-}{a-\gamma} - \frac{v^-}{1+\gamma}\right)\right]\right\} \leqslant W_n\,[1-\gamma\theta(\gamma)],$$

где

$$\theta(\gamma) \triangleq \min\left\{\frac{1-v_-}{1+\gamma} - \frac{v_-}{a-\gamma},\ \frac{1}{N-1}\left(\frac{v^-}{1+\gamma} - \frac{1-v^-}{a-\gamma}\right)\right\}. \tag{2.24}$$

Обозначим

$$\lambda \triangleq 1 - \gamma\theta(\gamma). \tag{2.25}$$

Поскольку при $\gamma \in (0, \bar{\gamma})$ в условиях данной теоремы (2.24), (2.25) следует, что $\lambda \in (0, 1)$, то

$$\mathsf{M}\{W_{n+1} \mid \mathscr{F}_n\} \leqslant \lambda W_n \leqslant W_n, \tag{2.26}$$

т. е. $\{W_n\}$ — неотрицательный супермартингал. Из (2.23), (2.26) вытекает, что $p_n(\alpha) \to p^*(\omega) \in [0, 1]$ при $n \to \infty$ с вероятностью 1. Кроме того, при всех $n = 1, 2, \ldots$

$$\mathsf{M}\{(1 - p_n(\alpha))^2\} \leqslant \mathsf{M}\{W_n\} \leqslant \mathsf{M}\{W_1\}\lambda^{n-1} \xrightarrow[n\to\infty]{} 0,$$

откуда следует, что $p^*(\omega) = 1$ почти наверное, а значит, $p_n(\alpha) \xrightarrow[n\to\infty]{} 1$ с вероятностью 1 и в среднеквадратическом с требуемой скоростью. Теорема доказана. ▲

Следствие. *В условиях теоремы* 2.5 *последовательность текущих средних потерь* $\{\Phi_n\}$ *с вероятностью* 1 *сходится к минимальным условным средним потерям* v_-, *причем*

$$\lim_{n\to\infty} n\,\mathsf{M}\{(\Phi_n - v_-)^2\} = \sigma_\alpha^2,$$

где α *— номер оптимального варианта* (см. (2.8)).

Доказательство. Пользуясь рекуррентным представлением для $d_n \triangleq (\Phi_n - v_-)^2$, полученным при доказательстве леммы 2.2, имеем

$$\mathsf{M}\{d_n\} = \left(1 - \frac{1}{n}\right)^2 \mathsf{M}\{d_{n-1}\} + \frac{2}{n}\left(1 - \frac{1}{n}\right) a_n + \frac{1}{n^2} b_n,$$

где

$$a_n \triangleq \mathsf{M}\{(\Phi_{n-1} - v_-)(\xi_n - v_-)\} = \mathsf{M}\{(\Phi_{n-1} - v_-)(\mathsf{M}\{\xi_n \mid \mathscr{F}_{n-1}\} - v_-)\} = \mathsf{M}\{(\Phi_{n-1} - v_-)(V(p_n) - v_-)\},$$

$$\mathscr{F}_n \triangleq \sigma(x_1, \xi_1; \ldots; x_{n-1}, \xi_{n-1}),$$

$$V(p) \triangleq \sum_{i=1}^{N} v_i p_i, \quad b_n \triangleq \mathsf{M}\{(\xi_n - v_-)^2\}.$$

Поскольку из теоремы 2.5, соотношений (2.23)—(2.27) и предположений П1, П2 (§ 2.1) вытекает оценка

$$\mathsf{M}\{(V(p_n) - v_-)^2\} \leqslant C\lambda^n, \quad C \in (0, \infty), \quad \lambda \in (0, 1),$$

то

$$|a_n| \leqslant \sqrt{2C\lambda^n(1+v_-^2)}, \quad n = 1, 2, \ldots$$

Кроме того,

$$b_n = \mathsf{M}\{\mathsf{M}\{(\xi_n - v_-)^2 \mid \mathcal{F}_{n-1}\} =$$

$$= \mathsf{M}\left\{\sum_{i=1}^{N} p_n(i)\, \mathsf{M}\{(\xi_n(x(i), \omega) - v_-)^2\} = \right.$$

$$= \sigma_\alpha^2 \mathsf{M}\{p_n(\alpha)\} + \sum_{i \neq \alpha} \left[(v_i - v_-)^2 + \sigma_i^2\right] \mathsf{M}\{p_n(i)\} \xrightarrow[n\to\infty]{} \sigma_\alpha^2$$

в силу теоремы 2.5 и теоремы Лебега [147]. Таким образом,

$$\mathsf{M}\{d_n\} = \left(1 - \frac{2 + o(1)}{n}\right)\mathsf{M}\{d_{n-1}\} + \frac{\sigma_\alpha^2 + o(1)}{n^2},$$

где $o(1)$ — бесконечно малая последовательность. Утверждение данного следствия вытекает отсюда в силу следствия из леммы П.1 и леммы 2.2. Следствие доказано. ▲

Величина λ, характеризующая порядок скорости сходимости, определяется выражениями (2.24) — (2.25).

Как следует из теоремы 2.5, область сходимости алгоритма (2.22) зависит от параметров v_-, v^-, N. Если эти параметры априори не известны, то выбор постоянной длины шага может не соответствовать условиям сходимости ($\gamma \in (0, \bar{\gamma})$). Возникает вопрос: будет ли алгоритм Льюса (2.22) решать поставленную задачу с вероятностью 1 при произвольном $\gamma \in (0, 1)$? В следующей теореме на него дается отрицательный ответ.

Теорема 2.6. *Пусть выполнены предположения* П1, П2 *из* § 2.1, $\xi_n \in \{0, 1\}$ $(n = 1, 2 \ldots)$, *последовательность векторов* $\{p_n\}$ *порождается алгоритмом Льюса* (2.22) *и*

$$v_- < v^- \leqslant \frac{3}{3 + 2a^{-1} - (1+a)^{-2}}.$$

Тогда при любом $\gamma \in (\tilde{\gamma}, 1]$,

$$\tilde{\gamma} \triangleq 2\,\frac{v^-(1 + a^{-1}) - 1}{1 - v^-[1 - (1+a)^{-2}]}$$

с положительной вероятностью $\{p_n\}$ *сходится при* $n \to \infty$ *к* e_i^N $(v_i > v_-)$.

Доказательство. Определим последовательность чисел $\{n_k\}$ рекуррентным соотношением

$$n_{k+1}=n_k+k+c, \quad k=1,\ 2,\ \ldots,$$

причем $n_1>1$ и $c\geqslant 0$ — некоторые целые числа, которые задаются ниже. Рассмотрим события

$$A_k \triangleq \{\omega \mid x_n = x(s),\ n = n_k, \ldots, n_{k+1}-1\},\ k = 0,1,\ldots, n_0 \triangleq 1,$$

$$B_k \triangleq \left\{\omega \left| \frac{1}{n_{k+1}-n_k} \sum_{t=n_k}^{n_{k+1}-1} \xi_t(x(s), \omega) < v^- + \delta \right.\right\},\ k = 1, 2, \ldots,$$

$$B_0 \triangleq \{\omega \mid \xi_t(x(s), \omega) = 0,\ t = 1, \ldots, n_1 - 1\}, \tag{2.27}$$

где s — номер неоптимального варианта, для которого $v_s=v^-$, $\delta>0$. Покажем сначала, что можно выбрать n_1, c и δ такими, что для почти всех

$$\omega \in \bigcap_{k=0}^{\infty} (A_k \cap B_k)$$

последовательность $\{p_n\}$ сходится к неоптимальной s-й чистой стратегии, т. е.

$$\lim_{n\to\infty} p_n(s) = 1. \tag{2.28}$$

Положим $q_n \triangleq 1 - p_n(s)$. В силу алгоритма (2.22) имеем

$$q_{n+1} \leqslant q_n - \gamma q_n(1-q_n)[1-(1+a^{-1})\xi_n(x(s),\ \omega)]. \tag{2.29}$$

Введем функцию

$$W(q) \triangleq \left(\frac{q}{1-q}\right)^{\mu}, \quad q \in (0,\ 1), \tag{2.30}$$

которая при $\mu \in (0,\ 1)$ монотонно возрастает вместе со своей второй производной.

Обозначим $W_n \triangleq W(q_n)$. Если предположить, что для некоторой последовательности $\{\varepsilon_k\}$,

$$0 < \varepsilon_k < \bar{\varepsilon} \triangleq \min\{a\gamma^{-1}, (1-\mu)[2(1+\gamma a^{-1})]^{-1}\}$$

$$q_n \leqslant \varepsilon_k, \quad n = n_k, \ \ldots, \ n_{k+1}-1, \quad k = 1,\ 2,\ \ldots, \tag{2.31}$$

то, используя монотонность $W''(q)$, на основе формулы Тейлора из (2.29), (2.30) получим при $\xi_n(x(s),\ \omega)=0$

$$W_{n+1}\leqslant W_n\left[1-\mu\gamma+\frac{1}{2}\gamma^2\mu(\mu-1+2q_n)\right]\leqslant$$
$$\leqslant W_n[1+\gamma\mu r_0(\gamma,\ \mu,\ \varepsilon_k)],$$
$$r_0(\gamma,\ \mu,\ \varepsilon)\triangleq-1+\frac{\gamma}{2}(\mu-1+2\varepsilon);$$

при $\xi_n(x(s),\ \omega)=1$

$$W_{n+1}\leqslant W_n\left[1+\mu\frac{\gamma}{a}+\frac{\gamma^2\mu}{2a^2}\left(\mu-1+2q_n\left[1+\frac{\gamma}{a}(1-q_n)\right]\right)\times\right.$$
$$\left.\times\frac{\left(1+\gamma a^{-1}(1-q_n)\right)^{\mu-2}}{\left(1-\gamma a^{-1}q_n\right)^{\mu+2}}\right]\leqslant W_n[1+\gamma\mu r_1(\gamma,\ \mu,\ \varepsilon_k)],$$
$$r_1(\gamma,\ \mu,\ \varepsilon)\triangleq\frac{\gamma}{2a^2}\left(1+\frac{\gamma}{a}\right)^{\mu-2}\left(\mu-1+2\varepsilon\left[1+\frac{\gamma}{a}(1-\varepsilon)\right]\right).$$

В последней оценке используется тот факт, что $\varepsilon_k<\bar{\varepsilon}$. Объединяя эти неравенства, приходим к соотношению

$$W_{n+1}\leqslant W_n[1+\gamma\mu r_0(\gamma,\ \mu,\ \varepsilon_k)]^{1-\xi_n(x(s),\omega)}\times$$
$$\times[1+\gamma\mu r_1(\gamma,\ \mu,\ \varepsilon_k)]^{\xi_n(x(s),\omega)}\leqslant$$
$$\leqslant W_n\exp\{[1-\xi_n(x(s),\ \omega)]\ln[1+\gamma\mu r_0(\gamma,\ \mu,\ \varepsilon_k)]+$$
$$+\xi_n(x(s),\ \omega)\ln[1+\gamma\mu r_1(\gamma,\ \mu,\ \varepsilon_k)]\}.$$

Отсюда и из неравенства $\ln(1+z)\leqslant z,\ z>-1$ вытекает для $n\geqslant n_k+1$ и $n\leqslant n_{k+1}-1$

$$W_n\leqslant W_{n_k}\exp\{\mu\gamma(n-n_k)[r_0(\gamma,\ \mu,\ \varepsilon_k)+$$
$$+\left(\frac{1}{n-n_k}\sum_{t=n_k}^{n-1}\xi_t\right)(r_1(\gamma,\ \mu,\ \varepsilon_k)-r_0(\gamma,\ \mu,\ \varepsilon_k))]\};\qquad(2.32)$$

для $n=n_k$ (учитывая, что $\omega\in B_k$)

$$W_{n_{k+1}}\leqslant W_{n_k}\exp\{\mu\gamma(k+c)h(\gamma,\ \mu,\ \varepsilon_k,\ \delta)\},\qquad(2.33)$$

где

$$h(\gamma,\ \mu,\ \varepsilon,\ \delta)\triangleq r_0(\gamma,\ \mu,\ \varepsilon)+\left(v^-+\delta\right)\{r_1(\gamma,\ \mu,\ \varepsilon)-r_0(\gamma,\ \mu,\ \varepsilon)\}_+,\qquad(2.34)$$

$$\{z\}_+\triangleq\begin{cases}z, & \text{при}\ z\geqslant0,\\ 0, & \text{при}\ z<0.\end{cases}$$

Поскольку из условия на v^- следует, что $h(\gamma, 0, 0, 0) < 0$ при $\gamma \in (\tilde{\gamma}, 1]$, то в силу непрерывности функции (2.34) существуют $\mu \in (0, 1)$, $\tilde{\varepsilon} \in (0, \bar{\varepsilon})$ и $\delta > 0$, при которых

$$\tilde{h} \triangleq \max_{\varepsilon \in [0, \tilde{\varepsilon}]} h(\gamma, \mu, \varepsilon, \delta) < 0 \tag{2.35}$$

для любых $\gamma \in (\tilde{\gamma}, 1]$. Из (2.32), учитывая, что при $n = n_k + 1, \ldots, n_{k+1} - 1$

$$\left| \frac{1}{n - n_k} \sum_{t=n_k}^{n-1} \xi_t(x(s), \omega) \right| \leqslant 1, \quad n - n_k < k + c,$$

получаем

$$W_n \leqslant W_{n_k} \exp\{\mu\gamma (k + c) H(\gamma, \mu, \varepsilon_k)\}, \tag{2.36}$$

где

$$H(\gamma, \mu, \varepsilon) \triangleq |r_0(\gamma, \mu, \varepsilon)| + |r_1(\gamma, \mu, \varepsilon) - r_0(\gamma, \mu, \varepsilon)|.$$

Положим

$$\tilde{H} \triangleq \max_{\varepsilon \in [0, \tilde{\varepsilon}]} H(\gamma, \mu, \varepsilon).$$

Тогда из (2.33)—(2.36), в предположении, что $\varepsilon_k \leqslant \tilde{\varepsilon}$ $(k = 1, 2, \ldots)$, следует

$$W_n \leqslant W_{n_1} \exp\left\{\mu\gamma \left[\tilde{h} \sum_{t=1}^{k} (t + c) + (k + c) \tilde{H}\right]\right\} \tag{2.37}$$

для всех $n = n_k, \ldots, n_{k+1} - 1$, $k = 1, 2, \ldots$

Определим теперь последовательность $\{\varepsilon_k\}$ следующим образом:

$$\frac{\varepsilon_k}{1 - \varepsilon_k} \triangleq \frac{\varepsilon_1}{1 - \varepsilon_1} \exp\left\{\gamma \left[\tilde{h}\left(\frac{k(k+1)}{2} + kc\right) + (k + c)\tilde{H}\right]\right\}, \tag{2.38}$$

$k \geqslant 2$, где $\varepsilon_1 \in (0, \tilde{\varepsilon})$; при этом если $c \geqslant \tilde{H}|\tilde{h}|^{-1}$, то $\varepsilon_{k+1} \leqslant \varepsilon_k$ $(k = 1, 2, \ldots)$. Такая последовательность $\{\varepsilon_k\}$ удовлетворяет сделанным предположениям.

Поскольку $\omega \in A_0 \cap B_0$, то из (2.29) получаем

$$q_{n_1} \leqslant q_{n_1 - 1}[1 - \gamma(1 - q_1)] \leqslant q_1 [1 - \gamma(1 - q_1)]^{n_1 - 1}.$$

Тогда из (2.37) при $k=1$ следует, что для всех $n = n_1, \ldots, n_2-1$

$$W_n = W(q_n) \leqslant$$
$$\leqslant W\left(q_1[1-\gamma(1-q_1)]^{n_1-1}\right)\exp\{\mu\gamma(1+c)(\tilde{h}+\tilde{H})\} \leqslant W(\varepsilon_1)$$

при достаточно большом n_1, откуда, в силу монотонности функции $W(q)$, $q_n \leqslant \varepsilon_1$, а значит, (2.31) выполняется при $k=1$. Справедливость (2.31) при $k \geqslant 2$ следует из (2.37), (2.38) и монотонности функции $W(q)$ (2.30).

Из (2.35), (2.38) получаем, что $\lim\limits_{k\to\infty}\varepsilon_k = 0$, откуда в силу (2.31) и равенства $q_n = 1 - p_n(s)$ вытекает справедливость (2.28).

Таким образом, установлено, что

$$\mathsf{P}\{\lim_{n\to\infty} p_n(s) = 1\} \geqslant \mathsf{P}\left\{\bigcap_{k=0}^{\infty}(A_k \cap B_k)\right\}. \qquad (2.39)$$

Используя формулу условных вероятностей, в силу предположений П1, П2 из § 2.1 получаем

$$\mathsf{P}\left\{\bigcap_{k=0}^{K}(A_k \cap B_k)\right\} =$$
$$= \prod_{t=0}^{K}\mathsf{P}\{B_t\} \times \mathsf{P}\left\{A_t \mid B_t \bigcap_{k=0}^{t-1}(A_k \cap B_k)\right\}, \quad K = 1, 2, \ldots \qquad (2.40)$$

По формуле Бернштейна [147] для $k \geqslant 1$

$$\mathsf{P}\{B_k\} = \mathsf{P}\left\{\frac{1}{n_{k+1}-n_k}\sum_{t=n_k}^{n_{k+1}-1}\xi_t(x(s), \omega) < v^- + \delta\right\} \geqslant$$
$$\geqslant 1 - \mathsf{P}\left\{\left|\frac{1}{n_{k+1}-n_k}\sum_{t=n_k}^{n_{k+1}-1}(\xi_t(x(s), \omega) - v^-)\right| \geqslant \delta\right\} \geqslant$$
$$\geqslant 1 - 2\exp\{-a\delta^2(n_{k+1}-n_k)\}, \quad a > 0. \qquad (2.41)$$

В силу (2.31) и неравенства $(1-z)^k \geqslant 1 - kz$, $z \in (0, 1)$

$$\mathsf{P}\left\{A_t \mid B_t \bigcap_{k=0}^{t-1}(A_k \cap B_k)\right\} =$$
$$= \prod_{n=n_t}^{n_{t+1}-1}\mathsf{P}\left\{x_n = x(s) \mid (x_k = x(s),\ t = \overline{1, n-1}) \cap \left(\bigcap_{k=0}^{t} B_k\right)\right\} \geqslant$$
$$\geqslant (1-\varepsilon_t)^{n_{t+1}-n_t} \geqslant 1 - \varepsilon_t(t+c). \qquad (2.42)$$

Подставляя (2.41), (2.42) в (2.40) и переходя к пределу при $K \to \infty$, из (2.39) получаем

$$\mathsf{P}\left\{\lim_{n\to\infty} p_n(s) = 1\right\} \geqslant \mathsf{P}\{B_0\}\,\mathsf{P}\{A_0 \mid B_0\} \times$$
$$\times \prod_{k=1}^{\infty} \left(1 - \varepsilon_k(k+c)\right)\left(1 - 2\exp\{-a\delta^2(k+c)\}\right) > 0,$$

поскольку $\mathsf{P}\{B_0\} > 0$, $\mathsf{P}\{A_0 \mid B_0\} \geqslant p^{n_1}(s) > 0$, а бесконечное произведение сходится к положительному числу в силу (2.38) и (2.35). Теорема доказана. ▲

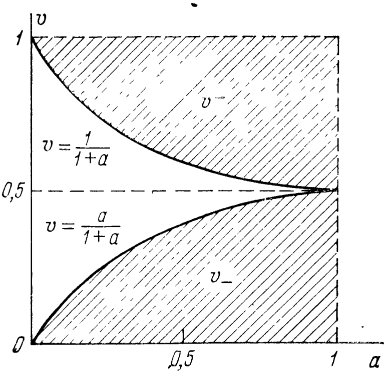

Рис. 9

Теорема 2.5 гарантирует сходимость алгоритма Льюса (2.22) к $p^* = e_\alpha^N$ лишь для таких задач, в которых условные средние потери v_- и v^- достаточно разделены между собой, а именно

$$v_- < \frac{a}{1+a}, \quad v^- > \frac{1}{1+a},$$
$$a \in (0, 1].$$

Соответствующая этим условиям область гарантируемой сходимости изображена на рис. 9. При $a = 1$ эта область становится наибольшей и алгоритм (2.22) упрощается:

$$p_{n+1} = p_n + \gamma p_n^{\mathrm{T}} e(x_n)[e(x_n) - p_n](1 - 2\xi_n), \qquad (2.43)$$
$$\gamma \in (0, 1], \quad n = 1, 2, \ldots$$

Этот алгоритм известен в литературе как *алгоритм Варшавского — Воронцовой* [15].

В результате проведенных на ЭВМ вычислений для различных значений v_-, v^-, a, удовлетворяющих условиям теоремы 2.5, были определены величины

$$\bar{\gamma} = \bar{\gamma}(a, v_-, v^-), \quad \gamma^* \triangleq \arg\min_{\gamma \in (0, \bar{\gamma})} \lambda(\gamma, a, v_-, v^-, N),$$
$$\lambda^* \triangleq \lambda(\gamma^*, a, v_-, v^-, N), \quad \tilde{\gamma} = \tilde{\gamma}(a, v_-, v^-).$$

Таблица 4

$N = 2$

v_-	v^-	a	$\bar{\gamma}$	γ^*	λ^*	$\tilde{\gamma}$
0,1	0,9	1,0	0,80	0,496	0,800	
		0,9	0,71	0,440	0,821	
		0,8	0,62	0,384	0,843	
		0,7	0,53	0,329	0,866	
		0,6	0,44	0,273	0,890	
		0,5	0,35	0,217	0,916	0,919
		0,4	0,26	0,156	0,942	0,751
		0,3	0,17	0,099	0,968	0,574
		0,2	0,08	0,045	0,990	0,389
0,1	0,8	1,0	0,60	0,336	0,900	
		0,9	0,52	0,292	0,915	
		0,8	0,44	0,246	0,930	
		0,7	0,36	0,202	0,947	
		0,6	0,28	0,151	0,962	0,935
		0,5	0,20	0,108	0,977	0,823
		0,4	0,12	0,065	0,990	0,692
		0,3	0,04	0,021	0,999	0,542
0,1	0,7	1,0	0,40	0,208	0,958	0,800
		0,9	0,33	0,172	0,968	0,821
		0,8	0,26	0,135	0,978	0,825
		0,7	0,19	0,099	0,986	0,810
		0,6	0,12	0,062	0,994	0,772
		0,5	0,05	0,025	0,999	0,710
0,1	0,6	1,0	0,20	0,100	0,990	0,400
		0,9	0,14	0,070	0,994	0,468
		0,8	0,08	0,040	0,998	0,523
		0,7	0,02	0,010	0,9999	0,563
0,2	0,9	1,0	0,60	0,336	0,900	
		0,9	0,52	0,291	0,915	
		0,8	0,44	0,246	0,931	
		0,7	0,36	0,202	0,947	
		0,6	0,28	0,151	0,962	
		0,5	0,20	0,108	0,977	0,919
		0,4	0,12	0,065	0,990	0,751
		0,3	0,04	0,021	0,999	0,574
0,2	0,8	1,0	0,60	0,336	0,900	
		0,9	0,52	0,291	0,915	
		0,8	0,44	0,246	0,931	
		0,7	0,36	0,202	0,947	
		0,6	0,28	0,151	0,962	0,936
		0,5	0,20	0,108	0,977	0,824
		0,4	0,12	0,065	0,990	0,692
		0,3	0,04	0,021	0,997	0,543

Продолжение табл. 4

v_-	v^-	a	$\bar{\gamma}$	γ^*	λ^*	$\tilde{\gamma}$
0,2	0,7	1,0	0,40	0,208	0,958	0,800
		0,9	0,33	0,172	0,968	0,821
		0,8	0,26	0,135	0,978	0,825
		0,7	0,19	0,099	0,986	0,810
		0,6	0,12	0,062	0,994	0,772
		0,5	0,05	0,025	0,999	0,710
0,2	0,6	1,0	0,20	0,100	0,990	0,400
		0,9	0,14	0,070	0,994	0,468
		0,8	0,08	0,040	0,998	0,523
		0,7	0,02	0,010	0,999	0,563
0,3	0,9	1,0	0,40	0,208	0,958	
		0,9	0,33	0,171	0,968	
		0,8	0,26	0,136	0,978	
		0,7	0,19	0,099	0,986	
		0,6	0,12	0,062	0,994	
		0,5	0,05	0,025	0,999	0,919
0,3	0,8	1,0	0,40	0,208	0,958	
		0,9	0,33	0,171	0,968	
		0,8	0,26	0,136	0,978	
		0,7	0,19	0,099	0,986	
		0,6	0,12	0,062	0,994	0,936
		0,5	0,05	0,025	0,999	0,824
0,3	0,7	1,0	0,40	0,208	0,958	0,800
		0,9	0,33	0,171	0,968	0,821
		0,8	0,26	0,136	0,978	0,825
		0,7	0,19	0,099	0,986	0,810
		0,6	0,12	0,062	0,994	0,772
		0,5	0,05	0,025	0,999	0,710
0,3	0,6	1,0	0,20	0,100	0,990	0,400
		0,9	0,14	0,070	0,994	0,468
		0,8	0,08	0,040	0,998	0,523
		0,7	0,02	0,010	0,9999	0,563
0,4	0,9	1,0	0,20	0,100	0,990	
		0,9	0,14	0,070	0,994	
		0,8	0,08	0,040	0,998	
		0,7	0,02	0,010	0,9999	
0,4	0,8	1,0	0,20	0,100	0,990	
		0,9	0,14	0,070	0,994	
		0,8	0,08	0,040	0,998	
		0,7	0,02	0,010	0,9999	

Продолжение табл. 4

v_{-}	v^{-}	a	$\bar{\gamma}$	γ^*	λ^*	$\tilde{\gamma}$
0,4	0,7	1,0	0,20	0,100	0,990	0,800
		0,9	0,14	0,070	0,994	0,821
		0,8	0,08	0,040	0,998	0,825
		0,7	0,02	0,010	0,9999	0,810
0,4	0,6	1,0	0,20	0,100	0,990	0,400
		0,9	0,14	0,070	0,994	0,468
		0,8	0,08	0,040	0,998	0,523
		0,7	0,02	0,010	0,9999	0,563

Таблица 5

$v_{-} = 0{,}1,\ v^{-} = 0{,}9$

N	a	$\bar{\gamma}$	γ^*	λ^*	$\tilde{\gamma}$
2	1,0	0,80	0,496	0,800	
	0,9	0,71	0,440	0,821	
	0,8	0,62	0,384	0,843	
	0,7	0,53	0,329	0,866	
	0,6	0,44	0,273	0,890	
	0,5	0,35	0,217	0,916	0,919
	0,4	0,26	0,156	0,942	0,751
	0,3	0,17	0,099	0,968	0,574
	0,2	0,08	0,045	0,990	0,389
5	1,0	0,80	0,496	0,950	
	0,9	0,71	0,440	0,955	
	0,8	0,62	0,384	0,961	
	0,7	0,53	0,329	0,966	
	0,6	0,44	0,273	0,973	
	0,5	0,35	0,217	0,979	0,919
	0,4	0,26	0,156	0,986	0,751
	0,3	0,17	0,099	0,992	0,574
	0,2	0,08	0,045	0,998	0,389
10	1,0	0,80	0,496	0,978	
	0,9	0,71	0,440	0,980	
	0,8	0,62	0,384	0,983	
	0,7	0,53	0,329	0,985	
	0,6	0,44	0,273	0,988	
	0,5	0,35	0,217	0,991	0,919
	0,4	0,26	0,156	0,994	0,751
	0,3	0,17	0,099	0,996	0,574
	0,2	0,08	0,045	0,999	0,389

Параметр $\bar{\gamma}$ характеризует максимально допустимое значение длины шага γ в алгоритме Льюса (2.22), при котором теоремой 2.5 гарантируется его сходимость со скоростью $\lambda^n(\gamma, a, v_-, v^-, N)$, а $\tilde{\gamma}$ — нижняя граница длины шага γ, при которой алгоритм с положительной вероятностью приводит к выбору неоптимального варианта (только для значений a, v^-, удовлетворяющих условиям теоремы 2.6). Результаты вычислений приведены в табл. 4 для $N=2$ и в табл. 5 (при $v_-=0{,}1$ и $v^-=0{,}9$) для трех значений количества возможных вариантов ($N=2$, 5, 10).

Полученные результаты показали, что во всех случаях при любом $\gamma \in (0, \bar{\gamma})$ и $a \in (0, 1)$

$$\lambda(\gamma, 1, v_-, v^-, N) < \lambda(\gamma^*, a, v_-, v^-, N),$$

т. е. значение параметра $a=1$ в алгоритме Льюса является оптимальным. Таким образом, в классе алгоритмов Льюса (2.22) оптимальным по скорости сходимости (в смысле оценки, даваемой теоремой 2.5) является алгоритм Варшавского — Воронцовой (2.43). Кроме того, из табл. 5 видно, что с ростом числа N возможных вариантов λ^* быстро приближается к единице, т. е. скорость сходимости значительно уменьшается.

§ 2.5. Алгоритм Буша — Мостеллера

Одним из наиболее известных алгоритмов, моделирующих процессы обучения и адаптации, является алгоритм Буша — Мостеллера

$$p_{n+1} = p_n + \gamma_n \left(e(x_n) - p_n + \frac{e^N - Ne(x_n)}{N-1} \xi_n \right), \qquad (2.44)$$

где $\xi_n \in \{0, 1\}$, $e^N \triangleq (1, 1, \ldots, 1) \in R^N$, $\gamma_n \in (0, 1)$ — длина шага, $p_1 \in S^N$, $n=1, 2, \ldots$. Впервые этот алгоритм рассматривался в монографии [12] при постоянной длине шага $\gamma_n=\gamma$ ($n=1, 2, \ldots$). Впоследствии в ряде работ [54, 78] он применялся для отыскания оптимальных вариантов в условиях неопределенности. Однако, как показывается ниже, алгоритм Буша — Мостеллера позволяет решить сформулированную в § 2.1 задачу лишь в вырожденном случае ($v_-=0$), а в других случаях — распределить предельные вероятности выбора вариантов $p_n(i)$ обратно пропорционально соответствующим условным средним потерям v_i $(i=\overline{1, N})$.

Теорема 2.7 [138]. *Пусть выполнены предположения* П1, П2 *из* § 2.1, *условие* (2.8), $\xi_n \in \{0, 1\}$ $(n=1, 2, \ldots)$,

$$\gamma_n \in (0, 1], \quad \sum_{n=1}^{\infty} \gamma_n = \infty,$$

а последовательность векторов $\{p_n\}$ *порождается алгоритмом Буша — Мостеллера* (2.44). *Тогда*

1) *если* $v_- = 0$, *то с вероятностью* 1 $p_n \to e_\alpha^N$ *при* $n \to \infty$;

2) *если* $v_- > 0$ *и* $\sum_{n=1}^{\infty} \gamma_n^2 < \infty$, *то с вероятностью* 1 $p_n \to p^0 \triangleq (p^0(1), \ldots, p^0(N))$ *при* $n \to \infty$, *где*

$$p_i^0 = \frac{1}{v_i}\left(\sum_{j=1}^{N} \frac{1}{v_j}\right)^{-1}, \quad i=\overline{1, N}.$$

Доказательство. Введем σ-алгебры $\mathscr{F}_n \triangleq \sigma(x_1, \xi_1; \ldots; x_{n-1}, \xi_{n-1})$. Если $v_- = 0$, то $V(p_n) \geqslant v^-(1-p_n(\alpha))$ и в силу алгоритма (2.44) для всех $n=1, 2, \ldots$ с учетом этого неравенства получаем

$$\mathsf{M}\{1-p_{n+1}(\alpha) \mid \mathscr{F}_n\} \leqslant \left(1-\frac{v^-}{N-1}\gamma_n\right)(1-p_n(\alpha)). \quad (2.45)$$

Используя лемму П.11, приходим к утверждению 1) данной теоремы. Если $v_- > 0$, то из (2.44) для всех $n=1, 2, \ldots$ следует

$$\mathsf{M}\{\| p_{n+1}-p^0\|^2 \mid \mathscr{F}_n\} \leqslant \|p_n - p^0\|^2 - \frac{2N}{N-1}\gamma_n \sum_{i=1}^{N} v_i\left[p_n(i)-p_i^0\right]^2 + 2\gamma_n^2,$$

где $\|\cdot\|$ — евклидова норма. Отсюда получаем

$$\mathsf{M}\{\| p_{n+1}-p^0\|^2 \mid \mathscr{F}_n\} \leqslant \left(1-\frac{2Nv_-}{N-1}\gamma_n\right)\|p_n - p^0\|^2 + 2\gamma_n^2. \quad (2.46)$$

Утверждение 2) теоремы вытекает теперь из леммы П.11. Теорема доказана. ▲

Следствие. *В условиях теоремы* 2.7

1) *при* $v_- = 0$ *и* $\gamma_n = \gamma \in (0, 1)$ *справедлива оценка*

$$\mathsf{M}\{p_n(\alpha)\} \geqslant 1-(1-p_1(\alpha))\left(1-\frac{\gamma v^-}{N-1}\right)^{n-1}, \quad n=1, 2, \ldots;$$

2) *при* $v_- > 0$ *и* $\gamma_n = \gamma(n+a)^{-1}$, $n = 1, 2, \ldots, \frac{N-1}{2Nv_-} < \gamma < 1 + a$ *справедлива оценка*

$$\varlimsup_{n\to\infty} n\mathsf{M}\{\|p_n - p^0\|^2\} \leqslant 2\left(\frac{2Nv_-}{N-1}\gamma - 1\right)^{-1}.$$

Доказательство вытекает из неравенств (2.45), (2.46) и леммы П.2 (случай а)).

Замечание. *В случае* $v_- = 0$ *последовательность текущих средних потерь* $\{\Phi_n\}$ *с вероятностью 1 сходится к нулю, причем*

$$\lim_{n\to\infty} n\mathsf{M}\{\Phi_n^2\} = 0.$$

Доказательство этого факта аналогично доказательству следствия 1 из теоремы 2.5, при этом $\sigma_\alpha = 0$.

Таким образом, алгоритм Буша — Мостеллера (2.44) решает задачу безусловной минимизации средних потерь лишь в вырожденном случае $v_- = 0$, при этом, как нетрудно показать, средние потери стремятся к нулю экспоненциально при $\gamma_n = \gamma \in (0, 1]$. В остальных случаях (когда $v_- > 0$) при $\gamma_n = \gamma(n+a)^{-1}$ текущие средние потери стремятся к $V(p_0) > v_-$ при $n \to \infty$.

§ 2.6. Проекционный алгоритм стохастической аппроксимации

Перейдем теперь к рассмотрению алгоритмов, предназначенных для решения задачи адаптивного выбора вариантов (§ 2.1) с небинарными потерями. Как отмечалось ранее, в этом случае применяются проекционные алгоритмы вида (2.12). Одним из таких алгоритмов является следующий:

$$p_{n+1} = \pi^N_{\varepsilon_{n+1}}\left\{p_n - \gamma_n \frac{\xi_n - \Delta}{e^{\mathrm{T}}(x_n)\,p_n} e(x_n)\right\}, \tag{2.47}$$

где γ_n — длина шага, Δ — числовой параметр, π^N_ε — оператор проектирования (2.10) на ε-симплекс (2.9), $\{\varepsilon_n\}$ — последовательность чисел из интервала $(0, 1/N)$, отделяющая компоненты вектора p_n от нуля. Введение этой последовательности имеет определенный физический смысл: на каждом шаге обеспечивается ненулевая вероятность выбора любого варианта $x(i) \in X$, что необходимо для накопления полной статистической информации о его качестве (т. е. о величине v_i). Заметим, что при $\Delta = 0$ алгоритм (2.47) совпадает с алгоритмом, рассмотренным в [72, 138]. В этом случае вектор движения в (2.47)

$$R_n = \frac{\xi_n}{e^{\mathrm{T}}(x_n)\, p_n} e(x_n)$$

является в среднем градиентом функции $V(p)$ (2.3):

$$\mathsf{M}\{R_n \mid p_n = p\} = \sum_{i=1}^{N} p_i \frac{\mathsf{M}\{\xi_n \mid x_n = x(i)\}}{e^{\mathrm{T}}(x(i))' p} e(x(i)) = \nabla V(p).$$

Следовательно, алгоритм (2.47) при $\Delta = 0$ является алгоритмом стохастической аппроксимации для задачи минимизации функции $V(p)$ на симплексе S^N. При $\Delta \neq 0$, как нетрудно показать, алгоритм (2.47) соответствует методу стохастической аппроксимации при минимизации функции

$$V_\Delta(p) \triangleq \sum_{i=1}^{N} (v_i - \Delta)\, p_i,$$

которая при $p \in S^N$ равна $V(p) - \Delta$.

Приведем теперь условия сходимости этого алгоритма. Обозначим

$$m_i(\Delta) \triangleq \sup_n \mathsf{M}\{(\xi_n - \Delta)^2 \mid x_n = x(i)\} = \sigma_i^2 + (v_i - \Delta)^2$$

$$(i = \overline{1, N}),$$

$$m(\Delta) \triangleq \sum_{i \neq \alpha} m_i(\Delta), \quad \delta_v \triangleq v^- - v_-. \tag{2.48}$$

Теорема 2.8. *Пусть выполнены предположения* П1, П2 *из* § 2.1, *условие* (2.8), $p_1 \in S^N_{\varepsilon_1}$, *а* $\{p_n\}$ — *последовательность векторов, порожденная алгоритмом* (2.47), *в котором*

$$\sum_{n=1}^{\infty} \gamma_n = \infty, \quad \gamma_n > 0, \quad \varepsilon_n \in (0, 1/N) \quad (n = 1, 2, \ldots), \quad \varepsilon_n \xrightarrow[n \to \infty]{} 0.$$

Тогда

1) *если*

$$\sum_{n=1}^{\infty}\left(\frac{\gamma_n^2}{\varepsilon_n}+|\varepsilon_{n+1}-\varepsilon_n|\right)<\infty,$$

то с вероятностью 1 $p_n \to e_\alpha^N$ *при* $n\to\infty$;

2) *если при* $n\to\infty$

$$\left(\gamma_n\varepsilon_n^{-1}+|\varepsilon_{n+1}-\varepsilon_n|\gamma_n^{-1}\right)\to 0,$$

то последовательность $\{p_n\}$ *сходится к* $p^*=e_\alpha^N$ *в среднеквадратическом.*

Доказательство. Введем σ-алгебры $\mathscr{F}_n \triangleq \triangleq \sigma(x_1, \xi_1; \ldots; x_{n-1}, \xi_{n-1})$ и величины $W_n \triangleq \|p_n - p_n^*\|^2$, где

$$p_n^* \triangleq (1-N\varepsilon_n)\, e_\alpha^N + \varepsilon_n e^N \in S_{\varepsilon_n}^N \tag{2.49}$$

— единственная (в силу условия (2.8)) точка минимума функции $V(p)$ (2.3) на ε-симплексе $S_{\varepsilon_n}^N$ (2.9).

Из алгоритма (2.47), свойств проектора (2.10) и неравенства треугольника для любого $n = 1, 2, \ldots$ получим

$$W_{n+1}\leqslant\left(\sqrt{W_n}+\|p_{n+1}^*-p_n^*\|\right)^2 - \\ -2\gamma_n\frac{\xi_n-\Delta}{e^{\mathrm{T}}(x_n)\,p_n}\,e^{\mathrm{T}}(x_n)\left(p_n-p_{n+1}^*\right)+\gamma_n^2\frac{(\xi_n-\Delta)^2}{\left(e^{\mathrm{T}}(x_n)\,p_n\right)^2}, \tag{2.49a}$$

откуда в силу предположений П1, П2

$$\mathsf{M}\{W_{n+1}|\mathscr{F}_n\}\leqslant W_n+2\sqrt{N(N-1)}\,|\varepsilon_{n+1}-\varepsilon_n|\sqrt{W_n}+ \\ +N^2(\varepsilon_{n+1}-\varepsilon_n)^2+2\|v\|N\gamma_n|\varepsilon_{n+1}-\varepsilon_n|+ \\ +\gamma_n^2\sum_{i=1}^{N}m_i(\Delta)/p_n(i)-2\gamma_n\left(V(p_n)-V(p_n^*)\right), \tag{2.50}$$

где $v \triangleq (v_1, \ldots, v_N)$. Так как для любого $z\in[0,\ 1-\delta]$, $\delta\in(0,\ 1)$ справедливо неравенство

$$(1-z)^{-1}\leqslant 1+z\delta^{-1}$$

и поскольку $p_n(\alpha) \geqslant \varepsilon_n > 0$, то

$$\frac{1}{p_n(\alpha)} = \frac{1}{1-(N-1)\varepsilon_n}\left[1-\left(1-\frac{p_n(\alpha)}{1-(N-1)\varepsilon_n}\right)\right]^{-1} \leqslant$$

$$\leqslant \frac{1}{1-(N-1)\varepsilon_n}\left[1+\frac{1-(N-1)\varepsilon_n - p_n(\alpha)}{\varepsilon_n}\right] \leqslant$$

$$\leqslant \frac{1}{1-N\varepsilon_n}\left(1+\frac{\sqrt{W_n}}{\varepsilon_n}\right).$$

Нетрудно также установить оценку

$$V(p_n) - V(p_n^*) \geqslant \frac{1}{2}\delta_v W_n,$$

где $\delta_v > 0$. Подставляя эти соотношения в (2.50′), получим

$$\mathsf{M}\{W_{n+1} \mid \mathscr{F}_n\} \leqslant (1-\delta_v\gamma_n) W_n + m(\Delta)\gamma_n^2\varepsilon_n^{-1} +$$

$$+ \left(2\sqrt{N(N-1)}\,|\varepsilon_{n+1}-\varepsilon_n| + \frac{m_\alpha(\Delta)}{1-N\varepsilon_n}\gamma_n^2\varepsilon_n^{-1}\right)\sqrt{W_n} +$$

$$+ C\,|\varepsilon_{n+1}-\varepsilon_n| + \frac{m_\alpha(\Delta)}{1-N\varepsilon_n}\gamma_n^2, \qquad (2.51)$$

где $C \triangleq N(1+2\|v\|\sup_n \gamma_n) < \infty$. Тогда

1) сходимость с вероятностью 1 вытекает из (2.51), леммы П.11, условий теоремы, (2.49) и условия $\sqrt{W_n} \leqslant \sqrt{2}$,

2) сходимость в среднеквадратическом вытекает из неравенства, полученного усреднением (2.51) с последующим применением неравенства Иенсена $\mathsf{M}\{\sqrt{W_n}\} \leqslant \sqrt{\mathsf{M}\{W_n\}}$ и леммы П.5 (при $b=d=0$). Теорема доказана. ▲

Для оценки *скорости сходимости* алгоритма (2.47) ограничимся следующими последовательностями $\{\gamma_n\}$ и $\{\varepsilon_n\}$:

$$\gamma_n \sim \gamma n^{-\varkappa}, \quad \varepsilon_n = \varphi(n^{-\nu}), \qquad (2.52)$$

где $\gamma > 0$; символ эквивалентности «$\sim$» двух последовательностей означает, что отношение их общих членов стремится к единице при $n \to \infty$; $\varphi(\cdot)$ — функция, определенная на отрезке $[0, 1]$ и удовлетворяющая условиям: 1) $\varphi(0) = 0$ и $0 < \varphi(z) < N^{-1}$ для всех $z \in (0, 1]$, 2) $\varphi(z)$ имеет в положительной окрестности нуля ограниченную производную порядка $2+[\nu^{-1}]$, где $[\cdot]$ — целая часть числа, 3) $\varphi'(+0) \triangleq \varepsilon > 0$. Например, этим условиям удовлетворяют последовательности вида

$$\gamma_n = \frac{\gamma}{(n+a)^{\varkappa}}, \quad \varepsilon_n = \frac{\varepsilon}{(n+b)^{\nu}} \quad (a, b > 0).$$

В соответствии с теоремой 2.8 сходимость алгоритма (2.47) в классе последовательностей (2.52) имеет место:

1) с вероятностью 1, если $0 < \nu < 2\varkappa - 1$ и $\varkappa \leqslant 1$;
2) в среднеквадратическом, если

$$0 < \nu < \varkappa \leqslant 1. \tag{2.53}$$

Теорема 2.9 [69]. *Пусть выполнены условия теоремы 2.8 и последовательности* $\{\gamma_n\}$ *и* $\{\varepsilon_n\}$ *удовлетворяют условиям* (2.52), (2.53), *где* $\varkappa$, ν, γ, ε *и функция* $\varphi(\cdot)$ *обладают указанными выше свойствами. Тогда для последовательности* $\{p_n\}$, *порожденной алгоритмом* (2.47) *справедливо неравенство*

$$\varlimsup_{n\to\infty} n^{\theta} \mathsf{M}\{\|p_n - p^*\|^2\} \leqslant d, \tag{2.54}$$

где

$$\theta \triangleq \min\{2\nu, \tau\},$$

$$\tau \triangleq \min\{\varkappa - \nu, 2(1 + \nu - \varkappa)\},$$

$$d \triangleq (\varepsilon\sqrt{N(N-1)}\,\chi(\tau \geqslant 2\nu) + \sqrt{D}\,\chi(\tau \leqslant 2\nu))^2, \tag{2.54a}$$

а D *определяется следующим образом:*

а) *если* $\varkappa < \nu + 2/3$ *и при* $\varkappa = 1$ *имеем* $\gamma\delta_v > 1 - \nu$, *то*

$$D = \frac{2\nu\varepsilon\sqrt{N(N-1)}\,\chi(\varkappa - \nu \geqslant 1/2) + \chi(\varkappa - \nu \leqslant 1/2)\, m(\Delta)\,\gamma^2\varepsilon^{-1}}{\gamma\delta_v + (\nu - 1)\,\chi(\varkappa = 1)};$$

б) *если* $\varkappa = \nu + 2/3$ *и при* $\varkappa = 1$ *имеем* $\gamma\delta_v > 1 - \nu$, *то*

$$D = \left(\frac{\nu\varepsilon\sqrt{N(N-1)} + [N(N-1)\,\nu^2\varepsilon^2 + \gamma(\delta_v + (\nu-1)\chi(\varkappa=1))m(\Delta)\gamma^2\varepsilon^{-1}]^{1/2}}{\gamma\delta_v + (\nu-1)\,\chi(\varkappa=1)}\right)^2;$$

в) *если* $\varkappa > \nu + 2/3$ *и при* $\varkappa = 1$ *имеем* $\gamma\delta_v > 2\nu$, *то*

$$D = \left(\frac{2\sqrt{N(N-1)}\,\nu\varepsilon}{\gamma\delta_v - 2\nu\,\chi(\varkappa = 1)}\right)^2.$$

Здесь δ_v, $m(\Delta)$ *задаются формулами* (2.48).

Доказательство. Усредним обе части неравенства (2.51) и воспользуемся неравенством Иенсена. Получим для всех $n = 1, 2, \ldots$

$$\mathsf{M}\{W_{n+1}\} \leqslant (1 - \delta_v \gamma_n)\,\mathsf{M}\{W_n\} + m(\Delta)\,\gamma_n^2 \varepsilon_n^{-1} + \\ + \left(2\sqrt{N(N-1)}\,|\varepsilon_{n+1} - \varepsilon_n| + \frac{m_\alpha(\Delta)}{1 - N\varepsilon_n}\,\gamma_n^2 \varepsilon_n^{-1}\right)\sqrt{\mathsf{M}\{W_n\}} + \\ + C\,|\varepsilon_{n+1} - \varepsilon_n| + \frac{m_\alpha(\Delta)}{1 - N\varepsilon_n}\,\gamma_n^2.$$

Из (2.52) и указанных свойств функции $\varphi(\cdot)$ следует

$$\varepsilon_n \sim \varepsilon n^{-\nu}, \quad |\varepsilon_{n+1} - \varepsilon_n| \sim \varepsilon \nu n^{-(1+\nu)}.$$

Используем теперь лемму П.2 при $r = 1/2$ и неравенство

$$\mathsf{M}\{\|p_n - p^*\|^2\} \leqslant (1 + \lambda)\,\mathsf{M}\{d_n\} + (1 + \lambda^{-1})\,\mathsf{M}\{\|p_n^* - p^*\|^2\},$$

справедливое при всех $\lambda \in (0, \infty)$ и $n = 1, 2, \ldots$ Получим неравенство вида (2.54), в котором d зависит, вообще говоря, от λ. Взятие инфинума d по $\lambda > 0$ завершает доказательство. Теорема доказана. ▲

Следствие 1. *При выполнении предположений теоремы* 2.9 *справедливо целевое условие* (2.4) *и неравенство*

$$\overline{\lim_{n\to\infty}}\, n^\theta \mathsf{M}\{(\Phi_n - v_-)^2\} \leqslant \frac{d}{(1 - \theta/2)^2} \sum_{i \neq \alpha} (v_i - v_-)^2. \qquad (2.55)$$

Это следует из неравенства

$$[V(p_n) - v_-]^2 = \left(\sum_{i \neq \alpha} (v_i - v_-)\, p_n(i)\right)^2 \leqslant \|p_n - p^*\| \sum_{i \neq \alpha} (v_i - v_-)^2$$

и леммы 2.2.

Теорема 2.9 позволяет определить *оптимальные параметры алгоритма* (2.47), т. е. такие величины $\varkappa^*$, ν^*, γ^*, ε^* и Δ^*, которые обеспечивают наибольшую скорость в смысле полученных верхних оценок (2.54), (2.55).

Следствие 2. *Если выполнены условия теоремы* 2.9, *то для параметра* θ, *входящего в соотношения* (2.54) *и* (2.55), *справедлива оценка* $\theta \leqslant 2/3$, *причем* $\theta = 2/3$ *лишь при*

$$\varkappa = \varkappa^* = 1, \quad \nu = \nu^* = 1/3; \qquad (2.56)$$

для этих значений $\varkappa$ *и* ν *минимум* d *в* (2.54), (2.55) *достигается при* $\gamma = \gamma^* = \frac{16}{9\delta_v}$, $\varepsilon = \varepsilon^* = \left(\frac{32m(\Delta^*)}{81N(N-1)\delta_v^2}\right)^{1/3}$, $\Delta = \Delta^* = \frac{1}{N-1}\sum_{i\neq\alpha} v_i$ *и равен*

$$d = d^* = 16N(N-1)\varepsilon^2 = \frac{128}{27}[N(N-1)]^{1/3}\left[\frac{m(\Delta^*)}{\delta_v^2}\right]^{2/3},$$

где

$$m(\Delta^*) = m(0) - (N-1)(\Delta^*)^2 = \sum_{i\neq\alpha}(\sigma_i^2+v_i^2) - \frac{1}{N-1}\left(\sum_{i\neq\alpha} v_i\right)^2.$$

Этот результат нетрудно получить, если учесть, что при выполнении (2.56) величина D в теореме 2.9 является корнем уравнения $\hat{c}(\varkappa)f = d_1 + a\sqrt{f}$ (см. лемму П.2), в котором

$$\hat{c}(\varkappa) = \gamma\delta_v - \frac{2}{3}, \quad a = \frac{2}{3}\varepsilon\sqrt{N(N-1)}, \quad d_1 = m(\Delta)\frac{\gamma^2}{\varepsilon},$$

и провести последовательно минимизацию величин $\sqrt{D}$ и d по γ и ε соответственно путем неявного дифференцирования обеих частей уравнения.

В силу следствий 1 и 2 для алгоритма (2.47) с оптимальными параметрами γ^*, ε^*, Δ^*, $\varkappa^*$ и ν^* справедливы неравенства (2.54), (2.55) с $\theta = 2/3$ и $d = d^*$.

Как видно из следствия 2, для определения оптимальных параметров γ^*, ε^*, Δ^* необходимо знать величины δ_v (2.48), $m(0)$ и $\sum_{i\neq\alpha} v_i$. Такая априорная информация о решаемой задаче, как правило, отсутствует. Однако часто бывают известны границы области принадлежности параметров δ_v, $m_i = m_i(0)$, v_i, т. е.

$$\begin{gathered}\delta_v \geqslant \delta_{\min} > 0, \quad m_i \leqslant m_{\max}, \\ v_{\min} \leqslant v_i \leqslant v_{\max} \quad (i = \overline{1, N}),\end{gathered} \tag{2.57}$$

где $\delta_{\min}$, $m_{\max}$, $v_{\min}$ и $v_{\max}$ — заданные величины, причем предполагается, что $m_{\max} > \max\{v_{\min}^2, v^{\max}\}$. В этом случае можно пользоваться такими значениями параметров γ^0, ε^0, Δ^0, которые гарантируют, что при условиях (2.57) величина d в (2.54), (2.55) не превзойдет некоторой константы d^0, причем за счет выбора других значений γ, ε, δ уменьшить d^0 нельзя. Такие значения параметров γ^0, ε^0, Δ^0 будем называть гарантирующими.

Обозначим через

$$d(\gamma, \varepsilon, \Delta; \delta_v, \mu, v) = \left(\varepsilon \sqrt{N(N-1)} + \right.$$
$$\left. + \frac{\varepsilon \sqrt{N(N-1)} + \sqrt{\varepsilon^2 N(N-1) + 3(3\gamma\delta_v - 2)\, m(\Delta)\, \gamma^2 \varepsilon^{-1}}}{3\gamma\delta_v - 2}\right)^2$$

величину d (2.54а) при $\varkappa = 1$, $\nu = 2/3$ как функцию γ, ε, Δ, δ_v, $\mu \triangleq (m_1, \ldots, m_N)$, $v \triangleq (v_1, \ldots, v_N)$. Допустим сначала, что

$$\text{либо } v_{\min} \geqslant 0, \quad \text{либо } v_{\max} \leqslant 0. \tag{2.58}$$

Тогда нетрудно получить равенство

$$\max_{\delta_v, \mu, v} \min_{\gamma, \varepsilon, \Delta} d(\gamma, \varepsilon, \Delta; \delta_v, \mu, v) = \min_{\gamma, \varepsilon, \Delta} \max_{\delta_v, \mu, v} d(\gamma, \varepsilon, \Delta; \delta_v, \mu, v) \triangleq d^0. \tag{2.59}$$

Здесь минимум берется при условии

$$\gamma > \frac{2}{3\delta_v}, \quad \varepsilon > 0, \tag{2.60}$$

а максимум — при условии (2.57). Внешние максимум и минимум в (2.59) достигаются соответственно при

$$\delta_v = \delta_{\min}, \quad m_i = m_{\max}, \quad v_i = v_{\pm}, \quad i = \overline{1, N}; \tag{2.61}$$

$$\gamma = \gamma^0 = \frac{16}{9\delta_{\min}}, \quad \varepsilon = \varepsilon^0 = \left[\frac{32\left(m_{\max} - v_{\pm}^2\right)}{81 N \delta_{\min}^2}\right]^{1/3}, \tag{2.62}$$

$$\Delta = \Delta^0 = v_{\pm},$$

где

$$v_{\pm} \triangleq \begin{cases} v_{\max} & \text{при } |v_{\max}| < |v_{\min}|, \\ v_{\min} & \text{при } |v_{\max}| \geqslant |v_{\min}|. \end{cases}$$

Таким образом, в случае (2.58) функция $d(\gamma, \varepsilon, \Delta; \delta_v, \mu, v)$ имеет седловую точку при условиях (2.57), (2.60) и, следовательно, формулы (2.62) определяют гарантирующие параметры γ^0, ε^0, Δ^0, а формулы (2.61) —

«наихудшую задачу», для которой

$$d^0 = \min_{\gamma,\varepsilon,\Delta} d(\gamma, \varepsilon, \Delta; \delta_v, \mu, v) = \frac{128\sqrt[3]{6N}}{27}(N-1)\left[\frac{m_{\max} - v_{\pm}^2}{\delta_{\min}^2}\right]^{2/3}. \tag{2.63}$$

Пусть теперь (2.58) не выполняется, т. е.

$$v_{\min} < 0 < v_{\max}. \tag{2.64}$$

Тогда вместо (2.59) получим для любого Δ (с учетом ограничений (2.57) и (2.60))

$$\max_{\delta_v,\mu,v} \min_{\gamma,\varepsilon} d(\gamma, \varepsilon, \Delta; \delta_v, \mu, v) =$$

$$= \min_{\gamma,\varepsilon} \max_{\delta_v,\mu,v} d(\gamma, \varepsilon, \Delta; \delta_v, \mu, v) \triangleq d^0(\Delta), \tag{2.65}$$

причем

$$\min_{\Delta} d^0(\Delta) = \frac{128\sqrt[3]{6N}}{27}(N-1)\left[\frac{m_{\max}}{\delta_{\min}^2}\right]^{2/3} \triangleq d^0 \tag{2.66}$$

и этот минимум достигается при

$$\Delta = \Delta^0 = 0. \tag{2.67}$$

Для этого значения Δ^0 внешние максимум и минимум в (2.65) достигаются при

$$\gamma = \gamma^0 = \frac{16}{9\delta_{\min}^2}, \quad \varepsilon = \varepsilon^0 = \left[\frac{32 m_{\max}}{81 N \delta_{\min}^2}\right]^{1/3}, \tag{2.68}$$

$$\delta_v = \delta_{\min}, \quad m_i = m_{\max}, \quad v_i \in [v_{\min}, v_{\max}], \quad i = \overline{2, N}. \tag{2.69}$$

Таким образом, в случае (2.64) формулы (2.67), (2.68) определяют гарантирующие (в смысле (2.65), (2.66)) параметры γ^0, ε^0, Δ^0, а формулы (2.69) — «наихудшие задачи», для которых $d^0 = \min\limits_{\gamma,\varepsilon,\Delta} d(\gamma, \varepsilon, \Delta; \delta_v, \mu, v)$, где d^0 определяется выражением (2.66).

Скорость сходимости последовательности текущих средних потерь $\{\Phi_n\}$ для алгоритма (2.47) с гарантирующими параметрами и произвольной задачи из класса (2.57) определяется неравенством (2.55) с $\theta = \theta^* = 2/3$ и $d = d^0$ (2.63) в случае (2.58) или (2.66) при условии (2.64).

Таким образом, максимальная асимптотическая скорость сходимости (в смысле оценок (2.54), (2.55)), т. е.

гарантируемая скорость убывания величины $\mathsf{M}\{\|p_n - p^*\|^2\}$, равна $[d^* + o(1)]n^{-2/3}$, где d^* зависит от числа возможных вариантов N, первых двух условных моментов потерь v_i и m_i для всех вариантов кроме оптимального, а также от параметра δ_v, характеризующего «обусловленность задачи». Оптимальные параметры также определяются этими величинами. Интересно отметить, что оптимальный параметр шага зависит лишь от δ_v.

В том случае, когда величины v_i, m_i и δ_v неизвестны, формулы для γ^*, ε^*, Δ^* имеют лишь теоретическое значение, так как использовать их невозможно. Но если известны границы (2.57), в которых лежат эти величины, то можно воспользоваться значениями γ^0, ε^0, Δ^0 параметров алгоритма, которые гарантируют максимальную скорость сходимости для «наихудшей задачи». Эти значения параметров задаются формулами (2.62) в случае (2.58), либо (2.67), (2.68) в случае (2.64), при этом гарантируемая асимптотическая скорость сходимости алгоритма равна $[d^0 + o(1)]n^{-2/3}$, где d^0 определяется либо (2.63), либо (2.66) соответственно.

Рассмотрим теперь вопрос о работоспособности алгоритма (2.47) в отсутствие условия (2.8) о единственности оптимального варианта.

Теорема 2.9А. *Пусть выполнены предположения* П1, П2 *из* § 2.1, $p_1 \in S^N_{\varepsilon_1}$, *а* $\{p_n\}$ — *последовательность векторов, порожденная алгоритмом* (2.47), *в котором числовые последовательности* $\{\gamma_n\}$, $\{\varepsilon_n\}$ *удовлетворяют условиям*

$$0 < \gamma_{n+1} \leqslant \gamma_n, \quad \varepsilon_n \in (0, 1/N) \quad (n = 1, 2, \ldots),$$

$$\lim_{n\to\infty} n\gamma_n = \infty, \quad \lim_{n\to\infty} \left(\varepsilon_n + \gamma_n^{-1} |\varepsilon_n - \varepsilon_{n+1}|\right) = 0,$$

$$\sum_{n=1}^{\infty} \frac{\gamma_n}{n\varepsilon_n} < \infty.$$

Тогда с вероятностью 1 *выполняется целевое условие* (2.4).

Доказательство. Рассмотрим величины W_n и σ-алгебры $\mathcal{F}_n$, введенные при доказательстве теоремы 2.8, где индекс α в (2.49) соответствует одному из возможных оптимальных вариантов, т. е. $v_\alpha = v_-$. Используя соотношение (2.49а), получаем при каждом

$n = 1, 2, \ldots$ с вероятностью 1

$$2\gamma_n \frac{\xi_n - \Delta}{e^{\mathrm{T}}(x_n)\, p_n} e^{\mathrm{T}}(x_n)\left(p_n - p_{n+1}^*\right) \leqslant$$
$$\leqslant W_n - W_{n+1} + C\,|\,\varepsilon_n - \varepsilon_{n+1}\,| + \gamma_n^2 \frac{(\xi_n - \Delta)^2}{\left(e^{\mathrm{T}}(x_n)\, p_n\right)^2},$$

где $C \in (0, \infty)$ — некоторая достаточно большая константа. Разделив на γ_n левую и правую части этого неравенства и вычислив средние арифметические полученных таким образом выражений, приходим после несложных преобразований с учетом П1, П2 к следующему соотношению:

$$\frac{2}{n}\sum_{t=1}^{n}[V(p_t) - v_-] \leqslant \frac{1}{n}\sum_{t=1}^{n} g_t\gamma_t^{-1} + \frac{1}{n}\sum_{t=1}^{n}\gamma_t h_t +$$
$$+ \frac{C}{n}\sum_{t=1}^{n}\left(\varepsilon_{t+1} + \gamma_t^{-1}\,|\,\varepsilon_t - \varepsilon_{t+1}\,|\right), \quad n = 1, 2, \ldots, \tag{2.69a}$$

где

$$g_t \triangleq W_t - W_{t+1} + g_t',$$
$$g_t' \triangleq 2\gamma_t\left[V(p_t) - V\left(p_{t+1}^*\right) - \frac{\xi_t - \Delta}{e^{\mathrm{T}}(x_t)\, p_t} e^{\mathrm{T}}(x_t)\left(p_t - p_{t+1}^*\right)\right],$$
$$h_t \triangleq \left(\frac{\xi_t - \Delta}{e^{\mathrm{T}}(x_t)\, p_t}\right)^2 \geqslant 0.$$

Поскольку

$$\mathsf{M}\,\{g_t' \mid \mathscr{F}_{t-1}\} \overset{\text{п.н.}}{=} 0, \quad t = 1, 2, \ldots,$$

то в условиях данной теоремы

$$\sum_{t=1}^{\infty} \mathsf{M}\,\{(g_t')^2\} < \infty.$$

Отсюда и из равномерной ограниченности величин W_n следует, что с вероятностью 1

$$\sup_n \left|\sum_{t=1}^{n} g_t\right| < \infty,$$

а значит, по лемме П.6 первое слагаемое правой части неравенства (2.69a) в пределе при $n \to \infty$ стремится к нулю.

Далее, в силу неотрицательности величин $\gamma_t h_t$ и сходимости ряда

$$\sum_{t=1}^{\infty} \frac{1}{t} \mathsf{M}\{\gamma_t h_t\} < \infty,$$

получаем с вероятностью 1

$$\sum_{t=1}^{\infty} \frac{1}{t} \gamma_t h_t < \infty$$

и, следовательно, по лемме П.8 второе слагаемое правой части неравенства (2.69а) стремится к нулю при $n \to \infty$.

Сходимость к нулю третьего слагаемого правой части неравенства (2.69а) вытекает из леммы П.7 и условий данной теоремы.

Итак, доказано, что с вероятностью 1

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \sum_{t=1}^{n} V(p_t) \leqslant v_-.$$

Но, как нетрудно показать, используя лемму П.13,

$$\lim_{n\to\infty} \left(\frac{1}{n} \sum_{t=1}^{n} V(p_t) - \Phi_n \right) \overset{\text{п.н.}}{=} 0,$$

поскольку с вероятностью 1 сходится ряд

$$\sum_{t=1}^{\infty} t^{-2} \mathsf{M}\{(V(p_t) - \xi_t)^2 \mid \mathscr{F}_{t-1}\} < \infty$$

и

$$V(p_t) \overset{\text{п.н.}}{=} \mathsf{M}\{\xi_t \mid \mathscr{F}_{t-1}\}.$$

Таким образом, с вероятностью 1

$$\overline{\lim_{n\to\infty}} \Phi_n \leqslant v_-.$$

Отсюда и из леммы 2.1, п. 2), следует выполнение с вероятностью 1 целевого условия (2.4). Теорема доказана. ▲

Для оценки порядка скорости достижения целевого условия (2.4) при использовании алгоритма (2.47) ограничимся последовательностями $\{\gamma_n\}$, $\{\varepsilon_n\}$ вида

$$\gamma_n = \gamma (n+a)^{-\varkappa}, \quad \varepsilon_n = \varepsilon (n+b)^{-\nu}, \tag{2.69б}$$

где параметры a, b, γ, ε, $\varkappa$, ν таковы, что выполняются условия теоремы 2.9А. В частности, параметры $\varkappa$, ν должны удовлетворять неравенствам

$$0<\nu<\varkappa<1. \tag{2.69в}$$

Теорема 2.9Б. *Пусть последовательности* $\{\gamma_n\}$, $\{\varepsilon_n\}$ *имеют вид* (2.69б) *и выполнены условия теоремы* 2.9А, *а значит, справедливы неравенства* (2.69в). *Тогда с вероятностью* 1 *при любом* $\delta\in(0,\varkappa-\nu)$

$$\Phi_n-v_-\leqslant O^*\left(n^{\delta-\theta}\right),\quad \theta\triangleq\min\{\nu,\varkappa-\nu,1-\varkappa\}.$$

Доказательство. Оценим порядки скорости сходимости к нулю при $n\to\infty$ правой части неравенства (2.69а). В силу леммы П.6 имеем

$$\left|\frac{1}{n}\sum_{t=1}^{n}g_t\gamma_t^{-1}\right|\leqslant 2\left(\sup_n\left|\sum_{t=1}^{n}g_t\right|\right)(n\gamma_n)^{-1}\overset{\text{п.н.}}{=}O^*\left(n^{\varkappa-1}\right),$$

поскольку (см. доказательство теоремы 2.9А)

$$\sup_n\left|\sum_{t=1}^{\infty}g_t\right|\overset{\text{п.н.}}{<}\infty.$$

Второе слагаемое неравенства (2.69а) оценивается аналогичным образом с учетом неотрицательности величин h_t:

$$\left|\frac{1}{n}\sum_{t=1}^{n}\gamma_t h_t\right|\leqslant\frac{1}{n^{1-\varphi}}\sum_{t=1}^{\infty}\frac{\gamma_t h_t}{t^{\varphi}}\overset{\text{п.н.}}{=}O^*\left(n^{\varphi-1}\right),$$

где $\varphi\triangleq1+\nu-\varkappa+\delta\in(1+\nu-\varkappa,1)$.

Присутствующий здесь бесконечный ряд неотрицательных величин сходится почти наверное, что следует из леммы П.9 и сходимости ряда условных математических ожиданий:

$$\sum_{t=1}^{\infty}\gamma_t t^{-\varphi}\mathsf{M}\{h_t\mid\mathcal{F}_{t-1}\}\overset{\text{п.н.}}{\leqslant}C\sum_{t=1}^{\infty}\gamma_t t^{-\varphi}\varepsilon_t^{-1}\leqslant C'\sum_{t=1}^{\infty}t^{-(1+\delta)}<\infty,$$

где C и C' — достаточно большие неотрицательные константы, а σ-алгебры $\mathcal{F}_t$ те же, что и в доказательстве теоремы 2.8.

Для третьего слагаемого правой части (2.69а) из (2.69б), (2.69в) имеем

$$\frac{C}{n}\sum_{t=1}^{n}\left(\varepsilon_{t+1}+\gamma_t^{-1}\,|\,\varepsilon_t-\varepsilon_{t+1}\,|\right)=O^*\left(n^{-\nu}\right).$$

Таким образом, из (2.69а) получаем

$$\frac{1}{n}\sum_{t=1}^{n}V(p_t)-v_-\leqslant O^*\left(n^{\delta-\theta}\right).$$

Но в силу леммы П.13А и условий данной теоремы

$$\frac{1}{n}\sum_{t=1}^{n}V(p_t)-\Phi_n=o\left(n^{\psi-1/2}\right)\qquad\forall\psi>0,$$

а так как из (2.69в) следует, что $\theta<1/2$, то, взяв достаточно малое $\psi>0$, получаем требуемую оценку. Теорема доказана. ▲

Следствие. *В условиях данной теоремы параметр* θ, *определяющий порядок скорости достижения целевого условия* (2.4) *при использовании алгоритма* (2.47), *не превосходит* 1/3, *причем* $\theta=1/3$ *лишь при* $\varkappa=\varkappa^*=2/3$, $\nu=\nu^*=1/3$.

Таким образом, параметры $\varkappa=\varkappa^*$ и $\nu=\nu^*$ являются оптимальными в смысле верхней оценки, даваемой теоремой 2.9Б, и обеспечивают скорость достижения целевого условия (2.4) порядка $O^*(n^{-1/3+\delta})$, где $\delta>0$ — сколь угодно малое число.

Итак, теоремы 2.9А и 2.9Б гарантируют работоспособность алгоритма (2.47) в смысле выполнения целевого условия (2.4) даже в случае неединственности оптимального варианта. При этом, однако, нельзя утверждать, что последовательность векторов p_n условных вероятностей выбора вариантов сходится к пределу. Тем не менее можно показать, что имеет место сходимость $\{p_n\}$ к множеству

$$\{p\,|\,p\in S^N,\quad V(p)=v_-\},$$

которое лишь при единственном оптимальном варианте $x(\alpha)$ состоит из одного вектора e_α^N.

§ 2.7. Алгоритм случайного поиска оптимального варианта

Рассмотрим еще одну процедуру вида (2.12), которая так же, как и алгоритм (2.47), решает задачу адаптивного выбора вариантов (§ 2.1) в случае небинарных потерь:

$$\begin{gathered}\tilde{p}_{n+1} = \pi^N_{\varepsilon_{n+1}}\{\tilde{p}_n - \gamma_n \xi_n q_n\},\\ p_n = \tilde{p}_n + 0{,}5\varepsilon_n\left(I - N^{-1}E\right)q_n,\end{gathered} \tag{2.70}$$

где $\varepsilon_n \in (0,\ N^{-1})$, $\gamma_n > 0$ $(n = 1,\ 2,\ \ldots)$, E — матрица $N \times N$, все элементы которой равны 1, $q_n \triangleq (q_{1n}, \ldots, q_{Nn})$ — случайные векторы с независимыми компонентами, такие, что $q_{in} \in [-1, 1]$, $\mathsf{M}\{q_{in}\} = 0$ при всех $i = \overline{1,N}$, $\mathsf{M}\{q_n q_n^{\mathrm{T}}\} = I\sigma_q^2 > 0$, I — единичная матрица, а $\{q_n\}$ — последовательность независимых случайных векторов. Поиск оптимального вектора $p^* = e_\alpha^N$ на симплексе S^N (2.9) по этому алгоритму осуществляется с помощью пробных шагов длиной $\varepsilon_n/2$ в направлении случайного вектора $(I - N^{-1}E)q_n$ с последующим использованием реализации ξ_n потерь за выбор некоторого варианта x_n в соответствии с условным распределением $p_n = (p_n(1), \ldots, p_n(N))$, т. е.

$$\mathsf{P}\{x_n = x(i) \mid x_1, \xi_1, q_1; \ldots; x_{n-1}, \xi_{n-1}, q_{n-1}\} \triangleq p_n(i),\ i = \overline{1, N}.$$

Идея использования поисковых алгоритмов для решения задачи адаптивного выбора не нова. Так, например, в [130] был предложен алгоритм поискового типа (в несколько отличном от (2.70) виде), причем его свойства изучались лишь в случае бинарных потерь. Покажем, что алгоритм (2.70) сохраняет работоспособность в общем случае небинарных потерь.

Теорема 2.10. *Пусть выполнены предположения* П1, П2 *из* § 2.1, *условия* (2.8), $\tilde{p}_1 \triangleq p_1 \in S^N_{\varepsilon_1}$, *а* $\{p_n\}$ — *последовательность векторов, порожденная алгоритмом* (2.70), *в котором* $\gamma_n > 0$, $\varepsilon_n \in (0,\ N^{-1})$ $(n = 1,\ 2,\ \ldots)$, $\lim\limits_{n\to\infty} \varepsilon_n = 0$, $\sum\limits_{n=1}^{\infty} \gamma_n \varepsilon_n = \infty$. *Тогда*

1) *если*

$$\sum_{n=1}^{\infty}\left(\gamma_n^2 + |\varepsilon_n - \varepsilon_{n+1}|\right) < \infty,$$

то $p_n \to e_\alpha^N$ *с вероятностью* 1 *при* $n \to \infty$;

2) *если при* $n \to \infty$

$$\left[\gamma_n \varepsilon_n^{-1} + (\varepsilon_n - \varepsilon_{n+1})^2 \varepsilon_n^{-1} \gamma_n^{-1}\right] \to 0,$$

то последовательность $\{p_n\}$ *сходится к* e_α^N *в среднеквадратическом.*

Доказательство. Введем σ-алгебры $\mathscr{F}_n \triangleq \sigma(x_1, \xi_1, q_1; \ldots; x_{n-1}, \xi_{n-1}, q_{n-1})$ и величины $W_n \triangleq \|\tilde{p}_n - p_n^*\|^2$, где, как и ранее,

$$p_n^* = \pi_{\varepsilon_n}^N \{p^*\} = (1 - N\varepsilon_n) e_\alpha^N + \varepsilon_n e^N \in S_{\varepsilon_n}^N, \quad p^* \triangleq e_\alpha^N.$$

Из алгоритма (2.70), свойств проектора (2.10) и неравенства треугольника для всех $n = 1, 2, \ldots$ получим

$$W_{n+1} \leqslant W_n - 2\gamma_n \xi_n \langle \tilde{p}_n - p_{n+1}^*, q_n \rangle + \gamma_n^2 \xi_n^2 \|q_n\|^2 + \\ + 2\|p_n^* - p_{n+1}^*\| \sqrt{W_n} + \|p_n^* - p_{n+1}^*\|^2.$$

Поскольку $\|p_n^* - p_{n+1}^*\| \leqslant N|\varepsilon_n - \varepsilon_{n+1}|$ и $\|q_n\|^2 \leqslant N$, то

$$W_{n+1} \leqslant W_n - 2\gamma_n \xi_n \langle \tilde{p}_n - p_n^*, q_n \rangle + 2\gamma_n |\varepsilon_n - \varepsilon_{n+1}| \sqrt{N^3} |\xi_n| + \\ + \gamma_n^2 N \xi_n^2 + 2N|\varepsilon_n - \varepsilon_{n+1}| \sqrt{W_n} + N^2 (\varepsilon_n - \varepsilon_{n+1})^2. \quad (2.71)$$

Заметим, что по неравенству Иенсена [147] и в силу предположений П1, П2

$$\mathsf{M}^2\{|\xi_n| \mid \mathscr{F}_n\} \leqslant \mathsf{M}\{\xi_n^2 \mid \mathscr{F}_n\} = \\ = \sum_{i=1}^N p_n(i) \mathsf{M}\{\xi_n^2 \mid x_n = x(i)\} \leqslant \sum_{i=1}^N p_n(i)(\sigma_i^2 + v_i^2) \leqslant \\ \leqslant \max_{i=\overline{1,N}} (\sigma_i^2 + v_i^2) \triangleq m, \quad (2.72)$$

$$\mathsf{M}\{\xi_n \langle \tilde{p}_n - p_n^*, q_n \rangle \mid \mathscr{F}_n\} = \langle \tilde{p}_n - p_n^*, \mathsf{M}\{\xi_n q_n \mid \mathscr{F}_n\} \rangle = \\ = \left\langle \tilde{p}_n - p_n^*, \mathsf{M}\left\{\sum_{i=1}^N p_n(i) v_i q_n \mid \mathscr{F}_n\right\}\right\rangle = \\ = \left\langle \tilde{p}_n - p_n^*, \mathsf{M}\left\{v^{\mathsf{T}}\left(\tilde{p}_n + \frac{\varepsilon_n}{2}(I - N^{-1}E) q_n\right) q_n \mid \mathscr{F}_n\right\}\right\rangle = \\ = \frac{\varepsilon_n}{2} \sigma_q^2 \langle \tilde{p}_n - p_n^*, v - N^{-1} e^N v^{\mathsf{T}} e^N \rangle = \frac{\varepsilon_n}{2} \sigma_q^2 [V(\tilde{p}_n) - V(p_n^*)]. \\ (2.73)$$

Усреднив (2.71) при условии $\mathscr{F}_n$ и использовав в получившемся неравенстве соотношения (2.72), (2.73), будем иметь

$$\mathsf{M}\{W_{n+1}\,|\,\mathscr{F}_n\} \leqslant W_n(1-\gamma_n\varepsilon_n C_0) + 2N\,|\,\varepsilon_n-\varepsilon_{n+1}|\sqrt{W_n}+\beta_n, \tag{2.74}$$

где

$$C_0 \triangleq \sigma_q^2\delta_v,$$
$$\beta_n \triangleq \gamma_n^2 Nm + 2\gamma_n\sqrt{N^3m}\,|\,\varepsilon_n-\varepsilon_{n+1}\,| + N^2(\varepsilon_n-\varepsilon_{n+1})^2.$$

Здесь, как и ранее, учтено, что

$$V(\tilde{p}_n) - V(p_n^*) \geqslant \frac{1}{2}\,\delta_v W_n, \quad \delta_v \triangleq v^- - v_- > 0.$$

1) Сходимость с вероятностью 1 вытекает из (2.74), леммы П.11, условий теоремы и того, что $\sqrt{W_n} \leqslant \sqrt{2}$.

2) Сходимость в среднеквадратическом вытекает из неравенства, получаемого усреднением (2.74) с последующим применением неравенства Иенсена [147]

$$\mathsf{M}\{\sqrt{W_n}\} \leqslant \sqrt{\mathsf{M}\{W_n\}}$$

и леммы П.5 (при $b = d = 0$). Теорема доказана. ▲

Оценим скорость сходимости алгоритма (2.70) в классе последовательностей $\{\gamma_n\}$, $\{\varepsilon_n\}$ вида (2.52). В этом случае в соответствии с теоремой 2.10 сходимость имеет место:

1) с вероятностью 1, если $\varkappa > 1/2$, $\nu > 0$, $\varkappa + \nu \leqslant 1$;

2) в среднеквадратическом, если $0 < \nu < \varkappa$, $\varkappa + \nu \leqslant 1$.

Из (2.74) и леммы П.2 получаем

$$\varlimsup_{n\to\infty} n^\theta \mathsf{M}\{\|p_n - p^*\|^2\} < \infty,$$

где $\theta \triangleq \min\{2\nu, \tau\}$, $\tau \triangleq \min\{\varkappa-\nu, 2+\nu-\varkappa\} = \varkappa - \nu$. Значение $\max\limits_{\varkappa,\nu}\theta \triangleq \theta^*$ при ограничениях $\varkappa > \nu > 0$, $\varkappa + \nu \leqslant 1$, обеспечивающих сходимость в среднеквадратическом, достигается при

$$\varkappa = \varkappa^* \triangleq 3/4, \quad \nu = \nu^* \triangleq 1/4, \tag{2.75}$$

т. е. параметры $\varkappa^*$ и ν^* (2.75) гарантируют максимальную асимптотическую скорость сходимости алгоритма

(2.70), при этом $\theta^* = 1/2$, т. е. порядок скорости сходимости равен $n^{-1/2}$.

Таким образом, поисковый алгоритм (2.70) обладает более низкой скоростью сходимости (предельно возможный порядок которой равен $n^{-1/2}$), чем проекционный алгоритм стохастической аппроксимации (2.47) (гарантирующий порядок сходимости $n^{-2/3}$). В связи с этим далее везде в случае небинарных потерь будем использовать только алгоритмы типа (2.47), получаемые по методу стохастической аппроксимации.

§ 2.8. Иерархический алгоритм без передачи информации нижнему уровню

Рассмотренные до сих пор алгоритмы решения задачи безусловной минимизации средних потерь формируют последовательность стохастических векторов $p_n = (p_n(1), \ldots \ldots, p_n(N)) \in S^N$, на основе которых производится рандомизированный выбор варианта x_n в соответствии с (2.5). Возможно, однако, организовать иерархический выбор вариантов, когда выбор осуществляется в несколько этапов. На первом этапе выделяется некоторое подмножество из множества возможных вариантов, на втором — подмножество из этого подмножества и т. д. Наконец, на последнем этапе выделяется конкретный вариант из подмножества, полученного на предыдущем этапе. Если заранее задаться определенным количеством этапов (уровней иерархии) и зафиксировать подмножества, которые могут выделяться на каждом этапе, а само выделение подмножеств производить случайным образом, то мы придем к рандомизированному иерархическому выбору вариантов, управлять которым можно путем изменения условных вероятностей выбора того или иного подмножества на каждом этапе (уровне иерархии). Алгоритмы изменения этих вероятностей и будем называть *иерархическими алгоритмами* выбора вариантов.

Отметим, что целесообразность иерархического выбора вариантов может обусловливаться наличием априорной информации об иерархической структуре задачи. Это означает, что априори известна такая иерархическая структура выбора вариантов, для которой условные средние потери, соответствующие каждому конкретному варианту, при переходе от одного варианта к другому «внутри» любого подмножества вариантов (выделяемого в соответствии с имеющейся структурой) изменяются значительно меньше, чем при переходе к любому варианту из другого подмножества (на том же уровне иерархии). Как будет показано ниже, учет такой априорной информации в алгоритмах адаптивного выбора вариантов позволяет ускорить их сходимость на начальных шагах процесса адаптации.

Кроме того, организация иерархического выбора вариантов целесообразна также при решении задач большой размерности. В этом случае упрощается практическая реализация процесса выбора.

Рассмотрим для простоты случай двухуровневой иерархии, причем будем считать, что множество возможных вариантов представлено в виде

$$X = \{(x^1(i),\ x^2(j) \mid i = \overline{1, N_1},\ j = \overline{1,\ N_2}\}. \tag{2.76}$$

Вариант, выбранный в момент времени t_n, будем обозначать (x_n^1, x_n^2),а соответствующие потери — $\xi_n = \xi_n(x_n^1, x_n^2, \omega)$. Предположения П1 и П2 из § 2.1 запишутся теперь следующим образом:

П1) для всех $n = 1, 2, \ldots$ совокупности $\{\xi_n(x^1(i), x^2(j), \omega) \mid i = \overline{1, N_1},\ j = \overline{1, N_2}\}$ и $\{\xi_t(x^1(i), x^2(j), \omega), x_k^1, x_k^2 \mid i = \overline{1, N_1}, j = \overline{1, N_2}, t = \overline{1, n-1}, k = \overline{1, n}\}$ независимы;

П2) при любых $i = \overline{1, N_1}$, $j = \overline{1, N_2}$ и всех $n = 1, 2, \ldots$

$$\mathsf{M}\{\xi_n(x^1(i), x^2(j), \omega)\} \triangleq v_{ij},$$

$$\sup_n \mathsf{M}\{(\xi_n(x^1(i), x^2(j), \omega) - v_{ij})^2\} \triangleq \sigma_{ij}^2 < \infty.$$

Как и раньше, будем предполагать также, что выполняется условие, аналогичное (2.28), а именно:

$$v_- \triangleq v_{\alpha\beta} < \min_{(i,j)\neq(\alpha,\beta)} v_{ij}. \tag{2.77}$$

Более того, для простоты будем считать, что для всякого $i = \overline{1, N_1}$

$$v_{i\beta_i} \triangleq \min_j v_{ij} < \min_{j\neq\beta_i} v_{ij}. \tag{2.78}$$

Пусть имеющаяся априорная информация о задаче позволяет сделать вывод о целесообразности двухуровневого иерархического выбора, при котором первоначально выбирается элемент x_n^1 из множества

$$X_1 \triangleq \{x^1(i) \mid i = \overline{1, N_1}\},$$

а затем, учитывая значение x_n^1, выбирается x_n^2 из множества

$$X_2 \triangleq \{x^2(j) \mid j = \overline{1, N_2}\}.$$

Обозначим $\{p_n\}$ и $\{p_n^i\}$ $(i = \overline{1, N_1})$ последовательности стохастических векторов со значениями в симплексах S^{N_1} и S^{N_2} соответственно, порождающие последовательность вариантов $\{(x_n^1, x_n^2)\}$, т. е.

$$\begin{gathered}\mathsf{P}\,\{x_n^1 = x^1(i) \mid x_k^1, x_k^2, \xi_k\,(k = \overline{1, n-1})\} \triangleq p_n(i),\\ \mathsf{P}\,\{x_n^2 = x^2(j) \mid x_n^1 = x^1(i);\, x_k^1, x_k^2, \xi_k\,(k = \overline{1, n-1})\} \triangleq p_n^i(j),\\ i = \overline{1, N_1}, \quad j = \overline{1, N_2}, \quad n = 1, 2, \ldots\end{gathered} \tag{2.79}$$

Тогда условная вероятность выбора варианта $(x^1(i), x^2(j))$ на n-м шаге равна

$$\mathsf{P}\,\{x_n^1 = x^1(i), x_n^2 = x^2(j) \mid x_k^1, x_k^2, \xi_k\,(k = \overline{1, n-1})\} = p_n(i)\, p_n^i(j). \tag{2.80}$$

Из (2.80) и леммы 2.2 непосредственно вытекают накладываемые на последовательности $\{p_n\}$ и $\{p_n^i\}$ $(i = \overline{1, N_1})$ условия, выполнение которых обеспечивает достижение целевого условия (2.4). Они в рассматриваемом случае формулируются следующим образом.

Лемма 2.3. *Пусть выполнены условия* П1 *и* П2 *из данного параграфа и для некоторого* $\tau \in (0, 1)$

$$\overline{\lim_{n\to\infty}}\, n^\tau \mathsf{M}\,\{[V(p_n, p_n^1, \ldots, p_n^{N_1}) - v_-]^2\} \triangleq C < \infty,$$

где функция средних потерь

$$V(p, p^1, \ldots, p^{N_1}) \triangleq \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} v_{ij}\, p(i)\, p^i(j) \tag{2.81}$$

определена для $p \triangleq (p(1), \ldots, p(N_1)) \in S^{N_1}$ *и* $p^i \triangleq (p^i(1), \ldots, p^i(N_2)) \in S^{N_2}$. *Тогда с вероятностью* 1 *выполняется целевое условие* (2.4), *причем*

$$\varlimsup_{n\to\infty} n^{\tau}\mathrm{M}\{(\Phi_n - v_-)^2\} \leqslant C\left(1-\frac{\tau}{2}\right)^{-2},$$

т. е. последовательность текущих средних потерь $\{\Phi_n\}$ *сходится к* v_- *также и в среднеквадратическом смысле со скоростью, не меньшей* $n^{-\tau}$.

Таким образом, учитывая, что $\min V\left(p, p^1, \ldots, p^{N_1}\right) = v_-$, где минимум берется по всем $p \in S^{N_1}$ и $p^i \in S^{N_2}$ $(i = \overline{1, N_1})$, приходим к следующему выводу: для достижения цели (2.4) следует при формировании последовательности составных векторов $\left(p, p_n^1, \ldots, p_n^{N_1}\right)$ стремиться к минимизации функции средних потерь (2.81) на множестве $S \triangleq S^{N_1} \times \underbrace{S^{N_2} \times \ldots \times S^{N_2}}_{N_2}$ (с произвольной степенной скоростью). Это позволяет применить метод стохастической аппроксимации для построения алгоритма адаптивного выбора вариантов, порождающего последовательность $\left\{\left(p_n, p_n^1, \ldots, p_n^{N_1}\right)\right\}$ с такими свойствами.

Нетрудно убедиться, учитывая (2.79), (2.80) и предположения П1 и П2 данного параграфа, что справедливы следующие равенства:

$$\mathrm{M}\left\{\frac{\xi_n}{p_n^{\mathrm{T}} e^{N_1}\left(x_n^1\right)} e^{N_1}\left(x_n^1\right) \mid p_n = p, p_n^1 = p^1, \ldots, p_n^{N_1} = p^{N_1}\right\} = \nabla_p V\left(p, p^1, \ldots, p^{N_1}\right),$$

$$\mathrm{M}\left\{\frac{\xi_n \chi\left(x_n^1 = x^1(i)\right)}{p_n^{i\mathrm{T}} e^{N_2}\left(x_n^2\right)} e^{N_2}\left(x_n^2\right) \mid p_n = p, p_n^1 = p^1, \ldots, p_n^{N_1} = p^{N_1}\right\} = \nabla_{p^i} V\left(p, p^1, \ldots, p^{N_1}\right) \quad (i = \overline{1, N_1}). \tag{2.82}$$

Отсюда, поступая по аналогии с § 2.6, получаем следующий алгоритм:

$$p_{n+1} = \pi_{\varepsilon_{n+1}}^{N_1}\left\{p_n - \gamma_n \frac{\xi_n - \Delta}{p_n^{\mathrm{T}} e^{N_1}\left(x_n^1\right)} e^{N_1}\left(x_n^1\right)\right\}, \tag{2.83}$$

$$p_{n+1}^{i} = \pi_{\varepsilon_{n+1}''}^{N_2} \left\{ p_n^i - \gamma_n'' \frac{\xi_n - \Delta_i}{p_n^{i\tau} e^{N_2}\left(x_n^2\right)} \chi\left(x_n^1 = x^1(i)\right) e^{N_2}\left(x_n^2\right) \right\} \tag{2.83}$$

$$\left(i = \overline{1, N_1}, \quad n = 1, 2, \ldots\right).$$

Здесь Δ, Δ_i — некоторые константы, γ_n' и γ_n'' — длины шагов на первом и втором уровнях иерархии соответственно, π_ε^N — оператор проектирования (2.10) на соответствующий ε-симплекс, $\varepsilon_n' \in \left(0, N_1^{-1}\right)$ и $\varepsilon_n'' \in \left(0, N_2^{-1}\right)$. Отметим, что при каждом $n = 1, 2, \ldots$ на втором уровне иерархии изменению подлежит лишь один из векторов p_n^i, который соответствует выбранному на первом уровне элементу x_n^1 (если последовательность ε_n'' монотонно не возрастает). Следующее утверждение устанавливает достаточные условия сходимости алгоритма (2.83) к точке минимума функции $V\left(p, p^1, \ldots, p^{N_1}\right)$ (2.81) на множестве S.

Теорема 2.11. *Пусть выполнены предположения* П1, П2 *данного параграфа, условия* (2.77), (2.78), $p_1 \in S_{\varepsilon_1'}^{N_1}$ и $p_1^i \in S_{\varepsilon_1''}^{N_2}$ $\left(i = \overline{1, N_1}\right)$, *а последовательность* $\left\{\left(p_n, p_n^1, \ldots, p_n^{N_1}\right)\right\}$ *порождается иерархическим алгоритмом* (2.83), *в котором*

$$\gamma_n', \gamma_n'' > 0; \qquad \varepsilon_n' \in \left(0, N_1^{-1}\right), \ \varepsilon_n'' \in \left(0, N_2^{-1}\right) \quad (n = 1, 2, \ldots);$$

$$\varepsilon_n', \varepsilon_n'' \xrightarrow[n\to\infty]{} 0; \qquad \sum_{n=1}^{\infty} \gamma_n' = \sum_{n=1}^{\infty} \gamma_n'' \varepsilon_n' = \infty.$$

Тогда

1) *если*

$$\sum_{n=1}^{\infty} \left[\frac{(\gamma_n')^2}{\varepsilon_n'} + \frac{(\gamma_n'')^2}{\varepsilon_n''} + \left|\varepsilon_n'' - \varepsilon_{n+1}''\right| + \left|\varepsilon_n' - \varepsilon_n''\right| \right] < \infty,$$

$$\overline{\lim_{n\to\infty}} \frac{\gamma_n'}{\gamma_n'' \varepsilon_n'} < \infty,$$

то с вероятностью 1 $p_n \to e_\alpha^{N_1}$, $p_n^i \to e_{\beta_i}^{N_2}$ $\left(i = \overline{1, N_1}\right)$ *при* $n \to \infty$, *где* α *и* β_i *определены в* (2.77), (2.78);

2) *если*

$$\lim_{n\to\infty}\left(\frac{\gamma_n'}{\varepsilon_n'}+\frac{\gamma_n''}{\varepsilon_n'\varepsilon_n''}+\frac{|\varepsilon_n'-\varepsilon_{n+1}'|}{\gamma_n'}+\frac{|\varepsilon_n''-\varepsilon_{n+1}''|}{\gamma_n''\varepsilon_n'}\right)=0,$$

то при $n\to\infty$ *сходимость* $p_n\to e_\alpha^{N_1}$ *и* $p_n^i\to e_{\beta_i}^{N_2}$ $(i=\overline{1,N_1})$ *имеет место в среднеквадратическом смысле.*

Доказательство аналогично доказательству теоремы 2.8 о сходимости проекционного алгоритма (2.47), поэтому приведем его в краткой форме. Введем σ-алгебры $\mathscr{F}_n\triangleq\sigma\left(x_k^1, x_k^2, \xi_k\,|\,k=\overline{1, n-1}\right)$ и величины $W_n^i\triangleq\|p_n^i-p_n^{i*}\|^2$ $(i=\overline{1,N_1})$, где

$$p_n^{i*}\triangleq\left(1-N_2\varepsilon_n''\right)e_{\beta_i}^{N_2}+\varepsilon_n''e^{N_2}\in S_{\varepsilon_n''}^{N_2} \tag{2.84}$$

— единственная (в силу условий (2.78)) точка минимума функции

$$V_i\left(p^i\right)\triangleq\sum_{j=1}^{N_2}v_{ij}p^i(j) \tag{2.85}$$

на ε-симплексе $S_{\varepsilon_n''}^{N_2}$. Отметим, что в (2.77), (2.78) $\beta_\alpha=\beta$.

В силу алгоритма (2.83), учитывая свойства проектора (2.10), предположения П1 и П2 данного параграфа и соотношения (2.78), (2.79), (2.82) и (2.84), нетрудно для произвольного $i=\overline{1, N_1}$ получить неравенство

$$\begin{aligned}\mathsf{M}\{W_{n+1}^i\,|\,\mathscr{F}_n\}\leqslant W_n^i+C_1\left|\varepsilon_{n+1}''-\varepsilon_n''\right|\sqrt{W_n^i}&+\\ +C_2\left(\varepsilon_n''-\varepsilon_{n+1}''\right)^2+C_3\gamma_n''\left|\varepsilon_n''-\varepsilon_{n+1}''\right|&+\\ +C_4\frac{(\gamma_n'')^2}{\varepsilon_n''}-2\gamma_n''\sum_{j=1}^{N_2}v_{ij}p_n(i)\left[p_n^i(j)-p_{n+1}^{i*}(j)\right]&.\end{aligned} \tag{2.86}$$

Здесь и далее C_k $(k=1, 2, \ldots)$ — положительные константы. Поскольку с учетом (2.85) и $p_n\in S_{\varepsilon_n'}^{N_1}$

$$\sum_{j=1}^{N_2}v_{ij}p_n(i)\left[p_n^i(j)-p_{n+1}^{i*}(j)\right]\geqslant\frac{1}{2}\delta_v^i\varepsilon_n'W_n'-C_5\left|\varepsilon_{n+1}''-\varepsilon_n''\right|, \tag{2.87}$$

где в силу (2.78)

$$\delta_v^i \triangleq \min_{j\neq\beta_i} v_{ij} - v_{i\beta_i} > 0,$$

то, учитывая лемму П.11 и очевидное неравенство $\sqrt{W_n^i} \leqslant \sqrt{2}$, получаем сходимость с вероятностью 1 $p_n^i \to e_{\beta_i}^{N_2}$ $(i=\overline{1, N_1})$ и рекуррентное неравенство

$$\mathsf{M}\{W_{n+1}^i\} \leqslant (1-\gamma_n''\varepsilon_n'\delta_v^i)\,\mathsf{M}\{W_n^i\} + C_1\,|\varepsilon_n'' - \varepsilon_{n+1}''|\sqrt{\mathsf{M}\{W_n^i\}} + \\ + C_6\,|\varepsilon_n' - \varepsilon_{n+1}'| + C_4\,(\gamma_n'')^2\,(\varepsilon_n'')^{-1}, \qquad (2.88)$$

из которого (см. лемму П.5) следует сходимость $p_n^i \xrightarrow[n\to\infty]{} e_{\beta_i}^{N_2}$ $(i=\overline{1, N_1})$ в среднеквадратическом смысле.

Отметим также, что при доказательстве сходимости с вероятностью 1 попутно получаем сходимость почти наверное рядов

$$\sum_{n=1}^{\infty} \gamma_n''\varepsilon_n'\,[V_i(p_n^i) - V_i(p_n^{i*})] < \infty \qquad (i=\overline{1, N_1}). \qquad (2.89)$$

Проведем теперь аналогичные рассуждения для величины

$$W_n \triangleq \|p_n - p_n^*\|^2, \text{ где } p_n^* \triangleq (1-N_1\varepsilon_n')\,e_\alpha^{N_1} + \varepsilon_n' e^{N_1} \in S_{\varepsilon_n'}^{N_2},$$

в результате чего получим неравенство

$$\mathsf{M}\{W_{n+1} \mid \mathcal{F}_n\} \leqslant W_n + C_1\,|\varepsilon_n' - \varepsilon_{n+1}'|\sqrt{W_n} + C_6\,|\varepsilon_{n+1}' - \varepsilon_n'| + \\ + C_2\,(\gamma_n')^2\,(\varepsilon_n')^{-1} - 2\gamma_n' \sum_{i=1}^{N_1}\sum_{j=1}^{N_2} v_{ij}\,[p_n(i) - p_{n+1}^*(i)]\,p_n^i(j). \qquad (2.90)$$

Из свойств функции (2.81) следует

$$\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} v_{ij}\,[p_n(i) - p_{n+1}^*(i)]\,p_n^i(j) \geqslant \\ \geqslant [V(p_n, p_n^1, \ldots, p_n^{N_1}) - V(p_n, p_n^{1*}, \ldots, p_n^{N_1*})] + \\ + \sum_{i=1}^{N_1} [p_n(i) - p_n^*(i)]\,[V_i(p_n^i) - V_i(p_n^{i*})] - C_7\,|\varepsilon_n' - \varepsilon_{n+1}'| \geqslant \\ \geqslant \frac{1}{2}\,\delta_n W_n - \sqrt{W_n}\sum_{i=1}^{N_1} [V_i(p_n^i) - V_i(p_n^{i*})] - C_7\,|\varepsilon_n' - \varepsilon_{n+1}'|,$$

где $\delta_n \triangleq \min_{i \neq \alpha} V_i(p_n^{i*}) - V_\alpha(p_n^{\alpha*}) \xrightarrow[n\to\infty]{} \min_{i\neq\alpha} v_{i\beta_i} - v_{\alpha\beta} > 0$ в силу (2.77), (2.78). С учетом этого из (2.90), используя (2.89), неравенство $\sqrt{W_n} \leqslant \sqrt{2}$ и лемму П.11, получаем сходимость $p_n \to e_\alpha^{N_1}$ с вероятностью 1 и рекуррентное неравенство

$$\mathsf{M}\{W_{n+1}\} \leqslant (1 - \gamma_n'\delta_n)\mathsf{M}\{W_n\} + C_2(\gamma_n')^2(\varepsilon_n')^{-1} + \\ + C_8|\varepsilon_n' - \varepsilon_{n+1}'| + C_9\gamma_n' \sum_{i=1}^{N_1} \sqrt{\mathsf{M}\{W_n^i\}}\sqrt{\mathsf{M}\{W_n\}}, \tag{2.91}$$

из которого по лемме П.5 следует сходимость $p_n \xrightarrow[n\to\infty]{} e_\alpha^{N1}$ в среднеквадратическом смысле. Теорема доказана. ▲

Используем полученные при доказательстве этой теоремы рекуррентные неравенства (2.88), (2.91) для оценки скорости сходимости иерархического алгоритма (2.83). Пусть последовательности $\{\gamma_n'\}$, $\{\gamma_n''\}$ и $\{\varepsilon_n'\}$, $\{\varepsilon_n''\}$ выбраны в соответствии с (2.52); например,

$$\begin{gathered} \gamma_n' = \gamma_1(n+a_1)^{-\varkappa_1}, \quad \gamma_n'' = \gamma_2(n+a_2)^{-\varkappa_2}, \\ \varepsilon_n' = \varepsilon_1(n+b_1)^{-\nu_1}, \quad \varepsilon_n'' = \varepsilon_2(n+b_2)^{-\nu_2}, \end{gathered} \tag{2.92}$$

где константы γ_k, a_k, $\varkappa_k$, ε_k, b_k, ν_k $(k=1, 2)$ таковы, что выполняются условия теоремы 2.11. В частности, для сходимости алгоритма (2.83) с вероятностью 1 следует обеспечить выполнение неравенств

$$\varkappa_2 + \nu_1 \leqslant \varkappa_1 \leqslant 1, \quad 2\varkappa_k - \nu_k > 1, \quad \nu_k > 0, \quad k = 1, 2, \tag{2.93}$$

а для сходимости в среднеквадратическом — неравенств

$$0 < \nu_1 < \varkappa_1 \leqslant 1, \quad \nu_1 + \nu_2 < \varkappa_2', \quad \varkappa_2 + \nu_1 \leqslant 1, \quad \nu_2 > 0. \tag{2.94}$$

Теорема 2.12. *Пусть последовательности* $\{\gamma_n'\}$, $\{\gamma_n''\}$, $\{\varepsilon_n'\}$, $\{\varepsilon_n''\}$, *выбранные в виде* (2.92), *удовлетворяют условиям теоремы* 2.11, *обеспечивающим сходимость иерархического алгоритма* (2.83) *в среднеквадратическом смысле, в частности — условиям* (2.94). *Тогда для алгоритма*

(2.83) для достаточно больших γ_1, γ_2

$$\varlimsup_{n\to\infty} n^{\theta}\mathsf{M}\left\{\left\|p_n - e_{\alpha}^{N_1}\right\|^2 + \sum_{i=1}^{N_1}\left\|p_n^i - e_{\beta_i}^{N_2}\right\|^2\right\} < \infty,$$
$$\theta \triangleq \min\{2\nu_1, 2\nu_2; \varkappa_2 - \nu_1 - \nu_2, 2(1 + \nu_2 - \nu_1 - \varkappa_2); \varkappa_1 - \nu_1, 2(1 + \nu_1 - \varkappa_1)\} > 0. \tag{2.95}$$

Доказательство проводится аналогично доказательству теоремы 2.9 (с той лишь разницей, что теперь определяется лишь порядок скорости сходимости) и основывается на использовании рекуррентных неравенств (2.88), (2.91) и леммы П.2.

Следствие 1. *При выполнении условий теоремы* 2.12 *достигается цель* (2.4) *и справедливо неравенство*

$$\varlimsup_{n\to\infty} n^{\theta}\mathsf{M}\{(\Phi_n - v_-)^2\} < \infty. \tag{2.96}$$

Этот результат следует из (2.95), того факта, что функция средних потерь $V(p, p^1, \ldots, p^{N_1})$ имеет константу Липшица, а также из леммы 2.3, в которой следует учесть, что при $p = e_{\alpha}^{N_1}$ и $p^{\alpha} = e_{\beta}^{N_2}$ эта функция достигает минимума на множестве S, равного v_-.

Следствие 2. *При выполнении условий теоремы* 2.12 *имеем в* (2.95), (2.96) $\theta \leqslant 2/5$, *причем* $\theta = 2/5$ *при следующих условиях*:

$$\varkappa_1 = \varkappa_1^* \in [3/5, 1], \quad \varkappa_2 = \varkappa_2^* = 4/5, \quad \nu_1 = \nu_1^* = \nu_2 = \nu_2^* = 1/5.$$

Этот результат нетрудно получить из выражения для θ (2.95), если провести максимизацию θ сначала по $\varkappa_1$, $\varkappa_2$ при фиксированных ν_1, ν_2, а затем и по ν_1, ν_2 (с учетом неравенств (2.94)).

В алгоритме (2.83) учитывается иерархичность структуры задачи посредством организации соответствующей структуры выбора вариантов. Его можно рассматривать и как алгоритм, составленный из нескольких процедур вида (2.47), используемых для перестройки стохастических векторов $p_n, p_n^1, \ldots, p_n^{N_1}$ и организованных в соответствующую иерархическую структуру, причем процедуры одного уровня иерархии не используют значений стохастических векторов другого уровня. В связи с этим алгоритм (2.83) будем называть *иерархическим алгоритмом без передачи информации нижнему уровню.*

§ 2.9. Иерархический алгоритм с передачей информации нижнему уровню

Перейдем теперь к изучению *иерархического алгоритма с передачей информации нижнему уровню*, в котором процедуры перестройки стохастических векторов $p_n^1, \ldots$ $\ldots, p_n^{N_1}$ нижнего уровня используют текущее значение стохастического вектора p_n верхнего уровня. Как будет показано, передача информации нижнему уровню приводит к увеличению скорости сходимости.

Пусть, как и в предыдущем параграфе, множество возможных вариантов имеет вид (2.76), выполняются предположения П1, П2 (§ 2.8) и условия (2.77), (2.78). Сформулированный в лемме 2.3 результат позволил использовать метод стохастической аппроксимации для получения алгоритма (2.83). Заметим, однако, что при этом не учитывалась в полной мере специфика функции средних потерь. В самом деле, из (2.81), (2.85) имеем

$$V\left(p, p^1, \ldots, p^{N_1}\right) = \sum_{i=1}^{N_1} p(i)\, V_i\left(p^i\right)$$

и, следовательно,

$$\min_S V\left(p, p^1, \ldots, p^{N_1}\right) = \min_{p \in S^{N_1}} \sum_{i=1}^{N_1} p(i) \min_{p^i \in S^{N_2}} V_i(p_i),$$

где $S \triangleq S^{N_1} \times \underbrace{S^{N_2} \times \ldots \times S^{N_2}}_{N_1}$. Отсюда видно, что в качестве условного среднего направления изменения стохастического вектора $p_n^i\ (i = \overline{1, N_1})$ на n-м шаге можно взять направление антиградиента функции $V_i(p^i) - \Delta_i$ (по аналогии с § 2.6), вычисленного в точке p_n^i. Нетрудно убедиться, что при сделанных предположениях

$$\mathsf{M}\left\{\frac{\xi_n \chi\left(x_n^1 = x^1(i)\right)}{\left[p_n^{\mathrm{T}} e^{N_1}\left(x_n^1\right)\right]\left[p_n^{i\mathrm{T}} e^{N_2}\left(x_n^2\right)\right]}\, e^{N_2}\left(x_n^2\right) \,\middle|\, p_n = p,\right.$$
$$\left. p_n^1 = p^1, \ldots, p_n^{N_1} = p^{N_1}\right\} = \nabla_{p^i} V_i\left(p^i\right) \qquad \left(i = \overline{1, N_1}\right). \tag{2.97}$$

Отсюда, учитывая также первое соотношение в (2.82), получаем следующий алгоритм:

$$p_{n+1} = \pi^{N_1}_{\varepsilon'_{n+1}} \left\{ p_n - \gamma'_n \frac{\xi_n - \Delta}{p_n^{\mathrm{T}} e^{N_1}(x_n^1)} e^{N_1}(x_n^1) \right\},$$

$$p_{n+1}^i = \pi^{N_2}_{\varepsilon''_{n+1}} \left\{ p_n^i - \gamma''_n \frac{(\xi_n - \Delta_i)\,\chi(x_n^1 = x^1(i))}{[p_n^{\mathrm{T}} e^{N_1}(x_n^1)]\,[p_n^{i\mathrm{T}} e^{N_2}(x_n^2)]} e^{N_2}(x_n^2) \right\} \tag{2.98}$$

$$(i = \overline{1, N_1}, \quad n = 1, 2, \ldots).$$

Здесь константы Δ и Δ_i, последовательности $\{\gamma'_n\}$, $\{\gamma''_n\}$, $\{\varepsilon'_n\}$ и $\{\varepsilon''_n\}$ и оператор проектирования π_ε^N имеют тот же смысл, что и в алгоритме (2.83). Установим условия сходимости алгоритма (2.98) к точке минимума функции $V(p, p^1, \ldots, p^{N_1})$ на множестве S.

Теорема 2.13. *Пусть выполнены предположения* П1, П2 *из* § 2.8 *и условия* (2.77), (2.78). *Пусть также* $p_1 \in S^{N_1}_{\varepsilon'_1}$ *и* $p_1^i \in S^{N_2}_{\varepsilon''_1}$ $(i = \overline{1, N_1})$, *а последовательность* $\{(p_n, p_n^1, \ldots, p_n^{N_1})\}$ *порождена иерархическим алгоритмом* (2.98), *в котором*

$$\gamma'_n, \gamma''_n > 0; \quad \varepsilon'_n \in (0, N_1^{-1}), \quad \varepsilon''_n \in (0, N_2^{-1}), \quad n = 1, 2, \ldots;$$

$$\varepsilon'_n, \varepsilon''_n \xrightarrow[n\to\infty]{} 0; \quad \sum_{n=1}^{\infty} \gamma'_n = \sum_{n=1}^{\infty} \gamma''_n = \infty.$$

Тогда

1) *если*

$$\sum_{n=1}^{\infty} \left[\frac{(\gamma'_n)^2}{\varepsilon'_n} + \frac{(\gamma''_n)^2}{\varepsilon'_n \varepsilon''_n} + |\varepsilon'_n - \varepsilon'_{n+1}| + |\varepsilon''_n - \varepsilon''_{n+1}| \right] < \infty,$$

$$\varlimsup_{n\to\infty} \gamma'_n (\gamma''_n)^{-1} < \infty,$$

то с вероятностью 1 $p_n \to e_\alpha^{N_1}$ *и* $p_n^i \to e_{\beta_i}^{N_2}$ $(i = \overline{1, N_1})$ *при* $n \to \infty$, *где* α *и* β_i *определены в* (2.78);

2) *если*

$$\lim_{n\to\infty} \left(\frac{\gamma'_n}{\varepsilon'_n} + \frac{\gamma''_n}{\varepsilon'_n \varepsilon''_n} + \frac{|\varepsilon'_n - \varepsilon'_{n+1}|}{\gamma'_n} + \frac{|\varepsilon''_n - \varepsilon''_{n+1}|}{\gamma''_n} \right) = 0,$$

то сходимость $p_n \to e_\alpha^{N_1}$ *и* $p_n^i \to e_{\beta_i}^{N_2}$ $(i = \overline{1, N_1})$ *имеет место в среднеквадратическом смысле.*

Доказательство полностью аналогично доказательству теоремы 2.11. Условимся через C_k ($k=1, 2, \ldots$) обозначать положительные константы. Введем σ-алгебры $\mathcal{F}_n$ и величины W_n^i ($i=\overline{1, N_1}$) и W_n так же, как они вводились при доказательстве теоремы 2.11. Тогда вместо (2.86) получим для алгоритма (2.98)

$$\mathsf{M}\{W_{n+1}^i \mid \mathcal{F}_n\} \leqslant W_n^i + C_1 |\varepsilon_{n+1}'' - \varepsilon_n''| \sqrt{W_n^i} + C_2 |\varepsilon_{n+1}'' - \varepsilon_n''| + \\ + C_3 \gamma_n'' |\varepsilon_{n+1}'' - \varepsilon_n''| + C_3 (\gamma_n'')^2 (\varepsilon_n' \varepsilon_n'')^{-1} - \\ - 2\gamma_n'' \sum_{j=1}^{N_2} v_{ij} \left[p_n^i(j) - p_{n+1}^{i*}(j)\right]. \quad (2.99)$$

Но

$$\sum_{j=1}^{N_2} v_{ij} \left[p_n^i(j) - p_{n+1}^{i*}(j)\right] \geqslant \frac{1}{2} \delta_v^i W_n^i - C_5 |\varepsilon_{n+1}'' - \varepsilon_n''|, \quad (2.100)$$

где δ_v^i имеет ту же величину, что и в (2.87). Поэтому из (2.99), (2.100), учитывая лемму П.11 и неравенство $\sqrt{W_n^i} \leqslant \sqrt{2}$, получаем сходимость $p_n^i \to e_{\beta_i}^{N_2}$ ($i=\overline{1, N_1}$) с вероятностью 1 и рекуррентное неравенство

$$\mathsf{M}\{W_{n+1}^i\} \leqslant (1 - \delta_v^i \gamma_n'') \mathsf{M}\{W_n^i\} + C_1 |\varepsilon_{n+1}'' - \varepsilon_n''| \sqrt{\mathsf{M}\{W_n^i\}} + \\ + C_6 |\varepsilon_{n+1}'' - \varepsilon_n''| + C_4 (\gamma_n'')^2 (\varepsilon_n' \varepsilon_n'')^{-1}, \quad (2.101)$$

из которого в силу леммы П.5 следует сходимость в среднеквадратическом. Кроме того, с вероятностью 1 сходятся ряды

$$\sum_{n=1}^{\infty} \gamma_n'' \left[V_i(p_n^i) - V_i(p_n^{i*})\right] < \infty \qquad (i = \overline{1, N_1}), \quad (2.102)$$

где функция $V_i(p^i)$ определена в (2.85).

Сходимость $p_n \to e_{\alpha}^{N_1}$ с вероятностью 1 и в среднеквадратическом доказывается в точности так же, как это делалось при доказательстве сходимости алгоритма (2.83) (теорема 2.11), с той лишь разницей, что вместо (2.89) используется (2.102). При этом также остается справедливым рекуррентное неравенство (2.91). Теорема доказана. ▲

Рассмотрим теперь такие последовательности $\{\gamma'_n\}$, $\{\gamma''_n\}$ и $\{\varepsilon'_n\}$, $\{\varepsilon''_n\}$, которые удовлетворяют условиям (2.52), например последовательности (2.92), где константы γ_k, a_k, $\varkappa_k$, ε_k, b_k, ν_k $(k=1, 2)$ таковы, что выполняются условия теоремы 2.13. В частности, для сходимости с вероятностью 1 алгоритма (2.98) должны выполняться неравенства

$$\varkappa_2 \leqslant \varkappa_1 \leqslant 1, \quad 0 < \nu_1 < 2\varkappa_1 - 1, \quad 0 < \nu_2 < 2\varkappa_2 - \nu_1 - 1, \tag{2.103}$$

а для сходимости в среднеквадратическом — неравенства

$$\varkappa_2 \leqslant 1, \quad 0 < \nu_1 < \varkappa_1 \leqslant 1, \quad 0 < \nu_2 < \varkappa_2 - \nu_1. \tag{2.104}$$

Используя рекуррентные неравенства (2.91), (2.101) и лемму П.2, нетрудно придти к следующему утверждению.

Теорема 2.14. *Пусть последовательности* $\{\gamma'_n\}$, $\{\gamma''_n\}$, $\{\varepsilon'_n\}$, $\{\varepsilon''_n\}$ *выбраны в виде* (2.92) *и удовлетворяются условия теоремы* 2.13, *обеспечивающие сходимость иерархического алгоритма* (2.98) *в среднеквадратическом смысле, в частности — условия* (2.104). *Тогда для алгоритма* (2.98) **с достаточно большими** γ_1, γ_2

$$\overline{\lim_{n\to\infty}}\, n^{\theta} \mathsf{M} \left\{ \| p_n - e_{\alpha}^{N_1} \|^2 + \sum_{i=1}^{N_1} \| p_n^i - e_{\beta_i}^{N_2} \|^2 \right\} < \infty, \tag{2.105}$$

$$\theta \triangleq \min\{2\nu_1, 2\nu_2, \varkappa_2 - \nu_1 - \nu_2, 2(1 + \nu_2 - \varkappa_1),$$
$$\varkappa_1 - \nu_1, 2(1 + \nu_1 - \varkappa_1)\} > 0.$$

Мы не приводим доказательства этой теоремы, поскольку оно (как и доказательство теоремы 2.12) полностью аналогично той части доказательства теоремы 2.9, где устанавливается порядок скорости сходимости проекционного алгоритма (2.47).

Следствие 1. *При выполнении условий теоремы* 2.14 *для алгоритма* (2.98) *достигается цель* (2.4) *и справедливо неравенство*

$$\overline{\lim_{n\to\infty}}\, n^{\theta} \mathsf{M} \{(\Phi_n - v_-)^2\} < \infty. \tag{2.106}$$

Этот результат следует из (2.105) точно так же, как для алгоритма (2.83) неравенство (2.96) вытекало из (2.95). Аналогично, и следующее утверждение получается так же, как было получено следствие 2 теоремы 2.12 из § 2.8.

Следствие 2. *При выполнении условий теоремы* 2.14 *имеем в* (2.105), (2.106) $\theta \leqslant 1/2$, *причем* $\theta = 1/2$ *при следующих условиях*:

$$\varkappa_1 = \varkappa_1^* \in [3/4,\ 1],\ \varkappa_2 = \varkappa_2^* = 1,\quad \nu_1 = \nu_1^* = \nu_2 = \nu_2^* = 1/4.$$

Таким образом, полученные в теоремах 2.12, 2.14 оценки скорости сходимости иерархических алгоритмов (2.83), (2.98) показывают, что максимальный порядок скорости сходимости алгоритма с передачей информации нижнему уровню (2.98) больше, чем у алгоритма без передачи информации нижнему уровню (2.83), но меньше, чем у проекционного (неиерархического) алгоритма (2.47).

В том случае, когда условия (2.77), (2.78) единственности оптимального варианта не выполняются, можно (по аналогии с теоремами 2.9А и 2.9Б из § 2.6) указать условия, при которых гарантируется достижение целевого условия (2.4) с некоторой скоростью при использовании алгоритмов (2.83) и (2.98).

Описанный выше подход к формированию проекционных алгоритмов адаптивного выбора вариантов применим и в случае многоуровневой иерархии. Можно показать, что для произвольного числа уровней иерархии L наибольшая гарантируемая скорость сходимости иерархического алгоритма без передачи информации нижнему уровню (типа (2.83)) имеет порядок $n^{-2/(2L+1)}$, а с передачей информации (типа (2.98)) — порядок $n^{-2/(L+2)}$. Следовательно, скорость сходимости падает с ростом L, при этом она повышается при передаче информации нижнему уровню. Однако это касается только скорости сходимости в асимптотике. На конечном интервале времени в зависимости от свойств решаемой задачи более эффективными могут быть как простейшие, так и иерархические алгоритмы. Так, при малом числе возможных вариантов N эффективнее, как правило, оказываются неиерархические автоматные алгоритмы. С ростом же N их эффективность падает, а целесообразность использования иерархических алгоритмов возрастает, особенно для задач с ярко выраженной иерархической структурой.

Для иллюстрации этого рассмотрим тестовую задачу с двухуровневой структурой [70], в которой $N_1 = N_2 = 4$, а величины потерь $\xi_n(x^1(i),\ x^2(j),\ \omega)$ распределены по нормальному закону со средними значениями и дисперсиями, определяемыми следующими матрицами:

$$\|v_{ij}\| = \begin{Vmatrix} -5,1 & -4,1 & -3,1 & -2,1 \\ -2,7 & -1,7 & -0,7 & 0,3 \\ -0,3 & 0,7 & 1,7 & 2,7 \\ 2,1 & 3,1 & 4,1 & 5,1 \end{Vmatrix},$$

$$\|\sigma_{ij}^2\| = \begin{Vmatrix} 0,5 & 1,0 & 2,0 & 3,0 \\ 4,0 & 1,5 & 2,5 & 3,5 \\ 3,5 & 2,5 & 1,5 & 4,0 \\ 3,0 & 2,0 & 1,0 & 0,5 \end{Vmatrix}.$$

Эта задача имеет двухуровневую иерархическую структуру, поскольку изменение j на единицу приводит к изменению v_{ij} на 1,0, а изменение i — к изменению v_{ij} на 2,4.

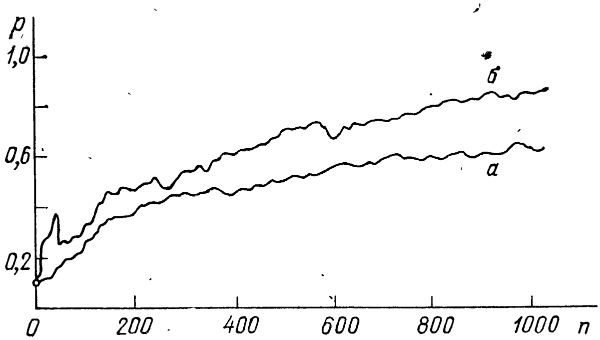

Рис. 10

На рис. 10 показано изменение условной вероятности выбора оптимального варианта $(x^1(1),\ x^2(1))$ для неиерархического алгоритма (2.47) (кривая *а*) и для иерархического алгоритма (2.98) (кривая *б*). Параметры выбирались следующими:

для алгоритма (2.47)

$$\gamma_n = \begin{cases} 0{,}001, & n \leqslant 50, \\ 0{,}001\,[1 + 0{,}01\,(n - 50)]^{-1}, & n > 50, \end{cases}$$

$$\varepsilon_n = \begin{cases} 0{,}04, & n \leqslant 50, \\ 0{,}04\,(n - 50)^{-1/3}, & n > 50, \end{cases}$$

$p_1(i) = 1/16 \qquad (i = \overline{1,16});$

для алгоритма (2.98)

$$4\gamma_n'' = \gamma_n' = \begin{cases} 0{,}01, & n \leqslant 50, \\ 0{,}01\,[1 + 0{,}1\,(n - 50)]^{-1}, & n > 50, \end{cases}$$

$$\varepsilon_n' = \varepsilon_n'' = \begin{cases} 0{,}05, & n \leqslant 50, \\ 0{,}05\,(n - 50)^{-1/4}, & n > 50, \end{cases}$$

$$p_1(i) = p_1^i(j) = 1/4 \quad (i, j = \overline{1,4}).$$

Из рисунка видно, что за 800 — 1000 шагов вероятность выбора оптимального варианта для иерархического алгоритма (2.98) достигла уровня 0,8, в то время как для неиерархического алгоритма (2.47) — лишь 0,6.

§ 2.10. Обсуждение результатов

Задача адаптивного выбора вариантов типа безусловной минимизации средних потерь может решаться различными рекуррентными алгоритмами, исследованными в этой главе.

Для решения задач с бинарными потерями могут быть использованы беспроекционные алгоритмы Нарендры — Шапиро, Льюса, Варшавского — Воронцовой. Наибольшей скоростью сходимости из этих алгоритмов обладает алгоритм Льюса. Однако же его сходимость обеспечивается не на всем классе рассматриваемых задач, причем наиболее широкий класс соответствует алгоритму Варшавского — Воронцовой, который является частной модификацией алгоритма Льюса. Отметим, что никакой другой алгоритм адаптивного выбора не обеспечивает достижения решения со скоростью геометрической прогрессии, как это имеет место для алгоритмов Льюса и Варшавского — Воронцовой. Обычной в этих задачах является значительно более медленная сходимость, имеющая порядок $n^{-\theta}$ $(0 < \theta < 1)$.

Свойства рекуррентных алгоритмов во многом зависят от вида последовательности $\{\gamma_n\}$ длины шага. Так, выбор постоянной длины шага $\gamma_n \equiv \gamma > 0$ в алгоритме Нарендры — Шапиро, как это и предлагалось делать его авторами, оставляет ненулевую вероятность неправильного решения поставленной задачи. Теорема 2.2 дает верхнюю и нижнюю оценки этой вероятности. Если же длину шага γ_n выбрать переменной, сходящейся определенным образом к нулю при $n \to \infty$, то алгоритм Нарендры — Шапиро обеспечит решение исходной задачи со степенной скоростью $n^{-\theta}$ $(0<\theta<1)$ (см. теоремы 2.3, 2.4 и следствие из теоремы 2.3).

Алгоритмы Льюса и Варшавского — Воронцовой порождают оптимальную стратегию при постоянной длине шага $\gamma_n \equiv \gamma \in (0,\ 1)$, если γ достаточно мало (теорема 2.5). Если же γ «слишком близко к единице», то алгоритм с положительной вероятностью не решает поставленную задачу (теорема 2.6). Верхняя граница $\bar{\gamma}$, которую не должна превосходить длина шага γ в соответствии с теоремой 2.5, зависит от параметров решаемой задачи, т. е. определяется априорной информацией. Если такой априорной информации в наличии нет, то выбор $\gamma \in (0,\ 1)$ наугад не может, таким образом, гарантировать с вероятностью 1, что задача будет решена.

Установлено, что алгоритм Буша — Мостеллера, который часто использовался различными авторами для решения задач дискретной оптимизации, поставленную задачу решает лишь в вырожденном случае, когда вероятность штрафа при выборе оптимального варианта равна нулю. В остальных случаях при уменьшении до нуля длины шага алгоритма γ_n удается лишь обеспечить распределение вероятностей выбора вариантов так, чтобы они были обратно пропорциональны величинам v_i соответствующих средних потерь (теорема 2.7).

Для решения задач с небинарными потерями, распределенными на $(-\infty,\ \infty)$, предназначены алгоритмы проекционного типа. Они более сложны в вычислительном отношении по сравнению с алгоритмами беспроекционными, так как для осуществления операции проектирования требуется на каждом шаге решать дополнительную задачу квадратичного программирования. Однако эта задача проста, и ее точное решение достигается за конечное число вычислительных операций. В приложении 2 приведена фортран-программа, реализующая оператор проектирования.

Сходимость исследованных в данной главе проекционных алгоритмов может быть обеспечена для задачи с произвольными значениями средних потерь v_i, причем со степенной скоростью, порядок которой также не зависит от решаемой задачи. Так, для одноуровневого, т. е. неиерахического алгоритма (2.47), в случае неединственности оптимального варианта может быть обеспечена скорость сходимости в среднеквадратическом порядка

$n^{-2/3}$, причем для обеспечения этого нет необходимости в дополнительной априорной информации; следует лишь обеспечить определенную скорость убывания последовательностей $\{\gamma_n\}$, $\{\varepsilon_n\}$, входящих в алгоритм. Показано, что в случае неединственности оптимального варианта алгоритм (2.47) гарантирует с вероятностью 1 достижение целевого условия (2.4) со скоростью порядка $n^{\delta-1/3}$, где δ — сколь угодно малое положительное число. Скорость сходимости иерархических проекционных алгоритмов имеет меньший порядок, который убывает с ростом числа уровней иерархии. Но на конечных шагах такие алгоритмы могут порождать правила выбора, более близкие к оптимальному правилу, чем дает неиерархический алгоритм.

Изучен также алгоритм, основанный на методе случайного поиска. Скорость сходимости в среднеквадратическом, которую может обеспечить этот алгоритм, имеет порядок $n^{-1/2}$, т. е. несколько меньший, чем обеспечивает проекционный неиерархический алгоритм (2.47).

Детально исследован вопрос о выборе параметров алгоритма (2.47), обеспечивающих наибольшую среднеквадратическую скорость сходимости в смысле полученного в теореме 2.9 явного выражения верхней оценки (2.54). Эти значения параметров имеют, однако, лишь теоретическое значение, поскольку для их использования необходимо наличие дополнительной априорной информации о задаче. Эта трудность преодолевается за счет использования «гарантирующих» параметров, для вычисления которых требуется значительно меньший объем априорной информации.

ГЛАВА 3

АЛГОРИТМЫ УСЛОВНОЙ МИНИМИЗАЦИИ

В этой главе дается постановка задачи условной минимизации средних потерь. Обсуждается возможность применения регуляризованных методов штрафных функций и множителей Лагранжа. Приводятся и исследуются алгоритмы адаптивного выбора вариантов, реализующие эти методы.

§ 3.1. Задача условной минимизации средних потерь

Пусть в последовательные моменты времени t_n $(n = 1, 2, \ldots)$ необходимо осуществлять выбор вариантов $x_n \in X \triangleq \{x(1), \ldots, x(N)\}$, а потери на интервале времени (t_n, t_{n+1}) характеризуются несколькими наблюдаемыми случайными величинами $\xi_n^j \triangleq \xi_n^j(x_n, \omega)$ $(j = \overline{0, m})$ (см. сноску на с. 40), каждая из которых удовлетворяет предположениям, аналогичным П1, П2, а именно:

П1′. Для любых $j = \overline{0, m}$ и $n = 1, 2, \ldots$ совокупности

$$\{\xi_n^j(x, \omega) \mid x \in X\}$$

и

$$\{\xi_t^l(x, \omega), x_k \mid x \in X, l = \overline{0, m}, t = \overline{1, n-1}, k = \overline{1, n}\}$$

независимы;

П2′. При любых $j = \overline{0, m}$, $i = \overline{1, N}$ и $n = 1, 2, \ldots$

$$\mathsf{M}\{\xi_n^j(x(i), \omega)\} \triangleq v_i^j,$$

$$\sup_n \mathsf{M}\{(\xi_n^j(x(i), \omega))^2\} < \infty.$$

Величины условных средних потерь v_i^j априори неизвестны.

Совокупность текущих средних потерь

$$\Phi_n^j \triangleq \frac{1}{n}\sum_{t=1}^{n} \xi_t^j, \quad j=\overline{0,m} \tag{3.1}$$

характеризует качество последовательности вариантов $x_1, \ldots, x_n$, выбранных к моменту времени t_n включительно. Поскольку каждая последовательность случайных величин $\{\xi_n^j\}$ удовлетворяет условиям леммы 2.1, то с вероятностью 1 множество предельных точек последовательности $\{\Phi_n^j\}$ содержится в отрезке $\left[\min\limits_{i=\overline{1,N}} v_i^j, \max\limits_{i=\overline{1,N}} v_i^j\right]$. Следовательно, верхний предел текущих средних потерь по каждому показателю $\overline{\lim\limits_{n\to\infty}}\, \Phi_n^j$ конечен с вероятностью 1. Отсюда вытекает корректность следующей задачи адаптивного выбора вариантов:

установить последовательность правил (т. е. стратегию), позволяющую в каждый момент времени t_n $(n = 1, 2, \ldots)$ на основе имеющейся совокупности данных

$$\left(x_1, \xi_1^0, \ldots, \xi_1^m\right), \ldots, \left(x_{n-1}, \xi_{n-1}^0, \ldots, \xi_{n-1}^m\right) \tag{3.2}$$

осуществлять выбор очередного варианта $x_n \in X \triangleq \triangleq \{x(1), \ldots, x(N)\}$ таким образом, чтобы сформированная последовательность вариантов $\{x_n\}$ обеспечила с вероятностью 1 достижение цели

$$\overline{\lim_{n\to\infty}}\, \Phi_n^0 \to \min_{\{x_n\}},$$

$$\overline{\lim_{n\to\infty}}\, \Phi_n^j \leqslant 0, \quad j=\overline{1,m}. \tag{3.3}$$

Здесь, естественно, предполагается, что множество допустимых последовательностей $\{x_n\}$ (удовлетворяющих неравенствам в (3.3)) не пусто. То, что правые части в этих неравенствах равны нулю, очевидно, не ограничивает общность данной постановки.

Докажем утверждение, которое устанавливает связь сформулированной здесь задачи со следующей задачей линейного программирования:

$$\sum_{i=1}^{N} p_i v_i^0 \triangleq V_0(p) \to \min_p,$$

$$\sum_{i=1}^{N} p_i v_i^j \triangleq V_j(p) \leqslant 0, \; j=\overline{1,m}, \quad p \in S^N, \tag{3.4}$$

где, как и ранее,

$$S^N \triangleq \left\{p = (p_1, \ldots, p_N) \,\Big|\, \sum_{i=1}^{N} p_i = 1,\ p_i \geqslant 0\ (i = \overline{1, N})\right\}. \quad (3.5)$$

Лемма 3.1. *Пусть выполнены предположения* П1′ *и* П2′ *данного параграфа. Задача* (3.3) *и задача линейного программирования* (3.4) *разрешимы одновременно, при этом с вероятностью* 1 *минимальное значение* $\overline{\lim\limits_{n\to\infty}}\, \Phi_n^0$ *в задаче* (3.3) *совпадает с минимумом* $V_0(p)$ *в задаче* (3.4).

Доказательство. В силу леммы 2.1 и предположений П1′, П2′ последовательность $\{x_n\}$ тогда и только тогда удовлетворяет неравенствам в (3.3), когда все предельные точки $p \in S^N$ последовательности $\{f_n\}$ векторов $f_n = (f_n(1), \ldots, f_n(N))$ (2.2) удовлетворяют системе неравенств в (3.4). Следовательно, если задача (3.3) разрешима, то разрешима и задача (3.4), причем с вероятностью 1 для всех допустимых в задаче (3.3) последовательностей $\{x_n\}$

$$\overline{\lim_{n\to\infty}}\, \Phi_n^0 \geqslant V_0(p^*), \quad (3.6)$$

где p^* — решение задачи (3.4). Пусть теперь разрешима задача (3.4). Для последовательности независимых случайных величин $\{x_n^*\}$, удовлетворяющих условию $\mathsf{P}\{x_n^* = x(i)\} = p_i^*$ $(i = \overline{1, N},\ n = 1, 2, \ldots)$, в силу закона больших чисел почти для всех ω $\lim\limits_{n\to\infty} \Phi_n^j = V_j(p^*)$ $(j = \overline{0, m})$. Отсюда и из (3.6) вытекает, что последовательность $\{x_n^*\}$ с вероятностью 1 доставляет минимум величине $\overline{\lim\limits_{n\to\infty}}\, \Phi_n^0$ на множестве допустимых в задаче (3.3) последовательностей вариантов. Лемма доказана. ▲

Поскольку условные средние потери v_i^j априори неизвестны, то невозможно определить решение p^* задачи линейного программирования (3.4) и реализовать последовательность $\{x_n^*\}$ вариантов, которая, как показано в доказательстве леммы 3.1, обеспечивает достижение цели (3.3). Тем не менее если, используя текущую информацию (3.2), осуществлять выбор очередного варианта x_n в соответствии с условными вероятностями

$$p_n(i) \triangleq \mathsf{P}\{x_n = x(i) \mid x_t, \xi_t^j\ (t = \overline{1, n-1},\ j = \overline{0, m})\}, \quad i = \overline{1, N} \quad (3.7)$$

которые при $n \to \infty$ достаточно быстро сходятся к соответствующим компонентам вектора p^*, то цель (3.3) также будет достигнута. Это утверждает следующая лемма.

Лемма 3.2. *Пусть выполнены предположения* П1′, П2′ *из настоящего параграфа и существует* $\tau \in (0, 1)$, *для которого*

$$\varlimsup_{n\to\infty} n^\tau \mathsf{M}\{[V_j(p_n) - V_j(p^*)]^2\} \triangleq C_j < \infty \ (j = \overline{0, m}),$$
$$p_n \triangleq (p_n(1), \ldots, p_n(N)). \tag{3.8}$$

Тогда с вероятностью 1 *выполняется целевое условие* (3.3) *и*

$$\varlimsup_{n\to\infty} n^\tau \mathsf{M}\left\{\left[\Phi_n^j - V_j(p^*)\right]^2\right\} \leqslant C_j \left(1 - \frac{\tau}{2}\right)^{-2}, \tag{3.9}$$

т. е. последовательности текущих средних потерь $\{\Phi_n^j\}$ $(j = \overline{0, m})$ *сходятся к* $V_j(p^*)$ (3.4) *также и в среднеквадратическом смысле со скоростью, не меньшей, чем* $n^{-\tau}$.

Доказательство аналогично доказательству леммы 2.2. Пусть $\mathcal{F}_n$ — σ-алгебра, порожденная величинами (3.2) и $d_n^j \triangleq \left[\Phi_n^j - V_j(p^*)\right]^2$ $(j = \overline{0, m}, n=1, 2, \ldots)$. Тогда в силу условий настоящей леммы

$$\mathsf{M}\{d_n^j \mid \mathcal{F}_{n-1}\} \leqslant \left(1 - \frac{1}{n}\right) d_{n-1}^j +$$
$$+ \frac{1}{n}[V_j(p_n) - V_j(p^*)]^2 + \frac{C}{n^2}, \quad C \in (0, \infty) \tag{3.10}$$

и

$$\mathsf{M}\{d_n^j\} \leqslant \left(1 - \frac{1}{n}\right)^2 \mathsf{M}\{d_{n-1}^j\} +$$
$$+ \frac{2}{n}\left(1 - \frac{1}{n}\right)\sqrt{\mathsf{M}\{d_{n-1}^j\}}\, \mathsf{M}^{1/2}\{[V_j(p_n) - V_j(p^*)]^2\} + \frac{C}{n^2}. \tag{3.11}$$

Из (3.10) в силу (3.8) и леммы П.11 вытекает, что $d_n^j \to 0$ с вероятностью 1 при $n \to \infty$. Из (3.11), условий (3.8), П2′ и леммы П.2 (случай в)) получаем (3.9). Лемма доказана. ▲

Таким образом, анализ и синтез алгоритмов формирования векторов $p_n = (p_n(1), \ldots, p_n(N))$ (3.7), обеспечивающих решение задачи условной минимизации средних потерь (3.3), можно осуществлять с позиций решения задачи линейного программирования (3.4) в условиях неопределенности. Будем далее предполагать, что задача (3.4) имеет (возможно, не единственное) решение p^*.

Среди методов решения *задач стохастического программирования*, к которым относится и задача (3.4), выделим метод штрафных функций и метод множителей Лагранжа [91]. Перейдем к построению и исследованию свойств рекуррентных алгоритмов, реализующих эти методы для решения задачи (3.4), а следовательно, и задачи (3.3).

§ 3.2. Регуляризованный метод штрафных функций

Пусть P^* — множество решений задачи (3.4) линейного программирования с ограничениями типа неравенств, причем

$$P^* \subseteq P_D \triangleq \{p \mid V_j(p) \leqslant 0 \ (j = \overline{1, m}), \ p \in S^N\}, \quad (3.12)$$

где P_D — непустое множество векторов p, допустимых указанными ограничениями. Хорошо известно [38], что задача (3.4) с ограничениями типа неравенств может быть сведена к некоторой задаче линейного программирования с ограничениями типа равенств за счет введения дополнительных переменных $u \triangleq (u_1, \ldots, u_m)$:

$$V_0(p) \to \min_{p,u}, \quad V_j(p) + u_j = 0, \quad (3.13)$$

$$u_j \geqslant 0 \ \ (j = \overline{1, m}), \quad p \in S^N.$$

Эквивалентность задач (3.4) и (3.13) очевидна. Применим теперь для решения задачи (3.13) метод штрафных функций в его регуляризованном варианте [92]. С этой целью рассмотрим задачу минимизации функции

$$L_{\mu,\delta}(p, u) \triangleq \mu v^{\mathrm{T}} p + \frac{\delta}{2} \| p \|^2 + \frac{1}{2} \| Vp + u \|^2 \to \min_{p,u} \quad (3.14)$$

на множестве

$$Z_\varepsilon \triangleq \{(p, u) \mid p \in S_\varepsilon^N, \ u \geqslant 0\}, \quad (3.15)$$

где $\mu > 0$ — «штрафной коэффициент», $\delta > 0$ и $\varepsilon \in [0, N^{-1})$ — параметры регуляризации, $V \triangleq \| v_i^j \|_{j=\overline{1,m}, \ i=\overline{1,N}}$ — матрица коэффициентов задачи (3.4) или (3.13) размером

$m \times N$, $v \triangleq (v_1^0, \ldots, v_N^0)$. В (3.15)

$$S_\varepsilon^N \triangleq \left\{ p = (p_1, \ldots, p_N) \,\middle|\, \sum_{i=1}^{N} p_i = 1,\ p_i \geqslant \varepsilon\ (i = \overline{1, N}) \right\}. \quad (3.16)$$

Наиболее часто метод штрафных функций применяется в несколько ином варианте: вместо (3.13) рассматривается задача минимизации штрафной функции $\tilde{L}_k(p, u) \triangleq v^{\mathrm{T}} p + k \| Vp + u \|^2$ при ограничениях $u_j \geqslant 0$ $(j = \overline{1, m})$, $p \in S^N$ и показывается, что при достаточно больших k решения этой и исходной (3.13) задач близки. Если разделить функцию $\tilde{L}_k(p, u)$ на k, то очевидно мы придем к функции типа (3.14) (при $\delta = 0$), для которой при малых коэффициентах $\mu = k^{-1}$ точка минимума также близка к множеству решений исходной задачи (3.13). Параметр $\delta > 0$ обеспечивает сильную выпуклость функции (3.14), тем самым как бы регуляризуя задачу и облегчая процесс нахождения точки ее минимума. Параметр $\varepsilon \in (0, N^{-1})$ в (3.15), как и в § 2.6, также регуляризует задачу.

Установим условия близости решений исходной (3.13) и модифицированной (3.14), (3.15) задач.

Лемма 3.3. *Пусть множество* P_D (3.12) *не пусто и выполнено условие Слейтера по отношению к ограничениям* $V_j(p) \leqslant 0$ $(j = \overline{1, m})$ *в задаче* (3.4), *т. е. существует точка* $\bar{p} \in P_D$, *для которой выполнены строгие неравенства*

$$V_j(\bar{p}) < 0, \quad j = \overline{1, m}.$$

Пусть, кроме того, $\overset{*}{p}_n \triangleq p^*(\mu_n, \delta_n, \varepsilon_n)$, $\overset{*}{u}_n \triangleq u^*(\mu_n, \delta_n, \varepsilon_n)$ — *решение задачи* (3.14), (3.15) *при* $\mu = \mu_n$, $\delta = \delta_n$, $\varepsilon = \varepsilon_n$ $(n = 1, 2, \ldots)$, *а числовые последовательности* $\{\mu_n\}$, $\{\delta_n\}$, $\{\varepsilon_n\}$ *таковы, что*

$$\mu_n > 0, \quad \delta_n > 0, \quad 0 < \varepsilon_n < N^{-1}, \quad n = 1, 2, \ldots$$

и при $n \to \infty$

$$\varepsilon_n \to 0,\ \frac{\delta_n}{\mu_n} \to 0,\ \frac{\mu_n^{3/2}}{\delta_n} \to 0,\ \frac{\varepsilon_n}{\delta_n} \to \zeta \geqslant 0.$$

Тогда $\overset{*}{p}_n \to p^{**}$, $\overset{*}{u}_n \to u^{**}$ *при* $n \to \infty$, *где* $p^{**} \in P^*$ — *решение задачи* (3.4) *с минимальной нормой* ($\|p^{**}\| \leqslant \|p^*\|\ \forall p^* \in P^*$), *а* $u^{**} = -Vp^{**}$.

Доказательство. Функция $L_{\mu,\delta}(p, u)$ (3.14) сильно выпукла, поскольку она является квадратичной и ее гессиан положительно определен:

$$H \triangleq \begin{Vmatrix} \nabla_p^2 L_{\mu,\delta}(p, u) & \nabla_p \nabla_u^{\text{T}} L_{\mu,\delta}(p, u) \\ \nabla_u \nabla_p^{\text{T}} L_{\mu,\delta}(p, u) & \nabla_u^2 L_{\mu,\delta}(p, u) \end{Vmatrix} > 0.$$

Последнее вытекает из теоремы 9.1.6 из [2] в силу положительной определенности следующих четырех матриц:

1) $\nabla_p^2 L_{\mu,\delta}(p, u) = V^{\text{T}}V + \delta E_p > 0,$

2) $\nabla_u^2 L_{\mu,\delta}(p, u) = E_u > 0,$

3) $\nabla_p^2 L_{\mu,\delta}(p, u) - \nabla_p \nabla_u^{\text{T}} L_{\mu,\delta}(p, u) [\nabla_u^2 L_{\mu,\delta}(p, u)]^{-1} \nabla_u \nabla_p^{\text{T}} L_{\mu,\delta}(p, u) = V^{\text{T}}V + \delta E_p - V^{\text{T}} E_u V = \delta E_p > 0,$

4) $\nabla_u^2 L_{\mu,\delta}(p, u) - \nabla_u \nabla_p^{\text{T}} L_{\mu,\delta}(p, u) \left[\nabla_p^2 L_{\mu,\delta}(p, u)\right]^{-1} \nabla_p \nabla_u^{\text{T}} L_{\mu,\delta}(p, u) = E_u - V\left(V^{\text{T}}V + \delta E_p\right)^{-1} V^{\text{T}} = E_u - \delta V V^{\text{T}} \left(E_u + \delta^{-1} V V^{\text{T}}\right)^{-1} = \delta\left(V V^{\text{T}} + \delta E_u\right)^{-1} > 0,$

где E_p, E_u — единичные матрицы $N \times N$ и $m \times m$ соответственно. При установлении соотношения 4) использовано матричное тождество

$$(A + BCB^{\text{T}})^{-1} = A^{-1} - A^{-1}BC(C + CB^{\text{T}}A^{-1}BC^{\text{T}})^{-1}CB^{\text{T}}A^{-1}. \tag{3.17}$$

Поскольку функция $L_{\mu,\delta}(p, u)$ сильно выпукла на замкнутом выпуклом множестве Z_ε (3.15), то оптимальный вектор $z_n^* \triangleq (p_n^*, u_n^*)$ единствен [38] и для любых $z \in Z_{\varepsilon_n}$ справедливо соотношение

$$0 \geqslant \nabla_z^{\text{T}} L_{\mu,\delta}(z_n^*)(z_n^* - z) = \mu_n v^{\text{T}}(p_n^* - p) + \delta_n (p_n^* - p)^{\text{T}} p_n^* + (p_n^* - p)^{\text{T}} V^{\text{T}} (V p_n^* + u_n^*) + (u_n^* - u)^{\text{T}} (V p_n^* + u_n^*). \tag{3.18}$$

После подстановки в (3.18) векторов

$$p \triangleq \varepsilon_n e^N + (1 - N\varepsilon_n) p^*, \; p^* \in P^*, \quad u \triangleq -Vp^*$$

и деления на $\mu_n > 0$, получим

$$0 \geqslant v^{\mathrm{T}}(p_n^* - p^*) + \varepsilon_n v^{\mathrm{T}}(e^N - Np^*) + \\ + \frac{\delta_n}{\mu_n}(p_n^* - p^*)^{\mathrm{T}} p_n^* - \varepsilon_n \frac{\delta_n}{\mu_n}(e^N - Np^*)^{\mathrm{T}} p_n^* + \\ + \frac{1}{\mu_n}\left(\|Vp_n^* + u_n^*\|^2 - \varepsilon_n (Ve^N - NVp^*)^{\mathrm{T}}(Vp_n^* + u_n^*)\right). \quad (3.19)$$

Из (3.18) при $p = p_n^*$ и $u = 0$ следует, что $\{u_n^*\}$ — ограниченная последовательность.

Пусть (p_∞^*, u_∞^*) — произвольный частичный предел последовательности $\{z_n^*\}$. Тогда, переходя в последнем неравенстве к пределу при $n \to \infty$, получим

$$0 \geqslant v^{\mathrm{T}}(p_\infty^* - p^*) + \overline{\lim_{n\to\infty}} \frac{1}{\mu_n}\|Vp_n^* + u_n^*\|^2, \quad (3.20)$$

откуда следует, что $p_\infty^* \in P^*$ и вектор $u_\infty^* = -Vp_\infty^*$ однозначно определяется вектором p_∞^*. Обозначим $\hat{p}_n$ проекцию вектора p_n^* на множество P_D и докажем, что для некоторого $C > 0$

$$\|p_n^* - \hat{p}_n\| \leqslant C\sqrt{\mu_n}, \quad n = 1, 2, \ldots \quad (3.21)$$

Из (3.20) следует, что векторы p_n^* удовлетворяют неравенствам

$$\|Vp_n^* + u_n^*\| \leqslant C_1\sqrt{\mu_n}, \quad C_1 \in (0, \infty),$$

и, следовательно,

$$Vp_n^* \leqslant C_1\sqrt{\mu_n}e^m - u_n^* \leqslant C_1\sqrt{\mu_n}e^m,$$
$$e^m \triangleq (1, \ldots, 1) \in R^m$$

(неравенство между векторами понимается покомпонентно). Поскольку

$$\|p_n^* - \hat{p}_n\|^2 \leqslant \max_{\substack{Vp \leqslant C_1\sqrt{\mu_n}e^m \\ p \in S^N}} \ \min_{\substack{Vq \leqslant 0 \\ q \in S^N}} \|p - q\|^2 \triangleq g(\mu_n),$$

то, сделав замену переменных

$$\tilde{p} \triangleq (1 - \nu_n)p + \nu_n\bar{p}, \quad \nu_n \triangleq \frac{C_1\sqrt{\mu_n}}{C_1\sqrt{\mu_n} - \max\limits_{j=\overline{1,m}} V_j(\bar{p})},$$

(в силу условия Слейтера $\nu_n \in (0, 1)$) и учитывая, что при этой замене множество $\{p \mid Vp \leqslant C_1\sqrt{\mu_n}e^m,\ p \in S^N\}$ отображается в множество $\{\tilde{p} \mid V\tilde{p} \leqslant 0,\ \tilde{p} \in S^N\}$, получим

$$g(\mu_n) \leqslant \max_{\substack{V\tilde{p} \leqslant 0 \\ \tilde{p} \in S^N}} \min_{\substack{Vq \leqslant 0 \\ q \in S^N}} \left\| \frac{1}{1-\nu_n}(\tilde{p} - \nu_n\bar{p}) - q \right\|^2 \leqslant$$

$$\leqslant \max_{\substack{V\tilde{p} \leqslant 0 \\ \tilde{p} \in S^N}} \left\| \frac{1}{1-\nu_n}(\tilde{p} - \nu_n\bar{p}) - \tilde{p} \right\|^2 \leqslant \frac{\nu_n^2}{(1-\nu_n)^2} \max_{\tilde{p}, \bar{p} \in S^N} \|\tilde{p} - \bar{p}\|^2 \leqslant$$

$$\leqslant C_2\nu_n^2, \quad C_2 \in (0, \infty),$$

что и доказывает (3.21).

Установим неравенство

$$0 \geqslant (p_\infty^* - p^*)^{\mathrm{T}} p_\infty^* \quad \forall p^* \in P^*. \tag{3.22}$$

Из (3.19), после деления на δ_n/μ_n, получаем неравенство

$$0 \geqslant \frac{\mu_n}{\delta_n} v^{\mathrm{T}}(p_n^* - \hat{p}) + \frac{\mu_n}{\delta_n} v^{\mathrm{T}}(\hat{p}_n - p^*) + \frac{\varepsilon_n}{\delta_n}\mu_n v^{\mathrm{T}}(e^N - Np^*) +$$
$$+ (p_n^* - p^*)^{\mathrm{T}} p_n^* - \varepsilon_n (e^N - Np^*)^{\mathrm{T}} p_n^* -$$
$$- \frac{\varepsilon_n}{\delta_n}(Ve^N - NVp^*)^{\mathrm{T}}(Vp_n^* + u_n^*).$$

Отсюда, учитывая соотношение $v^{\mathrm{T}}(\hat{p}_n - p_n^*) = V_0(\hat{p}_n) - V_0(p_n^*) \geqslant 0$ и оценку (3.21), получим после предельного перехода неравенство (3.22).

Поскольку (3.22) представляет собой необходимое и достаточное условие минимума сильно выпуклой функции $\|p^*\|^2$ на замкнутом выпуклом множестве P^*, то из произвольности предельной точки $\overset{*}{p}$ следует существование предела $\lim\limits_{n\to\infty} p_n^*$, имеющего минимальную норму на P^*. Теорема доказана. ▲

Установим теперь оценку скорости изменения решения задачи (3.14)—(3.15) $p_n^* \triangleq p^*(\mu_n, \delta_n, \varepsilon_n)$ и $u_n^* \triangleq u^*(\mu_n, \delta_n, \varepsilon_n)$. Эта оценка потребуется в дальнейшем при изучении свойств алгоритма адаптивного выбора вариантов, реализующего метод штрафных функций.

Лемма 3.4. *Пусть выполнены условия леммы 3.3. Тогда найдутся положительные константы C_i $(i = \overline{1, 4})$ и номер n_0, такие, что для любых n, $t \geqslant n_0$ справедлива* *оценка*

$$\|p_n^* - p_t^*\| + \|u_n^* - u_t^*\| \leqslant C_1 |\mu_n - \mu_t| + C_2 |\delta_n - \delta_t| + \\ + C_3 |\varepsilon_n - \varepsilon_t| + C_4 \left| \frac{\varepsilon_n}{\delta_n} - \frac{\varepsilon_t}{\delta_t} \right| \triangleq \varkappa_{nt}. \quad (3.23)$$

Доказательство. При каждом фиксированном наборе параметров μ, δ, ε решения $p^* = p^*(\mu, \delta, \varepsilon)$ и $u^* = u^*(\mu, \delta, \varepsilon)$, минимизирующие штрафную функцию $L_{\mu,\delta}(p, u)$ (3.14) на Z_ε (3.15), принадлежат одному из множеств:

$$R_1 \triangleq \left\{(p, u) \middle| \sum_{i=1}^{N} p_i = 1;\ p_i > \varepsilon\ (i = \overline{1, N});\ u_j > 0\ (j = \overline{1, m})\right\},$$

$$R_2 \triangleq \left\{(p, u) \middle| \sum_{i=1}^{N} p_i = 1;\ p_1 = \varepsilon;\ p_i > \varepsilon\ (i = \overline{2, N});\ u_j > 0\ (j = \overline{1, m})\right\},$$

. . . .

$$R_{N+1} \triangleq \left\{(p, u) \middle| \sum_{i=1}^{N} p_i = 1;\ p_i > \varepsilon\ (i = \overline{1, N-1});\ p_N = \varepsilon;\ u_j > 0\ (j = \overline{1, m})\right\},$$

$$R_{N+2} \triangleq \left\{(p, u) \middle| \sum_{i=1}^{N} p_i = 1;\ p_i > \varepsilon\ (i = \overline{1, N});\ u_1 = 0;\ u_j > 0\ (j = \overline{2, m})\right\},$$

. . . .

$$R_{N+m+1} \triangleq \left\{(p, u) \middle| \sum_{i=1}^{N} p_i = 1,\ p_i > \varepsilon\ (i = \overline{1, N});\ u_j > 0\ (j = \overline{1, m-1});\ u_m = 0\right\},$$

. . . .

$$R_{2^{N+m}-1} \triangleq \left\{(p, u) \middle| \sum_{i=1}^{N} p_i = 1;\ p_i = \varepsilon\ (i = \overline{2, N});\ u_j = 0\ (j = \overline{1, m})\right\}.$$

Для каждого из $R_l\left(l = \overline{1, 2^{N+m-1}}\right)$ зафиксируем множества индексов:

$$I(R_l) \triangleq \{i \mid p_i = \varepsilon,\ i = \overline{1, N},\ (p, u) \in R_l\},$$
$$J(R_l) \triangleq \{j \mid u_j = 0,\ j = \overline{1, m},\ (p, u) \in R_l\}.$$

Необходимые и достаточные условия оптимальности в задаче (3.14), (3.15) в том случае, когда $(p^*, u^*) \in R_l$, запишутся в виде

$$\mu v_i^0 + \delta p_i^* + \sum_{j=1}^{m} v_i^j \left(\sum_{k=1}^{N} v_k^j p_k^* + u_j^* \right) - \lambda_l = 0, \quad i \overline{\in} I(R_l);$$
$$\sum_{k=1}^{N} v_k^j p_k^* + u_j^* = 0,\ j \overline{\in} J(R_l); \quad \sum_{i=1}^{N} p_i^* = 1,$$

или в раскрытой форме

$$\mu v_i^0 + \delta p_i^* + \sum_{j=1}^{m} v_i^j \left(\sum_{k \overline{\in} I(R_l)} v_k^j p_k^* + \varepsilon \sum_{k \in I(R_l)} v_k^j \right) + \\ + \sum_{j \overline{\in} J(R_l)} v_i^j u_j^* - \lambda_l = 0, \quad i \overline{\in} I(R_l); \tag{3.24}$$
$$\sum_{k \overline{\in} I(R_l)} v_k^j p_k^* + \varepsilon \sum_{k \in I(R_l)} v_k^j + u_j^* = 0,\ j \overline{\in} J(R_l);$$
$$\sum_{i \overline{\in} I(R_l)} p_i^* = 1 - \varepsilon K_l,$$

где K_l — количество элементов множества $I(R_l)$, λ_l — множитель Лагранжа, соответствующий ограничению $\sum\limits_{i=1}^{N} p_i = 1$. Равенства (3.24) можно рассматривать как систему линейных алгебраических уравнений относительно неизвестных p^*, u^*, решая которую, находим, что p_i^* и u_j^* представимы в виде

$$\frac{\sum\limits_{r=0}^{N-K_l} \delta^r (c_r + d_r \mu + f_r \varepsilon)}{\sum\limits_{r=0}^{N-K_l} \delta^r h_r} \triangleq H(\mu, \delta, \varepsilon), \tag{3.25}$$

где c_r, d_r, f_r, h_r — некоторые константы, вообще говоря, различные для разных компонент $\overset{*}{p}_i$, $\overset{*}{u}_j$ и множеств R_l. Система уравнений (3.24) имеет единственное решение при любом $\delta > 0$, поскольку при этом функция $L_{\mu,\delta}(p, u)$ сильно выпукла и задача (3.14), (3.15) имеет единственное решение. Отсюда, в частности, следует, что полиномы по δ, стоящие в знаменателе (3.25), не имеют корней при $\delta > 0$.

Отметим, что начиная с некоторого момента времени n_0, все решения $(\overset{*}{p}_n\ \overset{*}{u}_n)$ задачи (3.14), (3.15) при $\mu = \mu_n$, $\delta = \delta_n$, $\varepsilon = \varepsilon_n$ могут принадлежать лишь таким множествам R_l, для которых полиномы по δ в знаменателе (3.25) могут иметь нулевой корень кратности не более единицы (при этом считается, что выражения (3.25) представляют собой несократимые отношения полиномов). Это следует из того, что последовательности $\{\overset{*}{p}_n\}$, $\{\overset{*}{u}_n\}$ ограничены и имеют предел в силу леммы 3.3, а также из свойств последовательностей $\{\mu_n\}$, $\{\delta_n\}$, $\{\varepsilon_n\}$: если $h_0 = 0$, то $c_0 = d_0 = 0$, так как иначе дробь (3.25) неограниченно возрастает при $n \to \infty$ и $\delta = \delta_n$, $\mu = \mu_n$, $\varepsilon = \varepsilon_n$. Таким образом, отношения полиномов (3.25) с указанными свойствами, начиная с некоторого номера n_0, удовлетворяют неравенству типа (3.23), т. е. для любых $n, t \geqslant n_0$

$$|H(\mu_n, \delta_n, \varepsilon_n) - H(\mu_t, \delta_t, \varepsilon_t)| \leqslant \varkappa_{nt},$$

откуда вытекает утверждение леммы. Лемма доказана. ▲

Следствие. *В условиях леммы* 3.4 *при любом* $n \geqslant n_0$ *справедливы неравенства*

$$\|\overset{*}{p}_n - \overset{*}{p}_{n+1}\| + \|\overset{*}{u}_n - \overset{*}{u}_{n+1}\| \leqslant C\Big(|\mu_n - \mu_{n+1}| + \\ + |\delta_n - \delta_{n+1}| + |\varepsilon_n - \varepsilon_{n+1}| + \left|\frac{\varepsilon_n}{\delta_n} - \frac{\varepsilon_{n+1}}{\delta_{n+1}}\right|\Big) \triangleq \varkappa_n,$$

$$\|\overset{*}{p}_n - p^{**}\| + \|\overset{*}{u}_n - u^{**}\| \leqslant \\ \leqslant C\left(\mu_n + \delta_n + \varepsilon_n + \left|\frac{\varepsilon_n}{\delta_n} - \zeta\right|\right) \triangleq \varkappa_n^0,$$

где C — некоторая положительная константа.

Эти неравенства вытекают из (3.23) при $t = n+1$ и $t \to \infty$ соответственно.

Леммы 3.3, 3.4 позволяют сделать вывод о том, что решение p^{**} исходной задачи (3.4), имеющее минимальную норму, может быть получено с некоторой точностью как решение вспомогательной задачи (3.14), (3.15); при этом точность решения исходной задачи повышается, если уменьшать до нуля соответствующим образом (как предписано в теореме 3.3) параметр штрафной функции μ и два параметра регуляризации δ и ε.

Наличие априорной неопределенности не позволяет определить векторы $p^*(\mu, \delta, \varepsilon)$, $u^*(\mu, \delta, \varepsilon)$ как функции параметров μ, δ, ε. Однако, используя текущую информацию, можно оценивать эти векторы с помощью какого-либо алгоритма, при этом естественно уменьшать значения μ, δ, ε в соответствии с условиями леммы 3.3 с тем, чтобы обеспечить сходимость получаемых оценок к (p^{**}, u^{**}). Рассмотрим один из таких алгоритмов.

§ 3.3. Алгоритм, реализующий метод штрафных функций

Рассмотрим следующий алгоритм адаптивного выбора вариантов:

$$p_{n+1} = \begin{cases} p_n & \text{при } n = 2t, \\ \pi^N_{\varepsilon_{n+1}}\{p_n - \gamma^p_n T_n\} & \text{при } n = 2t-1, \end{cases} \tag{3.26}$$

$$u_{n+1}(j) = \begin{cases} u_n(j) & \text{при } n = 2t, \\ \max\{0; u_n(j) - \gamma^u_n U_n(j)\} & \text{при } n = 2t-1, \end{cases}$$

где $t = 1, 2, \ldots,$ $j = \overline{1, m}$,

$$T_n \triangleq \frac{\mu_n}{2}\left[\xi^0_n e(x_n) + \xi^0_{n-1} e(x_{n-1})\right] + \frac{\delta_n}{2}\left[e(x_n) + e(x_{n-1})\right] + \\ + \frac{1}{2}\sum_{j=1}^{m}\left[\left(\xi^j_n + u_n(j)\right)\frac{\xi^j_{n-1}}{e^{\mathrm{T}}(x_{n-1})\,p_{n-1}}\,e(x_{n-1}) + \right. \\ \left. + \left(\xi^j_{n-1} + u_n(j)\right)\frac{\xi^j_n}{e^{\mathrm{T}}(x_n)\,p_n}\,e(x_n)\right], \tag{3.27}$$

$$U_n(j) \triangleq \frac{1}{2}\left(\xi^j_n + \xi^j_{n-1}\right) + u_n(j), \quad j = \overline{1, m}, \tag{3.28}$$

$\{\gamma^p_n\}$, $\{\gamma^u_n\}$ — неотрицательные числовые последовательности длин шагов. Отметим, что условные средние значения векторов движения T_n и U_n в этом алгоритме являются градиентами $\nabla_p L_{\mu_n,\delta_n}(p_n, u_n)$ и $\nabla_u L_{\mu_n,\delta_n}(p_n, u_n)$

штрафной функции $L_{\mu_n,\delta_n}(p, u)$. В самом деле,

$$\mathsf{M}\{T_n \mid p_n = p, u_n = u\} = \mu_n v + \delta_n p + V^{\mathrm{T}}(Vp + u), \tag{3.29}$$

$$\mathsf{M}\{U_n(j) \mid p_n = p, u_n = u\} = \sum_{i=1}^{N} v_i^j p(i) + u(j), j = \overline{1,m}, \tag{3.30}$$

где (p_n, u_n) — текущие оценки точки (p^*, u^*), являющейся решением задачи (3.13), измеримые относительно σ-алгебры

$$\mathscr{F}_n \triangleq \sigma\left(x_t, \xi_t^j \mid t = \overline{1, n-1}, j = \overline{0, m}\right). \tag{3.31}$$

Особенность алгоритма (3.26) состоит в том, что в нем параллельно с пересчетом вектора p_n, представляющего собой оценку p^{**}, осуществляется также перестройка вспомогательного вектора u_n, который является оценкой u^{**}. Вычисление новых значений этих векторов производится лишь при нечетных значениях $n = 2t - 1$ $(t = 1, 2, \ldots)$ по результатам двух последних наблюдений ξ_n^j, ξ_{n-1}^j $(j = \overline{0, m})$, поскольку матрица V входит в выражение $\nabla_p L_{\mu,\delta}(p, u)$ (3.29) квадратично. При этом варианты x_{n-1} и x_n выбираются на основе одного и того же распределения p_n независимо друг от друга. Перейдем к исследованию основных свойств алгоритма (3.26)—(3.28). Следующая теорема устанавливает достаточные условия сходимости $z_n \triangleq (p_n, u_n)$ к решению $z^{**} \triangleq (p^{**}, u^{**})$ задачи (3.13), в котором p^{**} имеет минимальную норму ($\|p^{**}\| \leqslant \|p^*\|$ $\forall p^* \in P^*$).

Теорема 3.1. *Пусть выполнены предположения* П1′, П2′ *из* § 3.1, *условия леммы* 3.3, $p_1 \in S_{\varepsilon_1}^N$, $u_1(j) \geqslant 0$ $(j = \overline{1, m})$, *а* $\{p_n\}$, $\{u_n\}$ — *последовательности векторов, порожденные алгоритмом* (3.26)—(3.28), *в котором*

$$\gamma_n^p > 0,\ \gamma_n^u > 0\ (n = 1, 2, \ldots),\ \lim_{n\to\infty} \frac{\gamma_{n+1}^u}{\gamma_n^u} = \lim_{n\to\infty} \frac{\gamma_{n+1}^p}{\gamma_n^p} = 1,$$

$$\sum_{n=1}^{\infty} \delta_n \gamma_n^p = \infty, \quad 0 < \varliminf_{n\to\infty} \frac{\gamma_n^u}{\gamma_n^p} \leqslant \varlimsup_{n\to\infty} \frac{\gamma_n^u}{\gamma_n^p} < \infty.$$

Тогда

1) *если*

$$\sum_{n=1}^{\infty}\left[\frac{\gamma_n^u\gamma_n^p}{\varepsilon_n}+\left|\frac{\gamma_{n+1}^p}{\gamma_{n+1}^u}-\frac{\gamma_n^p}{\gamma_n^u}\right|+\frac{\varkappa_n^2}{\gamma_n^p\delta_n}\right]<\infty,$$

$$\varkappa_n \triangleq |\mu_{n+1}-\mu_n|+|\delta_{n+1}-\delta_n|+|\varepsilon_{n+1}-\varepsilon_n|+\left|\frac{\varepsilon_{n+1}}{\delta_{n+1}}-\frac{\varepsilon_n}{\delta_n}\right|,$$

то последовательность векторов $z_n \triangleq (p_n, u_n)$ *с вероятностью* 1 *при* $n\to\infty$ *сходится к вектору* $z^{**}\triangleq(p^{**}, u^{**})$, *т. е.*

$$p_n \xrightarrow[n\to\infty]{\text{п.н.}} p^{**}, \quad u_n \xrightarrow[n\to\infty]{\text{п.н.}} u^{**}=-Vp^{**};$$

2) *если*

$$\left(\frac{\varkappa_n}{\gamma_n^p\delta_n}+\frac{\gamma_n^u}{\varepsilon_n\delta_n}+\frac{1}{\gamma_n^p\delta_n}\left|\frac{\gamma_{n+1}^p}{\gamma_{n+1}^u}-\frac{\gamma_n^p}{\gamma_n^u}\right|\right)\xrightarrow[n\to\infty]{} 0,$$

то последовательность векторов $z_n \triangleq (p_n, u_n)$ *сходится в среднеквадратическом при* $n\to\infty$ *к вектору* $z^{**}\triangleq(p^{**}, u^{**})$, *т. е.*

$$\mathsf{M}\{\|p_n-p^{**}\|^2+\|u_n-u^{**}\|^2\}\xrightarrow[n\to\infty]{} 0.$$

Доказательство. Вводя σ-алгебру событий $\mathcal{F}_n$ в соответствии с (3.31), в силу непрерывности решений задачи (3.14), (3.15) по параметрам μ, ε, δ и предположений данной теоремы, получаем, что для

$$W_n \triangleq \|p_n-p_n^*\|^2+\frac{\gamma_{n-1}^p}{\gamma_{n-1}^u}\|u_n-u_n^*\|^2$$

справедливо неравенство (при $n=2t-1$, $t=1, 2, \ldots$)

$$\mathsf{M}\{W_{n+1}\,|\,\mathcal{F}_n\}\leqslant(\|p_n-p_n^*\|+\|u_n^*-u_{n+1}^*\|)^2-$$
$$-2\gamma_n^p(p_n-p_{n+1}^*)^{\mathrm{T}}\mathsf{M}\{T_n\,|\,\mathcal{F}_n\}+(\gamma_n^p)^2\,\mathsf{M}\{\|T_n\|^2\,|\,\mathcal{F}_n\}+$$
$$+\frac{\gamma_n^p}{\gamma_n^u}(\|u_n-u_n^*\|+\|u_n^*-u_{n+1}^*\|)^2+\gamma_n^p\gamma_n^u\,\mathsf{M}\{\|U_n\|^2\,|\,\mathcal{F}_n\}-$$
$$-2\gamma_n^p(u_n-u_{n+1}^*)^{\mathrm{T}}\mathsf{M}\{U_n\,|\,\mathcal{F}_n\}\leqslant W_n+2\sqrt{W_n}S_{1n}+$$
$$+\bar{S}_{1n}-2\gamma_n^p\rho(p_n, u_n)+S_{2n}(p_n, u_n)+S_{3n}(p_n, u_n), \qquad (3.32)$$

где

$$S_{1n} \triangleq \| p^*_{n+1} - p^*_n \| + \frac{\gamma^p_n}{\gamma^u_n} \sqrt{\frac{\gamma^u_{n-1}}{\gamma^p_{n-1}}} \| u^*_{n+1} - u^*_n \|,$$

$$\overline{S}_{1n} \triangleq \| p^*_{n+1} - p^*_n \|^2 + \frac{\gamma^p_n}{\gamma^u_n} \| u^*_{n+1} - u^*_n \|^2,$$

$$S_{2n}(p_n, u_n) \triangleq 2\gamma^p_n \left[(p^*_{n+1} - p^*_n)^{\mathrm{T}} \mathsf{M}\{T_n \mid \mathcal{F}_n\} + \right.$$
$$\left. + (u^*_{n+1} - u^*_n)^{\mathrm{T}} \mathsf{M}\{U_n \mid \mathcal{F}_n\} \right],$$

$$S_{3n}(p_n, u_n) \triangleq (\gamma^p_n)^2 \mathsf{M}\{\| T_n \|^2 \mid \mathcal{F}_n\} +$$
$$+ \gamma^p_n \gamma^u_n \mathsf{M}\{\| U_n \|^2 \mid \mathcal{F}_n\} + \left| \frac{\gamma^p_n}{\gamma^u_n} - \frac{\gamma^p_{n-1}}{\gamma^u_{n-1}} \right| \| u_n - u^*_n \|^2,$$

$$\rho_n(p_n, u_n) \triangleq (p_n - p^*_n)^{\mathrm{T}} \mathsf{M}\{T_n \mid \mathcal{F}_n\} + (u_n - u^*_n)^{\mathrm{T}} \mathsf{M}\{U_n \mid \mathcal{F}_n\},$$

Оценим сверху входящие в (3.32) последовательности, используя следствие из леммы 3.4*):

$$\| u_n - u^*_n \|^2 \leqslant \frac{\gamma^u_{n-1}}{\gamma^p_{n-1}} W_n,$$

$$S_{1n} \leqslant C\varkappa_n \left(1 + \frac{\gamma^p_n}{\gamma^u_n} \sqrt{\frac{\gamma^u_{n-1}}{\gamma^p_{n-1}}} \right), \quad \overline{S}_{1n} \leqslant C\varkappa^2_n \left(1 + \frac{\gamma^p_n}{\gamma^u_n} \right),$$

$$S_{2n}(p_n, u_n) \leqslant C\gamma^p_n \varkappa_n (1 + \| u_n - u^*_n \|) \leqslant$$
$$\leqslant C\gamma^p_n \varkappa_n \left(1 + \sqrt{\frac{\gamma^u_{n-1}}{\gamma^p_{n-1}} W_n} \right),$$

$$S_{3n}(p_n, u_n) \leqslant C(\gamma^p_n)^2 \frac{\| u_n - u^*_n \|^2 + 1}{\varepsilon_n} + C\gamma^p_n \gamma^u_n (\| u_n - u^*_n \|^2 + 1) +$$
$$+ \left| \frac{\gamma^p_n}{\gamma^u_n} - \frac{\gamma^p_{n-1}}{\gamma^u_{n-1}} \right| \| u_n - u^*_n \|^2 \leqslant C \left[\frac{(\gamma^p_\eta)^2}{\varepsilon_n} + \gamma^p_n \gamma^u_n \right] +$$
$$+ CW_n \left[\frac{(\gamma^p_n)^2 \gamma^u_{n-1}}{\gamma^p_{n-1} \varepsilon_n} + \frac{\gamma^p_n}{\gamma^p_{n-1}} \gamma^u_n \gamma^u_{n-1} + \left| \frac{\gamma^p_n}{\gamma^u_n} - \frac{\gamma^p_{n-1}}{\gamma^u_{n-1}} \right| \frac{\gamma^u_{n-1}}{\gamma^p_{n-1}} \right].$$

---

*) Здесь и далее будем через C обозначать разные положительные константы, величины которых не влияют на конечный результат.

Для получения нижней оценки последовательности величин $\rho_n(p_n, u_n)$ воспользуемся выражениями (3.29), (3.30) условных средних векторов движения T_n и U_n. После простых преобразований получим

$$\rho_n(p_n, u_n) = L_{\mu_n,\delta_n}(p_n, u_n) - L_{\mu_n,\delta_n}(p_n^*, u_n^*) + \\ + \frac{\delta_n}{2}\|p_n - p_n^*\|^2 + \frac{1}{2}\|V[p_n - p_n^*] + (u - u_n^*)\|^2. \quad (3.33)$$

Поскольку функция $L_{\mu_n,\delta_n}(p, u)$ квадратична по обеим переменным p и u, то в силу (3.18) имеем

$$L_{\mu_n,\delta_n}(p_n, u_n) - L_{\mu_n,\delta_n}(p_n^*, u_n^*) \geqslant \frac{1}{2}\begin{pmatrix} p_n - p_n^* \\ u_n - u_n^* \end{pmatrix}^{\mathrm{T}} \times \\ \times \left\|\begin{array}{c|c} \nabla_p^2 L_{\mu_n,\delta_n}(p_n^*, u_n^*) & \nabla_p \nabla_u^{\mathrm{T}} L_{\mu_n,\delta_n}(p_n^*, u_n^*) \\ \hline \Delta_u \Delta_p^{\mathrm{T}} L_{\mu_n,\delta_n}(p_n^*, u_n^*) & \nabla_u^2 L_{\mu_n,\delta_n}(p_n^*, u_n^*) \end{array}\right\| \begin{pmatrix} p_n - p_n^* \\ u_n - u_n^* \end{pmatrix} = \\ = \frac{1}{2}\begin{pmatrix} p_n - p_n^* \\ u_n - u_n^* \end{pmatrix}^{\mathrm{T}} \left\|\begin{array}{c|c} \delta_n E_p + V^{\mathrm{T}} V & V^{\mathrm{T}} \\ \hline V & E_u \end{array}\right\| \begin{pmatrix} p_n - p_n^* \\ u_n - u_n^* \end{pmatrix},$$

где E_p, E_u — единичные матрицы размеров $N \times N$ и $m \times m$ соответственно. Подставив эту оценку в (3.33), приходим к следующим соотношениям:

$$\rho_n(p_n, u_n) \geqslant \begin{pmatrix} p_n - p_n^* \\ u_n - u_n^* \end{pmatrix}^{\mathrm{T}} \left\|\begin{array}{c|c} \delta_n E_p + V^{\mathrm{T}} V & V^{\mathrm{T}} \\ \hline V & E_u \end{array}\right\| \begin{pmatrix} p_n - p_n^* \\ u_n - u_n^* \end{pmatrix} = \\ = \begin{pmatrix} \bar p_n \\ \bar u_n \end{pmatrix}^{\mathrm{T}} \left\|\begin{array}{c|c} E_p & 0 \\ \hline 0 & \sqrt{\gamma_{n-1}^u/\gamma_{n-1}^p}\, E_u \end{array}\right\| \times \left\|\begin{array}{c|c} \delta_n E_p + V^{\mathrm{T}} V & V^{\mathrm{T}} \\ \hline V & E_u \end{array}\right\| \times \\ \times \left\|\begin{array}{c|c} E_p & 0 \\ \hline 0 & \sqrt{\gamma_{n-1}^u/\gamma_{n-1}^p}\, E_u \end{array}\right\| \begin{pmatrix} \bar p_n \\ \bar u_n \end{pmatrix} = \\ = \begin{pmatrix} \bar p_n \\ \bar u_n \end{pmatrix}^{\mathrm{T}} \left\|\begin{array}{c|c} \delta_n E_p + V^{\mathrm{T}} V & \sqrt{\gamma_{n-1}^u/\gamma_{n-1}^p}\, V^{\mathrm{T}} \\ \hline \sqrt{\gamma_{n-1}^u/\gamma_{n-1}^p}\, V & (\gamma_{n-1}^u/\gamma_{n-1}^p)\, E_u \end{array}\right\| \begin{pmatrix} \bar p_n \\ \bar u_n \end{pmatrix} \geqslant \\ \geqslant \frac{\delta_n \gamma_{n-1}^u/\gamma_{n-1}^p}{\mathrm{Spur}\,(V^{\mathrm{T}} V) + \delta_n + \sqrt{\gamma_{n-1}^u/\gamma_{n-1}^p}} (\|\bar p_n\|^2 + \|\bar u_n\|^2) \geqslant \\ \geqslant \frac{1}{2}\rho\delta_n \sqrt{\frac{\gamma_{n-1}^u}{\gamma_{n-1}^p}}\, W_n, \quad (3.34)$$

где $\bar{p}_n \triangleq p_n - p_n^*$, $\bar{u}_n \triangleq (u_n - u_n^*)\sqrt{\gamma_{n-1}^p/\gamma_{n-1}^u}$, ρ — достаточно малая положительная константа. Справедливость соотношений (3.34) устанавливается путем проверки условий положительной определенности блочной матрицы (см. теорему 9.1.6 из [2]) так же, как это делалось при доказательстве леммы 3.3. При этом учитывается, что $0 < \varliminf_{n\to\infty} \gamma_n^p/\gamma_n^p \leqslant \varlimsup_{n\to\infty} \gamma_n^u/\gamma_n^p < \infty$ и $V^{\mathsf{T}}V \leqslant \operatorname{Spur}(V^{\mathsf{T}}V)I$. После подстановки полученных оценок в (3.32) и простых преобразований с использованием условий данной теоремы приходим к следующему неравенству:

$$\mathrm{M}\{W_{n+1} \mid \mathscr{F}_n\} \leqslant W_n\left(1 - \rho\delta_n\sqrt{\gamma_n^u\gamma_n^p} + \bar{\nu}_n\right) + \bar{\beta}_n + \bar{\delta}_n\sqrt{W_n}, \tag{3.35}$$

где

$$\bar{\nu}_n \triangleq C\left(\frac{\gamma_n^u\gamma_n^p}{\varepsilon_n} + (\gamma_n^u)^2 + \left|\frac{\gamma_n^p}{\gamma_n^u} - \frac{\gamma_{n-1}^p}{\gamma_{n-1}^u}\right|\frac{\gamma_{n-1}^u}{\gamma_{n-1}^p}\right),$$

$$\bar{\beta}_n \triangleq C\left(\varkappa_n^2 + \gamma_n^p\varkappa_n + \frac{(\gamma_n^p)^2}{\varepsilon_n} + \gamma_n^p\gamma_n^u\right),$$

$$\bar{\delta}_n \triangleq C\varkappa_n.$$

Оценив последний член в (3.35):

$$\bar{\delta}_n\sqrt{W_n} = 2\left(\frac{\bar{\delta}_n}{2\sqrt{\gamma_n\delta_n}}\right)\left(\sqrt{\gamma_n\delta_nW_n}\right) \leqslant \frac{\bar{\delta}_n^2}{4\gamma_n\delta_n} + \gamma_n\delta_nW_n,$$

окончательно получим для $n = 2t-1$, $t = 1, 2, \ldots$

$$\mathrm{M}\{W_{n+1} \mid \mathscr{F}_n\} \leqslant W_n\left(1 - \rho\delta_n\sqrt{\gamma_n^u\gamma_n^p} + \bar{\nu}_n\right) + \bar{\beta}_n + \frac{\bar{\delta}_n^2}{4\gamma_n\delta_n}. \tag{3.36}$$

Утверждение данной теоремы вытекает из (3.36) и лемм П.11, П.5 (при $b = d = 0$) и 3.3. Теорема доказана. ▲

Приведем теперь условия, которые в силу теоремы 3.1 обеспечивают сходимость алгоритма (3.26)—(3.28) для последовательностей

$$\begin{gathered}\gamma_n^u \triangleq \frac{\gamma^u}{(n+a_u)^{\alpha_u}}, \quad \gamma_n^p \triangleq \frac{\gamma^p}{(n+a_p)^{\alpha_p}}, \\ \varepsilon_n \triangleq \frac{\varepsilon}{(n+b)^{\beta}}, \quad \mu_n \triangleq \frac{\mu\ln n}{(n+c)^{\rho}}, \quad \delta_n \triangleq \frac{\delta}{(n+d)^{\varkappa}}.\end{gathered} \tag{3.37}$$

Сходимость с вероятностью 1 будет иметь место в случае выполнения неравенств

$$
\begin{gathered}
0<\rho \leqslant \varkappa \leqslant \beta, \quad 2\varkappa<3\rho, \quad \alpha_u=\alpha_p \triangleq \alpha, \\
\alpha+\varkappa \leqslant 1, \quad 2\alpha>1+\beta,
\end{gathered} \tag{3.38}
$$

а сходимость в среднеквадратическом — в случае, если

$$
\begin{gathered}
0<\rho \leqslant \varkappa \leqslant \beta, \quad 2\varkappa<3\rho, \quad \alpha_u=\alpha_p \triangleq \alpha>\varkappa, \\
\alpha+\varkappa \leqslant 1, \quad \alpha>\beta+\varkappa.
\end{gathered} \tag{3.39}
$$

Кроме того, константы γ^u, γ^p, μ, ε, δ положительны, ε достаточно мало, а константы a_u, a_p, b, c, d неотрицательны.

Следующая теорема позволяет оценить порядок скорости сходимости алгоритма (3.26) — (3.28) для последовательностей (3.37).

Теорема 3.2. *Пусть выполняются условия теоремы* 3.1 *сходимости в среднеквадратическом алгоритма* (3.26) — (3.28) *с последовательностями* (3.37) (*в частности, условия* (3.39)).

Тогда при достаточно больших γ^u, γ^p

$$
\overline{\lim_{n\to\infty}}\, n^{\theta} \mathsf{M}\{\|p_n-p^{**}\|^2\}<\infty,
$$

где

$$
\begin{gathered}
\theta \triangleq \min\{2\rho, 2(1+\rho-\alpha-\varkappa), \quad \alpha-\varkappa, \tau\}, \\
\tau \triangleq \begin{cases} \min\{2(\beta-\varkappa), 1+\beta-2\varkappa-\alpha\}, & \beta>\varkappa, \\ 1, \quad \beta=\varkappa. \end{cases}
\end{gathered}
$$

Доказательство данной теоремы проводится аналогично доказательству теоремы 2.9 с использованием соотношения (3.35), следствия из леммы 3.3, леммы П.2 (при $r=1/2$) и следующих оценок, справедливых для последовательностей (3.37):

$$
|\mu_{n+1}-\mu_n| \leqslant Cn^{-(1+\rho)}\ln n, \quad |\varepsilon_{n+1}-\varepsilon_n| \leqslant Cn^{-(1+\beta)},
$$

$$
|\delta_{n+1}-\delta_n| \leqslant Cn^{-(1+\varkappa)}, \quad \left|\frac{\varepsilon_n}{\delta_n}-\frac{\varepsilon_{n+1}}{\delta_{n+1}}\right| \leqslant \begin{cases} Cn^{-(1+\beta-\varkappa)}, & \beta>\varkappa, \\ Cn^{-2}, & \beta=\varkappa, \end{cases}
$$

$$\left|\frac{\varepsilon_n}{\delta_n}-\zeta\right|\leqslant\begin{cases}Cn^{-(\beta-\varkappa)}, & \beta>\varkappa,\\ Cn^{-1}, & \beta=\varkappa,\end{cases}$$

$$\left|\frac{\gamma_{n+1}^p}{\gamma_{n+1}^u}-\frac{\gamma_n^p}{\gamma_n^u}\right|\frac{\gamma_n^u}{\gamma_n^p}\leqslant Cn^{-2},$$

Следствие 1. *При выполнении условий теоремы* 3.2 *удовлетворяется целевое условие* (3.3) *с вероятностью* 1 *и справедливы неравенства*

$$\varlimsup_{n\to\infty} n^{\theta}\mathrm{M}\left\{\left[\Phi_n^j-V_j(p^{**})\right]^2\right\}<\infty,\qquad (j=\overline{0,m})$$

Этот результат вытекает из леммы 3.2 и теоремы 3.2.

Теорема 3.2 позволяет также определить оптимальные параметры последовательностей (3.37), при которых максимален порядок θ скорости сходимости алгоритма (3.26)—(3.28).

Следствие 2. *При выполнении условий теоремы* 3.2

$$\theta\leqslant\frac{2}{5}\triangleq\theta^*,\tag{3.40}$$

причем максимальный порядок θ^* *достигается лишь при*

$$\alpha_p=\alpha_u\triangleq\alpha^*=\frac{4}{5},\quad \rho=\beta=\varkappa=\frac{1}{5}.\tag{3.41}$$

Доказательство данного следствия вытекает из выражения для θ в теореме 3.2 и цепочки неравенств

$$\theta\leqslant\min\{2\rho,2(1+\rho-\alpha-\beta),\alpha-2\beta,\alpha-\beta\}\leqslant$$
$$\leqslant\min\{2\beta,2(1-\alpha),\alpha-2\beta\}\leqslant\min\left\{2\beta,\frac{2}{3}(1-2\beta)\right\}\leqslant\frac{2}{5},$$

причем эти неравенства переходят в равенства лишь при выполнении (3.41).

Исследованный в этом параграфе алгоритм метода штрафных функций (3.26)—(3.28) позволяет решать задачи условной минимизации (3.3) со скоростью порядка $n^{-2/5}$. Изменение рандомизированных стратегий выбора p_n в этом алгоритме происходит лишь на нечетных тактах, т. е. по результатам двух актов выбора. Рассмотрим теперь другой подход и соответствующие ему алгоритмы, гарантирующие такой же порядок скорости сходимости и изменяющие правила выбора на каждом такте, т. е. в темпе поступления информации.

§ 3.4. Регуляризованный метод множителей Лагранжа

Известно [33, 38], что задача условной минимизации (3.4) эквивалентна задаче отыскания седловой точки функции Лагранжа

$$L(p, \lambda) \triangleq V_0(p) + \sum_{j=1}^{m} \lambda(j)\, V_j(p) \tag{3.42}$$

на множестве $S^N \times R_+^m$, где $R_+^m = \{\lambda = (\lambda(1), \ldots, \lambda(m)) \mid \lambda(j) \geqslant 0 (j = \overline{1, m})\}$, $\lambda = (\lambda(1), \ldots, \lambda(m))$. Для оптимальности вектора $p^* \in S^N$ (т. е. для того, чтобы этот вектор был решением задачи линейного программирования (3.4)) необходимо и достаточно, чтобы для некоторого вектора множителей Лагранжа $\lambda^* \in R_+^m$ и любых $p \in S^N$ и $\lambda \in R_+^m$ выполнялись неравенства *)

$$L(p^*, \lambda) \leqslant L(p^*, \lambda^*) \leqslant L(p, \lambda^*). \tag{3.43}$$

Непосредственное применение метода стохастической аппроксимации для построения рекуррентной процедуры поиска седловой точки (p^*, λ^*) функции Лагранжа (3.42) (т. е. использование в качестве векторов движения по p и λ реализаций градиентов функции $L(p, \lambda)$ по этим переменным) приводит к неустойчивому алгоритму [21, 39, 73]. Это связано с билинейностью функции Лагранжа (3.42). Качественной иллюстрацией возникающих при этом трудностей может служить следующий наглядный пример: градиентная процедура поиска седловой точки билинейной функции $f(x, y) = xy$, имеющая вид

$$x_n = x_{n-1} - \gamma_n y_{n-1},$$
$$y_n = y_{n-1} + \gamma_n x_{n-1},$$

порождает последовательность $\{x_n, y_n\}$, «удаляющуюся» от седловой точки $(x^*, y^*) = (0, 0)$, так как

$$x_n^2 + y_n^2 = (1 + \gamma_n^2)(x_{n-1}^2 + y_{n-1}^2).$$

*) Это означает, что задача (3.4) эквивалентна игре $\{S^N, R_+^m, L\}$ двух лиц с нулевой суммой, в которой первый игрок выбирает стратегию $p \in S^N$, второй — стратегию $\lambda \in R_+^m$ и первый игрок «платит» второму «сумму» $L(p, \lambda)$ [7, 86]. При этом оптимальные (чистые) стратегии первого игрока образуют множество решений задачи (3.4).

Один из возможных способов преодоления этих трудностей состоит в использовании метода регуляризации, согласно которому исходная задача заменяется последовательностью близких в некотором смысле задач с нужными свойствами [6, 123]. В нашем случае регуляризацию осуществим следующим образом: вместо (3.42) рассмотрим функцию

$$L_\delta(p, \lambda) \triangleq L(p, \lambda) + \frac{\delta}{2}(\|p\|^2 - \|\lambda\|^2), \tag{3.44}$$

зависящую от параметра регуляризации $\delta > 0$, а вместо множества S^N (3.5) — ε-симплекс S_ε^N, $\varepsilon \in [0, N^{-1})$ (2.9)*). Таким образом, регуляризованная функция Лагранжа (3.44) сильно выпукла по $p \in S_\varepsilon^N$ при любом $\lambda \in R_+^m$ и сильно вогнута по $\lambda \in R_+^m$ при любом $p \in S_\varepsilon^N$. Отсюда следует, что $L_\delta(p, \lambda)$ имеет седловую точку $(p^*(\varepsilon, \delta), \lambda^*(\varepsilon, \delta))$ на множестве $S_\varepsilon^N \times R_+^m$, причем единственную [7, 81]. Следующее утверждение устанавливает свойства этой седловой точки как функции параметров регуляризации ε и δ.

Лемма 3.5. *Пусть последовательности* $\{\varepsilon_n\}$, $\{\delta_n\}$ *таковы, что* $\varepsilon_n \in [0, N^{-1})$, $\delta_n > 0$ $(n = 1, 2, \ldots)$, $\lim\limits_{n\to\infty} \delta_n = 0$ *и существует предел* $\lim\limits_{n\to\infty} \varepsilon_n/\delta_n \triangleq \zeta$. *Тогда последовательность* $\{(p^*(\varepsilon_n, \delta_n), \lambda^*(\varepsilon_n, \delta_n))\}$ *сходится при* $n \to \infty$ *к седловой точке* (p^*, λ^*), *определяемой условием* (3.43) *и зависящей, вообще говоря, от* ζ.

Доказательство. Обозначим $p_n^* \triangleq p^*(\varepsilon_n, \delta_n)$, $\lambda_n^* \triangleq \lambda^*(\varepsilon_n, \delta_n)$. При любых $p \in S_{\varepsilon_n}^N$ и $\lambda \in R_+^m$ имеем $L_{\delta_n}(p_n^*, \lambda) \leqslant L_{\delta_n}(p, \lambda_n^*)$. Положим $\lambda = \lambda^*$ и

$$p = p^* - \varepsilon_n (Np^* - e^N) \in S_{\varepsilon_n}^N, \tag{3.45}$$

где $e^N \triangleq (1, 1, \ldots, 1)$. Тогда, учитывая (3.42) — (3.44),

*) Проведенная регуляризация означает, другими словами, переход от игры $\{S^N, R_+^m, L\}$ (см. предыдущую сноску) к регуляризованной игре $\{S_\varepsilon^N, R_+^m, L_\delta\}$ двух лиц с нулевой суммой с платежной функцией (3.44).

получим после элементарных преобразований

$$\|p_n^*\|^2 + \|\lambda_n^*\|^2 \leqslant$$
$$\leqslant \|p^*\|^2 + \|\lambda^*\|^2 - N\frac{2\varepsilon_n}{\delta_n}\left(L(p^*, \lambda_n^*) - L\left(\frac{e^N}{N}, \lambda_n^*\right)\right),$$

откуда следует ограниченность последовательностей $\{p_n^*\}$, $\{\lambda_n^*\}$. Выберем произвольную сходящуюся подпоследовательность $(p_{n_k}^*, \lambda_{n_k}^*) \to (\widehat{p}, \widehat{\lambda})$ при $k \to \infty$. Очевидно, $(\widehat{p}, \widehat{\lambda})$ — седловая точка функции Лагранжа (3.42) на множестве $S^N \times R_+^m$. Таким образом, в случае единственности этой точки лемма справедлива. В общем случае рассмотрим произвольную точку (p^*, λ^*), определяемую условием (3.43). Так как для любых $p \in S_{\varepsilon_n}^N$, $\lambda \in R_+^m$ и $n = 1, 2, \ldots$ справедливо неравенство

$$0 \geqslant \left(\nabla_p L_{\delta_n}(p_n^*, \lambda_n^*), p_n^* - p\right) - \left(\nabla_\lambda L_{\delta_n}(p_n^*, \lambda_n^*), \lambda_n^* - \lambda\right),$$

то, выбрав $\lambda = \lambda^*$ и p в соответствии с формулой (3.45), после несложных преобразований с учетом (3.43) получим

$$0 \geqslant \frac{N\varepsilon_n}{\delta_n}\left[L(p^*, \lambda_n^*) - L\left(\frac{e^N}{N}, \lambda_n^*\right) - L(\widetilde{p}_n, \lambda^*) + L\left(\frac{e^N}{N}, \lambda^*\right)\right] +$$
$$+ (p_n^*, p_n^* - p) + (\lambda_n^*, \lambda_n^* - \lambda),$$

где

$$\widetilde{p}_n \triangleq \frac{p_n^* - \varepsilon_n e^N}{1 - N\varepsilon_n} \in S^N.$$

Переходя к пределу на подпоследовательности $\{n_k\}$ при $k \to \infty$, получаем

$$0 \geqslant N\zeta\left[L\left(\frac{e^N}{N}, \lambda^*\right) - L\left(\frac{e^N}{N}, \widehat{\lambda}\right)\right] + (\widehat{p}, \widehat{p} - p^*) + (\widehat{\lambda}, \widehat{\lambda} - \lambda^*).$$

Отсюда следует, что седловая точка $(\widehat{p}, \widehat{\lambda})$ является единственной точкой минимума квадратичной функции

$$\frac{1}{2}(\|p\|^2 + \|\lambda\|^2) - \zeta N \sum_{j=1}^{m} \lambda(j)\, V_j\left(\frac{e^N}{N}\right) \qquad (3.46)$$

на множестве седловых точек (p^*, λ^*) (3.43). Поскольку это множество является выпуклым и замкнутым [7, 81], а $(\hat{p}, \hat{\lambda})$ — произвольная предельная точка последовательности $\{(p_n^*, \lambda_n^*)\}$, получаем

$$\lim_{n\to\infty} p_n^* = \hat{p}, \quad \lim_{n\to\infty} \lambda_n^* = \hat{\lambda}.$$

Лемма доказана. ▲

Из этой леммы следует, что при достаточно малых $\delta > 0$ и $\varepsilon > 0$ в качестве приближенного решения исходной (нерегуляризованной) задачи можно взять точку $(p^*(\varepsilon, \delta), \lambda^*(\varepsilon, \delta))$, для поиска которой возможно применение методов типа стохастической аппроксимации [21, 79]. Однако в нашем случае определение величин ε и δ, дающих достаточно хорошее (требуемое) приближение седловой точки функции Лагранжа (3.42), принципиально невозможно из-за отсутствия информации о величинах условных средних потерь v_i^j. Поэтому представляют особый интерес такие рекуррентные процедуры, в которых от шага к шагу значения параметров регуляризации ε и δ уменьшались бы, в пределе устремляясь к нулю. При определенных условиях в таких алгоритмах возможна сходимость к седловой точке нерегуляризованной функции Лагранжа (3.42).

Установим еще одно важное свойство седловой точки $(p^*(\varepsilon, \delta), \lambda^*(\varepsilon, \delta))$.

Лемма 3.6. *Пусть выполнены условия леммы* 3.5. *Тогда для некоторой положительной константы C и номера $n_0 \geqslant 1$ при всех n, $t \geqslant n_0$ справедливо неравенство*

$$\|p^*(\varepsilon_n, \delta_n) - p^*(\varepsilon_t, \delta_t)\| + \|\lambda^*(\varepsilon_n, \delta_n) - \lambda^*(\varepsilon_t, \delta_t)\| \leqslant$$
$$\leqslant C\left(|\varepsilon_n - \varepsilon_t| + |\delta_n - \delta_t| + \left|\frac{\varepsilon_n}{\delta_n} - \frac{\varepsilon_t}{\delta_t}\right|\right). \quad (3.47)$$

Доказательство. Определим множества

$$D_0 \triangleq \left\{(p, \lambda) \,\middle|\, \sum_{i=1}^{N} p(i) = 1\right\},$$

$$D_1(i_1, \ldots, i_s) \triangleq \{(p, \lambda) \mid p(i_k) = \varepsilon,\ k = \overline{1, s}\} \cap D_0,$$

$$D_2(j_1, \ldots, j_r) \triangleq \{(p, \lambda) \mid \lambda(j_k) = 0,\ k = \overline{1, r}\} \cap D_0,$$

$$D_3(i_1, \ldots, i_s; j_1, \ldots, j_r) \triangleq D_1(i_1, \ldots, i_s) \cap D_2(j_1, \ldots, j_r),$$

где $\varepsilon \in [0, N^{-1})$, $i_1, \ldots, i_s$ и $j_1, \ldots, j_r$ — всевозможные сочетания по s и r элементов из множеств $\{1, 2, \ldots, N\}$,

$\{1, 2, \ldots, m\}$ соответственно, $s=\overline{1, N-1}$, $r=\overline{1, m}$. Перенумеруем эти множества, обозначив их для удобства G_k, $k=1, 2, \ldots, 2^m(2^N-1)$ и каждому G_k поставим в соответствие задачу Z_k нахождения седловой точки функции $L_\delta(p, \lambda)$ (3.44) на множестве G_k, где $\delta>0$. Решение задачи Z_k будем обозначать $(p(Z_k), \lambda(Z_k))$. Заметим, что при каждом $\varepsilon\in[0, N^{-1})$ и $\delta>0$ точка $(p^*(\varepsilon, \delta), \lambda^*(\varepsilon, \delta))$ совпадает с решением одной из задач Z_k. Изучим зависимость точек $(p(Z_k), \lambda(Z_k))$ от параметров ε и δ.

Рассмотрим для определенности задачу Z_k, соответствующую множеству

$$D_3(1;1)\triangleq\left\{(p, \lambda)\,\middle|\,\sum_{i=1}^{N}p(i)=1,\ p(1)=\varepsilon,\ \lambda(1)=0\right\}.$$

Используя метод множителей Лагранжа для нахождения точки $(p(Z_k), \lambda(Z_k))$, приходим к системе линейных уравнений, которую можно записать следующим образом:

$$\begin{gathered}\nabla_p L(p, \lambda)+\delta p-\varkappa_0 e^N-\varkappa_1 e_1^N=0,\\ \nabla_\lambda L(p, \lambda)-\delta\lambda-\theta_1 e_1^m=0,\\ p^{\mathsf{T}}e^N=1,\quad p(1)=\varepsilon,\quad \lambda(1)=0.\end{gathered}\tag{3.48}$$

Отсюда следует, что зависимость компонент седловой точки $(p(Z_k), \lambda(Z_k))$ и множителей Лагранжа $\varkappa_0$, $\varkappa_1$, θ_1 от параметров ε, δ имеет вид отношения полиномов

$$\sum_{\tau=0}^{N+m}\delta^\tau(g_\tau+h_\tau\varepsilon)\Big/\sum_{\tau=0}^{N+m}a_\tau\delta^\tau,\tag{3.49}$$

причем полином, стоящий в знаменателе, равен детерминанту системы (3.48) и, следовательно, не имеет корней при $\delta>0$ (так как в этом случае функция $L_\delta(p, \lambda)$ имеет единственную седловую точку на множестве G_k, определяемую системой (3.48)). Пусть (после возможных сокращений) в этих отношениях полиномов наименьший общий их знаменатель имеет сомножитель δ^r, $r\geqslant0$. Покажем, что либо $r=0$, либо $r=1$. Для этого подставим найденное решение в систему (3.48), умножим скалярно первое уравнение на $p(Z_k)$, а второе на $-\lambda(Z_k)$ и сложим получившиеся равенства. Получим

$$\delta(\|p(Z_k)\|^2+\|\lambda(Z_k)\|^2)=\varkappa_0+\varkappa_1\varepsilon,$$

откуда следует, что либо $2r-1=r$ (тогда $r=1$), либо $2r-1=r-1$ (тогда $r=0$).

При $r=0$ седловая точка $(p(Z_k), \lambda(Z_k))$ как функция $\varepsilon \in [0, N^{-1})$ и $\delta > 0$ является липшицевой. При $r=1$ выражение (3.49) представимо в виде

$$\left(\frac{g_0}{\delta} + \sum_{\tau=1}^{N+m} g_\tau \delta^{\tau-1} + \frac{\varepsilon}{\delta} \sum_{\tau=0}^{N+m} h_\tau \delta^\tau\right) \Big/ \sum_{\tau=0}^{N+m-1} \tilde{a}_\tau \delta^\tau,$$

причем знаменатель не имеет корней при $\delta \geqslant 0$. Отсюда следует, что если $g_0 \neq 0$ хотя бы для одной из компонент вектора $(p(Z_k), \lambda(Z_k))$, то при $\varepsilon = \varepsilon_n$, $\delta = \delta_n$ и $n \geqslant n_0$ (где n_0 достаточно велико) точка $(p^*(\varepsilon_n, \delta_n), \lambda^*(\varepsilon_n, \delta_n))$ не может быть решением рассматриваемой задачи Z_k. Если же $g_0 = 0$ для всех компонент точки $(p(Z_k), \lambda(Z_k))$, то эта точка как функция $\varepsilon \in [0, N^{-1})$, $\delta > 0$ и отношения ε/δ является липшицевой.

Заметим теперь, что полученные выводы справедливы для каждой задачи Z_k. Но, как уже отмечалось, при каждом $n = 1, 2, \ldots$ точка $(p^*(\varepsilon_n, \delta_n), \lambda^*(\varepsilon_n, \delta_n))$ является решением одной из задач Z_k при $\varepsilon = \varepsilon_n$ и $\delta = \delta_n$. Отсюда и следует справедливость леммы. Лемма доказана. ▲

Следствие. *В условиях леммы* 3.6 *при* $n \geqslant n_0$ *справедливы неравенства*

$$\|p^*(\varepsilon_n, \delta_n) - p^*(\varepsilon_{n+1}, \delta_{n+1})\| + \|\lambda^*(\varepsilon_n, \delta_n) - \lambda^*(\varepsilon_{n+1}, \delta_{n+1})\| \leqslant$$
$$\leqslant C\left(|\varepsilon_n - \varepsilon_{n+1}| + |\delta_n - \delta_{n+1}| + \left|\frac{\varepsilon_n}{\delta_n} - \frac{\varepsilon_{n+1}}{\delta_{n+1}}\right|\right), \tag{3.50}$$

$$\|p^*(\varepsilon_n, \delta_n) - p^*\| + \|\lambda^*(\varepsilon_n, \delta_n) - \lambda^*\| \leqslant$$
$$\leqslant C\left(\varepsilon_n + \delta_n + \left|\frac{\varepsilon_n}{\delta_n} - \zeta\right|\right), \tag{3.51}$$

где $p^* = \lim_{n\to\infty} p^*(\varepsilon_n, \delta_n)$, $\lambda^* = \lim_{n\to\infty} \lambda^*(\varepsilon_n, \delta_n)$, $\zeta = \lim_{n\to\infty} \varepsilon_n/\delta_n$.

Эти неравенства получаются из (3.47) соответственно при $m = n+1$ и при $m \to \infty$ с учетом леммы 3.5.

Перейдем теперь к изучению алгоритмов адаптивного выбора вариантов, основанных на регуляризованном методе множителей Лагранжа.

§ 3.5. Алгоритм, реализующий метод множителей Лагранжа

Для формирования алгоритма отыскания седловой точки функции Лагранжа (3.42) воспользуемся методом стохастической аппроксимации. Учитывая замечания, сделанные в § 3.4, будем использовать регуляризованную функцию Лагранжа (3.44) с переменным параметром регуляризации $\delta = \delta_n$, где n — номер шага, причем $\delta_n \to 0$ при $n \to \infty$. Кроме того, заменим в $L(p, \lambda)$ (3.42) и $L_\delta(p, \lambda)$ (3.44) функцию $V_0(p)$ на функцию

$$V_0(p) - \Delta_0 = \sum_{i=1}^{N} (v_i^0 - \Delta_0)\, p(i), \quad p \in S^N,$$

а функции $V_j(p)$ $(j = \overline{1, m})$ представим в виде

$$V_j(p) = \sum_{i=1}^{N} (v_i^j - \Delta_j)\, p(i) + \Delta_j, \quad p \in S^N.$$

Очевидно, такое введение дополнительных параметров Δ_j не влияет на седловую точку функций Лагранжа (3.42), (3.44), но меняет градиенты этих функций по вектору p. В силу последнего обстоятельства (как мы уже видели в § 2.6) возникает возможность влиять на скорость сходимости алгоритмов, построенных по методу стохастической аппроксимации, за счет выбора параметров Δ_j $(j = \overline{0, m})$.

Итак, для функции

$$L_\delta(p, \lambda) \triangleq \sum_{i=1}^{N} (v_i^0 - \Delta_0)\, p(i) + \sum_{j=1}^{m} \lambda(j) \left[\sum_{i=1}^{N} (v_i^j - \Delta_j)\, p(i) + \Delta_j \right] + \frac{\delta}{2} (\| p \|^2 - \| \lambda \|^2) \tag{3.52}$$

при $p \in S_\varepsilon^N$ $(\varepsilon > 0)$, как нетрудно убедиться,

$$\nabla_p L_\delta(p, \lambda) = \mathsf{M} \left\{ \frac{e(x_n)}{p_n^{\mathrm{T}} e(x_n)} \left[\xi_n^0 - \Delta_0 + \sum_{j=1}^{m} \lambda_n(j)(\xi_n^j - \Delta_j) + \delta p_n^{\mathrm{T}} e(x_n) \right] \Big|\, p_n = p,\ \lambda_n = \lambda \right\}, \tag{3.53}$$

$$\frac{\partial L_\delta(p, \lambda)}{\partial \lambda(j)} = \mathsf{M} \{ \xi_n^j - \delta \lambda_n(j) \mid p_n = p,\ \lambda_n = \lambda \}, \quad j = \overline{1, m},$$

где векторы $p_n \triangleq (p_n(1), \ldots, p_n(N)) \in S_\varepsilon^N$ и $\lambda_n \triangleq (\lambda_n(1), \ldots, \lambda_n(m)) \in R_+^m$ — текущие оценки седловой точки (p^*, λ^*) функции Лагранжа (3.42) (измеримые относительно σ-алгебры $\mathscr{F}_n \triangleq \sigma\left(x_t, \xi_t^j \mid t=\overline{1, n-1}, j=\overline{0, m}\right)$. Отсюда, учитывая условия $p_n \in S_{\varepsilon_n}^N$ и $\lambda_n \in R_+^m$, путем введения операций проектирования, приходим к следующему алгоритму:

$$p_{n+1} = \pi_{\varepsilon_{n+1}}^N \left\{ p_n - \gamma_n \frac{e(x_n)}{p_n^{\mathrm{T}} e(x_n)} \left[\xi_n^0 - \Delta_0 + \right.\right.$$

$$\left.\left. + \sum_{j=1}^m \lambda_n(j)\left(\xi_n^j - \Delta_j\right) + \delta_n p_n^{\mathrm{T}} e(x_n) \right] \right\}, \quad (3.54)$$

$$\lambda_{n+1}(j) = \max\left\{0; \lambda_n(j) + \gamma_n\left(\xi_n^j - \delta_n \lambda_n(j)\right)\right\}, \quad j = \overline{1, m}.$$

Здесь на каждом шаге осуществляется проектирование соответствующих векторов на ε_{n+1}-симплекс и на множество R_+^m.

Теорема 3.3. *Пусть справедливы предположения* П1′, П2′ *из* § 3.1, $p_1 \in S_{\varepsilon_1}^N$, $\lambda_1 \in R_+^m$, *а* $\{(p_n, \lambda_n)\}$ — *последовательность векторов, порождаемая алгоритмом* (3.54), *в котором*

$$\gamma_n > 0, \quad \varepsilon_n \in (0, N^{-1}), \quad \delta_n > 0, \quad n = 1, 2, \ldots,$$

$$\lim_{n\to\infty} \delta_n = 0, \quad \sum_{n=1}^\infty \gamma_n \delta_n = \infty$$

и существует конечный предел $\lim\limits_{n\to\infty} \varepsilon_n/\delta_n \triangleq \zeta$. Тогда

1) *если*

$$\sum_{n=1}^\infty \left[\frac{\gamma_n^2}{\varepsilon_n} + |\varepsilon_n - \varepsilon_{n+1}| + |\delta_n - \delta_{n+1}| + \left| \frac{\varepsilon_n}{\delta_n} - \frac{\varepsilon_{n+1}}{\delta_{n+1}} \right| \right] < \infty,$$

то с вероятностью 1 $p_n \to p^*$, $\lambda_n \to \lambda^*$ *при* $n \to \infty$;

2) *если*

$$\lim_{n\to\infty} \left[\frac{\gamma_n}{\varepsilon_n \delta_n} + \frac{1}{\gamma_n \delta_n} \left(|\varepsilon_n - \varepsilon_{n+1}| + |\delta_n - \delta_{n+1}| + \right.\right.$$

$$\left.\left. + \left| \frac{\varepsilon_n}{\delta_n} - \frac{\varepsilon_{n+1}}{\delta_{n+1}} \right| \right) \right] = 0,$$

то последовательность $\{(p_n, \lambda_n)\}$ *сходится к седловой точке* (p^*, λ^*) *функции Лагранжа* (3.42) *в среднеквадратическом, причем* (p^*, λ^*) *зависит, вообще говоря, от* ζ.

Доказательство. Введем величины $W_n \triangleq \triangleq \|p_n - p_n^*\|^2 + \|\lambda_n - \lambda_n^*\|^2$, где (p_n^*, λ_n^*) — седловая точка функции $L_{\delta_n}(p, \lambda)$, на множестве $S_{\varepsilon_n}^N \times R_+^m$. В силу алгоритма (3.54), свойств проекторов на множества $S_{\varepsilon_{n+1}}^N$ и R_+^m, а также учитывая предположения П1′, П2′, равенства (3.53) и неравенство (3.50), получим для всякого $n = 1, 2, \ldots$

$$\mathsf{M}\{W_{n+1} \mid \mathscr{F}_n\} \leqslant W_n + C[\varkappa_n \sqrt{W_n} + \varkappa_n^2 + (1 + W_n)\gamma_n^2 \varepsilon_n^{-1}] - \\ - 2\gamma_n [L_{\delta_n}(p_n, \lambda_n^*) - L_{\delta_n}(p_n^*, \lambda_n) + 0{,}5\delta_n W_n],$$

где $\varkappa_n \triangleq |\varepsilon_n - \varepsilon_{n+1}| + |\delta_n - \delta_{n+1}| + |\varepsilon_n \delta_n^{-1} - \varepsilon_{n+1}\delta_{n+1}^{-1}|$, C — положительная константа. Отсюда, используя определение седловой точки и неравенство $2ab \leqslant a^2 + b^2$ для $a = \sqrt{\gamma_n \delta_n W_n}$, $b = C\varkappa_n(\gamma_n\delta_n)^{-1/2}$, приходим к соотношению

$$\mathsf{M}\{W_{n+1} \mid \mathscr{F}_n\} \leqslant (1 - 0{,}5\gamma_n\delta_n + C\gamma_n^2\varepsilon_n^{-1}) W_n + \\ + C[\varkappa_n^2 + \gamma_n^2\varepsilon_n^{-1} + C\varkappa_n^2(2\gamma_n\delta_n)^{-1}], \quad (3.55)$$

из которого следует сходимость с вероятностью 1 $(p_n, \lambda_n) \to (p^*, \lambda^*)$ при $n \to \infty$ (в силу соответствующих условий теоремы, лемм П.11 и 3.5). Усреднив (3.55), получим рекуррентное неравенство

$$\mathsf{M}\{W_{n+1}\} \leqslant (1 - 0{,}5\gamma_n\delta_n + C\gamma_n^2\varepsilon_n^{-1}) \mathsf{M}\{W_n\} + \\ + C[\varkappa_n^2 + \gamma_n^2\varepsilon_n^{-1} + C\varkappa_n^2(2\gamma_n\delta_n)^{-1}], \quad (3.56)$$

из которого в силу леммы П.5 (при $b = d = 0$), леммы 3.5 и соответствующих условий данной теоремы вытекает сходимость $(p_n, \lambda_n) \to (p^*, \lambda^*)$ при $n \to \infty$ в среднеквадратическом. Теорема доказана. ▲

Замечание. Седловая точка (p^*, λ^*), к которой в условиях данной теоремы сходится последовательность $\{(p_n, \lambda_n)\}$, порождаемая алгоритмом (3.54), является точкой минимума функции (3.46) на множестве седловых точек функции Лагранжа (3.42). Это непосредственно вытекает из доказательства леммы 3.5 и из данной теоремы.

Отметим, что поскольку явные выражения для константы C, фигурирующей в формулировке леммы 3.6, не установлены, то константа в рекуррентном неравенстве (3.56) также не определена, и, как следствие, не устанавливается точная величина скорости сходимости алгоритма (3.54) (как это было сделано в § 2.6 для проекционного алгоритма (2.47)). Неравенство (3.56) позволяет оценить порядок скорости сходимости алгоритма (3.54).

Рассмотрим для простоты последовательности $\{\gamma_n\}$, $\{\varepsilon_n\}$, $\{\delta_n\}$ вида

$$\gamma_n = \frac{\gamma}{(n+a)^{\alpha}}, \quad \varepsilon_n = \frac{\varepsilon}{(n+b)^{\beta}}, \quad \delta_n = \frac{\delta}{(n+c)^{\varkappa}}, \tag{3.57}$$

где параметры γ, ε, δ, a, b, c, α, β, $\varkappa$ таковы, что выполняются условия теоремы 3.3. В частности, для выполнения условий, обеспечивающих сходимость с вероятностью 1, параметры α, β, $\varkappa$ должны удовлетворять неравенствам

$$\beta \geqslant \varkappa > 0, \quad 2\alpha > 1 + \beta,$$
$$\alpha + \varkappa \leqslant 1; \tag{3.58}$$

аналогично, для сходимости в среднеквадратическом должны выполняться неравенства

$$\beta \geqslant \varkappa > 0, \quad \alpha > \beta + \varkappa,$$
$$\alpha + \varkappa \leqslant 1. \tag{3.59}$$

Теорема 3.4. *Пусть последовательности* $\{\gamma_n\}$, $\{\varepsilon_n\}$, $\{\delta_n\}$ *имеют вид* (3.57), *выполнены условия теоремы* 3.3, *гарантирующие сходимость алгоритма* (3.54) *в среднеквадратическом* (*в частности, справедливы неравенства* (3.59)), *и пусть при* $\alpha + \varkappa = 1$ *выполняется неравенство* $0{,}5\gamma\delta > 2\alpha - \beta - 1$. *Тогда*

$$\overline{\lim_{n\to\infty}}\, n^{\theta} \mathsf{M}\{\|p_n - p^*\|^2 + \|\lambda_n - \lambda^*\|^2\} < \infty, \tag{3.60}$$

где $\theta \triangleq \min\{2\varkappa, \alpha - 2\varkappa\}$ *при* $\beta = \varkappa$ *и* $\theta \triangleq \min\{2\varkappa, 2(\beta - \varkappa), \alpha - \beta - \varkappa\}$ *при* $\beta > \varkappa$.

Доказательство. Заметим, что в условиях данной теоремы выполняются соотношения (3.50), (3.51). Рекуррентное неравенство (3.56) в предположениях (3.57) можно записать в виде

$$\mathsf{M}\{W_{n+1}\} \leqslant [1 - 0{,}5\,\gamma\delta(1 + o(1))n^{-(\alpha+\varkappa)}]\mathsf{M}\{W_n\} + Cn^{-(2\alpha-\beta)},$$

где C — некоторая положительная константа. Используем теперь лемму П.2 (в которой параметр s возьмем достаточно большим, $r=1/2$, $a=0$). Получим

$$\overline{\lim_{n\to\infty}}\, n^{(\alpha-\beta-\varkappa)}\,\mathsf{M}\{W_n\}<\infty.$$

Отсюда, учитывая (3.51) и неравенство

$$\|p_n-p^*\|^2+\|\lambda_n-\lambda^*\|^2\leqslant 2W_n+2\left(\|p_n^*-p^*\|^2+\|\bar{\lambda}_n^*-\lambda^*\|^2\right),$$

получим требуемый результат. Теорема доказана. ▲

Из этой теоремы, леммы 3.2 и неравенств

$$[V_j(p)-V_j(p^*)]^2\leqslant\|p-p^*\|^2\sum_{i=1}^{N}\left(v_i^j\right)^2,\quad j=\overline{0,m}$$

вытекает, что алгоритм (3.54) решает задачу (3.3) условной минимизации средних потерь. Этот факт содержится в следующем утверждении.

Следствие 1. *При выполнении условий теоремы* 3.4 *с вероятностью* 1 *выполняется целевое условие* 3.3 *и справедливы неравенства*

$$\overline{\lim_{n\to\infty}}\, n^{\theta}\mathsf{M}\left\{\left[\Phi_n^j-V_j(p^*)\right]^2\right\}<\infty,\quad j=\overline{0,m}.$$

В следующем утверждении устанавливаются оптимальные параметры α, β, $\varkappa$, максимизирующие порядок скорости сходимости θ алгоритма (3.54), гарантируемый теоремой 3.4.

Следствие 2. *При выполнении условий теоремы* 3.4 *параметр* θ *в* (3.60) *удовлетворяет неравенству* $\theta\leqslant 2/5$, *причем* $\theta=2/5$ *только в случае*

$$\alpha=\alpha^*=4/5,\quad \beta=\beta^*=\varkappa=\varkappa^*=1/5.$$

Для получения этого результата достаточно максимизировать величину θ, определенную в теореме 3.4, по параметрам α, β, $\varkappa$ при ограничениях (3.59).

§ 3.6. Алгоритм с рандомизацией ограничений

В алгоритме (3.54) на каждом шаге необходимо вычислять или наблюдать реализации всех $m+1$ случайных величин ξ_n^j. Другими словами, на каждом шаге n учитывается текущая информация о значении целевой функции $V_0(p)$ и выполнении всех ограничений задачи (3.4) при

$p = p_{n-1}$. Однако часто, особенно при большом числе ограничений m, некоторые ограничения оказываются несущественными, т. е. их отбрасывание не изменяет решения p^* задачи (3.4). Информация о выполнении таких ограничений, следовательно, является лишней. В связи с этим представляет интерес получение такого алгоритма решения задачи (3.4), который с течением времени распознавал бы несущественные ограничения и реже использовал бы информацию об их выполнении.

Для построения такого алгоритма воспользуемся идеей рандомизации ограничений, которая появилась при рассмотрении задач математического программирования как задач векторной оптимизации [37] и в дальнейшем использовалась в [66, 67, 136]. Суть ее заключается в учете на каждом шаге текущей информации либо только о значении целевой функции (как бы решая при этом задачу безусловной минимизации), либо об одном из ограничений (как бы заботясь лишь о выполнении соответствующего неравенства). При этом решение о том, реализацию какой из случайных величин ξ_n^j необходимо наблюдать на n-м шаге и использовать для изменения вектора p_{n-1}, носит случайный характер. Вероятности выбора конкретных значений индекса j должны с течением времени изменяться так, чтобы обеспечить решение задачи (3.3).

Последовательность выбранных значений индекса j обозначим $\{j_n\}$. Выберем $x \in X$ и $j_n \in \{0, 1, \ldots, m\}$ на n-м шаге независимо друг от друга, а именно так, что

$$\mathsf{P}\{x_n = x(i), j_n = j \mid \mathcal{F}_n\} = \mathsf{P}\{x_n = x(i) \mid \mathcal{F}_n\}\,\mathsf{P}\{j_n = j \mid \mathcal{F}_n\}, \tag{3.61}$$

с условными распределениями

$$\mathsf{P}\{x_n = x(i) \mid \mathcal{F}_n\} \triangleq p_n(i) \qquad (i = \overline{1, N}), \tag{3.62}$$

$$\mathsf{P}\{j_n = j \mid \mathcal{F}_n\} \triangleq \begin{cases} \left[1 + \sum\limits_{k=1}^{m} \lambda_n(k)\right]^{-1}, & j = 0, \\ \lambda_n(j)\left[1 + \sum\limits_{k=1}^{m} \lambda_n(k)\right]^{-1}, & j = \overline{1, m}, \end{cases} \tag{3.63}$$

где

$$\mathcal{F}_n \triangleq \sigma\left(x_t, j_t, \xi_t^{j_t} \middle| t = \overline{1, n-1}\right),$$

а векторы $p_n \triangleq (p_n(1), \ldots, p_n(N)) \in S^N$ и $\lambda_n \triangleq (\lambda_n(1), \ldots \ldots, \lambda_n(m)) \in R_+^m$ — текущие оценки седловой точки

(p^*, λ^*) функции Лагранжа (3.42) (p_n и λ_n измеримы относительно $\mathscr{F}_n$).

После несложных преобразований выражений (3.53) с учетом (3.61) — (3.63) получим для $p \in S_\varepsilon^N$, $\varepsilon > 0$, $\lambda_j > 0$ $(j = \overline{1, m})$

$$\nabla_p L_\delta(p, \lambda) = \mathsf{M}\left\{\frac{e(x_n)}{p_n^{\mathrm{T}} e(x_n)}\left[\left(1 + \sum_{k=1}^{m} \lambda_n(k)\right)\left(\xi_n^{j_n} - \Delta_{j_n}\right) + \right.\right.$$
$$\left.\left. + \delta p_n^{\mathrm{T}} e(x_n)\right] \Big| p_n = p, \lambda_n = \lambda\right\}, \quad (3.64)$$

$$\frac{\partial}{\partial \lambda_j} L_\delta(p, \lambda) =$$
$$= \mathsf{M}\left\{\left(1 + \sum_{k=1}^{m} \lambda_n(k)\right)\left(\frac{\xi_n^{j_n} - \Delta_{j_n}}{\lambda_n(j_n)} - \delta\right)\delta_{jj_n} \Big| p_n = p, \lambda_n = \lambda\right\},$$
$$j = \overline{1, m},$$

где δ_{jj_n} — символ Кронекера. Отсюда, используя, как и ранее, метод стохастической аппроксимации, приходим к следующему алгоритму:

$$p_{n+1} = \pi_{\varepsilon_{n+1}}^N \left\{p_n - \gamma_n \frac{e(x_n)}{p_n^{\mathrm{T}} e(x_n)}\left[\left(1 + \sum_{k=1}^{m} \lambda_n(k)\right)\left(\xi_n^{j_n} - \Delta_{j_n}\right) + \right.\right.$$
$$\left.\left. + \delta_n p_n^{\mathrm{T}} e(x_n)\right]\right\}, \quad (3.65)$$

$$\lambda_{n+1}(j) = \max\left\{\varepsilon_{n+1}; \lambda_n(j) + \gamma_n\left(1 + \sum_{k=1}^{m} \lambda_n(k)\right) \times \right.$$
$$\left. \times \left(\frac{\xi_n^j - \Delta_j}{\lambda_n(j)} - \delta_n\right)\right\} \text{ при } j = j_n,$$

$$\lambda_{n+1}(j) = \lambda_n(j) \text{ при } j \neq j_n.$$

Поскольку изменению в (3.65) на каждом шаге подлежит не более одного множителя Лагранжа, величину $z_n \triangleq 1 + \sum_{k=1}^{m} \lambda_n(k)$ нетрудно пересчитать с помощью рекуррентной процедуры:

$$z_{n+1} = \begin{cases} z_n, & j_{n+1} = 0, \\ z_n - \lambda_n(j_n) + \lambda_{n+1}(j_{n+1}), & j_{n+1} \neq 0, \quad n = 1, 2, \ldots; \end{cases}$$

$$z_1 \triangleq 1 + \sum_{k=1}^{m} \lambda_1(k).$$

Отметим, что ε-регуляризация в (3.65) распространена теперь и на ту часть алгоритма, которая относится к λ: на каждом шаге n вектор λ_n принадлежит множеству $R^m_{\varepsilon_n} = \{\lambda \mid \lambda(k) \geqslant \varepsilon_n \ (k = \overline{1, m})\}$. Это необходимо для того, чтобы вероятности выбора любого ограничения в (3.4) были бы на каждом шаге положительными, что позволяет с ростом n набрать полную статистическую информацию о существенности каждого из ограничений. Как и ранее, δ-регуляризация, т. е. переход от функции Лагранжа (3.42) к регуляризованной функции (3.44), также имеет определенный физический смысл: она эквивалентна введению в алгоритм поиска седловой точки затухания, или демпфирования.

Все полученные в § 3.5 результаты исследования сходимости алгоритма (3.54), в котором рандомизация ограничений не проводилась, полностью распространяются и на алгоритм с рандомизацией ограничений (3.65). Это объясняется тем, что неравенство (3.55), на основе которого и были получены все упомянутые результаты, имеет место и для алгоритма (3.65) (следует лишь по-новому определить σ-алгебры $\mathscr{F}_n$, а именно так, как это сделано в данном параграфе). Поэтому заранее трудно отдать предпочтение одному из них. При малом числе ограничений m целесообразно использовать алгоритм (3.54). С ростом же m количество вычислений и число наблюдаемых величин ξ^j_n в этом алгоритме растет. Если при этом имеется априорная информация о том, что какие-то ограничения решаемой задачи (3.4) несущественны, то разумнее использовать алгоритм с рандомизацией ограничений: в этом алгоритме при достаточно больших n по мере сходимости λ_n к λ^* для улучшения приближений (p_n, λ_n) чаще используется информация о тех ограничениях, выполнение которых более существенно. Это следует из определения случайной величины j_n (3.63) и интерпретации множителей Лагранжа как чувствительностей оптимального значения целевой функции к нарушениям ограничений.

Отметим, что при использовании алгоритма с рандомизацией ограничений параллельно с основной задачей (т. е. определением вероятностей $p^*(i)$ выбора вариантов $x(i)$, доставляющих минимум функции $V_0(p)$ при выполнении ограничений (3.4)) решается вспомогательная задача адаптивного выбора вариантов, состоящая в определении вероятностей наблюдения реализаций ξ_n^j с целью определения существенности ограничений и использования этой информации в процессе поиска вектора p^*.

§ 3.7. Обсуждение результатов

В этой главе рассматривалась задача условной минимизации средних потерь, которые в данном случае характеризовались несколькими наблюдаемыми случайными показателями. Показано, что исходная задача условной минимизации средних потерь эквивалентна некоторой задаче линейного стохастического программирования, причем любая последовательность $\{p_n\}$ рандомизированных правил выбора, которая достаточно быстро сходится (в среднеквадратическом смысле) к решению этой задачи стохастического программирования, гарантирует решение исходной задачи выбора с той же скоростью. Установленная эквивалентность задач позволила обоснованно подойти к анализу и синтезу алгоритмов адаптивного выбора вариантов с позиции решения задачи линейного стохастического программирования. Подробно рассматривались два метода решения: метод штрафных функций и метод множителей Лагранжа.

Метод штрафных функций позволил подойти к решению исходной задачи условной минимизации средних потерь как к задаче безусловной минимизации некоторой регуляризованной штрафной функции при априорной неопределенности. Полученный алгоритм адаптивного выбора представляет собой стохастический вариант метода проекции градиента штрафной функции, где в качестве вектора движения используются текущие реализации градиента, а изменение длины шага алгоритма согласовано с изменением параметров штрафной функции. В соответствии с этим алгоритмом изменение рандомизированных правил выбора p_n происходит лишь через такт, при этом среднеквадратическая скорость решения задачи имеет порядок $n^{-2/5}$.

Другим методом, позволяющим синтезировать работоспособные алгоритмы условного выбора, является регуляризованный метод множителей Лагранжа, который устанавливает соответствие между исходной задачей и задачей отыскания седловой точки функции Лагранжа. Алгоритмы, построенные по этому методу, также обеспечивают среднеквадратическую скорость решения задачи, по порядку равную $n^{-2/5}$. Они осуществляют изменение правил выбора на каждом такте, что позволяет надеяться на более высокую величину скорости сходимости (хотя бы на начальных тактах). Кроме того, этот метод дает возможность синтезировать удобную модификацию приведенных алгоритмов, которая названа алгоритмом «с рандомизацией ограничений». Ее удобство состоит в том, что при большом числе ограничений на каждом такте алгоритма допускается использование информации лишь об одном из них. Отбрасывание остальной информации не приводит к снижению порядка скорости сходимости. Хотя, по-видимому, ухудшается сама величина скорости сходимости, зато значительно сокращается объем требуемой оперативной памяти в цифровой вычислительной машине, на которой может быть реализован данный метод.

По сравнению с задачей безусловной минимизации рассмотренная в этой главе задача условной минимизации является существенно более сложной. Это обстоятельство сказывается не только на усложнении самих алгоритмов адаптивного выбора вариантов, но и на скорости решения задачи, которая в данном случае имеет порядок $n^{-2/5}$ (в отличие от алгоритма (2.47) решения задачи безусловной минимизации, имеющего порядок скорости сходимости в среднеквадратическом $n^{-2/3}$).

ГЛАВА 4

ИГРОВЫЕ АЛГОРИТМЫ

Дается постановка игровой задачи адаптивного выбора вариантов, представляющей собой матричную игру в условиях априорной неопределенности. Подробно рассматривается игра двух лиц с нулевой суммой и диагонально выпуклая игра многих лиц; предлагается и исследуется (в смысле сходимости последовательности рандомизированных правил) алгоритм, предназначенный для решения таких задач. Обосновывается работоспособность (в критериальном смысле) упрощенного варианта этого алгоритма для общего случая матричной игры в условиях неопределенности.

§ 4.1. Игровая задача адаптивного выбора вариантов

Пусть множество возможных вариантов есть

$$X \triangleq X_1 \times X_2 \times \ldots \times X_l,$$

$$X_k \triangleq \{x^k(1), \ldots, x^k(N_k)\}, \qquad k = \overline{1, l},$$

т. е. каждый вариант $x \in X$ представляет собой вектор $x = (x^1, \ldots, x^l)$, k-я компонента которого x^k принимает N_k значений из соответствующего множества X_k. Пусть, кроме того, потери на интервале времени (t_n, t_{n+1}) характеризуются l наблюдаемыми случайными величинами $\xi_n^k \triangleq \xi_n^k(x_n, \omega)$ $(k = \overline{1, l})$ (см. сноску на с. 40), где $x_n \in X$ — вариант, выбранный в момент времени t_n $(n = 1, 2, \ldots)$, а ξ_n^k — потери системы за выбор k-й компоненты x_n^k варианта x_n, которые, вообще говоря, зависят от всех других компонент, т. е. в целом от варианта x_n.

Далее, пользуясь терминологией теории игр [78, 79], будем говорить, что выбор k-й компоненты вектора вариантов x_n осуществляется k-м *игроком* $(k = \overline{1, l})$, который за этот выбор несет потери ξ_n^k. X_k естественно называть множеством чистых стратегий, а $x_n^k \in X_k$ — чистой стратегией k-го игрока, реализованной в момент времени t_n *). Отметим, что текущие средние потери k-го игрока

$$\Phi_n^k = \Phi_n^k(\{x_n\}) \triangleq \frac{1}{n} \sum_{t=1}^{n} \xi_t^k, \quad k = \overline{1, l}; \quad n = 1, 2, \ldots \quad (4.1)$$

зависят от чистых стратегий, реализованных всеми игроками к моменту времени t_n включительно.

Сделаем следующие предположения:

П1″. Для всех $n = 1, 2, \ldots$ и $k = \overline{1, l}$ совокупности $\{\xi_n^k(x, \omega) \mid x \in X\}$ и $\{\xi_t^k(x, \omega), x_r \mid x \in X, t = \overline{1, n-1}, r = \overline{1, n}\}$ независимы;

П2″. При любых $k = \overline{1, l}$, $x = (x^1(i_1), \ldots, x^l(i_l)) \in X$ и $n = 1, 2, \ldots$

$$\mathsf{M}\{\xi_n^k(x, \omega)\} \triangleq v_{i_1,\ldots,i_l}^k, \quad \sup_n \mathsf{M}\{[\xi_n^k(x, \omega)]^2\} < \infty;$$

П3″. Для всех $x = (x^1(i_1), \ldots, x^l(i_l)) \in X$ и $n = 1, 2, \ldots$

$$\mathsf{P}\{x_n = x \mid x_t, \xi_t^k (t = \overline{1, n-1}, k = \overline{1, l})\} = \prod_{k=1}^{l} \mathsf{P}\{x_n^k = x^k(i_k) \mid x_t^k, \xi_t^k (t = \overline{1, n-1})\}.$$

При каждом $k = \overline{1, l}$ из предположений П1″, П2″ для $\xi_n(x, \omega) \triangleq \xi_n^k(x, \omega)$ вытекает справедливость предположений П1, П2 из § 2.1 и, следовательно, в силу леммы 2.1 с вероятностью 1 все предельные точки последовательности $\{\Phi_n^k\}$ лежат в отрезках

$$\left[\min_{i_1,\ldots,i_l} v_{i_1,\ldots,i_l}^k, \max_{i_1,\ldots,i_l} v_{i_1,\ldots,i_l}^k\right].$$

Предположение П3″ означает статистическую независимость выбора игроками своих чистых стратегий.

Это позволяет сформулировать игровую задачу адаптивного выбора вариантов следующим образом:

---

*) Не следует путать понятия «чистая стратегия» и «смешанная стратегия» игрока (принятые в теории игр) с понятием «стратегия выбора», или «стратегия управления», введенным в § 1.4.

игроки должны по наблюдениям своих потерь выбирать чистые стратегии так, чтобы сформированная последовательность вариантов $\{x_n\}$ при каждом $k=\overline{1, l}$ с вероятностью 1 удовлетворяла неравенству

$$\varlimsup_{n\to\infty}\left[\Phi_n^k(\{x_n\})-\Phi_n^k(\{\hat{x}_n^k\})\right]\leqslant 0 \tag{4.2}$$

для любых последовательностей вариантов $\{\hat{x}_n^k\}$,

$$\hat{x}_n^k \triangleq \left(x_n^1,\ldots,x_n^{k-1},\ \tilde{x}_n^k,\ x_n^{k+1},\ldots,\ x_n^l\right)\in X,$$

удовлетворяющих условию: для любых $x=(x^1(i_1),\ldots$ $\ldots,\ x^l(i_l))\in X$ и $n=1,\ 2,\ \ldots$

$$\begin{aligned} \mathsf{P}\,\{\hat{x}_n^k = x\,|\,\hat{x}_t^k,\ \xi_t^\nu(x_t,\omega)\ (\nu=\overline{1, l};\ \nu\neq k),\\ \xi_t^k(\hat{x}_t^k,\omega)\,(t=\overline{1, n-1})\} &=\\ =\mathsf{P}\,\{\tilde{x}_n^k = x^k(i_k)\,|\,\tilde{x}_t^k,\ \xi_t^k(\hat{x}_t^k,\omega)\,(t=\overline{1, n-1})\}&\times\\ \times\prod_{\substack{\nu=1\\ \nu\neq k}}^{l}\mathsf{P}\,\{x_n^\nu = x^\nu(i_\nu)\,|\,x_t^\nu,\ \xi_t^\nu(x_t,\omega)\,(t=\overline{1, n-1})\}. \end{aligned} \tag{4.3}$$

Выполнение условий (4.2), (4.3) для всех $k=\overline{1, l}$ означает, другими словами, что последовательность вариантов $\{x_n\}$ реализует равновесное по Нэшу поведение игроков (в смысле предельных значений текущих средних потерь (4.1)): каждому игроку не выгодно изменять последовательность его чистых стратегий при условии, что остальные игроки придерживаются равновесных стратегий. Условие (4.3) означает, что последовательности вариантов $\{\hat{x}_n^k\}$, получающиеся из $\{x_n\}$ заменой лишь k-х компонент x_n^k на $\hat{x}_n^k$ ($n=1,\ 2,\ \ldots$), также удовлетворяют условию независимости выбора, аналогичному ПЗ″.

Таким образом, поставленная задача состоит в том, что игроки, действуя независимо друг от друга, т. е. располагая лишь информацией о своих использованных чистых стратегиях x_n^k и полученных потерях ξ_n^k, должны в асимптотике научиться равновесному по Нэшу поведению.

Для формулировки условий разрешимости этой задачи рассмотрим тесно связанную с ней бескоалиционную матричную игру l лиц $\Gamma\triangleq\Big(X_k,\ \|v_{i_1,\ldots,i_l}^k\|\,(i_k=\overline{1, N_k},$

$k=\overline{1, l})\big)$. В игре Г k-й игрок выбирает чистую стратегию из множества $X_k=\{x^k(1), \ldots, x^k(N_k)\}$, не зная о выборе остальных игроков. Если при этом оказались выбранными стратегии $x^1(i_1), \ldots, x^l(i_l)$, то потери k-го игрока составляют $v^k_{i_1,\ldots,i_l}$ $(k=\overline{1, l})$. Обозначим через $p^k \triangleq \left(p^k_1,\ldots, p^k_{N_k}\right)$ *смешанную стратегию* k-го игрока, где $p^k_{i_k}$ — вероятность выбора этим игроком чистой стратегии $x^k(i_k)$ $(i_k=\overline{1, N_k})$. Таким образом, вектор p^k принадлежит симплексу S^{N_k}. Функцию средних потерь k-го игрока в игре Г обозначим

$$V^k(p) \triangleq \sum_{i_1,\ldots,i_l} v^k_{i_1,\ldots,i_l} \prod_{\nu=1}^{l} p^{\nu}_{i_\nu}, \qquad p \triangleq (p^1,\ldots, p^l), \tag{4.4}$$

величина которой равна математическому ожиданию потерь k-го игрока при реализации игроками смешанных стратегий $p^k \in S^{N_k}$ $(k=\overline{1, l})$. По теореме Нэша [83, 107] игра Г имеет хотя бы одну *точку равновесия* $p^* \triangleq \triangleq (p^{1*},\ldots,p^{l*})$ в смешанных стратегиях, определяемую условием: для всех $k=\overline{1, l}$ и $p^k \in S^{N_k}$ выполняется неравенство

$$V^k(p^*) \leqslant V^k(p^{1*}, \ldots, p^{(k-1)*}, p^k, p^{(k+1)*}, \ldots, p^{l*}). \tag{4.5}$$

Точка равновесия p^* является устойчивой в том смысле, что каждому игроку не выгодно отклоняться от своей смешанной стратегии p^{k*}, если остальные игроки придерживаются точки равновесия.

Пусть выбор элементов $x^k_n \in X_k$ осуществляется случайным образом с некоторыми условными вероятностями

$$p^k_n(i_k) \triangleq \mathrm{P}\left\{x^k_n = x^k(i_k) \,\middle|\, x^k_t, \xi^k_t\,(t=\overline{1, n-1})\right\}, \tag{4.6}$$

$i_k=\overline{1, N_k}$, $k=\overline{1, l}$, $n=1, 2, \ldots$ Свойства последовательности $\{x_n\}$ полностью определяются свойствами последовательностей векторов $p^k_n \triangleq \left(p^k_n(1),\ldots, p^k_n(N_k)\right) \in S^{N_k}$. Сформулируем условия, накладываемые на последовательности $\{p^k_n\}$ $(k=\overline{1, l})$, выполнение которых обеспечивает решение поставленной задачи.

Лемма 4.1. *Пусть выполнены предположения* П1″—П3″ *и для некоторого* $\tau \in (0, 1)$

$$\overline{\lim_{n\to\infty}} n^\tau \mathsf{M}\{\|p_n - p^*\|^2\} < \infty. \tag{4.7}$$

Тогда с вероятностью 1 *выполняется целевое условие* (4.2), (4.3) *и для всех* $k = \overline{1, l}$

$$\overline{\lim_{n\to\infty}} n^\tau \mathsf{M}\{[\Phi_n^k - V^k(p^*)]^2\} < \infty. \tag{4.8}$$

Доказательство. Определим для произвольного $k = \overline{1, l}$ последовательность величин

$$d_n^k(\{\hat{x}_n^k\}, \{\tilde{p}_n^k\}) \triangleq$$

$$\triangleq \left[\Phi_n^k(\{\hat{x}^k\}) - \frac{1}{n}\sum_{t=1}^n V^k(p^{1*}, \ldots, p^{(k-1)*}, \tilde{p}_t^k, p^{(k+1)*}, \ldots, p^{l*})\right]^2,$$

где

$$\tilde{p}_n^k \triangleq (\tilde{p}_n^k(1), \ldots, \tilde{p}_n^k(N_k)),$$

$$\tilde{p}_n^k(i_k) \triangleq \mathsf{P}\{\tilde{x}_n^k = x^k(i_k) \mid \tilde{x}_t^k, \xi_t^k(\hat{x}_t^k, \omega)\,(t = \overline{1, n-1})\},$$

и последовательность σ-алгебр

$$\hat{\mathscr{F}}_n \triangleq \sigma\left(\hat{x}_t^k, \xi_t^\nu(x_t, \omega)\,(\nu = \overline{1, l};\ \nu \neq k), \xi_t^k(\hat{x}_t^k, \omega)\,(t = \overline{1, n-1})\right).$$

Тогда, используя предположения П1″—П3″, условие (4.3), неравенство $2ab \leqslant a^2 + b^2$ и липшицевость функции $V^k(p)$, нетрудно получить для каждого $n = 1, 2, \ldots$ неравенство

$$\mathsf{M}\{d_n^k(\{\hat{x}_n^k\}, \{\tilde{p}_n^k\}) \mid \hat{\mathscr{F}}_{n-1}\} \leqslant$$

$$\leqslant \left(1 - \frac{1}{n}\right) d_{n-1}^k(\{\hat{x}_n^k\}, \{\tilde{p}_n^k\}) + \frac{C}{n^2} + \frac{C}{n}\|p_n - p^*\|^2, \tag{4.9}$$

где C — достаточно большая положительная константа. Из (4.9) в силу леммы П.11 и условия (4.7) получаем, что с вероятностью 1 при $n \to \infty$

$$d_n^k(\{\hat{x}_n^k\}, \{\tilde{p}_n^k\}) \to 0. \tag{4.10}$$

При $\tilde{x}_n^k \triangleq x_n^k$, $\tilde{p}_n^k \triangleq p^{k*}$ из (4.9), (4.10) следует

$$\lim_{n\to\infty} d_n^k(\{x_n\}, \{p^*\}) \overset{\text{п.н.}}{=} 0, \tag{4.11}$$

$$\mathsf{M}\{d_n^k(\{x_n\}, \{p^*\})\} \leqslant$$
$$\leqslant \left(1 - \frac{1}{n}\right)\mathsf{M}\{d_{n-1}^k(\{x_n\}, \{p^*\})\} + \frac{C}{n^2} + \frac{C}{n}\mathsf{M}\{\|p_n - p^*\|^2\}.$$

Из последнего неравенства и (4.7) в силу леммы П.2 ($a = 0$, $s = 2$) вытекает справедливость (4.8). Выполнение целевого условия (4.2), (4.3) следует из (4.10), (4.11). Действительно, учитывая, кроме того, (4.5), имеем

$$\overline{\lim_{n\to\infty}} \left[\Phi_n^k(\{x_n\}) - \Phi_n^k(\{\hat{x}_n^k\})\right] =$$
$$= \overline{\lim_{n\to\infty}} \frac{1}{n} \sum_{t=1}^{n} \left[V^k(p^*) - V^k\left(p^{1*}, \ldots\right.\right.$$
$$\left.\left.\ldots, p^{(k-1)*}, \tilde{p}_t^k, p^{(k+1)*}, \ldots, p^{l*}\right)\right] \leqslant 0.$$

Лемма доказана. ▲

Таким образом, при построении игровых алгоритмов адаптивного выбора вариантов за основу могут быть взяты рекуррентные процедуры типа (1.6) для нахождения точек равновесия p^* соответствующей матричной игры Г l лиц в смешанных стратегиях.

§ 4.2. Игра двух лиц с нулевой суммой в условиях неопределенности

Рассмотрим сначала наиболее простой случай игровой задачи адаптивного выбора вариантов — игру двух лиц с нулевой суммой в условиях неопределенности. В принятых нами терминах это означает следующее. Множество возможных вариантов есть

$$X \triangleq X_1 \times X_2,$$
$$X_k \triangleq \{x^k(1), \ldots, x^k(N_k)\}, \quad k = 1, 2.$$

Игра состоит из бесконечного числа повторяющихся партий. В партии с номером $n = 1, 2, \ldots$ k-й игрок выбирает чистую стратегию $x_n^k \in X_k$ ($k = 1, 2$) независимо от другого игрока (в смысле выполнения условия независимости выбора ПЗ″). Тем самым производится выбор варианта

$x_n = (x_n^1, x_n^2)$, которым определяются потери игроков $\xi_n^k = \xi_n^k(x_n, \omega)$ в данной партии. Предполагаем, что потери ξ_n^k являются случайными величинами с неизвестными статистическими характеристиками, удовлетворяющими условиям П1″—П3″ из § 4.1 (при $l = 2$). Кроме того, будем считать, что

$$v_{i_1 i_2} \triangleq \mathsf{M}\{\xi_n^1(x_n, \omega) \mid x_n = (x^1(i_1), x^2(i_2))\} = \\ = -\mathsf{M}\{\xi_n^2(x_n, \omega) \mid x_n = (x^1(i_1), x^2(i_2))\} \quad (4.12)$$

для всех $i_k = \overline{1, N_k}$, $k = 1, 2$, $n = 1, 2, \ldots$, т. е. «интересы» игроков в каждой партии противоположны.

В классической постановке матрица игры $\|v_{i_1 i_2}\|$ считается известной, что дает возможность игрокам заранее определить оптимальные смешанные стратегии p^{1*} и p^{2*}, соответствующие седловой точке *) функции потерь первого игрока (выигрышей второго игрока)

$$V(p^1, p^2) = \sum_{i_1, i_2} v_{i_1 i_2} p^1(i_1) p^2(i_2), \\ p^k \in S^{N_k}, \quad k = 1, 2. \quad (4.13)$$

Существование седловой точки, т. е. таких значений $p^{k*} \in S^{N_k}$ $(k = 1, 2)$, для которых

$$V(p^{1*}, p^2) \leqslant V(p^{1*}, p^{2*}) \leqslant V(p^1, p^{2*}) \quad (4.14)$$

при любых $p^k \in S^{N_k}$ $(k = 1, 2)$, гарантируется, как известно, теоремой фон Неймана [80].

В рассматриваемой здесь постановке матрица игры $\|v_{i_1 i_2}\|$ априори неизвестна, но в каждой разыгрываемой партии игроки получают информацию о величинах $v_{i_1 i_2}$ в виде реализаций потерь ξ_n^k $(k = 1, 2)$. Заметим, что при этом игрокам остается неизвестной чистая стратегия, выбранная противником в текущей партии.

С учетом леммы 4.1 игровая задача адаптивного выбора вариантов сводится в данном случае к построению такого алгоритма формирования последовательностей $\{p_n^k\}$ $(k = 1, 2)$ (стохастических векторов условных вероятно-

*) Для матричных игр двух лиц с нулевой суммой понятия «точка равновесия по Нэшу» и «седловая точка» идентичны [86].

стей выбора игроками чистых стратегий), для которого при $n \to \infty$ эти последовательности сходились бы достаточно быстро к седловой точке (p^{1*}, p^{2*}).

Для построения такого алгоритма воспользуемся, как и ранее, методом стохастической аппроксимации. Как уже указывалось в § 3.4, непосредственное использование этого метода для формирования алгоритма отыскания седловой точки билинейной функции приводит к неудаче, которую, однако, можно избежать, если предварительно регуляризовать задачу. Учитывая сказанное, введем параметр регуляризации $\delta > 0$ и регуляризованную функцию средних потерь первого игрока (выигрышей второго игрока)

$$V_\delta(p^1, p^2) \triangleq V(p^1, p^2) + \frac{\delta}{2}(\|p^1\|^2 - \|p^2\|^2). \quad (4.15)$$

Введем также параметр $\varepsilon \in [0, \hat{\varepsilon})$, $\hat{\varepsilon} \triangleq \min\{N_1^{-1}, N_2^{-1}\}$ и ε-симплексы $S_\varepsilon^{N_k}$ в соответствии с (2.9). Рассмотрим теперь игру, в которой k-й игрок выбирает чистую стратегию $x_n^k \in X_k$ на основе смешанной стратегии $p^k \in S_\varepsilon^{N_k}$, а проигрыш первого игрока, равный выигрышу второго, составляет величину $V_\delta(p^1, p^2)$. Существование седловой точки этой игры $(p^{1*}(\varepsilon, \delta), p^{2*}(\varepsilon, \delta))$ при любых фиксированных $\delta \geqslant 0$, $\varepsilon \in [0, \hat{\varepsilon})$ следует из теоремы Исоде — Никайдо [81, 86]. Поскольку при $\delta > 0$ функция $V_\delta(p^1, p^2)$ является строго выпукло-вогнутой, седловая точка $(p^{1*}(\varepsilon, \delta), p^{2*}(\varepsilon, \delta))$ в этом случае единственная. Нерегуляризованную игру (т. е. игру при $\varepsilon = \delta = 0$) будем для краткости называть исходной игрой. Следующие две леммы устанавливают связь регуляризованной игры с исходной и описывают важные свойства точки $(p^{1*}(\varepsilon, \delta), p^{2*}(\varepsilon, \delta))$ как функции (ε, δ).

Лемма 4.2. *Пусть последовательности* $\{\varepsilon_n\}$, $\{\delta_n\}$ *таковы, что*

$$\varepsilon_n \in (0, \hat{\varepsilon}), \quad \delta_n > 0, \quad n = 1, 2, \ldots,$$

$$\lim_{n\to\infty} \delta_n = 0, \quad \lim_{n\to\infty} \frac{\varepsilon_n}{\delta_n} \triangleq \zeta \in [0, \infty).$$

Тогда последовательность точек $(p^{1*}(\varepsilon_n, \delta_n), p^{2*}(\varepsilon_n, \delta_n))$ *сходится при* $n \to \infty$ *к седловой точке исходной игры (зависящей, вообще говоря, от* ζ).

Лемма 4.3. *Пусть последовательности* $\{\varepsilon_n\}$, $\{\delta_n\}$ *удовлетворяют условиям предыдущей леммы. Тогда найдутся положительные константы* C_1, C_2, C_3, *для которых*

$$\sum_{k=1}^{2} \| p^{k*}(\varepsilon_{n+1}, \delta_{n+1}) - p^{k*}(\varepsilon_n, \delta_n) \| \leqslant C_1 |\varepsilon_{n+1} - \varepsilon_n| + \\ + C_2 |\delta_{n+1} - \delta_n| + C_3 \left| \frac{\varepsilon_{n+1}}{\delta_{n+1}} - \frac{\varepsilon_n}{\delta_n} \right|$$

при всех $n = 1, 2, \ldots$

Доказательства этих двух лемм полностью аналогичны доказательствам лемм 3.5 и 3.6 (кроме того, в § 4.4 эти результаты доказываются для более общего случая $l \geqslant 2$), поэтому мы на них останавливаться не будем. Перейдем непосредственно к построению и исследованию алгоритма адаптивного выбора вариантов, предназначенного для решения поставленной в этом параграфе игровой задачи.

§ 4.3. Алгоритм адаптивного выбора вариантов в игре двух лиц с нулевой суммой

Следуя методу стохастической аппроксимации, найдем случайные векторы R_n^1 и R_n^2, условные средние значения которых совпадают с градиентами $\nabla_{p^1} V_\delta(p_n^1, p_n^2)$ и $\nabla_{p^2} V_\delta(p_n^1, p_n^2)$ платежной функции регуляризованной игры. Каждый из этих векторов должен определяться лишь теми величинами, которые доступны наблюдению соответствующему игроку. Поскольку в силу предположений П1″—П3″ из § 4.1 и соотношений (4.6), (4.12)—(4.15) выполняются равенства

$$\nabla_{p^k} V_\delta(p^1, p^2) = \\ = (-1)^{k+1} \mathsf{M} \left\{ \left(\frac{\xi_n^k}{e^{\mathrm{T}}(x_n^k)\, p_n^k} + \delta \right) e(x_n^k) \mid p_n^1 = p^1, p_n^2 = p^2 \right\} \tag{4.16}$$

для всех $k = 1, 2, p^k \in S_\varepsilon^{N_k}$, $\varepsilon \in (0, \hat{\varepsilon})$, $n = 1, 2, \ldots$, то в качестве этих векторов можно взять векторы вида

$$R_n^k \triangleq \left(\frac{\xi_n^k}{e^{\mathrm{T}}(x_n^k)\, p_n^k} + \delta \right) e(x_n^k), \quad k = 1, 2,$$

где $e(x_n^k)$ — орт, соответствующий выбранному элементу $x_n^k \in X_k$. Таким образом, мы приходим к следующему алгоритму адаптивного выбора вариантов:

$$p_{n+1}^k = \pi_{\varepsilon_{n+1}}^{N_k} \left\{ p_n^k - \gamma_n \left(\frac{\xi_n^k}{e^{\mathrm{T}}(x_n^k)\, p_n^k} + \delta_n \right) e(x_n^k) \right\}, \quad k = 1, 2. \tag{4.17}$$

Здесь $\pi_\varepsilon^{N_k}$ — оператор проектирования (2.10), $\{\gamma_n\}$, $\{\varepsilon_n\}$, $\{\delta_n\}$ — последовательности положительных чисел. Следующие два утверждения устанавливают условия сходимости алгоритма (4.17) и дают оценку скорости сходимости.

Теорема 4.1. *Пусть выполнены предположения* П1″ — П3″ *из* § 4.1 (*для* $l = 2$), $p_1^k \in S_{\varepsilon_1}^{N_k}$ ($k = 1, 2$), *а* $\{(p_n^1, p_n^2)\}$ — *последовательность составных векторов, порождаемая алгоритмом* (4.17), *в котором*

$$\gamma_n > 0, \quad \varepsilon_n \in (0, \hat{\varepsilon}), \quad \delta_n > 0 \quad n = 1, 2, \ldots,$$

$$\lim_{n\to\infty} \delta_n = 0, \quad \lim_{n\to\infty} \frac{\varepsilon_n}{\delta_n} \triangleq \zeta \in [0, \infty), \quad \sum_{n=1}^{\infty} \gamma_n \delta_n = \infty.$$

Тогда

1) *если*

$$\sum_{n=1}^{\infty} \left(\frac{\gamma_n^2}{\varepsilon_n} + |\varepsilon_{n+1} - \varepsilon_n| + |\delta_{n+1} - \delta_n| + \left| \frac{\varepsilon_{n+1}}{\delta_{n+1}} - \frac{\varepsilon_n}{\delta_n} \right| \right) < \infty,$$

то с вероятностью 1 *последовательность* $\{(p_n^1, p_n^2)\}$ *сходится при* $n \to \infty$ *к седловой точке* (p^{1*}, p^{2*}), *определяемой условием* (4.14);

2) *если при* $n \to \infty$

$$\left[\frac{\gamma_n}{\varepsilon_n \delta_n} + \frac{1}{\gamma_n \delta_n} \left(|\varepsilon_{n+1} - \varepsilon_n| + |\delta_{n+1} - \delta_n| + \left| \frac{\varepsilon_{n+1}}{\delta_{n+1}} - \frac{\varepsilon_n}{\delta_n} \right| \right) \right] \to 0,$$

то имеет место сходимость $(p_n^1, p_n^2) \to (p^{1*}, p^{2*})$ *в среднеквадратическом.*

З а м е ч а н и е. Седловая точка (p^{1*}, p^{2*}) в случае неединственности зависит, вообще говоря, от величины ζ (см. лемму 4.2).

Теорема 4.2. *Пусть в условиях теоремы 4.1 последовательности* $\{\gamma_n\}$, $\{\varepsilon_n\}$, $\{\delta_n\}$ *имеют вид*

$$\gamma_n = \frac{\gamma}{(n+a)^{\alpha}}, \quad \varepsilon_n = \frac{\varepsilon}{(n+b)^{\beta}}, \quad \delta_n = \frac{\delta}{(n+c)^{\varkappa}}, \quad n = 1, 2, \ldots,$$

где γ, ε, δ, a, b, c — *положительные константы, а параметры* α, β, $\varkappa$ *удовлетворяют неравенствам*

$$\alpha > \beta + \varkappa, \quad \beta \geqslant \varkappa > 0, \quad \alpha + \varkappa \leqslant 1. \tag{4.18}$$

Тогда алгоритм (4.17) *сходится в среднеквадратическом, причем при достаточно большом* γ

$$\overline{\lim_{n\to\infty}}\, n^{\theta} \mathrm{M} \left\{ \sum_{k=1}^{2} \| p_n^k - p^{k*} \|^2 \right\} < \infty,$$

где $p^{k*} \triangleq \lim\limits_{n\to\infty} p^{k*}(\varepsilon_n, \delta_n)$, $k = 1, 2$ (см. лемму 4.2) *и*

$$\theta \triangleq \begin{cases} \min\{2\varkappa, 2(\beta - \varkappa), \alpha - \beta - \varkappa\} & \text{при} \quad \beta \neq \varkappa, \\ \min\{\alpha - 2\beta, 2\beta\} & \text{при} \quad \beta = \varkappa. \end{cases}$$

Если вместо (4.18) *выполняются неравенства*

$$\beta \geqslant \varkappa > 0, \quad \alpha + \varkappa \leqslant 1, \quad 2\alpha - \beta > 1,$$

то алгоритм (4.17) *сходится с вероятностью* 1.

Следствие 1. *При выполнении условий теоремы* 4.2, *обеспечивающих сходимость в среднеквадратическом, в силу леммы* 4.1 *выполняется целевое условие* (4.2), (4.3) *и справедлива следующая оценка скорости сходимости текущих средних потерь* (4.1):

$$\overline{\lim_{n\to\infty}}\, n^{\theta} \mathrm{M} \{ [\Phi_n^k - V^k(p^{1*}, p^{2*})]^2 \} < \infty, \quad k = 1, 2.$$

Следствие 2. *В условиях следствия* 1 *теоремы* 4.2 *для параметра* θ, *характеризующего порядок скорости сходимости алгоритма* (4.17), *справедливо неравенство*

$$\theta \leqslant \frac{2}{5};$$

причем $\theta = 2/5$ *лишь при* $\alpha = \alpha^* = 4/5$, $\beta = \beta^* = \varkappa = \varkappa^* = 1/5$.

Мы не приводим доказательств этих утверждений, так как, с одной стороны, они полностью аналогичны доказательствам теорем 3.3, 3.4, а с другой стороны, полученные в этом параграфе результаты обобщаются далее на игровые задачи с произвольным числом игроков.

§ 4.4. Игра многих лиц

Пусть теперь множество возможных вариантов есть

$$X \triangleq X_1 \times \ldots \times X_l,$$

$$X_k \triangleq \{x^k(1), \ldots, x^k(N_k)\}, \quad k = \overline{1, l}.$$

Как и ранее, будем предполагать, что игра состоит из бесконечной последовательности повторяющихся партий и в n-й партии ($n = 1, 2, \ldots$) k-й игрок выбирает чистую стратегию $x_n^k \in X_k$ $(k = \overline{1, l})$ независимо от других игроков (в смысле П3″). Полученный при этом вариант $x_n = \left(x_n^1, \ldots, x_n^l\right)$ определяет потери игроков $\xi_n^k \triangleq \xi_n^k(x_n, \omega)$ в данной партии, которые являются случайными величинами с априори неизвестными статистическими характеристиками, удовлетворяющими условиям П1″—П3″ из § 4.1. В классической постановке задачи о бескоалиционной игре многих лиц условные средние потери игроков $v_{i_1 \ldots i_l}^k$ (см. § 4.1), соответствующие каждому варианту $(x^1(i_1), \ldots, x^l(i_l))$, известны заранее, что позволяет до начала игры найти равновесные по Нэшу смешанные стратегии $p^* \triangleq (p^{1*}, \ldots, p^{l*})$, определяемые условием (4.5).

Выбор игроками элементов $x^k(i_k) \in X_k$ в соответствии с равновесными смешанными стратегиями p^{k*} обеспечил бы выполнение целевого условия (4.2), (4.3) (см. лемму 4.1). В нашем случае каждому игроку известна лишь его чистая стратегия, которую он выбрал в текущей партии, а также полученные в результате потери ξ_n^k. По этой информации игроки должны с течением времени определить равновесные смешанные (в общем случае) стратегии выбора вариантов. В этом состоит суть рассматриваемого подхода.

Так как в частном случае — в игре двух лиц с нулевой суммой — при формировании игрового алгоритма приходится рассматривать регуляризованную игру, проведем и в общем случае регуляризацию игры. Пусть при смешанных стратегиях $p^k \in S_\varepsilon^{N_k}$ $(k = \overline{1, l})$ — средние потери k-го игрока составляют

$$V_\delta^k(p) \triangleq V^k(p) + \frac{\delta}{2}\|p^k\|^2, \tag{4.19}$$

где $\delta > 0$, $\varepsilon \in [0, \hat{\varepsilon})$ — параметры регуляризации, $\hat{\varepsilon} \triangleq \min\{N_k^{-1} \mid k = \overline{1, l}\}$, $p = (p^1, \ldots, p^l)$ — составной вектор смешанных стратегий. Тогда множество точек равновесия такой игры

$$P^*_{\varepsilon\delta} \triangleq \{p = (p^1, \ldots, p^l) \mid p^k \in S_\varepsilon^{N_k};$$
$$V^k_\delta(p) \leqslant w^k_\delta(p, q^k) \ \forall q^k \in S_\varepsilon^{N_k},\ k = \overline{1, l}\},$$

зависящее, вообще говоря, от ε и δ, не пусто [107]. Здесь

$$\begin{gathered} w^k_\delta(p, q^k) \triangleq w^k(p, q^k) + \frac{\delta}{2}\|q^k\|^2, \\ w^k(p, q^k) \triangleq V^k\ (p^1, \ldots, p^{k-1}, q^k, p^{k+1}, \ldots, p^l). \end{gathered} \tag{4.20}$$

Нерегуляризованную игру (т. е. игру при $\varepsilon = \delta = 0$) будем, как и прежде, называть *исходной игрой*. Рассмотрим сначала важный класс игр многих лиц — так называемые *диагонально выпуклые игры* [107].

Исходную матричную игру Г (§ 4.1) будем называть диагонально выпуклой, если существуют такие положительные числа $r_k > 0$ $(k = \overline{1, l})$, что для любых p, $q \in S \triangleq S^{N_1} \times \ldots \times S^{N_l}$ выполняется неравенство

$$W(p, q) \triangleq \sum_{k=1}^{l} r_k [V^k(p) - w^k(p, q^k) + V^k(q) - w^k(p, q^k)] \geqslant 0. \tag{4.21}$$

Простейшими примерами диагонально выпуклых матричных игр являются, очевидно, следующие: 1) игра, в которой потери каждого участника не зависят от стратегий, выбираемых остальными игроками (когда, другими словами, каждый из участников взаимодействует только со «своей природой»), 2) игра, в которой, наоборот, потери каждого участника не зависят от его собственных стратегий и определяются лишь стратегиями остальных игроков и, наконец, 3) матричная игра двух лиц с нулевой суммой. Забегая вперед, скажем, что любую диагонально выпуклую матричную игру можно представить как некую суперпозицию игр этих трех видов (см. далее лемму 4.6).

Сформулируем и докажем ряд утверждений, касающихся свойств диагонально выпуклых матричных игр.

Лемма 4.4. *Матричная игра l лиц удовлетворяет условию диагональной выпуклости* (4.21) *тогда и только тогда, когда существуют числа $r_k > 0$ $(k=\overline{1, l})$, такие, что $W(p, q)=0$ для всех p и q из множества*

$$\tilde{S} \triangleq \left\{ p=(p^1, \ldots, p^l) \;\middle|\; \sum_{i_k=1}^{N_k} p^k_{i_k} = 1 \ (k=\overline{1, l}) \right\}.$$

Доказательство. Если $W(p, q)=0$ для всех $p, q \in \tilde{S}$, то (4.21) тем более выполняется, так как $S \subset \tilde{S}$. Докажем обратное. Пусть дано, что игра диагонально выпукла. Выберем два произвольных сочетания индексов $I \triangleq (i_1, \ldots, i_l)$, $J \triangleq (j_1, \ldots, j_l)$, $i_k, j_k = \overline{1, N_k}$ $(k=\overline{1, l})$. Тогда при $p^k = e^{N_k}_{i_k}$, $q^k = e^{N_k}_{j_k}$ из (4.19)—(4.20) получаем

$$W_{I,J} \triangleq \sum_{k=1}^{l} r_k \left(v^k_{i_1 \ldots i_l} - v^k_{i_1 \ldots i_{k-1} j_k i_{k+1} \ldots i_l} + \right.$$
$$\left. + v^k_{j_1 \ldots j_l} - v^k_{j_1 \ldots j_{k-1} i_k j_{k+1} \ldots j_l} \right) \geqslant 0. \quad (4.22)$$

Рассмотрим теперь множество всевозможных пар (A, B) сочетаний индексов $A=(a_1, \ldots, a_l)$, $B=(b_1, \ldots, b_l)$, у которых либо $a_k = i_k$, $b_k = j_k$, либо $a_k = j_k$, $b_k = i_k$ $(k=\overline{1, l})$. Это множество обозначим $T(I, J)$. Очевидно,

$$\sum_{(A,B) \in T(I,J)} W_{A,B} = 0.$$

Но по доказанному $W_{A,B} \geqslant 0$ для любых A, B, поэтому $W_{A,B}=0$ для всех $(A, B) \in T(I, J)$ и, в частности, $W_{I,J}=0$. Учитывая произвольность I и J и представление

$$W(p, q) = \sum_{I,J} W_{I,J} \prod_{\nu=1}^{l} p^{\nu}_{i_\nu} q^{\nu}_{j_\nu},$$

которое с очевидностью следует из (4.4), (4.20)—(4.22) при $p, q \in \tilde{S}$, получим $W(p, q)=0$. Лемма доказана. ▲

Лемма 4.5. *Множество P^* точек равновесия по Нэшу* (4.5) *матричной игры l лиц, удовлетворяющей условию диагональной выпуклости* (4.21), *замкнуто и выпукло.*

Доказательство. Замкнутость P^* вытекает из непрерывности $V^k(p)$, $p \in S$ (4.4). Докажем выпуклость P^*. Пусть $p, q \in P^*$, $\lambda \in [0, 1]$. Рассмотрим $z(\lambda) = \lambda p + (1-\lambda) q \in S \subset \tilde{S}$. Тогда из леммы 4.4, условия

(4.21) и свойств функций $V^k(p)$ получим для любого $y \in S$ следующие соотношения:

$$\sum_{k=1}^{l} r_k \left[V^k(z(\lambda)) - w^k(z(\lambda), y^k)\right] =$$

$$= \sum_{k=1}^{l} r_k \left[w^k(y, z^k(\lambda)) - V^k(y)\right] =$$

$$= \lambda \sum_{k=1}^{l} r_k \left[w^k(y, p^k) - V^k(y)\right] +$$

$$+ (1-\lambda) \sum_{k=1}^{l} r_k \left[w^k(y, q^k) - V^k(y)\right] =$$

$$= \lambda \sum_{k=1}^{l} r_k \left[V^k(p) - w^k(p, y^k)\right] +$$

$$+ (1-\lambda) \sum_{k=1}^{l} r_k \left[V^k(q) - w^k(q, y^k)\right] \leqslant 0.$$

Последнее неравенство вытекает из определения точки Нэша (4.5). Для каждого $\nu = \overline{1, l}$ положим $y^k = z^k(\lambda)$ при $k \neq \nu$. Получим

$$r_\nu \left[V^\nu(z(\lambda)) - w^\nu(z(\lambda), y^\nu)\right] \leqslant 0 \quad \forall y^\nu \in S^{N_\nu}.$$

Так как все $r_\nu > 0$, получаем $z(\lambda) \in P^*$. Лемма доказана. ▲

Таким образом, свойства диагонально выпуклых матричных игр многих лиц схожи со свойствами игр двух лиц с нулевой суммой: в обоих случаях множество точек равновесия в смешанных стратегиях является замкнутым выпуклым множеством. Следующая лемма раскрывает структуру матричных игр многих лиц, удовлетворяющих условию диагональной выпуклости (4.21). Как мы увидим, важное место в этой структуре занимают игры двух лиц с нулевой суммой.

Лемма 4.6. *Матричная игра l лиц удовлетворяет условию диагональной выпуклости* (4.21) *тогда и только тогда, когда найдутся числа $r_k > 0$, $a^k_{i_k}$, $b^{k\nu}_{i_k i_\nu}$ и*

$$c^k_{i_1 \ldots i_{k-1} i_{k+1} \ldots i_l} \; (k, \nu = \overline{1, l}, \; k \neq \nu, \; i_\nu = \overline{1, N_\nu}, \; i_k = \overline{1, N_k}),$$

для которых при всех $p \in S$ *и* $k = \overline{1, l}$

$$V^k(p) = \sum_{i_k=1}^{N_k} a^k_{i_k} p^k_{i_k} + \frac{1}{r_k} \sum_{\substack{\nu=1 \\ \nu \neq k}}^{l} \sum_{i_k=1}^{N_k} \sum_{i_\nu=1}^{N_\nu} b^{k\nu}_{i_k i_\nu} p^k_{i_k} p^\nu_{i_\nu} + \\ + \sum_{\substack{i_1, \ldots, i_{k-1} \\ i_{k+1}, \ldots, i_l}} c^k_{i_1 \ldots i_{k-1} i_{k+1} \ldots i_l} \prod_{\substack{\nu=1 \\ \nu \neq k}}^{l} p^\nu_{i_\nu}, \tag{4.23}$$

причем $b^{k\nu}_{i_k i_\nu} + b^{\nu k}_{i_\nu i_k} = 0$ *для любых* k, $\nu = \overline{1, l}$, $k \neq \nu$, $i_k = \overline{1, N_k}$, $i_\nu = \overline{1, N_\nu}$.

Доказательство. Нетрудно убедиться, что из (4.23) следует (4.21). Установим теперь обратное. Пусть выполняется (4.21). Тогда в силу леммы 4.4 найдутся числа $r_k > 0$ $(k = \overline{1, l})$, для которых при всех p, $q \in \tilde{S}$

$$U(p, q) + U(q, p) = 0, \tag{4.24}$$

где

$$U(p, q) \triangleq \sum_{k=1}^{l} r_k \left[V^k(p) - w^k(p, q^k) \right]. \tag{4.25}$$

Поскольку p, $q \in \tilde{S}$, т. е.

$$\sum_{i_k=1}^{N_k} p_{i_k} = 1, \qquad \sum_{i_k=1}^{N_k} q^k_{i_k} = 1,$$

исключим в выражении для $U(p, q)$ переменные $p^k_{N_k}$ и $q^k_{N_k}$ $(k = \overline{1, l})$. Получим функцию переменных $\tilde{p} \triangleq (\tilde{p}^1, \ldots, \tilde{p}^l)$ и $\tilde{q} \triangleq (\tilde{q}^k, \ldots, \tilde{q}^l)$; где $\tilde{p}^k \triangleq (p^k_1, \ldots, p^k_{N_k-1})$, $\tilde{q}^k \triangleq (q^k_1, \ldots, q^k_{N_k-1})$. Обозначим эту функцию через $\tilde{U}(\tilde{p}, \tilde{q})$. Заметим, что $\tilde{U}(\tilde{p}, \tilde{q}) = U(p, q)$, поэтому из (4.24) получим

$$\tilde{U}(\tilde{p}, \tilde{q}) + \tilde{U}(\tilde{q}, \tilde{p}) = 0.$$

Отсюда и из (4.25), (4.4) и (4.20) следует, что $\tilde{U}(\tilde{p}, \tilde{q})$ является линейной функцией по переменным $\tilde{p}$ при любом фиксированном значении $\tilde{q}$ и также линейной по переменным $\tilde{q}$ при фиксированном $\tilde{p}$, причем ее выражение не содержит произведений вида $p^{\nu}_{i_\nu}\cdot q^{\nu}_{j_\nu}$, т. е.

$$\tilde{U}(\tilde{p}, \tilde{q}) = u_0 + \sum_{k=1}^{l}\sum_{i_k=1}^{N_k-1}\left(\alpha^k_{i_k}p^k_{i_k} - \overline{\alpha}^k_{i_k}q^k_{i_k}\right) + \\ + \sum_{k=1}^{l}\sum_{\substack{\nu=1\\ \nu\neq k}}^{l}\sum_{i_k=1}^{N_k-1}\sum_{i_\nu=1}^{N_\nu-1}\beta^{k\nu}_{i_k i_\nu}p^k_{i_k}q^{\nu}_{i_\nu}, \tag{4.26}$$

где u_0, $\alpha^k_{i_k}$, $\overline{\alpha}^k_{i_k}$, $\beta^{k\nu}_{i_k i_\nu}$ — некоторые числа. Но из определения функции $\tilde{U}(\tilde{p}, \tilde{q})$ следует, что

$$\tilde{U}(\tilde{p}, \tilde{p}) = 0 \quad \forall \tilde{p}. \tag{4.27}$$

Отсюда и из (4.26) получим

а) при $\tilde{p} = 0$:

$$u_0 = 0, \tag{4.28}$$

б) при $p^{\nu}_{i_\nu} = \delta_{i_\nu j_\nu}(i_\nu = \overline{1, N_\nu - 1})$, $\tilde{p}^k = 0\,(k\neq\nu, k = \overline{1, l})$:

$$\alpha^{\nu}_{j_\nu} = \overline{\alpha}^{\nu}_{j_\nu}\left(\nu = \overline{1, l},\ j_\nu = \overline{1, N_\nu - 1}\right), \tag{4.29}$$

в) при $p^{\nu}_{i_\nu} = \delta_{i_\nu j_\nu}(i_\nu = \overline{1, N_\nu - 1})$, $p^{\tau}_{i_\tau} = \delta_{i_\tau j_\tau}(i_\tau = \overline{1, N_\tau - 1})$, $\tilde{p}^k = 0\ (k\neq\nu,\ k\neq\tau,\ k = \overline{1, l})$:

$$\beta^{\nu\tau}_{j_\nu j_\tau} + \beta^{\tau\nu}_{j_\tau j_\nu} = 0 \tag{4.30}$$

$$(\nu\neq\tau,\ \nu = \overline{1, l},\ \tau = \overline{1, l},\ j_\nu = \overline{1, N_\nu - 1},\ j_\tau = \overline{1, N_\tau - 1}).$$

Здесь δ_{ij} — символ Кропекера. Положим теперь для произвольного $\nu = \overline{1, l}$ и $p\in S$ $\tilde{q}^{\nu} = 0$, $\tilde{q}^k = \tilde{p}^k$ $(k\neq\nu,\ k = \overline{1, l})$. Тогда из (4.25), (4.26), учитывая (4.27), (4.28) и равенство $\tilde{U}(\tilde{p}, \tilde{q}) = U(p, q)$, получим

$$r_\nu\left[V^{\nu}(p) - w^{\nu}\left(p, e^{\nu}_{N_\nu}\right)\right] = \\ = \sum_{i_\nu=1}^{N_\nu-1}\alpha^{\nu}_{i_\nu}p^{\nu}_{i_\nu} + \sum_{k=1}^{l}\sum_{\substack{t=1\\ t\neq k\\ t\neq\nu}}^{l}\sum_{i_k=1}^{N_k-1}\sum_{i_t=1}^{N_t-1}\beta^{kt}_{i_k i_t}p^k_{i_k}p^t_{i_t}.$$

Отсюда и из (4.4), (4.20) следует

$$V^{\nu}(p)=\frac{1}{r_\nu}\left(\sum_{i_\nu=1}^{N_\nu-1}\alpha^{\nu}_{i_\nu}p^{\nu}_{i_\nu}+\sum_{\substack{t=1\\t\neq\nu}}^{l}\sum_{i_\nu=1}^{N_\nu-1}\sum_{i_t=1}^{N_t-1}\beta^{\nu t}_{i_\nu i_t}p^{\nu}_{i_\nu}p^{t}_{i_t}\right)+$$
$$+\frac{1}{r_\nu}\sum_{\substack{k=1\\k\neq\nu}}^{l}\sum_{\substack{t=1\\t\neq k\\t\neq\nu}}^{l}\sum_{i_k=1}^{N_k-1}\sum_{i_t=1}^{N_t-1}\beta^{kt}_{i_k i_t}p^{k}_{i_k}p^{t}_{i_t}+\sum_{\substack{i_1,\ldots,i_l\\i_\nu=N_\nu}}v^{\nu}_{i_1,\ldots,i_l}\prod_{\substack{t=1\\t\neq\nu}}^{l}p^{t}_{i_t}. \tag{4.31}$$

Определим следующие величины:

$$a^k_{i_k}=\begin{cases}\dfrac{1}{r_k}\alpha^k_{i_k}, & i_k=\overline{1,N_k-1},\\ 0, & i_k=N_k,\end{cases}$$

$$b^{kt}_{i_k i_t}=\begin{cases}\beta^{kt}_{i_k i_t}, & i_k=\overline{1,N_k-1},\quad i_t=\overline{1,N_t-1},\\ 0, & i_k=N_k,\\ 0, & i_t=N_t,\end{cases}$$

$$c^k_{i_1\ldots i_{k-1}i_{k+1}\ldots i_l}=v^k_{i_1\ldots,i_{k-1}N_k i_{k+1}\ldots i_l}+\frac{1}{r_k}\sum_{\substack{\nu=1\\\nu\neq k}}^{l}\sum_{\substack{t=1\\t\neq k\\t\neq\nu}}^{l}b^{\nu t}_{i_\nu i_t},$$

где $k,\ t=\overline{1,\ l}$, $t\neq k$, $i_k=\overline{1,\ N_k}$, $i_t=\overline{1,\ N_t}$. Тогда из (4.31) с учетом нормировки

$$\sum_{i_k=1}^{N_k}p^k_{i_k}=1,\quad k=\overline{1,l}$$

получим (4.23) (с точностью до переобозначения индексов). Требуемое свойство чисел $b^{kt}_{i_k i_t}$ следует из (4.30) и их определения. Лемма доказана. ▲

Прокомментируем результат леммы 4.6. Первое слагаемое в выражении средних потерь k-го игрока (4.23) зависит только от смешанных стратегий этого игрока. Этот член можно трактовать как потери игрока в некоторой игре с природой. Второе слагаемое в (4.23) показывает, что k-й игрок участвует также с каждым другим игроком (с номером $t=\overline{1,\ l}$, $t\neq k$) в игре с нулевой суммой (матрица этой игры есть $\|b^{kt}_{i_k i_t}\|$). Наконец, третье слагаемое в (4.23) соответствует такой игре l лиц, в которой потери k-го игрока определяются лишь чужими чистыми стратегиями и не зависит от чистых стратегий самого этого игрока. Таким образом, лемма 4.6 утверждает, что диагонально выпуклая матричная игра Г многих лиц может быть представлена как определенная суперпозиция простейших диагонально выпуклых матричных игр: 1) игры с природой, 2) игры двух лиц с нулевой суммой и 3) игры многих лиц, в которой потери каждого игрока не зависят от его собственных чистых стратегий. Отсюда, в частности, следует, что матричная игра двух лиц с нулевой суммой является диагонально выпуклой; это, впрочем, легко установить непосредственно из (4.21).

Перейдем теперь к изучению свойств точки равновесия регуляризованной игры, соответствующей матричной диагонально выпуклой игре. Итак, пусть исходная матричная игра многих лиц удовлетворяет условию диагональной выпуклости (4.21). Из определения функций $V^k(p)$, $V^k_\delta(p)$ (4.4) и (4.19), условия (4.21) и леммы 4.4 следует, что для любых $p, q \in S_\varepsilon \triangleq S_\varepsilon^{N_1} \times \ldots \times S_\varepsilon^{N_l}$ справедливы соотношения

$$\sum_{k=1}^{l} r_k \langle \nabla_{p^k} V^k_\delta(p) - \nabla_{q^k} V^k_\delta(q), p^k - q^k \rangle = \delta \sum_{k=1}^{l} r_k \| p^k - q^k \|^2 \geqslant 0,$$

причем равенство нулю здесь возможно лишь при $p = q$ $(\delta > 0)$. В [107] такие игры названы *строго диагонально выпуклыми* и установлено, что они имеют точку равновесия, причем единственную. Таким образом, множество $P^*_{\varepsilon\delta}$ при $\delta > 0$ состоит из единственной точки, которую обозначим $p^*(\varepsilon, \delta) = (p^{1*}(\varepsilon, \delta), \ldots, p^{l*}(\varepsilon, \delta))$. Докажем, что эта точка как функция параметров ε и δ имеет такие же свойства, как и седловая точка регуляризованной матричной игры двух лиц с нулевой суммой (см. леммы 4.2, 4.3).

Лемма 4.7. *Пусть исходная матричная игра многих лиц является диагонально выпуклой* (4.21), *а числовые последовательности* $\{\varepsilon_n\}$, $\{\delta_n\}$ *удовлетворяют условиям леммы 4.2. Тогда последовательность точек равновесия* $p^*(\varepsilon_n, \delta_n)$ *регуляризованной игры сходится при* $n \to \infty$ *к точке равновесия исходной игры (эта точка зависит, вообще говоря, от значения* $\zeta \triangleq \lim_{n\to\infty} \varepsilon_n/\delta_n$).

Доказательство. Пусть p^* — частичный предел последовательности $\{p^*(\varepsilon_n, \delta_n)\}$, соответствующий подпоследовательности номеров $\{n_t\}$, т. е.

$$\lim_{t\to\infty} p^*\left(\varepsilon_{n_t}, \delta_{n_t}\right) = p^*.$$

Из свойств функций $V^\nu(p)$, $V^\nu_\delta(p)$ (4.4), (4.19) и определения точки равновесия по Нэшу с необходимостью следует, что p^* является точкой равновесия исходной матричной игры. Покажем теперь единственность частичного предела последовательности точек $\overset{*}{p}_n \triangleq p^*(\varepsilon_n, \delta_n)$. Так как для любого $p \in P^* \subseteq S$ (где P^* — множество точек равновесия исходной игры) точки

$$\overset{\nu}{p}_n \triangleq p^\nu - \varepsilon_n\left(N_\nu p^\nu - e^{N_\nu}\right)$$

лежат в соответствующих симплексах $S^{N_\nu}_{\varepsilon_n}$ $(\nu = \overline{1, l})$, то, используя свойства точки равновесия и условие диагональной выпуклости, получим

$$0 \geqslant \sum_{\nu=1}^{l} r_\nu \langle \nabla_{p^\nu} V^\nu_{\delta_n}(\overset{*}{p}_n), \overset{\nu*}{p}_n - \overset{\nu}{p}_n \rangle \geqslant$$

$$\geqslant \varepsilon_n \sum_{\nu=1}^{l} r_\nu N_\nu\Big[w^\nu\left(p, e^{N_\nu}/N_\nu\right) - w^\nu(p, \overset{\nu*}{p}_n) -$$

$$- w^\nu\left(\overset{*}{p}_n, e^{N_\nu}/N_\nu\right) + w^\nu(\overset{*}{p}_n, p^\nu)\Big] + \delta_n \sum_{\nu=1}^{l} r_\nu \langle \overset{\nu*}{p}_n, \overset{\nu*}{p}_n - \overset{\nu}{p}_n \rangle.$$

Разделив это неравенство на δ_n и переходя к пределу по подпоследовательности $\{n_t\}$ при $t \to \infty$, получим

$$0 \geqslant \zeta \sum_{\nu=1}^{l} r_\nu N_\nu\Big[w^\nu\left(p, e^{N_\nu}/N_\nu\right) - V^\nu(p) + V^\nu(p^*) -$$

$$- w^\nu\left(p^*, e^{N_\nu}/N_\nu\right)\Big] + \sum_{\nu=1}^{l} r_\nu \langle p^{\nu*}, p^{\nu*} - p^\nu \rangle. \quad (4.32)$$

Отметим, что здесь использовалось равенство $w^\nu(p^*, p^\nu) = V^\nu(p^*)$, $\nu = \overline{1, l}$ для любых p^*, $p \in P^*$, которое следует из условия диагональной выпуклости (4.21) и определения точки равновесия. Полученное неравенство, как это следует из проведенных рассуждений, выполняется для любых $p \in P^*$ и любых предельных точек p^* последовательности $\{\overset{*}{p}_n\}$. Предположим теперь, что эта последовательность имеет две различные предельные точки $\tilde{p}$ и $\hat{p}$. Как уже указывалось, имеем $\tilde{p}$, $\hat{p} \in P^*$, поэтому, используя (4.32) один раз при $p^* = \tilde{p}$ и $p = \hat{p}$, а другой раз при $p^* = \hat{p}$ и $p = \tilde{p}$ и складывая полученные таким образом два неравенства, будем иметь

$$0 \geqslant \sum_{\nu=1}^{l} r_\nu \| \tilde{p}^\nu - \hat{p}^\nu \|^2,$$

т. е. $\tilde{p} = \hat{p}$. Тем самым установлена единственность предельной точки p^* последовательности $\{p_n^*\}$, а значит, $\lim_{n\to\infty} p_n^* = p^*$. Лемма доказана. ▲

Лемма 4.8. *Пусть выполнены условия леммы 4.7. Тогда для некоторого $C > 0$ и момента времени $n_0 \geqslant 1$ при любых n, $t \geqslant n_0$ и $\nu = \overline{1, l}$ справедливо неравенство*

$$\| p^{\nu *}(\varepsilon_n, \delta_n) - p^{\nu *}(\varepsilon_t, \delta_t) \| \leqslant$$
$$\leqslant C \left(| \varepsilon_n - \varepsilon_t | + | \delta_n - \delta_t | + \left| \frac{\varepsilon_n}{\delta_n} - \frac{\varepsilon_t}{\delta_t} \right| \right). \quad (4.33)$$

Доказательство этой леммы во многом аналогично доказательству леммы 3.5, поэтому проведем его кратко, акцентируя внимание на отличительных моментах. Пусть (в силу леммы 4.6) функция средних потерь k-го игрока представлена в виде (4.23) ($k = \overline{1, l}$). Рассмотрим какую-либо задачу Z_i нахождения точки равновесия в игре l лиц, в которой функция потерь k-го игрока есть $V_\delta^k(p)$ (4.19), $p^k = \left(p_1^k, \ldots, p_{N_k}^k\right)$ — вектор его (чистых) стратегий, выбираемых из множества D_i^k; это множество получается заменой в определении ε-симплекса

$$S_\varepsilon^{N_k} = \left\{ p^k \,\middle|\, \sum_{i_k=1}^{N_k} p_{i_k}^k = 1,\ p_{i_k}^k \geqslant \varepsilon \ (i_k = \overline{1, N_k}) \right\}$$

ряда неравенств $p_{i_k}^k \geqslant \varepsilon$ равенствами $p_{i_k}^k = \varepsilon$ и отбрасыванием остальных неравенств (т. е. это одно из следующих множеств:

$$\left\{ p^k \,\middle|\, \sum_{i_k=1}^{N_k} p_{i_k}^k = 1 \right\}, \quad \left\{ p^k \,\middle|\, \sum_{i_k=1}^{N_k} p_{i_k}^k = 1,\ p_1^k = \varepsilon \right\},$$

$$\left\{p^k \,\middle|\, \sum_{i_k=1}^{N_k} p_{i_k}^k = 1,\ p_1^k = \varepsilon,\ p_{N_k}^k = \varepsilon\right\} \text{ и т.п.}).$$

Число таких задач равно $K = \prod_{k=1}^{l}\left(2^{N_k} - 1\right)$, т. е. $i = \overline{1, K}$. Учитывая (4.19), лемму 4.6 и используя метод множителей Лагранжа так же, как это делалось при доказательстве леммы 3.5, получим для определения точки равновесия $p^*(Z_i)$ и множителей Лагранжа $\lambda_r^*(Z_i)$ в задаче Z_i систему линейных алгебраических уравнений, определитель которой зависит от параметра δ и при $\delta > 0$ отличен от нуля (так как при $\delta > 0$ система имеет единственное решение в силу единственности точки равновесия $p^*(Z_i)$). Те уравнения этой системы, которые получаются приравниванием нулю частных производных функций Лагранжа по переменным $p_{i_k}^k$, можно записать в векторном виде следующим образом:

$$(B + \delta I)p + G^i\lambda^i = a, \tag{4.34}$$

где $p = (p^1, \ldots, p^l)$, $p^k = \left(p_1^k, \ldots, p_{N_k}^k\right)$ $(k = \overline{1, l})$, λ^i — вектор множителей Лагранжа в задаче Z_i, G^i и a — некоторые матрицы и вектор, значения и свойства которых нам не важны, I — единичная, а B — кососимметричная (в силу леммы 4.6) матрицы размером $N_0 \times N_0$, где $N_0 \triangleq \sum_{k=1}^{l} N_k$. Остальные уравнения системы суть ограничения, определяющие множества D_i^k $(k = \overline{1, l})$. Из вида этой системы непосредственно следует, что каждая компонента вектора $p^*(Z_i)$ как функция параметров (ε, δ) выражается соотношением вида

$$p_{i_k}^{k*}(Z_i) = \sum_{r=0}^{N_0} \delta^r (g_r + h_r\varepsilon) \Big/ \sum_{r=0}^{N_0} d_r\delta^r, \tag{4.35}$$

где коэффициенты g_r, h_r зависят также и от индексов k и i_k (для наших рассуждений эта зависимость не важна), а знаменатель как полином по δ не имеет корней при $\delta > 0$. Соответствующие задаче Z_i множители Лагранжа $\lambda_r^*(Z_i)$ представимы также в виде (4.35) с тем же полиномом в знаменателе. Заметим теперь, что при любых $\delta > 0$ и $\varepsilon \in (0, \hat{\varepsilon})$ точка равновесия $p^*(\varepsilon, \delta)$ регуляризованной игры совпадает с точкой равновесия $p^*(Z_i)$ одной из задач Z_i. Учитывая лемму 4.7, получим, что существует такой номер $n_0 \geqslant 1$, для которого при любом $n \geqslant n_0$ точка $p^*(\varepsilon_n, \delta_n)$ совпадает с точкой $p^*(Z_i)$ при $\varepsilon = \varepsilon_n$, $\delta = \delta_n$ только такой задачи Z_i, у которой в (4.35) либо $d_0 \neq 0$, либо $d_0 = g_0 = 0$, $d_1 \neq 0$. Но в обоих этих случаях выражение, стоящее в правой части (4.35), как функция параметров $\varepsilon \in (0, \hat{\varepsilon})$, $\delta > 0$ и их отношения ε/δ, имеет константу Липшица. Отсюда и следует (4.33). Лемма доказана. ▲

Следствие. *В условиях леммы 4.7 при $n \geqslant n_0$ справедливы неравенства*

$$\| p^{\nu *}(\varepsilon_n, \delta_n) - p^{\nu *}(\varepsilon_{n+1}, \delta_{n+1}) \| \leqslant$$

$$\leqslant C \left(|\varepsilon_n - \varepsilon_{n+1}| + |\delta_n - \delta_{n+1}| + \left| \frac{\varepsilon_n}{\delta_n} - \frac{\varepsilon_{n+1}}{\delta_{n+1}} \right| \right). \quad (4.36)$$

$$\| p^{\nu *}(\varepsilon_n, \delta_n) - p^* \| \leqslant C \left(\varepsilon_n + \delta_n + \left| \frac{\varepsilon_n}{\delta_n} - \zeta \right| \right), \quad (4.37)$$

где $p^* = \lim\limits_{n \to \infty} p^*(\varepsilon_n, \delta_n)$, $\zeta = \lim\limits_{n \to \infty} \varepsilon_n/\delta_n$, $C > 0$.

Неравенства (4.36), (4.37) получаются из (4.33) при $t = n + 1$ и $t \to \infty$ соответственно.

Перейдем теперь к обобщению алгоритма (4.17) на случай игры многих лиц и к изучению его свойств.

§ 4.5. Алгоритм адаптивного выбора вариантов в игре многих лиц (случай диагональной выпуклости)

Заметим сразу, что для функций $V_\delta^k(p)$ (4.19) в силу (4.6) и предположений П1″ — П3″ из § 4.1 справедливы равенства, аналогичные (4.16):

$$\nabla_{p^k} V_\delta^k(p) = \mathsf{M} \left\{ \left(\frac{\xi_n^k}{e^{\mathrm{T}}(x_n^k)\, \bar{p}_n^k} + \delta \right) e(x_n^k) \,\middle|\, p_n = p \right\} \quad (4.38)$$

для всех $k = \overline{1, l}$, $p^k \in S_\varepsilon^{N_k}$ $\left(\varepsilon \in (0, \hat{\varepsilon}),\quad \hat{\varepsilon} \triangleq \min\limits_{k=\overline{1,l}} N_k^{-1} \right)$, $n = 1, 2, \ldots$ Отсюда, используя, как и ранее, метод стохастической аппроксимации, приходим к следующему игровому алгоритму адаптивного выбора вариантов:

$$p_{n+1}^k = \pi_{\varepsilon_n}^{N_k}\left\{p_n^k - \gamma_n\left(\frac{\xi_n^k}{e^{\mathrm{T}}(x_n)\,p_n^k} + \delta_n\right)e\left(x_n^k\right)\right\}, \tag{4.39}$$

$$p_1^k \in S_{\varepsilon_1}^{N_k}\left(k = \overline{1, l}\right).$$

Здесь последовательности положительных чисел $\{\gamma_n\}$, $\{\varepsilon_n\}$, $\{\delta_n\}$ и проектор $\pi_\varepsilon^{N_k}$ имеют тот же смысл, что и в алгоритме (4.17). Установим условия сходимости алгоритма (4.39) и получим оценку скорости его сходимости.

Теорема 4.3. *Пусть исходная матричная игра l лиц диагонально выпукла, выполнены предположения* П1″ — П3″ *из* § 4.1, $p_1^k \in S_{\varepsilon_1}^{N_k}$ $(k = \overline{1, l})$, *а последовательность* $\{p_n\}$ *составных векторов* $p_n = \left(p_n^1, \ldots, p_n^l\right)$ *порождается алгоритмом* (4.39), *в котором числовые последовательности* $\{\gamma_n\}$, $\{\varepsilon_n\}$, $\{\delta_n\}$ *удовлетворяют условиям*

$$\gamma_n > 0, \quad \varepsilon_n \in (0, \hat{\varepsilon}), \quad \delta_n > 0, \quad n = 1, 2, \ldots,$$

$$\lim_{n\to\infty}\delta_n = 0, \quad \lim_{n\to\infty}\frac{\varepsilon_n}{\delta_n} = \zeta \in [0, \infty), \quad \sum_{n=1}^{\infty}\gamma_n\delta_n = \infty.$$

Тогда

1) *если, кроме того,*

$$\sum_{n=1}^{\infty}\left(\frac{\gamma_n^2}{\varepsilon_n} + |\varepsilon_n - \varepsilon_{n+1}| + |\delta_n - \delta_{n+1}| + \left|\frac{\varepsilon_{n+1}}{\delta_{n+1}} - \frac{\varepsilon_n}{\delta_n}\right|\right) < \infty,$$

то с вероятностью 1 *последовательность* $\{p_n\}$ *сходится при* $n \to \infty$ *к точке равновесия* p^* *исходной игры, определяемой условием* (4.5);

2) *если же при* $n \to \infty$

$$\left[\frac{\gamma_n}{\varepsilon_n\delta_n} + \frac{1}{\gamma_n\delta_n}\left(|\varepsilon_n - \varepsilon_{n+1}| + |\delta_n - \delta_{n+1}| + \left|\frac{\varepsilon_n}{\delta_n} - \frac{\varepsilon_{n+1}}{\delta_{n+1}}\right|\right)\right] \to 0,$$

то $p_n \to p^*$ *в среднеквадратическом.*

Доказательство. Введем σ-алгебры $\mathcal{F}_n \triangleq \sigma\{x_t, \xi_t^k \mid k = \overline{1, l}, t = \overline{1, n-1}\}$ и $W_n \triangleq \sum_{k=1}^{l} r_k \left\| p_n^k - p^{k*}(\varepsilon_n, \delta_n)\right\|^2$, где r_k — коэффициенты, фигурирующие в (4.21). Используя свойства проектора $\pi_\varepsilon^{N_k}$ (2.10) и предположения П1″—П3″ из § 4.1, а также (4.6) и неравенства (4.36), получим в силу алгоритма (4.39) для всякого $n \geqslant 1$

$$\mathsf{M}\{W_{n+1}|\mathcal{F}_n\} \leqslant W_n + C_1\varkappa_n\sqrt{W_n} + C_2\varkappa_n^2 + \\ + C_3\gamma_n^2\varepsilon_n^{-1} - 2\gamma_n\sum_{k=1}^{l} r_k\left[V^k(p_n) - w^k(p_n, p^{k*}(\varepsilon_n, \delta_n)) + \right. \\ \left. + \delta_n\langle p_n^k, p_n^k - p^{k*}(\varepsilon_n, \delta_n)\rangle\right],$$

где $\varkappa_n \triangleq |\varepsilon_n - \varepsilon_{n+1}| + |\delta_n - \delta_{n+1}| + |\varepsilon_n/\delta_n - \varepsilon_{n+1}/\delta_{n+1}|$, функции w^k определены в (4.20), C_k $(k = 1, 2, 3)$ — положительные константы. Из этого неравенства, условия диагональной выпуклости (4.21) и ограниченности W_n $\left(W_n \leqslant 2\sum_{k=1}^{l} r_k, n = 1, 2, \ldots\right)$ следует, что

$$\mathsf{M}\{W_{n+1}/\mathcal{F}_n\} \leqslant (1 - 0{,}5\,\gamma_n\delta_n)W_n + \sqrt{2}C_1\varkappa_n + \\ + C_2\varkappa_n^2 + C_3\gamma_n^2\varepsilon_n^{-1}. \quad (4.40)$$

Отсюда в силу соответствующих условий данной теоремы, лемм П.11 и 4.7, следует сходимость $p_n \to p^*$ при $n \to \infty$ с вероятностью 1. Усредним левую и правую части неравенства (4.40). Получим для всех $n = 1, 2, \ldots$

$$\mathsf{M}\{W_{n+1}\} \leqslant (1 - 0{,}5\,\gamma_n\delta_n)\mathsf{M}\{W_n\} + C_3\gamma_n^2\varepsilon_n^{-1} + C_4\varkappa_n, \quad (4.41)$$

где $C_4 \triangleq \sqrt{2}C_1 + C_2\sup\limits_n \varkappa_n < \infty$. Из этого рекуррентного неравенства в силу лемм П.5, 4.7 и следствия из леммы 4.8 вытекает сходимость $p_n \to p^*$ при $n \to \infty$ в среднеквадратическом. Теорема доказана. ▲

Замечание. Из доказательства последней теоремы и леммы 4.7 следует, что точка равновесия p^* в общем случае зависит от значения $\zeta = \lim\limits_{n\to\infty} \varepsilon_n/\delta_n$.

Для оценки порядка скорости сходимости алгоритма (4.39) ограничимся последовательностями $\{\gamma_n\}$, $\{\varepsilon_n\}$, $\{\delta_n\}$ следующего вида:

$$\gamma_n = \frac{\gamma}{(n+a)^\alpha}, \quad \varepsilon_n = \frac{\varepsilon}{(n+b)^\beta}, \quad \delta_n = \frac{\delta}{(n+c)^\varkappa}, \quad (4.42)$$

где параметры a, b, c, γ, ε, δ, α, β, $\varkappa$ таковы, что выполняются условия теоремы 4.3. В частности, для выполнения условий сходимости с вероятностью 1 алгоритма (4.39) параметры α, β, $\varkappa$, определяющие порядок убывания величин γ_n, ε_n, δ_n (4.42), должны удовлетворять неравенствам

$$\beta \geqslant \varkappa > 0, \quad \alpha + \varkappa \leqslant 1, \quad 2\alpha - \beta > 1. \qquad (4.43)$$

Аналогично, для сходимости алгоритма (4.39) в среднеквадратическом должны выполняться неравенства

$$\beta \geqslant \varkappa > 0, \quad \alpha + \varkappa \leqslant 1, \quad \alpha > \beta + \varkappa. \qquad (4.44)$$

Теорема 4.4. *Пусть последовательности* $\{\gamma_n\}$, $\{\varepsilon_n\}$, $\{\delta_n\}$ *имеют вид* (4.42), *выполнены условия теоремы* 4.3, *касающиеся сходимости алгоритма* (4.39) *в среднеквадратическом (в частности, справедливы неравенства* (4.44)) *и пусть при* $\alpha + \varkappa = 1$ *выполняется также неравенство* $0{,}5\gamma\delta > 2\alpha - \beta - 1$. *Тогда*

$$\overline{\lim_{n\to\infty}}\, n^{\theta} \mathsf{M}\left\{\sum_{k=1}^{l} \| p_n^k - p^{k*} \|^2\right\} < \infty, \qquad (4.45)$$

где $p^* \triangleq \lim_{n\to\infty} p^*(\varepsilon_n, \delta_n)$ (см. лемму 4.7) *и*

$$\theta \triangleq \begin{cases} \min\{2\varkappa, 2(\beta - \varkappa), \alpha - \beta - \varkappa\} & \textit{при} \quad \beta \neq \varkappa, \\ \min\{\alpha - 2\beta, 2\beta\} & \textit{при} \quad \beta = \varkappa. \end{cases}$$

Доказательство. В условиях данной теоремы справедливы соотношения (4.36), (4.37), а рекуррентное неравенство (4.41) можно записать следующим образом:

$$\mathsf{M}\{W_{n+1}\} \leqslant \\ \leqslant [1 - 0{,}5\gamma\delta n^{-(\alpha+\varkappa)}(1 + o(1))]\, \mathsf{M}\{W_n\} + Cn^{-(2\alpha-\beta)},$$

где $C > 0$ $(n = 1, 2, \ldots)$. Применяя теперь лемму П.2 (в которой параметр s следует выбрать достаточно большим и положить $r = 1/2$, $a = 0$), получим

$$\overline{\lim_{n\to\infty}}\, n^{\alpha-\beta-\varkappa} \mathsf{M}\{W_n\} < \infty.$$

Используя теперь неравенство

$$\| p_n^k - p^{k*} \|^2 \leqslant 2 \| p_n^k - p^{k*}(\varepsilon_n, \delta_n) \|^2 + 2 \| p^{k*}(\varepsilon_n, \delta_n) - p^{k*} \|^2$$

и учитывая (4.37), (4.42) и (4.44), получим (4.45). Теорема доказана. ▲

Из этой теоремы и леммы 4.1 вытекает следующее утверждение.

Следствие 1. *В условиях теоремы 4.4 с вероятностью 1 выполняется целевое условие* (4.2), (4.3) *и справедливы неравенства*

$$\varlimsup_{n\to\infty} n^{\theta}\mathsf{M}\left\{\left(\Phi_n^k - V^k(p^*)\right)^2\right\} < \infty, \quad k = \overline{1, l}. \tag{4.46}$$

Полученные результаты позволяют найти оптимальные параметры α, β, $\varkappa$, максимизирующие порядок скорости сходимости алгоритма (4.39), который гарантируется теоремой 4.4 и соотношением (4.46).

Следствие 2. *В условиях теоремы 4.4 параметр* θ, *определяющий порядок скорости сходимости алгоритма* (4.39) (*в смысле соотношений* (4.45), (4.46)), *удовлетворяет неравенству* $\theta \leqslant 2/5$, *причем* $\theta = 2/5$ *только при* $\alpha = \alpha^* = 4/5$, $\beta = \beta^* = \varkappa = \varkappa^* = 1/5$.

Результат этого следствия получается максимизацией параметра θ по параметрам α, β, $\varkappa$ при условиях (4.44).

§ 4.6. Алгоритм адаптивного выбора вариантов в игре многих лиц (общий случай)

Теоремы 4.3, 4.4 устанавливают работоспособность алгоритма (4.39) лишь при выполнении условия диагональной выпуклости (4.21), т. е. для довольно узкого класса игровых задач. Вспомним, что решение исходной игровой задачи адаптивного выбора вариантов, сформулированной в § 4.1, было получено (с учетом леммы 4.1) путем построения рекуррентного алгоритма пересчета условных вероятностей выбора (4.6), сходящихся к точке равновесия по Нэшу p^{k*} ($k = \overline{1, l}$) в соответствующей матричной игре Г (§ 4.1). Однако сходимость $p_n \to p^*$ при $n \to \infty$ не является необходимым условием для достижения цели (4.2), (4.3), поэтому δ-регуляризация задачи не является обязательной. Покажем, что работоспособность алгоритма (4.39) (в смысле обеспечения целевого условия (4.2), (4.3)) при $\delta_n \equiv 0$ имеет место и для недиагонально выпуклых игр. При этом, однако, будем требовать, чтобы последовательности вариантов $\hat{x}^k = (x_n^1, \ldots$ $\ldots, x_n^{k-1}, \tilde{x}_n^k, x_n^{k+1}, \ldots, x_n^l) \in X$ в (4.2), (4.3) удовлетворяли дополнительному требованию — стационарности условных вероятностей выбора чистых стратегий $\tilde{x}_n^k \in X$:

$$\mathsf{P}\{\tilde{x}_n^k = x^k(i_k) \mid \tilde{x}_t^k, \xi_t^k(\hat{x}_t^k, \omega)(t = \overline{1, n-1})\} \triangleq$$
$$\triangleq p^k(i_k) = \text{const}, \quad n = 1, 2, \ldots, \quad k = \overline{1, l}, \quad i_k = \overline{1, N_k}. \tag{4.47}$$

Хотя это требование несколько сужает постановку задачи, данную в § 4.1, оно, тем не менее, всегда явно или неявно фигурирует во всех известных постановках игровых задач адаптивного выбора вариантов.

Рассмотрим алгоритм (4.39) при $\delta_n = 0$ $(n = 1, 2, \ldots)$:

$$\begin{gathered} p_{n+1}^k = \pi_{\varepsilon_{n+1}}^{N_k}\left\{p_n^k - \gamma_n \frac{\xi_n^k}{e^{\mathrm{T}}\left(x_n^k\right) p_n^k} e\left(x_n^k\right)\right\} \\ p_1 \in S_{\varepsilon_1}^{N_k}, \quad k = \overline{1, l}. \end{gathered} \tag{4.48}$$

Следующее утверждение устанавливает условия, при выполнении которых порождаемая этим алгоритмом последовательность вариантов $\{x_n\}$ обеспечивает достижение цели (4.2), (4.3), (4.47).

Теорема 4.5. *Пусть выполнены предположения* П1″ — П3″ *из* § 4.1, *а последовательности* $\{p_n^k\}$ $(k = \overline{1, l})$ *векторов условных вероятностей* (4.6) *порождаются алгоритмом* (4.48) *с произвольными начальными значениями* $p_1^k \in S_{\varepsilon_1}^{N_k}$ *и числовыми последовательностями* $\{\gamma_n\}$, $\{\varepsilon_n\}$, *удовлетворяющими условиям*:

$$\begin{gathered} 0 < \gamma_{n+1} \leqslant \gamma_n, \quad \varepsilon_n \in \left(0, \max_{k=\overline{1,l}} N_k^{-1}\right), \quad n = 1, 2, \ldots, \quad \varliminf_{n\to\infty} n\gamma_n = \infty, \\ \lim_{n\to\infty}\left(\varepsilon_n + \frac{|\varepsilon_n - \varepsilon_{n+1}|}{\gamma_n}\right) = 0, \quad \sum_{n=1}^{\infty} \frac{\gamma_n}{n\varepsilon_n} < \infty. \end{gathered} \tag{4.49}$$

Тогда с вероятностью 1 *целевое условие* (4.2), (4.3), (4.47) *выполняется для всех* $\tilde{p}^k \triangleq (\tilde{p}^k(1), \ldots, \tilde{p}^k(N_k)) \in S^{N_k}$ $(k = \overline{1, l})$.

Доказательство. Покажем сначала, что для любых $k = \overline{1, l}$ и $\tilde{p}^k \in S^{N_k}$ с вероятностью 1

$$\varlimsup_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} [V^k(p_t) - w^k(p_t, \tilde{p}^k)] \leqslant 0, \tag{4.50}$$

где функции V^k и w^k заданы соотношениями (4.4) и (4.20). Определим векторы

$$q_n^k \triangleq (1-\varepsilon_n N_k)\tilde{p}^k + \varepsilon_n e^{N_k} \in S_{\varepsilon_n}^{N_k}, \tag{4.51}$$

где $e^{N_k} \triangleq (1, \ldots, 1)$. В силу алгоритма (4.47) имеем при каждом $n = 1, 2, \ldots$

$$\| p_{n+1}^k - q_{n+1}^k \|^2 \leqslant \| p_n^k - q_n^k \|^2 + C_1 | \varepsilon_n - \varepsilon_{n+1} | - \\ - 2\gamma_n \left\langle p_n^k - q_{n+1}^k, \frac{\xi_n^k e\left(x_n^k\right)}{e^{\mathrm{T}}\left(x_n^k\right) p_n^k} \right\rangle + \gamma_n^2 \left(\frac{\xi_n^k}{e^{\mathrm{T}}\left(x_n^k\right) p_n^k} \right)^2. \tag{4.52}$$

Здесь и далее C_ν $(\nu = 1, 2, \ldots)$ — положительные константы. Введем σ-алгебры $\mathscr{F}_n \triangleq \sigma\left(x_t^k, \xi_t^k \mid k = \overline{1, l}, t = \overline{1, n}\right)$. Поскольку в силу предположений П1″ — П3″ из § 4.1 и равенств (4.4), (4.6) имеют место соотношения

$$\mathsf{M}\left\{ \left\langle p_n^k - q_{n+1}^k, \frac{\xi_n^k e\left(x_n^k\right)}{e^{\mathrm{T}}\left(x_n^k\right) p_n^k} \right\rangle \middle| \mathscr{F}_{n-1} \right\} = V^k(p_n) - w^k(p_n, q_{n+1}^k), \tag{4.52a}$$

$$\mathsf{M}\left\{ \left(\frac{\xi_n^k}{e^{\mathrm{T}}\left(x_n^k\right) p_n^k} \right)^2 \middle| \mathscr{F}_{n-1} \right\} \leqslant C_2 \varepsilon_n^{-1},$$

то из (4.52) с учетом (4.51) получаем

$$\| p_{n+1}^k - q_{n+1}^k \|^2 \leqslant \| p_n^k - q_n^k \|^2 + C_1 | \varepsilon_n - \varepsilon_{n-1} | - \\ - 2\gamma_n \left[V^k(p_n) - w^k(p_n, q^k) \right] + C_3 \gamma_n \varepsilon_{n+1} + 2\gamma_n g_n' + \gamma_n^2 g_n'',$$

где

$$g_n' \triangleq V^k(p_n) - w^k(p_n, q_{n+1}^k) - \left\langle p_n^k - q_{n+1}^k, \frac{\xi_n^k e\left(x_n^k\right)}{e^{\mathrm{T}}\left(x_n^k\right) p_n^k} \right\rangle,$$

$$g_n'' \triangleq \left(\frac{\xi_n^k}{e^{\mathrm{T}}\left(x_n^k\right) p_n^k} \right)^2 \geqslant 0.$$

Отсюда

$$\frac{2}{n} \sum_{t=1}^{n} \left[V^k(p_t) - w^k(p_t, q^k) \right] \leqslant \frac{1}{n} \sum_{t=1}^{n} \frac{g_t}{\gamma_t} + \\ + \frac{1}{n} \sum_{t=1}^{n} \gamma_t g_t'' + \frac{C_4}{n} \sum_{t=1}^{n} \left(\varepsilon_{t+1} + \gamma_t^{-1} | \varepsilon_t - \varepsilon_{t+1} | \right), \tag{4.53}$$

где

$$g_t \triangleq \| p_t^k - q_t^k \|^2 - \| p_{t+1}^k - q_{t+1}^k \|^2 + 2\gamma_t g_t'.$$

В условиях данной теоремы

$$\sum_{t=1}^{\infty} \mathsf{M}\,\{(\gamma_t g_t')^2\} < \infty, \quad \sum_{t=1}^{\infty} \mathsf{M}\left\{\frac{1}{t}\,\gamma_t g_t''\right\} < \infty,$$

поэтому (что следует, например, из леммы П.9) с вероятностью 1 ряды

$$\sum_{t=1}^{\infty} \gamma_t g_t' \quad \text{и} \quad \sum_{t=1}^{\infty} \frac{1}{t}\,\gamma_t g_t''$$

сходятся. Значит, с вероятностью 1

$$\sup_n \left| \sum_{t=1}^{n} g_t \right| < \infty.$$

Воспользовавшись леммами П.6—П.8, получим из (4.53) неравенство (4.50).

Повторяя рассуждения, сделанные при доказательстве леммы 4.1, нетрудно показать, что последовательности величин

$$\begin{aligned} d_n^k(\{\hat{x}_n^k\}) &\triangleq \left[\Phi_n^k(\{\hat{x}_n^k\}) - \frac{1}{n} \sum_{t=1}^{n} w^k(p_t, \tilde{p}^k) \right]^2, \\ d_n^k(\{x_n\}) &\triangleq \left[\Phi_n^k(\{x_n\}) - \frac{1}{n} \sum_{t=1}^{n} V^k(p_t) \right]^2 \end{aligned} \tag{4.54}$$

с вероятностью 1 при $n \to \infty$ сходятся к нулю для любой последовательности вариантов $\{\hat{x}_n^k\}$, удовлетворяющей условию (4.3). Отсюда следует, что

$$\overline{\lim_{n\to\infty}} \left[\Phi_n^k(\{x_n\}) - \Phi_n^k(\{\hat{x}_n^k\}) \right] =$$

$$= \overline{\lim_{n\to\infty}} \frac{1}{n} \sum_{t=1}^{n} \left[V^k(p_t) - w^k(p_t, \tilde{p}^k) \right] \leqslant 0$$

в силу (4.50). Из произвольности $k = \overline{1, l}$ и $\tilde{p}^k \in S^{N_k}$ вытекает утверждение теоремы. Теорема доказана. ▲

Для оценки порядка скорости достижения целевого условия (4.2), (4.3), (4.47) при использовании алгоритма (4.48) ограничимся последовательностями $\{\gamma_n\}$, $\{\varepsilon_n\}$ вида

$$\gamma_n = \frac{\gamma}{(n+a)^\alpha}, \quad \varepsilon_n = \frac{\varepsilon}{(n+b)^\beta}, \tag{4.55}$$

где параметры a, b, γ, ε, α, β таковы, что выполняются условия теоремы 4.5. В частности, параметры α, β должны удовлетворять неравенствам

$$0 < \beta < \alpha < 1. \tag{4.56}$$

Теорема 4.6. *Пусть последовательности* $\{\gamma_n\}$, $\{\varepsilon_n\}$ *имеют вид* (4.55) *и выполнены условия теоремы* 4.5, *а значит, справедливы неравенства* (4.56). *Тогда с вероятностью* 1 *при каждом* $k=\overline{1, l}$ *и произвольном* $\nu \in (0, \alpha-\beta)$

$$\Phi_n^k(\{x_n\}) - \Phi_n^k(\{\hat{x}_n^k\}) \leqslant O^*_\cdot(n^{\nu-\theta}) \tag{4.57}$$

для любых последовательностей $\{\hat{x}_n^k\}$ *из* (4.2), (4.47), *где*

$$\theta \triangleq \min\{\beta, \alpha-\beta, 1-\alpha\}. \tag{4.58}$$

Доказательство основывается на использовании неравенства (4.53) и оценки скорости сходимости к нулю с вероятностью 1 последовательностей величин $d_n^k(\{\hat{x}_n^k\})$ и $d_n^k(\{x_n\})$ (4.54). В силу леммы П.6 первое слагаемое правой части (4.53) оценивается следующим образом:

$$\left|\frac{1}{n}\sum_{t=1}^{n} g_t\gamma_t^{-1}\right| \leqslant 2\left(\sup_n\left|\sum_{t=1}^{n} g_t\right|\right)(n\gamma_n)^{-1} \overset{\text{п.н.}}{=} O^*(n^{\alpha-1}), \tag{4.59}$$

поскольку, как было установлено при доказательстве теоремы 4.5,

$$\sup_n\left|\sum_{t=1}^{n} g_t\right| \overset{\text{п.н.}}{<} \infty.$$

Аналогичным образом оцениваем второе слагаемое правой части неравенства (4.53), учитывая неотрицательность величин g_t'' $(t=\overline{1,n})$:

$$\left|\frac{1}{n}\sum_{t=1}^{n} \gamma_t g_t''\right| \leqslant \frac{1}{n^{1-\varphi}}\sum_{t=1}^{\infty}\frac{g_t''\gamma_t}{t^\varphi} \overset{\text{п.н.}}{=} O^*(n^{\varphi-1}), \tag{4.60}$$

где

$$\varphi \triangleq 1 + \beta - \alpha + \nu \in (1 + \beta - \alpha, 1); \tag{4.61}$$

сходимость почти наверное бесконечного ряда неотрицательных величин в (4.60) следует по лемме П.9 из сходимости ряда условных математических ожиданий (см. (4.52а)):

$$\sum_{t=1}^{\infty} \gamma_t t^{-\varphi} \mathsf{M}\{g_t'' \mid \mathcal{F}_{t-1}\} \overset{\text{п.н.}}{\leqslant} C_2 \sum_{t=1}^{\infty} \gamma_t t^{-\varphi} \varepsilon_t^{-1} \leqslant$$

$$\leqslant C_2 \sum_{t=1}^{\infty} t^{-(1+\nu)} < \infty. \tag{4.62}$$

Оценка третьего слагаемого правой части неравенства (4.53) вытекает непосредственно из (4.55):

$$\frac{C_4}{n} \sum_{t=1}^{n} \left(\varepsilon_{t+1} + \frac{\gamma_t}{\varepsilon_t} + \frac{|\varepsilon_t - \varepsilon_{t+1}|}{\gamma_t} \right) = O^*(n^{-\rho}), \tag{4.63}$$

где $\rho = \min\{\beta, \alpha - \beta\}$.

Наконец, оценка величин $d_n^k(\{\hat{x}_n^k\})$ вытекает из леммы П.13А:

$$d_n^k(\{\hat{x}_n^k\}) = o(n^{\varkappa - 1}) \quad \forall \varkappa > 0, \tag{4.64}$$

а для $d_n^k(\{x_n\})$ оценка получается из (4.64) при $\hat{x}_n^k \equiv x_n$

Требуемая оценка (4.57), (4.58) следует из (4.53), (4.54) и соотношений (4.59)—(4.64) (при этом в (4.64) величину $\varkappa$ надо взять достаточно малой). Теорема доказана. ▲

Используя результат теоремы 4.6, определим оптимальные параметры алгоритма (4.48), т. е. такие величины α^*, β^*, которые обеспечивают наибольший порядок скорости достижения целевого условия (4.2), (4.3), (4.47) в смысле оценки (4.57), (4.58).

Следствие. *В условиях теоремы* 4.6 *параметр* θ, *определяющий порядок скорости сходимости верхней оценки* (4.57) *к нулю, не превосходит* 1/3, *причем* $\theta = 1/3$ *лишь при*

$$\alpha = \alpha^* = \frac{2}{3}, \quad \beta = \beta^* = \frac{1}{3}. \tag{4.65}$$

§ 4.7. Обсуждение результатов

Сформулированная в этой главе игровая задача адаптивного выбора вариантов эквивалентна бескоалиционной матричной игре многих лиц в условиях априорной неопределенности (т. е. игроки априори не знают матрицы потерь, а в бесконечно повторяющейся последовательности разыгрываемых партий наблюдают лишь реализации своих выигрышей). Лемма 4.1 устанавливает, что для достижения целевого условия (4.2), (4.3) (т. е. решения поставленной задачи) достаточно обеспечить сходимость почти наверное последовательности $\{p_n\}$ составных векторов (условных вероятностей выбора чистых стратегий игроков (4.6)) к точке равновесия по Нэшу p^*. В стремлении обеспечить эту сходимость и состоит идея использования методов регуляризации и стохастической аппроксимации для формирования игровых алгоритмов адаптированного выбора вариантов.

Полученные результаты, касающиеся асимптотических свойств алгоритмов (4.17) (для игры двух лиц с нулевой суммой) и (4.39) (для диагонально выпуклой игры многих лиц), устанавливают достаточные условия сходимости и позволяют оценить скорость сходимости (теоремы 4.1—4.4) последовательности $\{p_n\}$ к p^*, а следовательно, и скорость решения поставленной задачи (следствия 1 из теорем 4.2, 4.3). Показано, что может быть обеспечен порядок скорости сходимости в среднеквадратическом, равный $n^{-2/5}$.

Поскольку класс диагонально выпуклых матричных игр чрезвычайно узок, перечисленные результаты имеют фактически лишь теоретическое значение. Поэтому в § 4.6 эти результаты распространены на общий случай. Введенное при этом условие (4.47), сужающее постановку задачи, данную в § 4.1, является, тем не менее, естественным. Оно означает стационарность условных вероятностей выбора чистых стратегий $\tilde{x}_n^k$, фигурирующих в целевом условии (4.2), (4.3), и присутствует во всех известных постановках игровых задач адаптивного выбора вариантов [14, 117].

Теорема 4.5 дает достаточные условия сходимости игрового алгоритма (4.48) в смысле достижения цели (4.2), (4.3), (4.47). При этом последовательность $\{p_n\}$, порождаемая этим алгоритмом, не обязательно имеет предел. Это связано с тем, что в (4.48) отсутствует δ-регуляризация (алгоритм (4.48) совпадает с алгоритмом (4.39) при $\delta_n \equiv 0$), наличие которой соответствует наличию демпфирования, или затухания, что наглядно демонстрируется примером, рассмотренным в § 4.3. Оценка скорости достижения почти наверное целевого условия (4.2), (4.3), (4.47) получена в теореме 4.6 (см. (4.55) — (4.58)), причем максимальная скорость имеет порядок $n^{\nu-1/3}$ (где ν — сколь угодно малое положительное число) и достигается при выборе параметров алгоритма (4.48) в соответствии с (4.55), (4.65).

ГЛАВА 5

АЛГОРИТМЫ ВЫБОРА ВАРИАНТОВ В ЗАДАЧЕ АДАПТИВНОГО УПРАВЛЕНИЯ КОНЕЧНЫМИ ОДНОРОДНЫМИ МАРКОВСКИМИ ЦЕПЯМИ

Рассматривается задача адаптивного управления конечной марковской цепью в условиях неопределенности. В предположении однородности (стационарности характеристик) цепи исследуются предельные значения текущих средних потерь и формулируется задача адаптивного управления на классе однородных конечных управляемых марковских цепей. Для класса регулярных цепей предлагается рекуррентный алгоритм управления и устанавливаются достаточные условия, при которых выполняется целевое условие, т. е. достигается с вероятностью 1 минимальное предельное значение текущих средних потерь. Этот результат далее обобщается на класс связных цепей при более жестких ограничениях на параметры алгоритма, а затем показывается его работоспособность при тех же ограничениях и на классе произвольных цепей, а именно обеспечение с вероятностью 1 минимаксного целевого условия.

§ 5.1. Основные понятия и предположения

Пусть необходимо управлять некоторой системой, имеющей конечное множество состояний $Z \triangleq \{z(1), \ldots z(K)\}$ и конечное множество управляющих воздействий $X \triangleq \{x(1), \ldots, x(N)\}$. Переход системы из состояния $z_n \in Z$ в состояние $z_{n+1} \in Z$ осуществляется в момент t_{n+1} и зависит от управления $x_n \in X$, реализованного в момент t_n $(n = 1, 2, \ldots)$. Значение управляющего воздействия x_n определяет тем самым тот или иной вариант функционирования системы на интервале (t_n, t_{n+1}), поэтому, как и ранее, x_n можно называть вариантом, а множество X — множеством возможных вариантов.

Пусть также качество функционирования управляемой системы на интервале времени (t_n, t_{n+1}) зависит как от состояния z_n самой системы, так и от выбранного варианта x_n и характеризуется случайной величиной $\xi_n = \xi_n(z_n, x_n, \omega)$, имеющей смысл потерь на этом интервале (см. сноску на с. 40). Величины ξ_n, а также состояния системы z_n доступны наблюдению.

Будем далее считать, что выполнены следующие предположения:

П1‴. Для всех $n = 1, 2, \ldots$ совокупности $\{\xi_n(z, x, \omega) \mid z \in Z, x \in X\}$ и $\{\xi_t(z, x, \omega), z_k, x_k \mid z \in Z, \quad x \in X, \quad t = \overline{1, n-1}, \; k = \overline{1, n}\}$ независимы;

П2‴. При любых $z(i)$, $x(l)$ $(i = \overline{1, K}, \; l = \overline{1, N})$ и $n = 1, 2, \ldots$

$$\mathsf{M}\{\xi_n(z(i), x(l), \omega)\} \triangleq v_{il}, \tag{5.1}$$

$$\sup_{n,i,l} \mathsf{M}\{\xi_n^2(z(i), x(l), \omega)\} \triangleq C_0 < \infty; \tag{5.2}$$

П3‴. Для всех $n = 1, 2, \ldots$ при каждых $i, j = \overline{1, K}$, $l = \overline{1, N}$

$$\mathsf{P}\{z_{n+1} = z(j) \mid z_n = z(i), \; x_n = x(l), \; \mathcal{F}_n\} \triangleq \pi_{ij}^l, \tag{5.3}$$

где $\mathcal{F}_n \triangleq \sigma\{z_n, z_t, x_t, \xi_t \mid t = \overline{1, n-1}\}$.

Предположение П1‴ означает статистическую независимость потерь $\xi_n(z, x, \omega)$ для фиксированных состояний $z \in Z$ и вариантов управления $x \in X$ от предыстории управляемого процесса, а предположение П2‴ — стационарность этих потерь в широком смысле и равномерную по времени ограниченность их второго момента. Эти предположения полностью аналогичны соответствующим предположениям, сделанным в предшествующих главах при изучении других задач адаптивного выбора вариантов.

Предположение П3‴ касается свойств управляемой системы и означает, что при любом фиксированном варианте (управлении) $x_n = x(l)$ переход системы из одного состояния в другое осуществляется таким же образом, как если бы система представляла собой однородную марковскую цепь с переходной матрицей $\pi^l \triangleq \|\pi_{ij}^l\|_{i,j=\overline{1,K}}$.

При сделанных предположениях рассматриваемая управляемая система представляет собой *однородную конечную управляемую марковскую цепь* [118].

Будем говорить, что для некоторой управляемой конечной однородной марковской цепи из класса цепей $M(K, N)$, имеющих K состояний и N возможных управляющих воздействий, в момент времени t_n определено *правило* x_n *выбора* варианта (управления) $x \in X$, если задана функция (возможно, случайная) $x_n(\omega;\ z_n;\ z_t,\ x_t,\ \xi_t \,|\, t=\overline{1,\ n-1})$ со значениями в X, не зависящая от будущего. Такое правило является информационно допустимым в том смысле, что значение управляющего воздействия x_n при почти всех ω однозначно определяется предысторией процесса, доступной наблюдению. Если правила выбора варианта управления определены для всех $n = 1, 2, \ldots$, то будем говорить, что задана *последовательность правил выбора вариантов* (управлений) или, иначе, *стратегия управления.*

В силу предположений П1‴ и П3‴ условные вероятности перехода цепи в момент времени t_n можно представить в виде

$$\mathsf{P}\{z_{n+1} = z(j) \,|\, z_n = z(i),\ \mathcal{F}_n\} = \\ = \sum_{l=1}^{N} \mathsf{P}\{z_{n+1} = z(j) \,|\, z_n = z(i),\ x_n = x(l),\ \mathcal{F}_n\} \times \\ \times \mathsf{P}\{x_n = x(l) \,|\, z_n = z(i),\ \mathcal{F}_n\} = \sum_{l=1}^{N} \pi_{ij}^{l} d_n^{il}, \quad (5.4)$$

где $\mathcal{F}_n \triangleq \sigma(z_n;\ z_t,\ x_t,\ \xi_t \,|\, t = \overline{1,\ n-1})$ и

$$d_n^{il} \triangleq \mathsf{P}\{x_n = x(l) \,|\, z_n = z(i),\ \mathcal{F}_n\} \quad (5.5)$$

— условная вероятность выбора варианта (управления) $x(l)$ в момент времени t_n в состоянии $z(i)$, которая порождается правилами выбора вариантов $x_1, \ldots, x_{n-1}$. Последовательности $\{d_n^{il}\}$ условных вероятностей выбора полностью определяются стратегией управления.

С другой стороны, последовательности матриц $\{d_n\}$, $d_n \triangleq \|d_n^{il}\|_{\substack{i=\overline{1,K} \\ l=\overline{1,N}}}$ можно поставить в соответствие последовательность рандомизированных правил выбора вариантов, т. е. стратегию управления. Таким образом, стратегию управления можно отождествить с последовательностью $\{d_n\}$ случайных матриц d_n, измеримых при каждом n относительно $\mathcal{F}_n$ в соответствии с (5.5). Класс всех стратегий $\{d_n\}$ будем обозначать Σ, а подкласс стационарных стратегий, удовлетворяющих с вероятностью 1 условию $d_n^{il} = d^{il}$ $(i = \overline{1,\ K},\ \ l = \overline{1,\ N})$,— символом Σ_s, т. е. $\{d\} \in \Sigma_s \subset \Sigma$. В классе Σ_s будем выделять подкласс Σ_s^+ невырожденных стационарных стратегий, для которых

$d^{il} > 0$ $(i = \overline{1, K},\ l = \overline{1, N})$. Таким образом,

$$\Sigma_S^+ \subset \Sigma_S \subset \Sigma.$$

При фиксированной стационарной стратегии $\{d\} \in \Sigma_S$ управляемая марковская цепь превращается в обычную однородную марковскую цепь с переходной матрицей

$$\Pi(d) = \left\| \sum_{l=1}^{N} \pi_{ij}^{l} d^{il} \right\|_{i,j=\overline{1,K}} \triangleq \| \pi_{ij}(d) \|_{i,j=\overline{1,K}}. \tag{5.6}$$

Различным стратегиям $\{d\} \in \Sigma_S$ соответствует, вообще говоря, различная структура однородной марковской цепи, которая определяется структурой матрицы $\Pi(d)$. Рассмотрим подробнее вопрос о возможных структурах матриц вероятностей переходов однородной цепи. Пользуясь терминологией, принятой в теории однородных марковских цепей [112], будем называть состояние цепи $z(i) \in Z$ *несущественным*, если из $z(i)$ возможен переход в некоторое состояние $z(j) \in Z$, не допускающий возвращения в $z(i)$. Все остальные состояния назовем *существенными*. Существенные состояния $z(i)$, $z(j) \in Z$ называются *сообщающимися*, если вероятности перехода цепи из одного состояния в другое за конечное число шагов отличны от нуля. Очевидно, что если $z(i)$ сообщается с $z(j)$, а $z(j)$ сообщается с $z(k)$, то $z(i)$ сообщается с $z(k)$. Поэтому множество Z всех состояний однородной конечной цепи Маркова распадается на классы $Z^{(1)}, \ldots, Z^{(L)}$ и $Z^{(0)}$ так, что состояния, принадлежащие одному из классов $Z^{(l)}$ $(l = \overline{1, L})$,— сообщающиеся, $Z^{(0)}$ — класс несущественных состояний. Таким образом, попав однажды в одно из состояний класса $Z^{(l)}$ $(l = \overline{1, L})$, система никогда уже не выйдет за пределы этого класса состояний, а класс состояний $Z^{(0)}$ система с вероятностью 1 покинет за конечное число шагов и никогда в него не вернется. В итоге приходим к представлению множества Z состояний однородной конечной цепи в виде объединения непересекающихся множеств

$$Z = Z^{(1)} \cup \ldots \cup Z^{(L)} \cup Z^{(0)},$$

где $Z^{(l)}$ $(l = \overline{1, L})$ — связные компоненты, а $Z^{(0)}$ — множество несущественных состояний. Соответственно, матрица $\Pi(d)$ $(\{d\} \in \Sigma_S)$ вероятностей переходов такой цепи может быть представлена (соответствующей перенумерацией состояний) в «каноническом» виде:

$$
\Pi(d) = \begin{Vmatrix}
\Pi^{(1)}(d) & 0 & \dots & 0 & 0 \\
0 & \Pi^{(2)}(d) & \dots & 0 & 0 \\
\dots & \dots & \dots & \dots & \dots \\
0 & 0 & \dots & \Pi^{(L)}(d) & 0 \\
\Pi_1^{(0)}(d) & \Pi_2^{(0)}(d) & \dots & \Pi_L^{(0)}(d) & \Pi^{(0)}(d)
\end{Vmatrix},
$$

где $\Pi^{(l)}(d)$ — квадратная стохастическая матрица, соответствующая состояниям l-й связной компоненты $Z^{(l)}$ $(l = \overline{1, L})$, $\Pi_l^{(0)}(d)$ — неотрицательная прямоугольная матрица, строки которой изменяются в пределах состояний класса $Z^{(0)}$ несущественных состояний, а столбцы — в пределах состояний связной компоненты $Z^{(l)}$, $\Pi^{(0)}(d)$ — квадратная матрица (не стохастическая), соответствующая состояниям класса $Z^{(0)}$.

В [112] установлено также, что подматрицы $\Pi^{(l)}(d)$ $(l = \overline{1, L})$ являются неразложимыми, т. е. не могут быть (путем перенумерации состояний из $Z^{(l)}$) представлены в виде

$$
\Pi^{(l)}(d) = \begin{Vmatrix} Q & R \\ S & T \end{Vmatrix},
$$

где Q и T — квадратные матрицы, а R и S — прямоугольные подматрицы, хотя бы одна из которых ненулевая. В свою очередь, множество состояний, принадлежащих одной связной компоненте $Z^{(l)}$ $(l = \overline{1, L})$, может быть разбито на подмножества $Z^{(1l)}, \dots, Z^{(r_l l)}$ $(r_l \geqslant 1)$, чередующиеся в последовательных переходах системы по детерминированной схеме. Такому разбиению соответствует представление

$$
\Pi^{(l)}(d) = \begin{Vmatrix}
0 & \Pi_{12}^{(l)}(d) & 0 & \dots & 0 & 0 \\
0 & 0 & \Pi_{23}^{(l)}(d) & \dots & 0 & 0 \\
\dots & \dots & \dots & \dots & \dots & \dots \\
0 & 0 & 0 & \dots & \Pi_{r_l-1, r_l}^{(l)}(d) & 0 \\
\Pi_{r_l, 1}^{(l)}(d) & 0 & 0 & \dots & 0 & 0
\end{Vmatrix}.
$$

Неразложимую матрицу $\Pi^{(l)}(d)$, приводимую к такому виду, называют *циклической матрицей индекса* r_l. Стохастическую матрицу $\Pi^{(l)}(d)$ называют *примитивной*, если $r_l = 1$.

Отметим также важный в дальнейшем факт, касающийся свойств матрицы $[I - \Pi^{(l)}(d)]$, где I — единичная матрица той же размерности K_l, что и $\Pi^{(l)}(d)$ $(l = \overline{1, L})$:

$$\operatorname{rang}[I - \Pi^{(l)}(d)] = K_l - 1.$$

Этот результат является следствием теоремы Романовского [112]. Отсюда, в частности, следует, что система уравнений

$$p^{(l)} = [\Pi^{(l)}(d)]^{\mathrm{T}} p^{(l)}, \quad p^{(l)} \in R^{K_l}, \quad \sum_{i=1}^{K_l} p_i^{(l)} = 1$$

имеет единственное решение. При $r_l = 1$ это решение характеризует предельное распределение вероятностей состояний в l-м подклассе $Z^{(l)}$ $(l = \overline{1, L})$, не зависящее от начального распределения. Отметим также, что однородная марковская цепь (при некоторой стационарной стратегии $\{d\} \in \Sigma_s$) называется *регулярной*, если все ее состояния существенны и образуют одну связную компоненту $(L = 1)$ с одним циклическим подклассом $(r_1 = 1)$. Таким образом, для регулярных цепей существует и единственно финальное (при $n \to \infty$) распределение состояний цепи.

Перейдем теперь снова к рассмотрению однородных управляемых марковских цепей и связанных с ними понятий. Будем называть управляемую цепь *регулярной*, если при любой стационарной стратегии $\{d\} \in \Sigma_S$ соответствующая однородная марковская цепь регулярна в указанном выше смысле. Управляемая цепь называется *связной*, если для любой упорядоченной пары состояний $z(\alpha_1)$, $z(\alpha_T) \in Z$ найдется хотя бы одна детерминированная стратегия $\{x_n(z_n;\ z_t,\ x_t \mid t = \overline{1,\ n-1})\}$, для которой при некотором n оказывается, что

$$\mathsf{P}\{z_n = z(\alpha_T) \mid z_1 = z(\alpha_1); \\ x_s(z_s;\ z_t,\ x_t \mid t = \overline{1, s-1}),\ s = \overline{1, n-1}\} > 0.$$

Приведем важный результат, который установлен В. Г. Сраговичем [118], разъясняющий смысл понятия связности управляемой марковской цепи.

Лемма 5.1. *Для того чтобы управляемая марковская цепь из класса цепей $M(K, N)$ была связной, необходимо и достаточно, чтобы существовала такая стационарная невырожденная стратегия управления $\{d\} \in \Sigma_S^+$, что матрица цепи $\Pi(d)$ неразложима.*

Доказательство. 1) Необходимость. Пусть управляемая марковская цепь связна, т. е. для любых состояний цепи $z(\alpha_1)$, $z(\alpha_T) \in Z$ найдутся промежуточные состояния $z(\alpha_2)$, ..., $z(\alpha_{T-1}) \in Z$ и управления (варианты) $x(l_1)$, ..., $x(l_{T-1}) \in X$, такие, что

$$\pi^{l_1}_{\alpha_1 \alpha_2} > 0, \ \ldots, \ \pi^{l_{T-1}}_{\alpha_{T-1} \alpha_T} > 0.$$

Поскольку $(\Pi(d))_{ij} = \sum_{l=1}^{N} \pi^l_{ij} d^{il}$, то для любых стационарных невырожденных рандомизированных стратегий $\{d\} \in \Sigma^+_S$ вероятность перехода из состояния $z(\alpha_1)$ в состояние $z(\alpha_T)$ положительна, так как положительны вероятности переходов из $z(\alpha_i)$ в $z(\alpha_{i+1})$ $(i = \overline{1, T-1})$. В силу конечности цепи получим, что при такой стратегии $\{d\} \in \Sigma^+_S$ все состояния сообщающиеся, а потому матрица $\Pi(d)$ соответствующей цепи неразложима.

2) Достаточность. Пусть теперь существует такая стратегия $\{d\} \in \Sigma^+_S$, что $\Pi(d)$ неразложима. А это означает снова, что для любых состояний $z(\alpha_1)$, $z(\alpha_T) \in Z$ найдется цепочка состояний $z(\alpha_2)$, ..., $z(\alpha_{T-1})$, которые переходят последовательно одно в другое (можно считать, что в цепочке нет повторений) при соответствующих детерминированных управлениях. Но это и характеризует связность управляемой цепи. Лемма доказана. ▲

Замечание. В п. 1) доказан более сильный, чем сформулированный в лемме 5.1, результат, а именно: из связности управляемой марковской цепи с необходимостью вытекает неразложимость матрицы $\Pi(d)$ при любой невырожденной стационарной стратегии $\{d\} \in \Sigma^+_S$.

Поскольку элементы матрицы $\Pi(d)$ (5.6) представляют собой линейные относительно d функции, то ясно, что для любой невырожденной стационарной стратегии $\{d\} \in \Sigma^+_S$ распределение нулей в матрице $\Pi(d)$ одинаково. Отсюда следует, что при управлении произвольной цепью из класса $M(K, N)$ для любой стратегии $\{d\} \in \Sigma^+_S$ структура соответствующей однородной марковской цепи одинакова, т. е. все множество состояний Z однозначно представляется в виде объединения непересекающихся классов $Z^{(1)}_+, \ldots, Z^{(L)}_+$ сообщающихся состояний (так называемых связных компонент) и класса $Z^{(0)}_+$ несущественных состояний [112]. При вырождении стратегии $\{d\}$

(т. е. при занулении некоторых элементов матрицы d) часть элементов матрицы $\Pi(d)$ может обратиться в нуль, что может привести к изменению структуры множеств $Z_+^{(l)}(l=\overline{0, L})$: каждое из этих множеств может «развалиться» на связные компоненты и множество несущественных состояний. Кроме того, из сказанного выше следует, что при любой стратегии управления $\{d_n\}\in\Sigma$ невозможно появление ненулевых элементов в матрицах $\Pi(d_n)$ на тех местах, где были нули в $\Pi(d)$ при $\{d\}\in$ $\in\Sigma_S^+$.

Справедлива следующая лемма о соответствующем разбиении множества Ω элементарных исходов на непересекающиеся подмножества.

Лемма 5.2. *Для любой управляемой марковской цепи из класса цепей $M(K, N)$ при любом распределении вероятностей начального состояния и при любой стратегии управления $\{d_n\}\in\Sigma$ (5.5) существует разбиение множества Ω элементарных исходов на непересекающиеся подмножества Ω_l $(l=\overline{0, L})$, такие, что для почти всех $\omega\in\Omega_l$ существует конечный момент $n_l=n_l(\omega)$, что $z_n\in Z_+^{(l)}\forall n\geqslant n_l(\omega)$.*

Доказательство. Пусть $\Omega_0\subset\Omega$ — подмножество элементарных исходов, для которых $z_n\in Z_+^{(0)}$ при всех $n=1, 2, \ldots$ Рассмотрим множество $\overline{\Omega}_0=\Omega\backslash\Omega_0$. Тогда, очевидно, для почти всех $\omega\in\overline{\Omega}_0$ найдется конечный момент $n_l(\omega)$, такой, что процесс z_n попадет (при первом же выходе из $Z_+^{(0)}$) в одну из связных компонент $Z_+^{(l)}$. Если же $z_1\in Z_+^{(l)}$ $(l=\overline{1, L})$, то $n_l(\omega)=1$. Докажем, что на этих ω процесс уже никогда не выйдет из $Z_+^{(l)}$. Имеем

$$\mathsf{P}\left\{z_{n_l+1}\overline{\in} Z_+^{(l)}, z_{n_l}\in Z_+^{(l)}\right\}=$$
$$=\mathsf{M}\left\{\chi\left(z_{n_l+1}\overline{\in} Z_+^{(l)}, z_{n_l}\in Z_+^{(l)}\right)\right\}=$$
$$=\mathsf{M}\left\{\mathsf{M}\left\{\chi\left(z_{n_l+1}\overline{\in} Z_+^{(l)}\right)\chi\left(z_{n_l}\in Z_+^{(l)}\right)\middle|\mathscr{F}_{n_l}\right\}\right\}=$$
$$=\mathsf{M}\left\{\chi\left(z_{n_l}\in Z_+^{(l)}\right)\mathsf{M}\left\{\chi\left(z_{n_l+1}\overline{\in} Z_+^{(l)}\right)\middle|\mathscr{F}_{n_l}\right\}\right\},$$

где

$$\mathscr{F}_n\triangleq\sigma(z_t, d_t\,|\,t=\overline{1, n_l}).$$

Но

$$\chi\left(z_{n_l} \in Z_+^{(l)}\right) \mathsf{M}\left\{\chi\left(z_{n_l+1} \overline{\in} Z_+^{(l)}\right) \middle| \mathcal{F}_{n_l}\right\} =$$
$$= \chi\left(z_{n_l} \in Z_+^{(l)}\right) \sum_{z(j) \overline{\in} Z_+^{(l)}} \sum_{z(i) \in Z_+^{(l)}} \sum_{k=1}^{N} \pi_{ij}^k d_{n_l}^{ik} \overset{\text{п.н.}}{=} 0,$$

поскольку, как отмечалось выше,

$$\sum_{k=1}^{N} \pi_{ij}^k d_n^{ik} = 0 \quad \forall z(i) \in Z_+^{(l)}, \quad z(j) \overline{\in} Z_+^{(l)}$$

при любых стратегиях $\{d_n\} \in \Sigma$. А следовательно,

$$\mathsf{P}\left\{z_{n_l+1} \overline{\in} Z_+^{(l)},\ z_{n_l} \in Z_+^{(l)}\right\} \overset{\text{п.н.}}{=} 0.$$

Лемма доказана. ▲

Обратимся теперь снова к исходной управляемой системе, свойства которой сформулированы в предположениях П1‴ — П3‴. Рассмотрим произвольную стратегию управления $\{d_n\} \in \Sigma$ (5.5). Она порождает последовательность вариантов $\{x_n\}$ и потерь $\{\xi_n\}$. Качество функционирования управляемой системы к моменту времени t_n будем характеризовать величиной текущих средних потерь

$$\Phi_n \triangleq \frac{1}{n} \sum_{t=1}^{n} \xi_t, \tag{5.7}$$

а качество стратегии управления $\{d_n\} \in \Sigma$ при ее реализации на элементарном исходе $\omega \in \Omega$ будем определять по наибольшей предельной величине последовательности текущих средних потерь

$$\Phi = \Phi(\{d_n\}, \omega) \triangleq \overline{\lim_{n\to\infty}} \Phi_n. \tag{5.8}$$

Хотя Φ зависит также от начального состояния $z_1 \in Z$, это далее специально не оговаривается.

Поскольку величины v_{il} и π_{ij}^l $(i, j = \overline{1, K},\ l = \overline{1, N})$ априори неизвестны, то задачу адаптивного управления рассматриваемой системой естественно поставить как задачу формирования по доступным наблюдениям такой последовательности вариантов (управлений), которая в определенном вероятностном смысле минимизирует величину предельных средних потерь Φ (5.8). Для того, чтобы строго сформулировать эту задачу, перейдем к изучению свойств последовательности $\{\Phi_n\}$, порожденной произвольной стратегией $\{d_n\} \in \Sigma$.

§ 5.2. Свойства предельных средних потерь

При дальнейшем анализе управляемых конечных марковских цепей полезным оказывается следующее утверждение об асимптотическом представлении текущих средних потерь посредством линейной комбинации текущих частот выбора соответствующих вариантов (управлений).

Лемма 5.3. *В предположениях* П1‴, П2‴ *при любых стратегиях управления* $\{d_n\} \in \Sigma$ (5.5) *и любых распределениях вероятностей начального состояния* $z_1 \in Z$ *справедливо соотношение*

$$\lim_{n\to\infty}\left(\Phi_n - \sum_{i=1}^{K}\sum_{l=1}^{N} v_{il}\theta_n^{il}\right) \overset{\text{п.н.}}{=} 0, \tag{5.9}$$

где

$$\theta_n^{il} \triangleq \frac{1}{n}\sum_{t=1}^{n} \chi\left(z_t = z(i),\ x_t = x(l)\right) \tag{5.10}$$

— *текущие частоты выбора вариантов (управлений)* $x(l)$ $(l = \overline{1, N})$ *в состояниях* $z(i)$ $(i = \overline{1, K})$ *к моменту* $n = 1, 2, \ldots$ *включительно.*

Доказательство. Представим Φ_n в следующем виде:

$$\Phi_n = \sum_{i=1}^{K}\sum_{l=1}^{N}\frac{1}{n}\sum_{t=1}^{n}\xi_t\left(z(i),\ x(l), \omega\right)\chi\left(z_t = z(i),\ x_t = x(l)\right) = \sum_{i=1}^{K}\sum_{l=1}^{N}\hat{v}_n^{il}\hat{\theta}_n^{il}, \tag{5.11}$$

где

$$\hat{v}_n^{il} \triangleq \frac{\sum\limits_{t=1}^{n}\xi_t\left(z(i),\ x(l), \omega\right)\chi\left(z_t = z(i),\ x_t = x(l)\right)}{1 + \sum\limits_{t=1}^{n}\chi\left(z_t = z(i),\ x_t = x(l)\right)}, \tag{5.12}$$

$$\hat{\theta}_n^{il} \triangleq \theta_n^{il} + \frac{1}{n}, \quad i = \overline{1, K}, \quad l = \overline{1, N}, \ n = 1, 2, \ldots \tag{5.13}$$

Очевидно, что

$$\lim_{n\to\infty}\left(\hat{\theta}_n^{il} - \theta_n^{il}\right) = 0. \tag{5.14}$$

Покажем, что также при всех $i=\overline{1, K}$, $l=\overline{1, N}$ имеет место асимптотическое равенство

$$\lim_{n\to\infty} \hat{v}_n^{il} = \begin{cases} v_{il} \text{ при } \omega \in \Omega_{il}^{\infty}, \\ 0 \text{ при } \omega \overline{\in} \Omega_{il}^{\infty}, \end{cases} \tag{5.15}$$

где $\Omega_{il}^{\infty} \subseteq \Omega$ — множество элементарных исходов, на которых бесконечно много раз выбирается вариант $x(l)$ в состоянии $z(i)$, т. е.

$$\Omega_{il}^{\infty} \triangleq \left\{ \omega \,\middle|\, \sum_{t=1}^{\infty} \chi(z_t = z(i),\, x_t = x(l)) = \infty \right\}. \tag{5.16}$$

Для этого представим $\hat{v}_n^{il}$ (5.12) в рекуррентном виде:

$$\hat{v}_n^{il} = \hat{v}_{n-1}^{il}\left(1 - \frac{\chi(z_n = z(i),\, x_n = x(l))}{1 + \sum\limits_{t=1}^{n} \chi(z_t = z(i),\, x_t = x(l))}\right) + \\ + \frac{\chi(z_n = z(i),\, x_n = x(l))}{1 + \sum\limits_{t=1}^{n} \chi(z_t = z(i),\, x_t = x(l))}\, \xi_n(z(i),\, x(l),\, \omega).$$

Тогда в силу предположений П1‴, П2‴ для величин

$$\Delta_n^{il} \triangleq \hat{v}_n^{il} - v_{il} \quad (i=\overline{1, K},\ l=\overline{1, N},\ n=1, 2, \ldots)$$

с вероятностью 1 справедливы неравенства

$$\mathsf{M}\left\{(\Delta_n^{il})^2 \,\middle|\, \hat{\mathscr{F}}_{n-1}\right\} \leqslant \\ \leqslant (\Delta_{n-1}^{il})^2 - 2\Delta_{n-1}^{il}\gamma_n^{il}\left(\hat{v}_{n-1}^{il} - \mathsf{M}\left\{\xi_n(z(i),\, x(l),\, \omega) \,\middle|\, \hat{\mathscr{F}}_{n-1}\right\}\right) + \\ + C_0(\gamma_n^{il})^2 = (\Delta_{n-1}^{il})^2(1 - 2\gamma_n^{il}) + C_0(\gamma_n^{il})^2, \tag{5.17}$$

где

$$\gamma_n^{il} \triangleq \frac{\chi(z_n = z(i),\, x_n = x(l))}{1 + \sum\limits_{t=1}^{n} \chi(z_t = z(i),\, x_t = x(l))},$$

$$\hat{\mathscr{F}}_n \triangleq \sigma\{z_{n+1},\, x_{n+1},\, z_t,\, x_t,\, \xi_t \mid t=\overline{1, n}\}.$$

Для почти всех $\omega \in \Omega_{il}^{\infty}$ (5.16)

$$\sum_{n=1}^{\infty} \gamma_n^{il} = \sum_{k=2}^{\infty} k^{-1} = \infty, \quad \sum_{n=1}^{\infty} (\gamma_n^{il})^2 = \sum_{k=2}^{\infty} k^{-2} < \infty. \tag{5.18}$$

Это позволяет воспользоваться леммой П.9, в силу которой из соотношений (5.17), (5.18) получим, что $\Delta_n^{il} \xrightarrow[n\to\infty]{} 0$ для почти всех $\omega \in \Omega_{il}^{\infty}$ (5.16), т. е. верно (5.15). Но из (5.17), (5.15) с учетом представления (5.11)—(5.13) вытекает утверждение леммы. Лемма доказана. ▲

Для однородных управляемых марковских цепей при любой стратегии управления $\{d_n\} \in \Sigma$ на тех случайных траекториях процесса $\{z_n\}$, на которых ряд состояний цепи реализуется бесконечно много раз, справедлив так называемый *эргодический принцип*, позволяющий в асимптотике заменять некоторые средние по времени характеристики процесса их математическими ожиданиями. Более точно этот факт отражает следующая лемма.

Лемма 5.4. *Если выполнено предположение* П3‴, *то при любой стратегии управления* $\{d_n\} \in \Sigma$ (5.5) *и любых распределениях вероятностей начального состояния* $z_1 \in Z$ *для почти всех* ω, *таких, что*

$$\omega \in \left\{\omega \in \Omega \,\middle|\, \sum_{t=1}^{\infty}{}' \chi(z_t = z(i)) = \infty\right\} \triangleq \Omega_i, \quad i = \overline{1, K} \tag{5.19}$$

справедливо равенство

$$\lim_{n\to\infty}\left[\frac{\sum\limits_{t=1}^{n} \chi(z_t = z(i),\ z_{t+1} = z(j))}{1 + \sum\limits_{t=1}^{n} \chi(z_t = z(i))} - \pi_{ij}(\tilde{d}_n)\right] \overset{\text{п.н.}}{=} 0, \tag{5.20}$$

где

$$\pi_{ij}(\tilde{d}_n) \triangleq \sum_{l=1}^{N} \pi_{ij}^{l} \tilde{d}_n^{il}, \qquad i, j = \overline{1, K}, \tag{5.21}$$

$$\tilde{d}_n^{il} \triangleq \frac{\sum\limits_{t=1}^{n} \chi(z_t = z(i),\ x_t = x(l))}{1 + \sum\limits_{t=1}^{n} \chi(z_t = z(i))}, \quad l = \overline{1, N}. \tag{5.22}$$

Доказательство. Определив для всех $i, j = \overline{1, K}$ и $n = 1, 2, \ldots$ величины

$$\tilde{\Delta}_{n+1}^{ij} \triangleq \left[\sum_{t=1}^{n} \chi(z_t = z(i),\ z_{t+1} = z(j))\right] \Bigg/ \left(1 + \sum_{t=1}^{n} \chi(z_t = z(i))\right) - \pi_{ij}(\tilde{d}_n),$$

получим

$$\tilde{\Delta}_{n+1}^{ij} = \tilde{\Delta}_n^{ij} - \\ - \beta_n^i \left[\tilde{\Delta}_n^{ij} - \left(\chi(z_{n+1} = z(j)) - \sum_{l=1}^{N} \pi_{ij}^l \chi(x_n = x(l)) \right) \right],$$

где

$$\beta_n^i \triangleq \frac{\chi(z_n = z(i))}{1 + \sum_{t=1}^{n} \chi(z_t = z(i))}.$$

Тогда с учетом (5.3) и (5.5) получаем соотношение

$$\mathbf{M}\{(\tilde{\Delta}_{n+1}^{ij})^2 | \mathcal{F}_n\} \overset{\text{п.н.}}{\leqslant} (\tilde{\Delta}_n^{ij})^2 - \\ - 2\beta_n^i \tilde{\Delta}_n^{ij} \left[\tilde{\Delta}_n^{ij} - \sum_{s=1}^{K} \chi(z_n = z(s)) \left(\sum_{l=1}^{N} \pi_{sj}^l d_n^{sl} - \sum_{l=1}^{N} \pi_{sj}^l d_n^{sl} \right) \right] + \\ + 4(\beta_n^i)^2 \overset{\text{п.н.}}{=} (\tilde{\Delta}_n^{ij})^2 (1 - 2\beta_n^i) + 4(\beta_n^i)^2,$$

откуда в силу леммы П.9 следует, что

$$\tilde{\Delta}_n^{ij} \xrightarrow[n\to\infty]{\text{п.н.}} 0, \quad i, j = \overline{1, K}.$$

Лемма доказана. ▲

Доказанные выше леммы 5.3, 5.4 позволяют установить важное свойство текущих средних потерь, которое конкретизирует их структуру в асимптотике.

Лемма 5.5. *Если выполнены предположения* П1‴ — П3‴, *то при любой стратегии управления* $\{d_n\} \in \Sigma$ (5.5) *и любых распределениях вероятностей начального состояния* $z_1 \in Z$ *с вероятностью* 1

$$\lim_{n\to\infty} \left(\Phi_n - \sum_{i=1}^{K} \sum_{l=1}^{N} v_{il} \tilde{d}_n^{il} \tilde{p}_n(i) \right) = 0, \tag{5.23}$$

причем величины $\tilde{d}_n^{il}$ *определены в* (5.22), *а* $\tilde{p}_n(i)$ *являются компонентами стохастического вектора* $\tilde{p}_n$

$$\tilde{p}_n \in S^K \triangleq \left\{ p \mid p(i) \geqslant 0,\ i = \overline{1, K},\ \sum_{i=1}^{K} p(i) = 1 \right\} \tag{5.24}$$

и удовлетворяют с вероятностью 1 *соотношению*

$$\lim_{n\to\infty}\left[\tilde{p}_n(j)-\sum_{i=1}^{K}\tilde{p}_n(i)\,\pi_{ij}(\tilde{d}_n)\right]=0, \tag{5.25}$$

где величины $\pi_{ij}(\tilde{d}_n)$ *определены в* (5.21).

Доказательство. В силу предположений П1‴, П2‴ по лемме 5.3 имеем с вероятностью 1 (см. (5.9))

$$\Phi_n=\sum_{i=1}^{K}\sum_{l=1}^{N}v_{il}\theta_n^{il}+o(1), \tag{5.26}$$

где $o(1)$ — случайная бесконечно малая (почти наверное) последовательность, а θ_n^{il} определяются формулой (5.10). Заметим, что θ_n^{il} можно представить в виде

$$\theta_n^{il}=\tilde{d}_n^{il}\tilde{p}_n(i),\quad i=\overline{1,K},\ l=\overline{1,N}, \tag{5.27}$$

где $\tilde{d}_n^{il}$ определены равенством (5.22), а

$$\tilde{p}_n(i)=\frac{1}{n}\left(1+\sum_{t=1}^{n}\chi(z_t=z(i))\right). \tag{5.28}$$

Для доказательства теоремы достаточно теперь показать, что в (5.27) вектор $\tilde{p}_n$ (5.28) можно заменить на вектор $\tilde{p}_n(\tilde{d}_n)$, определенный в (5.24), (5.25). Представим $\tilde{p}_n(j)$ в виде

$$\tilde{p}_n(j)=\frac{1}{n}\sum_{t=2}^{n}\chi(z_t=z(j))\sum_{i=1}^{K}\chi(z_{t-1}=z(i))+\\+\frac{1}{n}[1+\chi(z_1=z(j))],$$

или, иначе,

$$\tilde{p}_n(j)=\sum_{i=1}^{K}\frac{1}{n}\left[1+\sum_{t=1}^{n-1}\chi(z_t=z(i))\right]\times\\\times\frac{\sum_{t=1}^{n-1}\chi(z_t=z(i),\ z_{t+1}=z(j))}{1+\sum_{t=1}^{n-1}\chi(z_t=z(i))}+O(n^{-1}),\quad j=\overline{1,K}. \tag{5.29}$$

Для почти всех $\omega\in\overline{\Omega}_i$ (5.19) из (5.28) следует, что $\tilde{p}_n(i)\xrightarrow[n\to\infty]{}0\left(i=\overline{1,K}\right)$. Для остальных $\omega\in\Omega_i$ в силу предположения П3‴ по лемме 5.4 получаем

$$\lim_{n\to\infty}\left[\tilde p_n(j)-\sum_{i=1}^{K}\frac{1}{n}\left[1+\sum_{t=1}^{n-1}\chi(z_t=z(i))\right]\pi_{ij}(\tilde d_n)\right]=0$$

и, значит, с вероятностью 1

$$\lim_{n\to\infty}\left[\tilde p_n(j)-\sum_{i=1}^{K}\tilde p_n(i)\,\pi_{ij}(\tilde d_n)\right]=0,$$

т. е. $\tilde p_n$ удовлетворяет в асимптотике соотношению (5.25). Лемма доказана. ▲

Лемма 5.5 позволяет установить следующий фундаментальный для данного раздела результат.

Теорема 5.1. *Если выполнены предположения* П1‴ — П3‴, *то при любой стратегии управления* $\{d_n\}\in\Sigma$ (5.5) *произвольной конечной марковской цепью из класса* $M(K, N)$ *и любом распределении вероятностей начального состояния* $z_1\in Z$ *для почти всех* $\omega\in\Omega_l$ $(l=\overline{1, L})$ *справедливы неравенства*

$$\max_{\{d\}\in\Sigma_s}\Phi^{(l)}_{\max}(d)\geqslant\varlimsup_{n\to\infty}\Phi_n\geqslant\varliminf_{n\to\infty}\Phi_n\geqslant\min_{\{d\}\in\Sigma_s}\Phi^{(l)}_{\min}(d),\tag{5.30}$$

где

$$\Phi^{(l)}_{\max}(d)\triangleq\max_{p^{(l)}\in Q^{(l)}(d)}\sum_{i\in Z^{(l)}_{+}}\sum_{k=1}^{N}v_{ik}d^{ik}p^{(l)}_i,\tag{5.31}$$

$$\Phi^{(l)}_{\min}(d)\triangleq\min_{p^{(l)}\in Q^{(l)}(d)}\sum_{i\in Z^{(l)}_{+}}\sum_{k=1}^{N}v_{ik}d^{ik}p^{(l)}_i,\tag{5.32}$$

множество $Q^{(l)}(d)$ *является множеством всех решений* $p^{(l)}$ *системы уравнений*

$$p^{(l)}=[\Pi^{(l)}(d)]^{\mathrm{T}}p^{(l)},\quad e^{\mathrm{T}}p^{(l)}=1\tag{5.33}$$

($\Pi^{(l)}(d)$ — *стохастическая матрица, соответствующая l-й связной компоненте* $Z^{(l)}_{+}$, *а множества* Ω_l *определены в лемме* 5.2).

Доказательство. Как отмечалось в лемме 5.5, с вероятностью 1

$$\varlimsup_{n\to\infty}\sum_{i=1}^{K}\sum_{l=1}^{N}v_{il}\tilde d^{il}_n\tilde p_n(i)\geqslant\varlimsup_{n\to\infty}\Phi_n\geqslant\varliminf_{n\to\infty}\Phi_n\geqslant$$
$$\geqslant\varliminf_{n\to\infty}\sum_{i=1}^{K}\sum_{l=1}^{N}v_{il}\tilde d^{il}_n\tilde p_n(i),\tag{5.34}$$

где $\tilde{d}_n^{il}$ и $\tilde{p}_n(i)$ определены соотношениями (5.22) и (5.28) соответственно, при этом для почти всех $\omega \in \Omega_l$ (см. лемму 5.2)

$$\begin{gathered}\tilde{p}_n(i) \xrightarrow[n\to\infty]{} 0, \quad z(i) \,\overline{\in}\, Z_+^{(l)}, \\ \tilde{p}_n(j) - \sum_{z(i)\in Z_+^{(l)}} \pi_{ij}(\tilde{d}_n)\,\tilde{p}_n(i) \xrightarrow[n\to\infty]{} 0, \quad z(i) \in Z_+^{(l)}.\end{gathered} \tag{5.35}$$

Следовательно, из (5.35) получаем, что для почти всех $\omega \in \Omega_l$

$$\sum_{i=1}^{K}\sum_{k=1}^{N} v_{ik}\tilde{d}_n^{ik}\tilde{p}_n(i) - \sum_{z(i)\in Z_+^{(l)}}\sum_{k=1}^{N} v_{ik}\tilde{d}_n^{ik}\tilde{p}_n(i) \xrightarrow[n\to\infty]{} 0.$$

Для произвольного $\omega \in \Omega_l$ рассмотрим последовательность моментов $\{n_k\}$, $k = 1, 2, \ldots$, такую, что последовательности $\{\tilde{d}_{n_k}\}$ и $\{\tilde{p}_{n_k}^{(l)}\}$ соответственно сходятся к предельным значениям d и $p^{(l)}$, где $\tilde{p}_n^{(l)}$ — вектор, составленный из компонент $\tilde{p}_n(i)$, $z(i) \in Z_+^{(l)}$. Из (5.35) следует, что $p^{(l)} \in Q^{(l)}(d)$ (5.33), а из (5.34) и последнего соотношения получаем (5.30). Теорема доказана. ▲

После того как в теореме 5.1 установлены возможные границы предельных точек последовательности $\{\Phi_n\}$ текущих средних потерь, появилась возможность строго сформулировать задачу адаптивного управления произвольной конечной однородной марковской цепью, которая и реализуется в следующем параграфе.

§ 5.3. Постановка задачи адаптивного управления конечной марковской цепью

Теорема 5.1 позволяет сформулировать задачу адаптивного управления конечной однородной цепью Маркова следующим образом:

найти такую стратегию управления $\{d_n\} \in \Sigma$ (5.5), которая для всякой марковской цепи из класса $M(K, N)$ реализовала бы такую последовательность вариантов (управлений) $\{x_n\}$ и соответствующую ей последовательность потерь $\{\xi_n\}$, что с вероятностью 1 выполнялось бы целевое условие

$$\overline{\lim_{n\to\infty}}\, \Phi_n \leqslant \Phi^*, \tag{5.36}$$

где

$$\Phi^* \triangleq \min_{\{d\}\in\Sigma_S} \max_{k=\overline{1,L}} \min_{p^{(k)}\in Q^{(k)}(d)} \sum_{i\in Z_+^{(k)}} \sum_{l=1}^{N} v_{il} d^{il} p_i^{(k)} \tag{5.37}$$

(множества $Z_+^{(k)}, Q^{(k)}(d)$ и векторы $p^{(k)}$ определены в теореме 5.1).

В следующих ниже параграфах будет предложен и исследован рекуррентный алгоритм решения поставленной задачи, который порождает последовательность невырожденных рандомизированных правил выбора вариантов (управлений), т. е. $d_n^{il} > 0, i = \overline{1, K}, l = \overline{1, N}, n = 1, 2, \ldots$ с вероятностью 1. Поэтому далее потребуется результат следующей теоремы.

Теорема 5.2. *Для любой управляемой системы, удовлетворяющей предположениям* П1‴ — П3‴, *справедливо соотношение*

$$\Phi_* \triangleq \inf_{\{d\}\in\Sigma_S^+} \max_{k=\overline{1,L}} \sum_{i\in Z_+^{(k)}} \sum_{l=1}^{N} v_{il} d^{il} p_i^{(k)} = \Phi^*, \tag{5.38}$$

где векторы $p^{(k)}$ определены в теореме 5.1 *и при $\{d\}\in\Sigma_S^+$ являются единственными решениями системы уравнений* (5.33).

Доказательство. На множестве

$$d\in D \triangleq \left\{ d \,\middle|\, d^{il} \geqslant 0,\quad \sum_{l=1}^{N} d^{il} = 1\ (i = \overline{1, K},\ l = \overline{1, N}) \right\}$$

рассмотрим функции

$$V^{(k)}(d) \triangleq \min_{p^{(k)}\in Q^{(k)}(d)} \sum_{i\in Z_+^{(k)}} \sum_{l=1}^{N} v_{il} d^{il} p_i^{(k)},\quad k = \overline{1, L}. \tag{5.39}$$

Если $d\in \operatorname{int} D$, т. е. стратегия $\{d\}\in\Sigma_S^+$, то по теореме 5.1 матрица каждой подцепи $\Pi^{(k)}(d)$ $(k = \overline{1, L})$ неразложима, а значит, в силу теоремы Романовского [111] множество $Q^{(k)}(d)$ состоит из единственной точки, при этом $p^{(k)}$ непрерывны по d, а следовательно, $V^{(k)}(d)$ также непрерывны по $d\in \operatorname{int} D$.

Пусть теперь $d \in \partial D \triangleq D \setminus \operatorname{int} D$, т. е. стратегия $\{d\}$ вырождена. При этом переходная матрица подцепи $\Pi^{(k)}(d)$ $(k = \overline{1, L})$ может стать разложимой и, в общем случае, может быть представлена в виде

$$\Pi^{(k)}(d) = \left\| \begin{matrix} \pi_1^{(k)} & & & & \\ & \pi_2^{(k)} & & 0 & \\ 0 & & \ddots & & \\ & & & \pi_{L_k}^{(k)} & \\ R_1^{(k)} & R_2^{(k)} & \ldots & R_{L_k}^{(k)} & R^{(k)} \end{matrix} \right\|. \tag{5.40}$$

Матрицы $\pi_i^{(k)}$ $(i = \overline{1, L_k})$ являются неразложимыми квадратными стохастическими матрицами, которые соответствуют эргодическим подклассам $Z_i^{(k)}$ состояний на множестве $Z^{(k)}$. Матрица $R^{(k)}$ соответствует несущественным состояниям, а матрицы $R_i^{(k)}$ $(i = \overline{1, L_k})$ характеризуют вероятности переходов из несущественных состояний в состояния из эргодических подклассов. Вообще говоря, различным вырожденным стратегиям $\{d\}$ могут соответствовать различные структуры переходной матрицы (5.40). Множество $Q^{(k)}(d)$ состоит из векторов $p^{(k)}$, удовлетворяющих соотношениям

$$p^{(k)} = [\Pi^{(k)}(d)]^{\mathrm{T}} p^{(k)}, \quad e^{\mathrm{T}} p^{(k)} = 1$$

и в силу (5.40) представимых в виде

$$p^{(k)} = (p^1, \ldots, p^{L_k}, 0), \tag{5.41}$$

где векторы p^i являются единственными (в силу неразложимости матриц $\pi_i^{(k)}(d)$) стационарными распределениями состояний цепи из i-го эргодического подкласса $Z_i^{(k)}$, т. е. p^i удовлетворяют соотношениям

$$p^i = [\pi_i^{(k)}(d)]^{\mathrm{T}} p^i, \tag{5.42}$$

$$e^{\mathrm{T}} p^i = \alpha_i \geqslant 0, \quad i = \overline{1, L_k}, \quad \sum_{i=1}^{L_k} \alpha_i = 1. \tag{5.43}$$

Задавшись произвольными $\delta \in (0, 1)$ и $\{d^+\} \in \Sigma_S^+$, рассмотрим стационарную невырожденную стратегию $\{d_\delta\} \in \Sigma_S^+$, определенную формулой

$$d_\delta \triangleq (1 - \delta) d + \delta d^+. \tag{5.44}$$

Стратегия $\{d_\delta\}$ получена регуляризацией стратегии $\{d\}$. Поскольку $d_\delta \in \operatorname{int} D$, то решение $p(d_\delta)$ системы уравнений

$$p(d_\delta) = [\Pi^{(k)}(d_\delta)]^{\mathrm{T}} p(d_\delta), \quad e^{\mathrm{T}} p(d_\delta) = 1 \tag{5.45}$$

единственно при любом $\delta > 0$. Решение же любой системы

$$\begin{aligned} &\tilde{p}^i(d_\delta) = [\pi_\delta^{(k)}(d_\delta)]^{\mathrm{T}} \tilde{p}^i(d_\delta), \quad i = \overline{1, L_k}, \\ &e^{\mathrm{T}} \tilde{p}^i(d_\delta) = \alpha_i \end{aligned} \tag{5.46}$$

единственно не только при $\delta > 0$, но и при $\delta = 0$, поскольку матрицы $\pi_i^k(d_\delta)$ неразложимы при любых $\delta \geqslant 0$ (см. (5.42)). Из (5.45) имеем

$$p^i(d_\delta) = [\pi_i^{(k)}(d_\delta)]^{\mathrm{T}} p^i(d_\delta) + O(\delta), \quad i = \overline{1, L_k},$$

откуда с учетом (5.46) получаем, что

$$\Delta^i(d_\delta) \triangleq p^i(d_\delta) - \tilde{p}^i(d_\delta) = [\pi_i^{(k)}(d_\delta)]^{\mathrm{T}} \Delta^i(d_\delta) + O(\delta),$$

и при этом в силу (5.43)

$$e^{\mathrm{T}} \Delta^i(d_\delta) = 1 - \sum_{z(i) \equiv Z_i^{(l)}} p_j(d_\delta) - \alpha_i = 0.$$

Но так как $\pi_i^{(k)}(d_\delta) = \pi_i^{(k)}(d) + O(\delta)$, а $\pi_i^{(k)}(d)$ — неразложимы, то решение системы уравнений относительно $\Delta^i(d_\delta)$ $(i = \overline{1, L_k})$

$$\begin{aligned} &\Delta^i(d_\delta) = [\pi_i^{(k)}(d)]^{\mathrm{T}} \Delta^i(d_\delta) + O(\delta), \\ &e^{\mathrm{T}} \Delta^i(d_\delta) = 0 \end{aligned}$$

единственно и $\Delta^i(d_\delta) = O(\delta) \underset{\delta \to 0}{\longrightarrow} 0$. Следовательно, вектор $p(d_\delta)$ (5.45) при $\delta \to 0$ стремится к вектору

$$(p^1(d), \ldots, p^{L_k}(d), 0),$$

поскольку $[p^i(d_\delta) - \tilde{p}^i(d_\delta)] \underset{\delta \to 0}{\longrightarrow} 0$ и $\tilde{p}^i(d_\delta) \underset{\delta \to 0}{\longrightarrow} p^i(d)$ $(i = \overline{1, L_k})$, а вектор $p^{L_k+1}(d_\delta) \underset{\delta \to 0}{\longrightarrow} 0$ в силу условия нормировки в (5.45) и условия (5.44).

Таким образом, имеет место представление

$$p(d_\delta) = p(d) + \Delta(\delta, d^+, d),$$

$$e^{\mathrm{T}} \Delta(\delta, d^+, d) = 0, \quad p(d) \in Q^{(k)}(d), \quad \Delta(\delta, d^+, d) \underset{\delta \to 0}{\longrightarrow} 0,$$

откуда в силу непрерывности функции $V^{(l)}(d)$ на $\operatorname{int} D$ вытекает соотношение

$$\inf_{d\in \operatorname{int} D} V^{(k)}(d) = \min_{d\in D} V^{(k)}(d), \quad k=\overline{1, L},$$

а значит, верно соотношение (5.38). Теорема доказана. ▲

Прежде чем переходить к рассмотрению алгоритмов, решающих поставленную задачу, остановимся на вопросе о том, во что переходит целевое условие (5.36), (5.37) в случае связной управляемой марковской цепи. Для этого класса цепей в силу леммы 5.1 при любой стационарной невырожденной стратегии $\{d\}\in\Sigma_S^+$ все множество состояний Z представляет собой одну связную компоненту $Z_+^{(1)}$ $(L=1)$ и, таким образом, $Z_+^{(0)}=\varnothing$ (несущественные состояния отсутствуют). Частным случаем таких цепей являются *эргодические управляемые цепи*, для которых это условие выполняется для любой (возможно и вырожденной) стационарной стратегии $\{d\}\in\Sigma_S$, при этом число циклических подклассов $r\geqslant 1$ и может изменяться при вырождении стратегии управления $\{d\}$. Еще более узким классом цепей являются регулярные управляемые марковские цепи, для которых $r=1$ при любых стратегиях $\{d\}\in\Sigma_S$.

Из теорем 5.1 и 5.2 следует, что для любой связной (в частности, для эргодической или регулярной) управляемой марковской цепи при любой стратегии управления $\{d_n\}\in\Sigma$ с вероятностью 1 выполняется неравенство

$$\varlimsup_{n\to\infty}\Phi_n \overset{\text{п.н.}}{\geqslant} \Phi^*, \tag{5.47}$$

причем

$$\Phi^* = \inf_{\{d\}\in\Sigma_S^+}\sum_{i=1}^{K}\sum_{k=1}^{N} v_{ik}d^{ik}p_i(d), \tag{5.48}$$

где $p_i(d)$ — компоненты вектора $p(d)$, удовлетворяющего системе уравнений (5.33). Таким образом, *задача адаптивного управления конечной марковской цепью, которая в общем случае состоит в обеспечении целевого неравенства* (5.36), *для связных цепей может быть сформулирована как задача минимизации предельных сред-*

них потерь, т. е. достижения с вероятностью 1 *величины потерь* Φ^* (5.48).

Перейдем теперь к решению задачи адаптивного управления классом конечных регулярных марковских цепей.

§ 5.4. Алгоритм адаптивного управления регулярной марковской цепью

Для класса регулярных управляемых марковских цепей задача адаптивного управления, как показано выше, может быть поставлена как задача минимизации предельных средних потерь $\Phi = \overline{\lim_{n\to\infty}} \Phi_n$, причем минимум Φ равен Φ^* (5.48). Поскольку в данном случае при любой (в том числе и вырожденной) стационарной стратегии $\{d\} \in \Sigma_s$ цепь имеет единственное финальное распределение вероятностей состояний $p_i(d)$ $(i = \overline{1, K})$, представляющее собой непрерывную функцию d, то

$$\Phi^* = \min_{d\in D} V(d), \tag{5.49}$$

где

$$V(d) = \sum_{i=1}^{K} \sum_{l=1}^{N} v_{il} d^{il} p_i(d), \tag{5.50}$$

вектор-столбец $p(d) = (p_1(d), \ldots, p_K(d))^{\mathrm{T}}$ определяется соотношением

$$p(d) = \Pi^{\mathrm{T}}(d) p(d), \quad p(d) \in S^K, \tag{5.51}$$

а D — множество стохастических матриц

$$D \triangleq \left\{ d \mid d^{il} \geqslant 0, \quad \sum_{l=1}^{N} d^{il} = 1 \; (i = \overline{1, K}, \; l = \overline{1, N}) \right\}. \tag{5.52}$$

Заметим, что минимум Φ^* величин Φ достигается при любой стационарной стратегии $\{d\} \in \Sigma_s$, $d = d^* \in D$, минимизирующей функцию $V(d)$ (5.50). В условиях полной информации о величинах v_{il} $(i = \overline{1, K}, \; l = \overline{1, N})$ и матрицах вероятностей переходов цепи $\| \pi_{ij}^{l} \|_{i,j=\overline{1,K}}^{l=\overline{1,N}}$ задача минимизации функции $V(d)$ (5.50) при ограничениях (5.51), (5.52) является задачей нелинейного программирования. Построение адаптивных алгоритмов типа стохастической аппроксимации, предназначенных непосредственно для минимизации этой функции по доступным реализациям, представляет собой достаточно сложную задачу, а получающиеся при этом алгоритмы являются громоздкими и трудно исследуемыми. Однако существует простая замена переменных, переводящая эту задачу в задачу линейного программирования. В самом деле, в новых переменных c^{il}, введенных по формуле

$$c^{il} \triangleq d^{il} p_i(d), \quad i = \overline{1, K}, \quad l = \overline{1, N}, \tag{5.53}$$

задача минимизации функции $V(d)$ (5.50), (5.51) на множестве D (5.52) записывается в следующем виде:

$$\tilde{V}(c) = \sum_{i=1}^{K} \sum_{l=1}^{N} v_{il} c^{il} \to \min_{c \in \mathbf{C}}, \tag{5.54}$$

где

$$\mathbf{C} = \mathbf{C}_\varepsilon|_{\varepsilon=0},$$

$$\mathbf{C}_\varepsilon \triangleq \Big\{ c = \|c^{il}\| \,\Big|\, c^{il} \geqslant \varepsilon, \quad \sum_{i=1}^{K} \sum_{l=1}^{N} c^{il} = 1,$$

$$\sum_{l=1}^{N} c^{jl} = \sum_{i=1}^{K} \sum_{l=1}^{N} \pi_{ij}^{l} c^{il} \left(j, i = \overline{1, K}, l = \overline{1, N}\right) \Big\}. \tag{5.55}$$

В силу регулярности цепи все компоненты вектора $p(d)$ положительны при любых $d \in D$ [112] и представляют собой непрерывные функции на D (5.52). Таким образом, из (5.53) в силу ограниченности и замкнутости множества D получаем

$$\sum_{l=1}^{N} c^{il} = p_i(d) \geqslant \min_{i=\overline{1,K}} \min_{d \in D} p_i(d) \triangleq c_- > 0 \quad \forall c \in \mathbf{C} \tag{5.56}$$

и

$$d^{il} = c^{il} \Big/ \sum_{k=1}^{N} c^{ik} \quad (i = \overline{1, K}, \, l = \overline{1, N}). \tag{5.57}$$

Соотношения (5.53), (5.57) устанавливают взаимооднозначное соответствие между множествами $\mathbf{C}$ и D, поэтому задачи минимизации функций $V(d)$, $d \in D$ и $\tilde{V}(c)$, $c \in \mathbf{C}$ эквивалентны. Последняя задача (5.54), (5.55), являющаяся задачей линейного программирования, может служить основой для построения алгоритмов адаптивного управления регулярными марковскими цепями.

Рассмотрим рекуррентный алгоритм минимизации функции $\tilde{V}(c)$, $c \in \mathbf{C}$, построенный по методу проекции градиента [31, 91]. При известных v_{il}, π^l_{ij} $(i, j = \overline{1, K}, l = \overline{1, N})$ этот алгоритм имеет вид

$$c_{n+1} = \pi^{\mathbf{C}}_{\varepsilon_{n+1}} \left\{ c_n - \gamma^c_{n+1} \frac{\partial \widetilde{V}(c_n)}{\partial c} \right\},$$

$$c_n = \left\| c_n^{\alpha\beta} \right\|_{\substack{\alpha=\overline{1,K},\\ \beta=\overline{1,N}}}, \quad \frac{\partial \widetilde{V}(c_n)}{\partial c} \triangleq \left\| \frac{\partial V(c_n)}{\partial c^{\alpha\beta}} \right\|_{\substack{\alpha=\overline{1,K},\\ \beta=\overline{1,N}}}, \tag{5.58}$$

$$\alpha = \overline{1, K}, \ \beta = \overline{1, N}, \ n = 1, 2, \ldots,$$

где $\pi^{\mathbf{C}}_{\varepsilon}$ — оператор проектирования на множество $\mathbf{C}_\varepsilon$ (5.55).

В условиях неполной информации о величинах v_{il} и π^l_{ij}, когда наблюдаются лишь их реализации, воспользуемся стохастическим аналогом алгоритма (5.58). Для его построения необходимо иметь в своем распоряжении

1) реализации градиента $\partial \widetilde{V}(c_n)/\partial c$,

2) оценки $\widehat{\pi}^l_{ij}$ величин π^l_{ij}, входящих в оператор проектирования $\pi^{\mathbf{C}}_{\varepsilon}$.

Поскольку в предположении стационарности распределения вероятностей состояний марковской цепи с учетом (5.53)

$$\frac{\partial \widetilde{V}(c_n)}{\partial c^{\alpha\beta}} = v_{\alpha\beta} = \mathbf{M} \left\{ \frac{\xi_n \chi(z_n = z(\alpha),\ x_n = x(\beta))}{d^{\alpha\beta} p_\alpha(d)} \right\} =$$

$$= \mathbf{M} \left\{ \frac{\xi_n \chi(z_n = z(\alpha),\ x_n = x(\beta))}{c^{\alpha\beta}} \right\}, \quad \alpha = \overline{1, K}, \quad \beta = \overline{1, N},$$

то реализацией $A_n^{\alpha\beta}$ величины $\partial \widetilde{V}(c_n)/\partial c^{\alpha\beta}$ в момент времени n является случайная величина

$$A_n^{\alpha\beta} \triangleq \frac{\xi_n \chi(z_n = z(\alpha),\ x_n = x(\beta))}{c_n^{\alpha\beta}}, \tag{5.59}$$

причем при $\varepsilon_n > 0$ $(n = 1, 2, \ldots)$ имеем $c_n^{\alpha\beta} \geqslant \varepsilon_n > 0$.

Оценки величин π^l_{ij} будем строить по рекуррентным формулам оценивания средних значений:

$$(\widehat{\pi}^l_{ij})_{n+1} = (\widehat{\pi}^l_{ij})_n -$$

$$- \frac{\chi(z_n = z(i), x_n = x(l))}{s^{il}_{n+1}} \left[(\widehat{\pi}^l_{ij})_n - \chi(z_{n+1} = z(j)) \right],$$

$$s^{il}_{n+1} = s^{il}_n + \chi(z_n = z(i), \quad x_n = x(l)), \tag{5.60}$$

$$s_1^{il} = 0, \quad i, j = \overline{1, K}, \quad l = \overline{1, N}.$$

Принимая во внимание также и тот факт, что марковская цепь является динамической системой, будем изменять правила d_n выбора вариантов (управлений) (5.5) не на каждом такте $n = 1, 2, \ldots$ поступления информации, а лишь на некоторых фиксированных тактах времени n_k ($k = 1, 2, \ldots, n_1 = 1, n_{k+1} > n_k$). Такое «замораживание» правил выбора управления позволяет накапливать более точную информацию о направлении дальнейшего движения в пространстве матриц c, близком к направлению антиградиента функции $V(c)$ (5.54). Итак, принимая во внимание эту возможность «замораживания» и используя соотношения (5.59), (5.60), запишем окончательно рекуррентный алгоритм адаптивного управления регулярной конечной марковской цепью в следующем виде:

$$c_{k+1} = \pi_{\varepsilon_{k+1}}^{\widehat{C}} \{c_k - \gamma_k A_{n_{k+1}}\}, \tag{5.61}$$

$$k = 1, 2, \ldots, n_{k+1} > n_k, \; n_1 = 1,$$

где элементы матрицы

$$A_{n_{k+1}} \triangleq \frac{1}{n_{k+1} - n_k} \sum_{t=n_k}^{n_{k+1}-1} \frac{\xi_t e(z_t) e^{\mathrm{T}}(x_t)}{e^{\mathrm{T}}(z_t) c_k e(x_t)}$$

вычисляются рекуррентно по формуле

$$A_{n+1}^{\alpha\beta} = A_n^{\alpha\beta} - \frac{1}{n - n_k + 1}\left(A_n^{\alpha\beta} - \frac{\xi_n \chi(z_n = z(\alpha),\; x_n = x(\beta))}{c_k^{\alpha\beta}}\right), \tag{5.62}$$

$$\alpha = \overline{1, K}, \; \beta = \overline{1, N}, \; n = \overline{n_k, n_{k+1} - 1},$$

правила выбора вариантов (управлений) d_n имеют компоненты

$$d_n^{\alpha\beta} = c_k^{\alpha\beta} \Big/ \sum_{s=1}^{N} c_k^{\alpha s}, \quad n = \overline{n_k, n_{k+1} - 1}, \tag{5.63}$$

а $\pi^{\widehat{C}}_{\varepsilon_{k+1}}$ — оператор проектирования на множество

$$\widehat{C}_{\varepsilon_{k+1}} = \Big\{ \|c^{il}\| \,|\, c^{il} \geqslant \varepsilon_{k+1}, \quad \sum_{i=1}^{K}\sum_{l=1}^{N} c^{il} = 1,$$

$$\sum_{l=1}^{N} c^{jl} = \sum_{i=1}^{K}\sum_{l=1}^{N} \left(\widehat{\pi}^{l}_{ij}\right)_{n_{k+1}-1} c^{il} \; (i = \overline{1, K},\ l = \overline{1, N})\Big\}, \qquad (5.64)$$

обладающий свойством

$$\left\| \pi^{\widehat{C}}_{\varepsilon} \{c\} - \widetilde{c} \right\| \leqslant \| c - \widetilde{c} \| \quad \forall c \in R^{K\times N}, \quad \widetilde{c} \in \widehat{C}_{\varepsilon}.$$

Начальные условия $c_1 \in \widehat{C}_{\varepsilon_1}$ и $\left(\widehat{\pi}^{\beta}_{\alpha j}\right)_1 \geqslant 0$, $\sum_{j=1}^{K} \left(\widehat{\pi}^{\beta}_{\alpha j}\right)_1 = 1$ $(\alpha,\ j = \overline{1, K},\ \beta = \overline{1, N})$ произвольны.

Последовательности $\{\gamma_k\}$, $\{\varepsilon_k\}$ и $\{n_k\}$ являются параметрами алгоритма (5.61). Их следует выбирать таким образом, чтобы обеспечить решение поставленной задачи.

Прежде чем перейти к доказательству работоспособности этого алгоритма, установим условия, при которых гарантируется сходимость с вероятностью 1 оценок $\left(\widehat{\pi}^{l}_{ij}\right)_n$ (5.60) к их истинным значениям π^{l}_{ij}.

Лемма 5.6. *Пусть выполняются предположения* П1‴ — П3‴, *варианты (управления) выбираются в соответствии с алгоритмом* (5.61) — (5.64) *и существует такая последовательность положительных чисел* $\{g_k\}$, *что*

$$\sum_{k=1}^{\infty} (k g_k)^{-1} < \infty, \quad \sum_{k=1}^{\infty} \left(\frac{\Delta n_k}{n_k}\right)^2 < \infty,$$

$$\varlimsup_{k\to\infty} \frac{\Delta n_k \sqrt{k g_k}}{\sum_{t=1}^{k} \varepsilon_t \Delta n_t} < \infty, \quad \Delta n_k \triangleq n_{k+1} - n_k \xrightarrow[k\to\infty]{} \infty. \qquad (5.65)$$

Тогда для любой регулярной управляемой марковской цепи и любом распределении вероятностей начального состояния $z_1 \in Z$ *с вероятностью* 1 *для всех* $i = \overline{1, K}$, $l = \overline{1, N}$ *выполняются соотношения*

$$\varliminf_{n\to\infty} s^{il}_{n+1} \Big/ \sum_{t=1}^{t(n)} \varepsilon_t \Delta n_t > 0, \qquad (5.66)$$

и при любом $\rho \in [0, 1)$ и $j = \overline{1, K}$

$$(\hat{\pi}^l_{ij})_n - \pi^l_{ij} \overset{\text{п.н.}}{=\!=} o\left(\left(\sum_{t=1}^{t(n)} \varepsilon_t \Delta n_t\right)^{-\rho/2}\right), \tag{5.67}$$

где (см. (5.60))

$$s^{il}_{n+1} = \sum_{\tau=1}^{n} \chi(z_\tau = z(i),\ x_\tau = x(l)), \tag{5.68}$$

$$t(n) = \{k:\ n_k \leqslant n < n_{k+1},\quad k = 1, 2, \ldots\}. \tag{5.69}$$

Доказательство. Введем частоты

$$S^{il}_{n+1} \triangleq \frac{1}{n} s^{il}_{n+1},\quad i = \overline{1, K},\quad l = \overline{1, N}. \tag{5.70}$$

Найдем условия, при которых они асимптотически эквивалентны величинам

$$\overline{S}^{il}_{n+1} \triangleq \frac{1}{n} \sum_{\tau=1}^{n} \mathsf{M}\{\chi(z_\tau = z(i),\ x_\tau = x(l) | \hat{\mathscr{F}}_{t(\tau)}\},$$

где

$$\hat{\mathscr{F}}_{t(\tau)} \triangleq \sigma\left(z_{n'(\tau)};\ z_s, x_s, \xi_s \mid s = \overline{1, n'(\tau) - 1}\right), \tag{5.71}$$

$$n'(\tau) \triangleq \{n_k:\ n_k \leqslant \tau < n_{k+1},\quad k = 1, 2, \ldots\}, \tag{5.72}$$

т. е. с вероятностью 1

$$\lim_{n\to\infty} \left(S^{il}_n - \overline{S}^{il}_n\right) = 0. \tag{5.73}$$

Справедливы соотношения

$$S^{il}_{n+1} - \overline{S}^{il}_{n+1} = \Delta S^{il}_n + O^*(\Delta n_k / n_k),$$

$$|\Delta S^{il}_{n+1}| = \frac{n}{n+1} |\Delta S^{il}_n| \leqslant |\Delta S^{il}_n| \quad \forall\, (n+1) \neq n'(n+1),$$

где

$$\Delta S^{il}_n \triangleq \frac{1}{n} \sum_{t=1}^{t(n)} \sum_{\tau=n_t}^{n_{t+1}-1} \overline{\chi}^{il}_\tau,$$

$$\overline{\chi}^{il}_\tau \triangleq \chi(z_\tau = z(i), x_\tau = x(l)) -$$
$$- \mathsf{M}\{\chi(z_\tau = z(i), x_\tau = x(l)) \mid \hat{\mathscr{F}}_{t(\tau)}\},$$

т. е. на интервалах времени $\overline{n_k,\ n_{k+1} - 1}$ (где $t(n) = k$, $n = \overline{n_k,\ n_{k+1} - 1}$) последовательность ΔS^{il}_n монотонно не возрастает. Следовательно, для асимптотической эквивалентности S_{n+1}^{il} и $\overline{S}_{n+1}^{il}$ достаточно доказать, что в условиях леммы $\Delta S_{n_k} \underset{k\to\infty}{\longrightarrow} 0$. Имеем

$$\Delta S_{n_k}^{il} = \frac{1}{n_k}\sum_{t=1}^{k}\sum_{\tau=n_t}^{n_{t+1}-1}\overline{\chi}_\tau^{il} = \frac{n_{k-1}}{n_k}\Delta S_{n_{k-1}}^{il} + \frac{1}{n_k}\sum_{\tau=n_k}^{n_{k+1}-1}\overline{\chi}_\tau^{il},$$

откуда

$$\mathsf{M}\left\{\left(\Delta S_{n_k}^{il}\right)^2 \middle| \mathscr{F}_{n_{k-1}}\right\} \overset{\text{п.н.}}{=\!=}$$

$$\overset{\text{п.н.}}{=} \left(1-\frac{\Delta n_k}{n_k}\right)^2\left(\Delta S_{n_{k-1}}^{il}\right)^2 + \mathsf{M}\left\{\left(\frac{1}{n_k}\sum_{\tau=n_k}^{n_{k+1}-1}\overline{\chi}_\tau^{il}\right)^2 \middle| \mathscr{F}_{n_{k-1}}\right\} \leqslant$$

$$\leqslant \left[1 - 2\frac{\Delta n_k}{n_k}\left(1 + O^*\left(\frac{\Delta n_k}{n_k}\right)\right)\right](\Delta S_{n_{k-1}}^{il})^2 + \left(\frac{\Delta n_k}{n_k}\right)^2,$$

где $\mathscr{F}_{n_k} \overset{\Delta}{=} \widehat{\mathscr{F}}_k = \sigma\left(z_{n_k};\ z_s,\ x_s,\ \xi_s \,\middle|\, s = \overline{1,\ n_k - 1}\right)$. Из последнего неравенства в силу леммы П.9 и вытекает, что $\Delta S_{n_k}^{il} \overset{\text{п.н.}}{\underset{k\to\infty}{\longrightarrow}} 0$. Итак, доказано соотношение (5.73).

Пользуясь (5.5) и тем, что условные вероятности d_n^{il} выбора управлений «заморожены» (в силу алгоритма (5.60)—(5.64)) на интервале $\overline{n_k,\ n_{k+1} - 1}$, получим оценку снизу:

$$\overline{S}_{n+1}^{il} = \frac{1}{n}\sum_{\tau=1}^{n}\mathsf{M}\left\{d_\tau^{il}\chi(z_\tau = z(i)) \middle| \widehat{\mathscr{F}}_{t(\tau)}\right\} \geqslant$$

$$\geqslant \frac{1}{n}\sum_{t=1}^{t(n)} d_{n_t}^{il}\sum_{\tau=n_t}^{n_{t+1}-1}\mathsf{M}\left\{\chi(z_\tau = z(i)) \middle| \widehat{\mathscr{F}}_t\right\} \geqslant$$

$$\geqslant \frac{1}{n}\sum_{t=1}^{t(n)}\varepsilon_t \Delta n_t \theta_t^i, \tag{5.74}$$

где

$$\theta_t^i \overset{\Delta}{=} \frac{1}{\Delta n_t}\sum_{\tau=n_t}^{n_{t+1}-1}\mathsf{M}\left\{\chi(z_\tau = z(i)) \middle| \widehat{\mathscr{F}}_t\right\}. \tag{5.75}$$

Поскольку на каждом интервале времени $\overline{n_t,\ n_{t+1} - 1}$ правила выбора вариантов (управлений) d_n не изменяются, то управляемая система представляет собой регулярную однородную марковскую цепь с переходной матрицей $\Pi(d_{n_t})$. Это позволяет воспользоваться теоремой 4 из [106] (с. 153), из которой с учетом (5.56), (5.67) для всех $\tau \in n_t,\ n_{t+1}-1$ следует неравенство

$$\left| \mathsf{M}\{\chi(z_\tau = z(i)) \,|\, \widehat{\mathscr{F}}_t\} - \sum_{s=1}^{N} c_t^{is} \right| \leqslant$$
$$\leqslant o(1) + D_1(d_{n_t}) \exp\{-D_2(d_{n_t})(\tau - n_t)\}, \quad (5.76)$$

причем в силу регулярности управляемой марковской цепи для любых $d \in D$ (5.52)

$$D_1(d) \leqslant \bar{D}_1 < \infty, \quad D_2(d) \geqslant \bar{D}_2 > 0. \quad (5.77)$$

Используя (5.56), (5.76) и (5.77) при оценке членов в (5.75), получим

$$\theta_t^i = \frac{1}{\Delta n_t} \sum_{\tau=n_t}^{n_{t+1}-1} \left[\mathsf{M}\{\chi(z_\tau = z(i)) \,|\, \widehat{\mathscr{F}}_t\} - \sum_{s=1}^{N} c_t^{is} \right] + \sum_{s=1}^{N} c_t^{is} \geqslant$$
$$\geqslant c_- - \frac{\bar{D}_1}{\Delta n_t} \sum_{\tau=n_t}^{n_{t+1}-1} \exp\{-\bar{D}_2(\tau - n_t)\} + o(1) =$$
$$= c_- - \left| O\left(\frac{1}{\Delta n_t}\right) \right| + o(1) \geqslant \frac{1}{2} c_- > 0 \quad (5.78)$$

для всех $t \geqslant T$ (T — некоторое достаточно большое число). Но тогда с учетом (5.78) из (5.74) будем иметь с вероятностью 1

$$\bar{S}_{n+1}^{il} \geqslant \frac{c_-}{2n} \sum_{t=T}^{t(n)} \varepsilon_t \Delta n_t. \quad (5.79)$$

Оценим теперь скорость сближения величин $S_{n_t}^{il}$ и $\bar{S}_{n_t}^{il}$ при $t \to \infty$. Определим $\Delta_{n_t}^{il} \triangleq S_{n_t}^{il} - \bar{S}_{n_t}^{il}$, для которых справедливы следующие рекуррентные соотношения:

$$\Delta_{n_k}^{il} = \frac{k}{n_k} \frac{1}{k} \sum_{t=1}^{k} r_t^{il} = \left(1 - \frac{1}{k}\right) \eta_k \Delta_{n_{k-1}}^{il} + n_k^{-1} r_k^{il}, \quad (5.80)$$

где

$$\eta_k \triangleq \frac{k}{k-1}\frac{n_{k-1}}{n_k}, \tag{5.81}$$

$$r_t^{il} \triangleq \sum_{\tau=n_t}^{n_{t+1}-1} \left[\chi\left(z_\tau = z(i), x_\tau = x(l)\right) - \right.$$
$$\left. - \mathsf{P}\left\{z_\tau = z(i), x_\tau = x(l) \mid \widehat{\mathscr{F}}_{t(\tau)}\right\}\right].$$

Отсюда

$$\mathsf{M}\left\{\left(\Delta_{n_k}^{il}\right)^2 \mid \widehat{\mathscr{F}}_{k-1}\right\} \overset{\text{п.н.}}{=} \left(1-\frac{1}{k}\right)^2 \eta_k^2 \left(\Delta_{n_{k-1}}^{il}\right)^2 +$$
$$+ n_k^{-2}\mathsf{M}\left\{\left(r_k^{il}\right)^2 \mid \widehat{\mathscr{F}}_{k-1}\right\} \leqslant \left(1-\frac{1}{k}\right)^2 \eta_k^2 \left(\Delta_{n_{k-1}}^{il}\right)^2 + \left(n_k^{-1}\Delta n_k\right)^2 \tag{5.82}$$

или

$$\begin{gathered}\mathsf{M}\left\{W_{n_k}^{il} \mid \widehat{\mathscr{F}}_{k-1}\right\} \overset{\text{п.н.}}{\leqslant} \left(1-\frac{1}{k}\right)^2 W_{n_{k-1}}^{il} + \left(k^{-1}\Delta n_k\right)^2, \\ W_{n_k}^{il} \triangleq k^{-1} n_k \left(\Delta_{n_k}^{il}\right)^2.\end{gathered} \tag{5.83}$$

Используя теперь лемму П.13Б, получаем из (5.83), (5.65)

$$W_k^{il} \overset{\text{п.н.}}{=} o\left(\nu_k^{-1}\right),$$

или

$$\Delta_{n_k}^{il} \overset{\text{п.н.}}{=} o\left(n_k^{-1}\sqrt{k g_k}\,\Delta n_k\right). \tag{5.84}$$

Следовательно, с учетом (5.72), (5.79) и (5.84) имеем

$$s_{\tau+1}^{il} = \sum_{n=1}^{\tau} \chi\left(z_n = z(i), x_n = x(l)\right) \geqslant$$
$$\geqslant \sum_{n=1}^{n'(\tau)} \chi\left(z_n = z(i), x_n = x(l)\right) = n'(\tau)\, S_{n'(\tau)}^{il} =$$
$$= n'(\tau)\left(\bar{S}_{n'(\tau)}^{il} + \Delta_{n'(\tau)}^{il}\right) \overset{\text{п.н.}}{\geqslant}$$
$$\overset{\text{п.н.}}{\geqslant} n'(\tau)\left[\frac{c_-}{2n'(\tau)} \sum_{t=1}^{t(\tau)} \varepsilon_t \Delta n_t + o\left(\frac{\sqrt{t(\tau)\, g_{t(\tau)}}}{n'(\tau)}\left(n''(\tau) - n'(\tau)\right)\right)\right],$$

где $n''(\tau) = \{n_{k+1} : n_k \leqslant \tau < n_{k+1},\ k = 1, 2, \ldots\}$, откуда в силу (5.56) и условий (5.65) следует, что для любых $\alpha = \overline{1, K}$, $\beta = \overline{1, N}$ с вероятностью 1

$$s_{\tau+1}^{il} \geqslant \sum_{t=1}^{t(\tau)} \varepsilon_t \Delta n_t \left(\frac{1}{2}c_- + o(1)\right) \xrightarrow[\tau\to\infty]{} \infty. \tag{5.85}$$

Из (5.85) вытекает требуемое неравенство (5.66). Установим теперь (5.67). Поскольку в силу (5.85) $s_n^{il} \xrightarrow[n\to\infty]{\text{п.н.}} \infty$, то $s_n^{il} > 0$, начиная с некоторого номера n_0, а следовательно, оценки $(\hat{\pi}_{ij}^l)_n$ существуют, и, кроме того,

$$(\Delta\pi_{ij}^l)_n \triangleq (\hat{\pi}_{ij}^l)_n - \pi_{ij}^l =$$

$$= \frac{1}{s_n^{il}} \sum_{t=1}^{n} \chi(z_t = z(i), x_t = x(l)) \left[\chi(z_{t+1} = z(j)) - \pi_{ij}^l\right] =$$

$$= \left(1 - \frac{\chi(z_n = z(i), x_n = x(l))}{s_n^{il}}\right)(\Delta\pi_{ij}^l)_{n-1} + $$

$$+ \frac{\chi(z_n = z(i), x_n = x(l))}{s_n^{il}}\left(\chi(z_{n+1} = z(j)) - \pi_{ij}^l\right).$$

Поскольку s_n^{il} и $(\hat{\pi}_{ij}^l)_{n-1}$ измеримы относительно $\hat{\mathscr{F}}_{n-1} \triangleq \sigma(x_t, z_t \mid t = \overline{1, n})$; то с учетом (5.3)

$$\mathsf{M}\left\{(s_n^{il})^\rho (\Delta\pi_{ij}^l)_n^2 \mid \hat{\mathscr{F}}_{n-1}\right\} \overset{\text{п.н.}}{\leqslant}$$

$$\overset{\text{п.н.}}{\leqslant} \left[1 - \frac{(2-\rho)}{s_n^{il}} \chi(z_n = z(i), x_n = x(l))\right]\left[(s_{n-1}^{il})^\rho (\Delta\pi_{ij}^l)_{n-1}^2\right] +$$

$$+ \frac{2\chi(z_n = z(i), x_n = x(l))}{(s_n^{il})^{2-\rho}}.$$

Так как с вероятностью 1

$$\sum_{n=n_0}^{\infty} \frac{\chi(z_n = z(i), x_n = x(l))}{s_n^{il}} = \sum_{k=1}^{\infty} \frac{1}{k} = \infty,$$

$$\sum_{n=n_0}^{\infty} \frac{\chi(z_n = z(i), x_n = x(l))}{(s_n^{il})^{2-\rho}} = \sum_{k=1}^{\infty} \frac{1}{k^{2-\rho}} < \infty \quad \forall \rho < 1,$$

то в силу леммы П.9 отсюда следует, что при $\rho \in [0, 1)$

$$\lim_{n\to\infty} (s_n^{il})^{\rho/2} (\Delta\pi_{ij}^l)_n \overset{\text{п.н.}}{=} 0,$$

а значит, (5.67) установлено. Лемма доказана. ▲

Следствие. *Если в лемме* 5.6 *последовательности* $\{\varepsilon_k\}$, $\{n_k\}$ *таковы, что*

$$\varepsilon_k \sim k^{-\beta}, \quad n_k \sim k^{\varkappa}, \quad \varkappa > \beta, \quad 0 \leqslant \beta < \frac{1}{2}, \tag{5.86}$$

то для всех $i=\overline{1, K}$ *и* $l=\overline{1, N}$ *с вероятностью* 1 *справедливо соотношение*

$$\varliminf_{n\to\infty} \frac{s_n^{il}}{n^{1-\beta/\varkappa}} > 0. \tag{5.87}$$

Доказательство. Из (5.86) следует, что

$$\sum_{t=1}^{k} \varepsilon_t \Delta n_t \sim \int_1^k t^{-\beta} t^{\varkappa-1} dt \sim k^{\varkappa-\beta} \xrightarrow[k\to\infty]{} \infty$$

и условия (5.65) выполняются, например, для $g_k = k^\rho$, $0<\rho\leqslant 1-2\beta$. И так как $t(n)\sim n^{1/\varkappa}$, то

$$\sum_{t=1}^{t(n)} \varepsilon_t \Delta n_t \sim n^{1-\beta/\varkappa},$$

откуда в силу (5.66) вытекает (5.87). Следствие доказано. ▲

Следующая теорема устанавливает достаточные условия работоспособности алгоритма (5.61)—(5.64).

Теорема 5.3. *Пусть выполнены предположения* П1‴—П3‴, *а параметры алгоритма* (5.61)—(5.64) $\{\gamma_k\}$, $\{\varepsilon_k\}$, $\{n_k\}$ *удовлетворяют условиям*

$$0<\varepsilon_k \xrightarrow[k\to\infty]{} 0, \quad \gamma_k>0,\; k n_k^{-1} \xrightarrow[k\to\infty]{} 0, \quad \Delta n_k \triangleq n_{k+1}-n_k \xrightarrow[k\to\infty]{} \infty,$$

$$\lim_{k\to\infty} \gamma_k^{-1}\left[n_k^{-1}\Delta n_k + |\varepsilon_{k+1}-\varepsilon_k| + \Delta n_k \left(\sum_{t=1}^{k} \varepsilon_t \Delta n_t\right)^{-1}\right] = 0,$$

$$\gamma_{k+1}^{-1}\Delta n_{k+1} \geqslant \gamma_k^{-1}\Delta n_k, \quad \sum_{k=1}^{\infty}\left[\frac{\gamma_k \Delta n_k}{\varepsilon_k n_k} + \left(\frac{\Delta n_k}{n_k}\right)^2\right] < \infty$$

и существует последовательность положительных чисел $\{h_k\}$ *такая, что*

$$\sum_{k=1}^{\infty} h_k^2 < \infty, \quad \varlimsup_{k\to\infty} \Delta n_k \left(h_k \sum_{t=1}^{k} \varepsilon_t \Delta n_t\right)^{-1} < \infty.$$

Тогда при любых начальных $c_1 \in \widehat{C}_{\varepsilon_1}$ *и* $z_1 \in Z$ *для произвольной регулярной управляемой марковской цепи, функционирующей в соответствии с алгоритмом* (5.61)—(5.64), *с вероятностью* 1 *достигается равенство в* (5.47), *т. е. предельные текущие средние потери* Ф (5.8) *с ве-*

роятностью 1 *достигают минимально возможного значения* Φ^* (5.48).

Доказательство. Пусть $\tilde{c} \in \mathbf{C}$ — произвольная точка множества $\mathbf{C}$ (5.55), а $\tilde{c}_k$ — ее проекция на множество $\hat{\mathbf{C}}_{\varepsilon_k}$ (5.64). Тогда, используя свойства оператора проектирования $\pi^{\hat{\mathbf{C}}}_{\varepsilon_k}$ и равномерную по n и ω ограниченность множества $\hat{\mathbf{C}}_{\varepsilon_k}$ (5.64), получим

$$\|c_{k+1} - \tilde{c}_{k+1}\|^2 \triangleq$$
$$\triangleq \sum_{\alpha,\beta} (c^{\alpha\beta}_{k+1} - \tilde{c}^{\alpha\beta}_{k+1})^2 \leqslant \sum_{\alpha,\beta} \left(c^{\alpha\beta}_k - \gamma_k A^{\alpha\beta}_{n_{k+1}} - \tilde{c}^{\alpha\beta}_{k+1}\right)^2 =$$
$$= \|c_k - \tilde{c}_{k+1}\|^2 - 2\gamma_k \sum_{\alpha,\beta} (c^{\alpha\beta}_k - \tilde{c}^{\alpha\beta}_{k+1}) A^{\alpha\beta}_{n_{k+1}} + \gamma_k^2 \|A_{n_{k+1}}\|^2 \leqslant$$
$$\leqslant \|c_k - \tilde{c}_k\|^2 - 2\gamma_k \sum_{\alpha,\beta} (c^{\alpha\beta}_k - \tilde{c}^{\alpha\beta}_k) A^{\alpha\beta}_{n_{k+1}} +$$
$$+ \gamma_k^2 \|A_{n_{k+1}}\|^2 + K_1 \|\tilde{c}_k - \tilde{c}_{k+1}\| (1 + \gamma_k \|A_{n_{k+1}}\|),$$
$$n = 1, 2, \ldots, \quad K_1 = \mathrm{const} \in (0, \infty).$$

Суммируя эти неравенства по k с весами $\Delta n_k/(2\gamma_k)$, получаем соотношение

$$\frac{1}{n_{k+1} - 1} \sum_{t=1}^{k} \Delta n_t \sum_{\alpha,\beta} (c^{\alpha\beta}_t - \tilde{c}^{\alpha\beta}_t) A^{\alpha\beta}_{n_{t+1}} \leqslant$$
$$\leqslant \frac{1}{2(n_{k+1} - 1)} \sum_{t=1}^{k} \gamma_t^{-1} \Delta n_t (\|c_t - \tilde{c}_t\|^2 - \|c_{t+1} - \tilde{c}_{t+1}\|^2) +$$
$$+ \frac{1}{2(n_{k+1} - 1)} \sum_{t=1}^{k} \gamma_t^{-1} \Delta n_t \Big[\gamma_t^2 \|A_{n_{t+1}}\|^2 +$$
$$+ K_1 \|\tilde{c}_t - \tilde{c}_{t+1}\| (1 + \gamma_t \|A_{n_{t+1}}\|)\Big]. \quad (5.88)$$

Преобразуем левую часть неравенства (5.88), используя выражение для матрицы $A_{n_{t+1}}$ в (5.61):

$$\frac{1}{n_{k+1} - 1} \sum_{t=1}^{k} \sum_{\alpha,\beta} (c^{\alpha\beta}_t - \tilde{c}^{\alpha\beta}_t) \sum_{s=n_t}^{n_{t+1}-1} \frac{\xi_s \chi(z_s = z(\alpha), x_s = x(\beta))}{e^{\mathrm{T}}(z_s) c_t e(x_s)} =$$
$$= \frac{1}{n_{k+1} - 1} \sum_{\tau=1}^{n_{k+1}-1} \theta_\tau, \quad (5.89)$$

где $t(\tau)$ определено в (5.69), а

$$\theta_\tau \triangleq \sum_{\alpha,\beta} \left(c_{t(\tau)}^{\alpha\beta} - \tilde{c}_{t(\tau)}^{\alpha\beta}\right) \frac{\xi_\tau \chi\left(z_\tau = z(\alpha),\, x_\tau = x(\beta)\right)}{e^{\mathrm{T}}(z_\tau)\, c_{t(\tau)} e(x_\tau)}. \tag{5.90}$$

Заметим, что среднее арифметическое величин θ_τ асимптотически эквивалентно среднему арифметическому их условных математических ожиданий $\mathsf{M}\{\theta_\tau \mid \mathscr{F}_\tau\}$, т.е.

$$\frac{1}{n}\sum_{\tau=1}^{n} \theta_\tau \overset{\text{п.н.}}{=} \frac{1}{n}\sum_{\tau=1}^{n} \mathsf{M}\{\theta_\tau \mid \mathscr{F}_\tau\} + o(1), \tag{5.91}$$

$$\mathscr{F}_\tau \triangleq \sigma\left(z_\tau;\, \xi_s, z_s, x_s \mid s = \overline{1, \tau-1}\right).$$

Это следует из леммы П.13, (5.56) и условий данной теоремы, поскольку в этих условиях сходится ряд

$$\sum_{\tau=1}^{\infty} \tau^{-2} \sum_{\alpha,\beta} \mathsf{M}\left\{\left(c_{t(\tau)}^{\alpha\beta} - \tilde{c}_{t(\tau)}^{\alpha\beta}\right)^2 \frac{\xi_\tau^2 \chi\left(z_\tau = z(\alpha),\, x_\tau = x(\beta)\right)}{\left(c_{t(\tau)}^{\alpha\beta}\right)^2} \,\middle|\, \mathscr{F}_\tau\right\} \leqslant$$

$$\leqslant \text{const} \sum_{\tau=1}^{\infty} \tau^{-2} \left(c_{t(\tau)}^{\alpha\beta} \sum_{l=1}^{N} c_{t(\tau)}^{\alpha l}\right)^{-1} \leqslant C(\omega) \sum_{\tau=1}^{\infty} \tau^{-2} \varepsilon_{t(\tau)}^{-1} \leqslant$$

$$\leqslant C(\omega) \sum_{t=1}^{\infty} \sum_{s=n_t}^{n_{t+1}-1} s^{-2} \varepsilon_t^{-1} \leqslant C(\omega) \sum_{t=1}^{\infty} \varepsilon_t^{-1} \int_{n_t-1}^{n_{t+1}-1} x^2 dx =$$

$$= C(\omega) \sum_{t=1}^{\infty} \varepsilon_t^{-1} \left(\frac{1}{n_t - 1} - \frac{1}{n_{t+1} - 1}\right) < \infty. \tag{5.92}$$

Последний ряд сходится, так как $\sum_{k=1}^{\infty} \gamma_k \Delta n_k \varepsilon_k^{-1} n_k^{-1} < \infty$ и $n_{k+1}\gamma_k \xrightarrow[k\to\infty]{} \infty$. Вычислим $\mathsf{M}\{\theta_\tau \mid \mathscr{F}_\tau\}$, используя П1‴, П2‴:

$$\mathsf{M}\{\theta_\tau \mid \mathscr{F}_\tau\} \overset{\text{п.н.}}{=} \sum_{\alpha,\beta} \left(c_{t(\tau)}^{\alpha\beta} - \tilde{c}_{t(\tau)}^{\alpha\beta}\right) \left(\sum_{s=1}^{N} c_{t(\tau)}^{\alpha s}\right)^{-1} \chi\left(z_\tau = z(\alpha)\right) \times$$

$$\times \mathsf{M}\left\{\xi_\tau \mid \hat{\mathscr{F}}_\tau, x_\tau = x(\beta)\right\} =$$

$$= \sum_{\alpha,\beta} \left(c_{t(\tau)}^{\alpha\beta} - \tilde{c}_{t(\tau)}^{\alpha\beta}\right) \left(\sum_{s=1}^{N} c_{t(\tau)}^{\alpha s}\right)^{-1} \chi\left(z_\tau = z(\alpha)\right) v_{\alpha\beta},$$

откуда получаем

$$\frac{1}{n}\sum_{\tau=1}^{n} \mathsf{M}\{\theta_\tau \mid \mathscr{F}_\tau\} \overset{\text{п.н.}}{=} \frac{1}{n}\sum_{\tau=1}^{n} \sum_{\alpha,\beta} v_{\alpha\beta} \chi\left(z_\tau = z(\alpha)\right) d_{n'(\tau)}^{\alpha\beta} -$$

$$- \sum_{\alpha,\beta} v_{\alpha\beta} \tilde{c}^{\alpha\beta} - r_{1n}, \tag{5.93}$$

где

$$r_{1n} \triangleq \sum_{\alpha,\beta} v_{\alpha\beta}\left[\frac{1}{n}\sum_{\tau=1}^{n}\tilde{c}_{t(\tau)}^{\alpha\beta}\chi(z_\tau = z(\alpha))\left(\sum_{s=1}^{N} c_{t(\tau)}^{\alpha s}\right)^{-1} - \tilde{c}^{\alpha\beta}\right]. \quad (5.94)$$

Таким образом, с учетом (5.89), (5.91), (5.93) и (5.94) соотношение (5.88) с вероятностью 1 эквивалентно следующему:

$$\frac{1}{n_{k+1}-1}\sum_{\tau=1}^{n_{k+1}-1}\sum_{\alpha,\beta} v_{\alpha\beta}\chi(z_\tau = z(\alpha))\, d_{n'(\tau)}^{\alpha\beta} - \sum_{\alpha,\beta} v_{\alpha\beta}\tilde{c}^{\alpha\beta} \leqslant$$

$$\leqslant \tilde{r}_{1n_k} + r_{2n_k} + r_{3n_k} + o(1), \quad (5.95)$$

где

$$\tilde{r}_{1n_k} \triangleq r_{1n}\big|_{n=n_{k+1}-1},$$

$$r_{2n_k} \triangleq \frac{1}{2(n_{k+1}-1)}\sum_{t=1}^{k}\Delta n_t\gamma_t^{-1}\left(\|c_t - \tilde{c}_t\|^2 - \|c_{t+1} - \tilde{c}_{t+1}\|^2\right), \quad (5.96)$$

$$r_{3n_k} \triangleq \frac{1}{2(n_{k+1}-1)}\sum_{t=1}^{k}\Delta n_t\gamma_t^{-1}\Big[\gamma_t^2\|A_{n_{t+1}}\|^2 + $$

$$+ K_1\|\tilde{c}_t - \tilde{c}_{t+1}\|\left(1 + \gamma_t\|A_{n_{t+1}}\|\right)\Big]. \quad (5.97)$$

Покажем, что правая часть неравенства (5.95) стремится к нулю при $k \to \infty$ с вероятностью 1.

а) Докажем, что $\tilde{r}_{1n_k} \xrightarrow[k\to\infty]{\text{п.н.}} 0$. Представим r_{1n} в виде

$$r_{1n} = r'_{1n} + r''_{1n}, \quad (5.98)$$

где

$$r'_{1n} \triangleq \sum_{\alpha,\beta} v_{\alpha\beta}\frac{1}{n}\sum_{\tau=1}^{n}\left(\tilde{c}_{t(\tau)}^{\alpha\beta} - \tilde{c}^{\alpha\beta}\right)\chi(z_\tau = z(\alpha))\left(\sum_{s=1}^{N} c_{t(\tau)}^{\alpha s}\right)^{-1},$$

$$r''_{1n} \triangleq \sum_{\alpha,\beta} v_{\alpha\beta}\tilde{c}^{\alpha\beta}\frac{1}{n}\sum_{\tau=1}^{n}\left[\chi(z_\tau = z(\alpha))\left(\sum_{s=1}^{N} c_{t(\tau)}^{\alpha s}\right)^{-1} - 1\right]. \quad (5.99)$$

Условия леммы 5.6 выполняются при $h_k \triangleq 1/\sqrt{kg_k}$, а следовательно, $(\hat{\pi}_{\alpha j}^{\beta})_n \xrightarrow[n\to\infty]{\text{п.н.}} \pi_{\alpha j}^{\beta}$ (α, $j = \overline{1, K}$, $\beta = \overline{1, N}$). Кроме того, $\varepsilon_k \to 0$ при $k \to \infty$. Поэтому $\tilde{c}_k^{\alpha\beta} \to \tilde{c}^{\alpha\beta}$, откуда с учетом (5.56) и леммы П.7 следует, что $r'_{1n} \xrightarrow[n\to\infty]{\text{п.н.}} 0$. Докажем теперь сходимость $\tilde{r}''_{1n_k} \triangleq r''_{1n}\big|_{n=n_{k+1}-1}$ к нулю. Имеем

$$\frac{1}{n_{k+1}-1}\sum_{\tau=1}^{n_{k+1}-1}\chi(z_\tau = z(\alpha))\left(\sum_{s=1}^{N} c_{t(\tau)}^{\alpha s}\right)^{-1} =$$

$$= \frac{1}{n_{k+1}-1}\sum_{t=1}^{k}\left(\sum_{s=1}^{N} c_t^{\alpha s}\right)^{-1}\sum_{\tau=n_t}^{n_{t+1}-1}\chi(z_\tau = z(\alpha)) \overset{\text{п.н.}}{=}$$

$$\overset{\text{п.н.}}{=} \frac{1}{n_{k+1}-1}\sum_{t=1}^{k}\left(\sum_{s=1}^{N} c_t^{\alpha s}\right)^{-1}\sum_{\tau=n_t}^{n_{t+1}-1}\mathsf{P}\{z_\tau = z(\alpha)\,|\,\mathcal{F}_{n_t}\} + o(1). \tag{5.100}$$

Это следует из леммы П.13, (5.56) и условий теоремы, поскольку сходится ряд

$$\sum_{t=1}^{\infty} n_t^{-2}\mathsf{M}\left\{\left(\sum_{s=1}^{N} c_t^{\alpha s}\right)^{-2}\left[\sum_{\tau=n_t}^{n_{t+1}-1}(\chi(z_t = z(\alpha)) -\right.\right.$$

$$\left.\left. - \mathsf{P}\{z_\tau = z(\alpha)\,|\,\mathcal{F}_{n_t}\})\right]^2\,\Big|\,\mathcal{F}_{n_t}\right\} \overset{\text{п.н.}}{\leqslant} C(\omega)\sum_{t=1}^{\infty} n_t^{-2}(\Delta n_t)^2 \overset{\text{п.н.}}{<} \infty.$$

С учетом (5.100) имеем

$$\tilde{r}''_{1n_k} \overset{\text{п.н.}}{=} \sum_{\alpha,\beta} v_{\alpha\beta}\tilde{c}^{\alpha\beta}\frac{1}{n_{k+1}-1}\sum_{t=1}^{k}\left[\left(\sum_{s=1}^{N} c_t^{\alpha s}\right)^{-1} \times\right.$$

$$\left. \times \sum_{\tau=n_t}^{n_{t+1}-1}\mathsf{P}\{z_\tau = z(\alpha)\,|\,\mathcal{F}_{n_t}\} - \Delta n_t\right] + o(1).$$

Используя (5.76), (5.77), получим

$$\tilde{r}''_{1n_k} \overset{\text{п.н.}}{\leqslant} C(\omega)\frac{1}{n_{k+1}-1}\sum_{t=1}^{k}\sum_{\tau=n_t}^{n_{t+1}-1}\exp\{-\overline{D}_2(\tau - n_t)\} + o(1) \leqslant$$

$$\leqslant C_1(\omega)\,k\,(n_{k+1}-1)^{-1} + o(1) \xrightarrow[k\to\infty]{\text{п.н.}} 0$$

в силу условий данной теоремы. Итак, $\tilde{r}_{1n_k} \xrightarrow[k\to\infty]{\text{п.н.}} 0$.

б) Рассмотрим r_{2n_k}. Имеем

$$\left|\sum_{t=1}^{k}\left(\|c_t-\tilde{c}_t\|^2-\|c_{t+1}-\tilde{c}_{t+1}\|^2\right)\right| = \\ = \left|\|c_1-\tilde{c}_1\|^2-\|c_{k+1}-c_{k+1}\|^2\right| \leqslant \text{const} < \infty,$$

откуда в силу условий данной теоремы и леммы П.6 $\lim\limits_{k\to\infty} r_{2n_k} \overset{\text{п.н.}}{=} 0.$

в) Рассмотрим r_{3n_k}. Используя лемму 5.6, а также алгоритм (5.60), легко установить, что

$$\left|(\hat{\pi}^{\beta}_{\alpha j})_n-(\hat{\pi}^{\beta}_{\alpha j})_{n+1}\right| \leqslant \left(s^{\alpha\beta}_{n+1}\right)^{-1} \overset{\text{п.н.}}{\leqslant} C(\omega)\left(\sum_{t=1}^{t(n)}\varepsilon_t\Delta n_t\right)^{-1},$$

$$C(\omega) \overset{\text{п.н.}}{\in} (0,\infty).$$

Нетрудно также показать, используя (5.64), что

$$\|\tilde{c}_t-\tilde{c}_{t+1}\| \leqslant K_1\left(|\varepsilon_t-\varepsilon_{t+1}|+\|\hat{\pi}_{n_t}-\hat{\pi}_{n_{t+1}}\|\right),$$

$$K_1 \in (0,\infty),$$

где $\hat{\pi}_{n_k} \triangleq \left\|(\hat{\pi}^{\beta}_{\alpha j})_{n_k}\right\|^{\beta=\overline{1,N}}_{\alpha,j=\overline{1,K}}$, откуда вытекает неравенство

$$\|\tilde{c}_k-\tilde{c}_{k+1}\| \overset{\text{п.н.}}{\leqslant} C(\omega)\left[|\varepsilon_k-\varepsilon_{k+1}|+\Delta n_k\left(\sum_{t=1}^{k}\varepsilon_t\Delta n_t\right)^{-1}\right]. \tag{5.101}$$

Воспользовавшись неравенствами $2ab\leqslant a^2+b^2$, $\left(\sum\limits_{i=1}^{m}a_i\right)^2 \leqslant m\sum\limits_{i=1}^{m}a_i^2$ и учитывая (5.101), запишем оценку для r_{3n_k} в виде

$$r_{3n_k} \leqslant \frac{K_2}{(n_{k+1}-1)}\sum_{t=1}^{k}\Delta n_t\gamma_t^{-1}\left(\gamma_t^2\|A_{n_{t+1}}\|^2+\|\tilde{c}_t-\tilde{c}_{t+1}\|\right) \leqslant \\ \leqslant \frac{K_2}{n_{k+1}-1}\sum_{t=1}^{k}\left[\gamma_t\sum_{s=n_t}^{n_{t+1}-1}\frac{\xi_s^2}{[e^{\mathrm{T}}(z_s)c_s e(x_s)]^2}+\right. \\ \left.+C(\omega)\Delta n_t\gamma_t^{-1}\left(|\varepsilon_t-\varepsilon_{t+1}|+\Delta n_t\left(\sum_{s=1}^{t}\varepsilon_s\Delta n_s\right)^{-1}\right)\right] = \\ = \frac{K_2}{n_{k+1}-1}\sum_{\tau=1}^{n_{k+1}-1}\theta'_\tau+\frac{K_2C(\omega)}{n_{k+1}-1}\sum_{t=1}^{k}\theta''_t, \tag{5.102}$$

где

$$\theta'_\tau \triangleq \gamma_{t(\tau)}\xi_\tau^2\left[e^{\mathrm{T}}(z_\tau)\,c_\tau e(x_\tau)\right]^{-2},$$
$$\theta''_t \triangleq \Delta n_t\gamma_t^{-1}\left[\,|\,\varepsilon_t-\varepsilon_{t+1}\,|+\Delta n_t\left(\sum_{s=1}^{t}\varepsilon_s\Delta n_s\right)^{-1}\right].$$

Средние арифметические неотрицательных величин θ'_τ в (5.102) стремятся к нулю с вероятностью 1. Это следует из леммы П.8, (5.56) и условий данной теоремы, поскольку

$$\sum_{\tau=1}^{\infty}\tau^{-1}\mathsf{M}\left\{\theta'_\tau\,|\,\mathscr{F}_{n'(\tau)}\right\}\overset{\text{п.н.}}{\leqslant} C(\omega)\sum_{\tau=1}^{\infty}\tau^{-1}\gamma_{t(\tau)}\varepsilon_{t(\tau)}^{-1}\leqslant$$
$$\leqslant C(\omega)\sum_{k=1}^{\infty}\gamma_k\varepsilon_k^{-1}\sum_{\tau=n_k}^{n_{k+1}-1}\tau^{-1}\leqslant C(\omega)\sum_{k=1}^{\infty}\gamma_k\varepsilon_k^{-1}\,\frac{n_{k+1}-n_k}{n_k}<\infty$$

и, значит, $\sum\limits_{\tau=1}^{\infty}\tau^{-1}\theta'_\tau<\infty$, согласно лемме П.9. Последнее слагаемое в (5.102) стремится к нулю по лемме Тёплица П.7 и условиям данной теоремы. Таким образом, $r_{3n_k}\xrightarrow[k\to\infty]{\text{п. н.}}0.$

Переходя к пределу в (5.95) и учитывая при этом результаты, доказанные в пп. а), б), в), получим

$$\overline{\lim_{k\to\infty}}\,\frac{1}{n_{k+1}-1}\sum_{\tau=1}^{n_{k+1}-1}\sum_{\alpha,\beta}v_{\alpha\beta}\chi\left(z_\tau=z(\alpha)\right)d_{n'(\tau)}^{\alpha\beta}\overset{\text{п.н.}}{\leqslant}\sum_{\alpha,\beta}v_{\alpha\beta}\tilde{c}^{\alpha\beta}. \tag{5.103}$$

Но в силу леммы П.13 текущие средние потери Φ_n асимптотически эквивалентны $\overline{\Phi}_n$, т. е. $\Phi_n\overset{\text{п.н.}}{=}\overline{\Phi}_n+o(1)$, где $\overline{\Phi}_n\triangleq\frac{1}{n}\sum\limits_{\tau=1}^{n}\mathsf{M}\left\{\xi_\tau\,|\,\mathscr{F}_\tau\right\}=\frac{1}{n}\sum\limits_{\tau=1}^{n}\sum\limits_{\alpha,\beta}v_{\alpha\beta}\chi\left(z_\tau=z(\alpha)\right)d_{n'(\tau)}^{\alpha\beta}$, поскольку в соответствии с (5.56) и условиями данной теоремы сходится с вероятностью 1 ряд

$$\sum_{\tau=1}^{\infty}\tau^{-2}\mathsf{M}\left\{\xi_\tau^2\,|\,\mathscr{F}_\tau\right\}\leqslant C(\omega)\sum_{\tau=1}^{\infty}\tau^{-2}\left(c_{t(\tau)}^{\alpha\beta}\sum_{\alpha,\beta}c_{t(\tau)}^{\alpha\beta}\right)^{-1}\overset{\text{п.н.}}{\leqslant}$$
$$\overset{\text{п.н.}}{\leqslant}C(\omega)\sum_{\tau=1}^{\infty}\tau^{-2}\varepsilon_{t(\tau)}^{-1}\leqslant C(\omega)\sum_{k=1}^{\infty}\varepsilon_k^{-1}\,\frac{\Delta n_k}{n_k n_{k+1}}<\infty.$$

И так как $\overline{\Phi}_n$ для каждого n можно представить в виде

$$\overline{\Phi}_n = \frac{n_k}{n}\overline{\Phi}_{n_k} + \frac{1}{n}\sum_{\tau=n_k}^{n}\sum_{\alpha,\beta} v_{\alpha\beta}\chi(z_\tau = z(\alpha))\, d_{n_k}^{\alpha\beta},$$

где $k = t(n)$, а $n_k = n'(n)$, то в силу условий данной теоремы

$$\left|\overline{\Phi}_n - \overline{\Phi}_{n_k}\right| \leqslant \frac{n-n_k}{n}\overline{\Phi}_{n_k} + \text{const}\,\frac{\Delta n_k}{n} \leqslant \text{const}\,\frac{\Delta n_k}{n_k} \to 0$$

при $n \to \infty$ (а следовательно, $k \to \infty$). Отсюда и из (5.103), (5.50) вытекает, что

$$\overline{\lim_{n\to\infty}}\,\Phi_n = \overline{\lim_{k\to\infty}}\,\overline{\Phi}_{n_k} \leqslant \sum_{\alpha,\beta} v_{\alpha\beta}\tilde{c}^{\alpha\beta} = V(\tilde{d}), \tag{5.104}$$

где $\tilde{d}^{\alpha\beta} \triangleq \tilde{c}^{\alpha\beta}\Big/\sum\limits_{s=1}^{N}\tilde{c}^{\alpha s}$, $\alpha = \overline{1, K}$, $\beta = \overline{1, N}$, $\tilde{d} = \left\|\tilde{d}^{\alpha\beta}\right\|_{\substack{\alpha=\overline{1,K}\\ \beta=\overline{1,N}}}$.

Из (5.104) в силу произвольности $\tilde{d} \in D$ вытекает утверждение теоремы. Теорема доказана. ▲

Условия данной теоремы выделяют класс последовательностей $\{\gamma_k\}$, $\{\varepsilon_k\}$, $\{n_k\}$, гарантирующих работоспособность алгоритма (5.61)—(5.64) адаптивного управления регулярной марковской цепью в смысле достижения равенства в (5.47). К этому классу последовательностей относятся, например, следующие:

$$\begin{gathered}\gamma_k = \gamma k^{-\nu}, \quad \varepsilon_k = \varepsilon k^{-\theta}, \quad n_k = [k^{\varkappa}] \qquad (\gamma, \varepsilon > 0),\\ 0 < \theta < \nu < 1 - \theta, \quad 1 < \varkappa.\end{gathered} \tag{5.105}$$

С какой скоростью происходит при этом достижение равенства в (5.47)? На этот вопрос отвечает следующая теорема.

Теорема 5.4. *Пусть в условиях теоремы* 5.3 *последовательности* $\{\gamma_k\}$, $\{\varepsilon_k\}$, $\{n_k\}$ *удовлетворяют соотношениям* (5.105). *Тогда при любых* $c_1 \in \hat{C}_{\varepsilon_1}$ *и* $z_1 \in Z$ *для произвольной регулярной марковской цепи, функционирующей в соответствии с алгоритмом* (5.61)—(5.64), *с вероятностью* 1 *при всяком* $\delta > 0$

$$\begin{aligned}&\Phi_n - \Phi^* \leqslant o\left(n^{\delta-\varphi_1}\right) + O^*\left(n^{-\varphi_2}\right),\\ &\varphi_1 \triangleq \varkappa^{-1}\min\{(\varkappa-\theta)/2, \nu-\theta\},\\ &\varphi_2 \triangleq \varkappa^{-1}\min\{1-\theta-\nu, \theta, \varkappa-1\},\end{aligned}$$

т. е. равенство в (5.47) *достигается со скоростью, либо сколь угодно близкой по порядку к* $n^{-\varphi_1}$ *при* $\varphi_1 \leqslant \varphi_2$, *либо равной по порядку* $n^{-\varphi_2}$ *при* $\varphi_2 < \varphi_1$.

Доказательство. Оценим скорости стремления к нулю каждого из четырех членов, входящих в правую часть неравенства (5.95). При этом учтем, что последний из них равен, как следует из (5.91),

$$o(1) = \frac{1}{n_{k+1}-1} \sum_{\tau=1}^{n_{k+1}-1} (\theta_\tau - \mathsf{M}\{\theta_\tau | \mathscr{F}_\tau\}) \triangleq r_{4n_k},$$

где θ_τ определены в (5.90). Применяя лемму П.13А, с учетом (5.56), получим при любом $\delta > 0$

$$r_{4n_k} \overset{\text{п.н.}}{=} o\left(n_k^{\delta - \frac{\varkappa-\theta}{2\varkappa}}\right),$$

поскольку для $\eta_n \triangleq n^{1-\theta/\varkappa-2\delta}$ с вероятностью 1

$$\sum_{n=1}^{\infty} n^{-2}\eta_n \mathsf{M}\{(\theta_n - \mathsf{M}\{\theta_n | \mathscr{F}_n\})^2 | \mathscr{F}_n\} \leqslant$$

$$\leqslant \sum_{n=1}^{\infty} n^{-2}\eta_n \mathsf{M}\{\theta_n^2 | \mathscr{F}_n\} =$$

$$= \sum_{n=1}^{\infty} n^{-2}\eta_n \sum_{\alpha,\beta} \chi(z_n = z(\alpha)) (c_{t(n)}^{\alpha\beta} - \tilde{c}_{t(n)}^{\alpha\beta})^2 \times$$

$$\times \left(c_{t(n)}^{\alpha\beta} \sum_{s=1}^{N} c_{t(n)}^{\alpha s}\right)^{-1} \mathsf{M}\{\xi_n^2(z(\alpha), x(\beta), \omega)\} \leqslant$$

$$\leqslant C_1(\omega) \sum_{n=1}^{\infty} n^{-2}\eta_n \varepsilon_{t(n)}^{-1} \leqslant C_2(\omega) \sum_{n=1}^{\infty} n^{-1-\delta} < \infty$$

и

$$\lim_{n\to\infty} \left(\frac{\eta_n}{\eta_{n-1}} - 1\right) n = 1 - \theta/\varkappa - 2\delta < 2.$$

Здесь учтено, что из (5.105) следует неравенство

$$\varepsilon_{t(n)} \geqslant \varepsilon(1 + n'(n))^{-\theta/\varkappa} \geqslant \varepsilon(1+n)^{-\theta/\varkappa}.$$

Через C_i и $C_i(\omega)$ условимся обозначать константы, а также конечные с вероятностью 1 случайные величины.

Оценим теперь $\tilde{r}_{1n_k}$. Используя (5.98), (5.99), свойства оператора проектирования на множество $\hat{C}_{\varepsilon_k}$ (5.64) и оценок $\widehat{\pi}_n$ (5.67), получаем

$$|r'_{1n}| \overset{\text{п.н.}}{\leqslant}$$

$$\overset{\text{п.н.}}{\leqslant} C_3 n^{-1} \sum_{\tau=1}^{n} \|\widetilde{c}_{t(\tau)} - \widetilde{c}\| \leqslant C_4 n^{-1} \sum_{k=1}^{t(n)} \Delta n_k \left(\varepsilon_k + \|\widehat{\pi}_{n_k} - \pi\|\right) \leqslant$$

$$\leqslant C_5 n^{-\theta/\varkappa} + o\left(n^{-\frac{\varkappa-\theta}{2\varkappa}+\delta}\right) \quad \forall \delta > 0.$$

Для оценки $\widetilde{r}''_{1n_k}$ можно снова применить лемму П.13А. Учитывая выкладки, проделанные в п. а) доказательства теоремы 5.3, будем иметь

$$\left|\widetilde{r}''_{1n_k}\right| \overset{\text{п.н.}}{\leqslant} C_6 k^{1-\varkappa} + o\left(k^{-1/2+\delta}\right),$$

поскольку для $\eta_k \triangleq k^{3-2\varkappa-2\delta}$, $\delta > 0$ с вероятностью 1 сходится ряд

$$\sum_{t=1}^{\infty} t^{-2} \eta_t \mathsf{M}\left\{\left(\sum_{s=1}^{N} c_t^{\alpha s}\right)^{-2} \left[\sum_{\tau=n_t}^{n_{t+1}-1} \left(\chi(z_\tau = z(\alpha)) - \right.\right.\right.$$

$$\left.\left.\left. - \mathsf{P}\{z_\tau = z(\alpha) \,|\, \mathscr{F}_{n_t}\}\right)\right]^2 \,\middle|\, \mathscr{F}_{n_t}\right\} \overset{\text{п.н.}}{\leqslant} C_7(\omega) \sum_{t=1}^{\infty} t^{-(1+2\delta)} < \infty$$

и

$$\lim_{k\to\infty} \left(\frac{\eta_k}{\eta_{k-1}} - 1\right) k = 3 - 2\varkappa - 2\delta < 2.$$

Но из (5.105) следует, что $k \geqslant n_k^{1/\varkappa}$ и $\varkappa > 1$, поэтому

$$\left|\widetilde{r}''_{1n_k}\right| \overset{\text{п.н.}}{\leqslant} C_6 n_k^{(1-\varkappa)/\varkappa} + o\left(n_k^{-(1/2-\delta)/\varkappa}\right).$$

Таким образом, из (5.98) при любом $\delta > 0$ следует неравенство

$$\left|\widetilde{r}_{1n_k}\right| \leqslant o\left(n_k^{-\frac{\varkappa-\theta}{2\varkappa}+\delta}\right) + C_5 n_k^{-\theta/\varkappa} + C_6 n_k^{\frac{1-\varkappa}{\varkappa}}.$$

Оценка скорости убывания r_{2n_k} (5.96) производится с помощью леммы П.6, из которой с учетом (5.105) и пункта б) доказательства теоремы 5.3 получаем

$$|r_{2n_k}| \overset{\text{п.н.}}{\leqslant} \frac{C_7}{n_{k+1} - 1} \frac{\Delta n_k}{\gamma_k} \leqslant C_8 n_k^{-\frac{1-\nu}{\varkappa}}.$$

Наконец, для r_{3n_k} (5.97) из (5.102) и леммы П.13Б вытекает неравенство

$$r_{3n_k} \overset{\text{п.н.}}{\leqslant} C_9 n_k^{\frac{\theta+\nu-1}{2}} + o\left(n_k^{\delta-\frac{\nu-\theta}{\varkappa}}\right) \quad \forall\delta>0,$$

так как с учетом (5.56) сходится ряд

$$\sum_{n=1}^{\infty} n^{-1} n^{\frac{\nu-\theta}{\varkappa}} \mathsf{M}\left\{\gamma_{t(n)}\xi_n^2\left[e^{\mathrm{T}}(z_n)\, c_n e(x_n)\right]^{-2} \mid \mathscr{F}_n\right\} \overset{\text{п.н.}}{\leqslant}$$

$$\overset{\text{п.н.}}{\leqslant} C_{10}(\omega) \sum_{n=1}^{\infty} n^{-1-\delta} < \infty$$

и

$$\lim_{n\to\infty}\left[\frac{(n+1)^{\nu/\varkappa-\delta}}{n^{\nu/\varkappa-\delta}} - 1\right] n = \frac{\nu}{\varkappa} - \delta < 1.$$

Итак, используя полученные оценки, находим в условиях (5.105) для любых $\tilde{c} \in \mathbf{C}$ и $\delta > 0$

$$\frac{1}{n_{k+1}-1} \sum_{\tau=1}^{n_{k+1}-1} \sum_{\alpha,\beta} v_{\alpha\beta}\chi(z_\tau = z(\alpha))\, d_{n'(\tau)}^{\alpha\beta} - \sum_{\alpha,\beta} v_{\alpha\beta}\tilde{c}^{\alpha\beta} \overset{\text{п.н.}}{\leqslant}$$

$$\overset{\text{п.н.}}{\leqslant} o\left(n_k^{\delta-\frac{\varkappa-\theta}{2\varkappa}}\right) + C_6 n_k^{\frac{1-\varkappa}{\varkappa}} + C_9 n_k^{(\theta+\nu-1)/\varkappa} + o\left(n_k^{\delta-\frac{\nu-\theta}{2}}\right).$$

Поскольку задача минимизации функций $\tilde{V}(c)$ (5.54) и $V(d)$ (5.50) на множествах $\mathbf{C}$ (5.55) и D (5.52) соответственно эквивалентны, получаем

$$\sum_{\alpha,\beta} v_{\alpha\beta}\tilde{c}^{\alpha\beta} = \tilde{V}(\tilde{c}) \geqslant \min_{d\in D} V(d) = \Phi^*.$$

Кроме того, из леммы П.13А (см. также доказательство теоремы 5.3 после (5.103)) вытекает оценка

$$\overline{\Phi}_n \triangleq \frac{1}{n} \sum_{\tau=1}^{n} \sum_{\alpha,\beta} v_{\alpha\beta}\chi(z_\tau = z(\alpha))\, d_{n'(\tau)}^{\alpha\beta} =$$

$$= \Phi_n + o\left(n^{-1/2+\delta}\right) \quad \forall\delta>0.$$

Но, как показано при доказательстве теоремы 5.3,

$$\left|\overline{\Phi}_n - \Phi_n\right| \leqslant \text{const}\,\frac{\Delta n_k}{n_k}.$$

А поскольку в условиях (5.105) $t(n) = O^*(n^{1/\varkappa})$ и, следовательно,

$$\left|\overline{\Phi}_n - \Phi_{n_k}\right| \leqslant O^*\left(n^{-1/\varkappa}\right), \quad k = t(n),$$

то окончательно приходим к следующей оценке:

$$\Phi_n - \Phi^* \leqslant C_9 n^{-\frac{1-\theta-\nu}{\varkappa}} + o\left(n^{\delta - \frac{\varkappa-\theta}{2\varkappa}}\right) + C_6 n^{-\frac{\varkappa-1}{\varkappa}} + \\ + C_5 n^{-\frac{\theta}{\varkappa}} + o\left(n^{\delta-\frac{\nu-\theta}{\varkappa}}\right) + o\left(n^{\delta-1/2}\right)$$

при любом $\delta > 0$, что и требовалось доказать. ▲

Следствие. *Наибольший порядок скорости достижения равенства в* (5.47), *гарантируемый теоремой* 5.4, *равен* $o(n^{\delta-1/5})$, *т. е.*

$$\Phi_n - \Phi^* \overset{\text{п.н.}}{\leqslant} o\left(n^{\delta-1/5}\right) \quad \forall \delta > 0$$

и достигается при

$$\nu = \nu^* \triangleq \frac{1}{2}, \quad \theta = \theta^* \triangleq \frac{1}{4}, \quad \varkappa = \varkappa^* = \frac{5}{4}.$$

Таким образом, параметры ν^*, θ^* и $\varkappa^*$ являются оптимальными в смысле верхней оценки, даваемой теоремой 5.4, и обеспечивают скорость достижения равенства в (5.47) порядка $o(n^{\delta-1/5})$, где δ — сколь угодно малое число.

Замечание. В условиях полной априорной информации о характеристиках π_{ij}^l и v^{il} управляемой марковской цепи оптимальная стационарная стратегия $\{d^*\}$ может быть вычислена заранее (до начала управления), например, по формуле (5.57) посредством решения задачи линейного программирования (5.54). При этом, реализуя стратегию $\{d^*\}$, получим, что средние потери Φ_n достигают с вероятностью 1 при $n \to \infty$ своего предельного значения Φ^* со скоростью порядка $o\left(n^{\delta-1/2}\right)$ $\forall \delta > 0$. Действительно, в силу леммы П.13А нетрудно показать, что

$$\Phi_n \overset{\text{п.н.}}{=} \sum_{i=1}^{K} \sum_{l=1}^{N} v_{il} \frac{1}{n} \sum_{t=1}^{n} \chi\left(z_t = z(i), x_t = x(l)\right) + o\left(n^{\delta-1/2}\right) \overset{\text{п.н.}}{=} \\ \overset{\text{п.н.}}{=} \sum_{i=1}^{K} \sum_{l=1}^{N} v_{il} (d^*)^{il} S_n(i) + o\left(n^{\delta-1/2}\right),$$

где

$$S_n(i) \triangleq \frac{1}{n} \sum_{t=1}^{n} \chi(z_t = z(i)) \overset{\text{п.н.}}{=}$$

$$\overset{\text{п.н.}}{=} \frac{1}{n} \sum_{t=1}^{n} \sum_{j=1}^{K} \chi(z_{t-1} = z(j)) \sum_{l=1}^{N} \pi_{ji}^{l}(d^*)^{il} + o\left(n^{\delta-1/2}\right).$$

Отсюда для векторов $S_n \triangleq (S_n(1), \ldots, S_n(K))^{\mathrm{T}}$ получим рекуррентное соотношение

$$S_n = \Pi^{\mathrm{T}}(d^*) S_{n-1} + o(n^{\delta-1/2}) e^K.$$

Поскольку стохастическая матрица $\Pi(d^*)$ соответствует регулярной марковской цепи, система уравнений

$$S = \Pi^{\mathrm{T}}(d^*) S, \quad (e^K)^{\mathrm{T}} S = 1$$

имеет единственное решение $p(d^*)$, причем

$$[\Pi^{\mathrm{T}}(d^*)]^n S_1 = p(d^*) + O(\lambda^n) e^K,$$

где $\lambda \in (0, 1)$ [106, с. 153]. Используя это, будем иметь

$$S_n = p(d^*) + \left[\Pi^{\mathrm{T}}(d^*)\right]^{n-1} S_1 - p(d^*) +$$

$$+ \sum_{t=1}^{n-1} \left[\Pi^{\mathrm{T}}(d^*)\right]^{n-1-t} e^K o\left(t^{\delta-1/2}\right) =$$

$$= p(d^*) + \left[O\left(\lambda^n\right) + O^*\left(\sum_{t=1}^{n} \lambda^{n-t} t^{\delta-1/2}\right)\right] e^K =$$

$$= p(d^*) + o\left(n^{\delta-1/2}\right) e^K.$$

Учитывая (5.49) — (5.52), получаем

$$\Phi_n \overset{\text{п.н.}}{=} \Phi^* + o\left(n^{\delta-1/2}\right).$$

Таким образом, отсутствие полной априорной информации в задаче адаптивного управления регулярной марковской цепью понижает порядок скорости сходимости текущих средних потерь Φ_n к оптимальному значению Φ^* с $o(n^{\delta-1/2})$ до, по крайней мере, $o(n^{\delta-1/5})$.

§ 5.5. Алгоритм адаптивного управления связной марковской цепью

Рассмотрим теперь вопрос о работоспособности алгоритма (5.61) — (5.64) в более общей ситуации, а именно когда управляемая марковская цепь является связной (см. § 5.1). Подробное рассмотрение класса связных цепей имеет особое значение, поскольку именно этот случай является ключевым при исследовании алгоритмов адаптивного управления произвольными марковскими цепями из класса $M(K, N)$.

Отличие свойств управляемых связных цепей от свойств регулярных сказывается на двух основных моментах доказательства работоспособности алгоритма (5.61) — (5.64):

1) в случае связных цепей не выполняется соотношение (5.56), т. е. $c_- = 0$; при этом, однако, отделимость $\sum\limits_{s=1}^{N} c_k^{\alpha s}$ $(\alpha = \overline{1, K})$ от нуля осуществляется чисто алгоритмически за счет присутствия в (5.61) оператора проектирования $\pi_{\varepsilon_k}^{\widehat{C}}$ на ε_k-симплекс (5.64), что при $\varepsilon_k > 0$ приводит к неравенству

$$\sum_{s=1}^{N} c_k^{\alpha s} \geqslant N\varepsilon_k > 0, \quad \alpha = \overline{1, K}; \tag{5.106}$$

2) при «замораживании» управлений $\left(d_n = d_{n_k}\right)$ на интервалах $n_k, \ldots, n_{k+1} - 1$, когда марковская цепь становится однородной, в случае связной цепи не выполняется соотношение (5.76), (5.77), гарантирующее экспоненциальное во времени приближение вектора распределения вероятностей состояний к финальному вектору, имеющему компоненты $\sum\limits_{s=1}^{N} c_k^{is}$ $\left(i = \overline{1, K}\right)$.

Поскольку в случае связной цепи при любой стационарной невырожденной стратегии управления $\{d\} \in \Sigma_S^+$ соответствующая переходная матрица $\Pi(d)$ цепи является неразложимой (см. § 5.1), то существует единственное стационарное распределение $p_i\{d\} = \sum\limits_{s=1}^{N} c^{is}$ $(i = \overline{1, K})$ вероятностей состояний этой цепи (5.51), (5.56). Оказывается, что и в этом случае имеет место аналог оценки (5.76), учитывающий возможность существования в цепи нескольких циклических подклассов. Это сформулировано в следующей лемме.

Лемма 5.7. *Пусть управляемая цепь из класса $M(K, N)$ является связной и применяется стационарная невырожденная стратегия управления $\{d\} \in \Sigma_S^+$, такая, что*

$$d^{il} \geqslant \varepsilon > 0, \quad i = \overline{1, K}, \quad l = \overline{1, N}. \tag{5.107}$$

Тогда при достаточно малом ε для любых $i = \overline{1, K}$ и $z \in Z$ справедлива оценка

$$\left| \frac{1}{r} \sum_{l=0}^{r-1} p_i(d, z, n+l) - \sum_{s=1}^{N} c^{is} \right| \leqslant K^{5/2} \left[1 - C_1 \varepsilon^{r(K-r)}\right]^{\frac{n-r+1}{r(K-r)} - 1} \tag{5.108}$$

где $C_1 > 0$, r — число циклических подклассов однородной марковской цепи с матрицей переходов $\Pi(d)$, $p_i(d, z, n)$ — вероятность перехода этой цепи из состояния $z \in Z$ в состояние $z(i) \in Z$ за n тактов $(i = \overline{1, K})$, а $c \in \mathrm{C}$ и $d \in D$ связаны соотношением (5.57).

Доказательство. Заметим, что матрица с элементами $p_i(d, z(j), n)$, где $i = \overline{1, K}$ и $j = \overline{1, K}$ — номера строки и столбца соответственно, равна n-й степени матрицы переходов $\Pi(d)$, т. е. $[\Pi(d)]^n = \| p_i(d, z(j), n) \|_{i,j=\overline{1,K}}$. Обозначим через $A(d)$ стохастическую матрицу, удовлетворяющую уравнению

$$A(d)\Pi(d) = A(d). \tag{5.109}$$

Поскольку в условиях данной леммы уравнения (5.51) имеют единственное решение, представляющее собой стохастический вектор-столбец $p(d)$ с компонентами $\sum_{s=1}^{N} c^{is}$ (5.56), то в силу (5.109) матрица $A(d)$ будет иметь одинаковые строки, равные $p^{\mathrm{T}}(d)$. Таким образом, для обоснования (5.108) достаточно доказать при больших n оценку

$$\left\| \frac{1}{r} \sum_{l=0}^{r-1} [\Pi(d)]^{n+l} - A(d) \right\| \leqslant K^{5/2} \left[1 - C_1 \varepsilon^{r(K-r)}\right]^{\frac{n-r+1}{r(K-r)} - 1}, \tag{5.110}$$

где $\|\cdot\|$ — евклидова норма матрицы.

Из условий данной леммы, как уже отмечалось выше, следует, что матрица $\Pi(d)$ соответствует однородной эргодической марковской цепи, т. е. цепи, все множество состояний Z которой составляет один эргодический класс, содержащий r циклических подклассов. Поэтому (после соответствующей перенумерации состояний) переходная матрица $\Pi(d)$ может быть представлена в виде

$$\Pi(d) = \begin{Vmatrix} 0 & \Pi_{12}(d) & 0 & \dots & 0 & \dots & 0 \\ 0 & 0 & \Pi_{23}(d) & \dots & 0 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \Pi_{r1}(d) & 0 & 0 & \dots & 0 & \dots & 0 \end{Vmatrix},$$

где $\Pi_{12}(d)$, ..., $\Pi_{r1}(d)$ — стохастические матрицы. При этом r-я степень матрицы $\Pi(d)$ имеет блочно-диагональную структуру

$$[\Pi(d)]^r = \begin{Vmatrix} \tilde{\Pi}_1(d) & 0 & \dots & 0 \\ 0 & \tilde{\Pi}_2(d) & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & \tilde{\Pi}_r(d) \end{Vmatrix}, \tag{5.111}$$

где $\tilde{\Pi}_l(d)$, $l=\overline{1, r}$ — стохастические квадратные матрицы, определяющие регулярные марковские подцепи (множествами состояний которых являются соответствующие циклические подклассы всей цепи). Отсюда следует, что существует предел

$$\lim_{k\to\infty} [\Pi(d)]^{kr} = A_0(d), \tag{5.112}$$

а значит,

$$\lim_{k\to\infty} \frac{1}{r}\sum_{l=0}^{r-1} [\Pi(d)]^{kr+l} = A_0(d)\frac{1}{r}\sum_{l=0}^{r-1}[\Pi(d)]^l = A(d).$$

Последнее равенство доказано в [41]. Обозначим целую часть $[n/r] \triangleq k_n$ и остаток $l_n = n - k_n r$. Учитывая это и (5.109), получим

$$\frac{1}{r}\sum_{l=0}^{r-1}[\Pi(d)]^{n+l} - A(d) =$$

$$= \left([\Pi(d)]^{k_n r} - A_0(d)\right)\left(\frac{1}{r}\sum_{l=0}^{r-1}[\Pi(d)]^l\right)\Pi^{l_n}(d),$$

откуда, используя свойства евклидовой нормы матриц (в частности, то, что норма стохастической $K \times K$-матрицы не превосходит $\sqrt{K}$), приходим к неравенству

$$\left\| \frac{1}{r} \sum_{l=0}^{r-1} [\Pi(d)]^{n+l} - A(d) \right\| \leqslant \sqrt{K} \left\| [\Pi(d)]^{k_n r} - A_0(d) \right\|.$$

Из (5.111), (5.112) следует, что матрица $A_0(d)$ имеет такую же блочно-диагональную структуру, что и матрица $[\Pi(d)]^r$, а l-й ее блок равен пределу $\lim\limits_{k \to \infty} [\tilde{\Pi}_l(d)]^k$. Применяя теперь теорему 4 из [106, с. 153] для оценки скорости сходимости степеней матриц $\tilde{\Pi}_l(d)$, получаем неравенство

$$\left\| \frac{1}{r} \sum_{l=0}^{r-1} [\Pi(d)]^{n+l} - A(d) \right\| \leqslant K^{5/2} \max_{l=\overline{1,r}} (1 - \rho_l(d))^{\frac{k_n}{\varkappa_l} - 1}, \tag{5.113}$$

где $\varkappa_l$ — наименьшее целое положительное число k, при котором матрица $[\tilde{\Pi}_l(d)]^k$ не имеет нулевых элементов, а $\rho_l(d) \in (0, 1)$ — коэффициент эргодичности l-й регулярной марковской подцепи (с переходной матрицей $\tilde{\Pi}_l(d)$).

Для $\varkappa_l$ имеем очевидную оценку сверху:

$$\varkappa_l \leqslant K - r, \tag{5.114}$$

так как $\varkappa_l$ меньше, чем число состояний l-й подцепи. Кроме того,

$$k_n = \frac{n - l_n}{r} \geqslant \frac{n - (r-1)}{r}. \tag{5.115}$$

Осталось оценить снизу $\rho_l(d)$. Из [36] имеем

$$\rho_l(d) \geqslant \max_j \min_i \left([\tilde{\Pi}_l(d)]^{\varkappa_l} \right)_{ij}, \tag{5.116}$$

где минимум и максимум берутся по тем значениям i и j, для которых $z(i)$ и $z(j)$ являются элементами множества состояний рассматриваемой марковской подцепи. С учетом этого и (5.6) имеем

$$\left([\tilde{\Pi}_l(d)]^{\varkappa_l} \right)_{ij} = \left([\Pi(d)]^{r\varkappa_l} \right)_{ij} \geqslant (N\varepsilon)^{r\varkappa_l} \left([\Pi(\bar{d})]^{r\varkappa_l} \right)_{ij}, \tag{5.117}$$

где $\bar{d} = \|\bar{d}^{ij}\|_{\substack{i=\overline{1,K}\\ j=\overline{1,N}}}$, $\bar{d}^{ij} \triangleq 1/N$ (т. е. $\bar{d}$ определяет правило выбора вариантов $x(1), \ldots, x(N)$ с одинаковыми вероятностями $1/N$ независимо от того, в каком состоянии находилась цепь). В последнем неравенстве учтено, что для элементов матрицы $\Pi(d)$ выполняется соотношение

$$\pi_{ij}(d) = \sum_{l=1}^{N} \pi_{ij}^{l} d^{il} \geqslant \varepsilon \sum_{l=1}^{N} \pi_{ij}^{l}.$$

В силу связности управляемой марковской цепи получаем, что для указанных i и j

$$\left([\Pi(\bar{d})]^{r\varkappa_l}\right)_{ij} > 0.$$

Таким образом, из (5.113)—(5.117) следует (5.110). Лемма доказана. ▲

Итак, экспоненциальная сходимость к стационарному распределению при отсутствии регулярности имеет место не для векторов $p(d, z_1, n)$, как в регулярном случае, а для их среднего арифметического за r тактов.

Таким образом, при отказе от условия регулярности управляемой марковской цепи и замене его условием связности следует учитывать две отмеченные выше особенности, что приводит к необходимости использования соотношений (5.106), (5.108) вместо (5.56), (5.76) соответственно. Это, естественно, сказывается на большей жесткости требований, предъявляемых к параметрам алгоритма управления (5.61)—(5.64).

Перейдем теперь к обоснованию работоспособности этого алгоритма в случае связной управляемой марковской цепи. Обобщим сначала лемму 5.6.

Лемма 5.8. *Пусть в формулировке леммы* 5.6 *вместо условий* (5.65) *выполняются следующие условия*:

$$\sum_{k=1}^{\infty}\left[(kg_k)^{-1} + \left(\frac{\Delta n_k}{n_k}\right)^2\right] < \infty, \quad \overline{\lim_{k\to\infty}} \frac{\Delta n_k \sqrt{kg_k}}{\sum\limits_{t=1}^{k} \varepsilon_t^2 \Delta n_t} < \infty, \tag{5.118}$$

$$n_k \varepsilon_k^{K^2/4} \xrightarrow[k\to\infty]{} \infty, \quad \varepsilon_k \Delta n_k \xrightarrow[k\to\infty]{} \infty,$$

$$n_k^{-1} \varepsilon_k^{-K^2/4} |\ln \varepsilon_k| \xrightarrow[k\to\infty]{} 0.$$

Тогда для любой связной управляемой марковской цепи из класса $M(K, N)$ и любом распределении вероятностей начального состояния $z_1 \in Z$ с вероятностью 1 для всех $i = \overline{1, K}$, $l = \overline{1, N}$ выполняются соотношения (5.67) *и*

$$\varliminf_{n\to\infty} s_{n+1}^{il} \Big| \sum_{t=1}^{t(n)} \varepsilon_t^2 \Delta n_t > 0 \tag{5.119}$$

и при любом $\rho \in [0, 1)$ *и* $j = \overline{1, K}$

$$(\hat{\pi}_{ij}^l)_n - \pi_{ij}^l \overset{\text{п.н.}}{=} o\left(\left(\sum_{t=1}^{t(n)} \varepsilon_t^2 \Delta n_t\right)^{-\rho/2}\right),$$

где s_{n+1}^{il}, $t(n)$ *определены в* (5.68), (5.69).

Доказательство. До соотношения (5.75) включительно доказательство леммы 5.6 здесь полностью повторяется. Оценим снизу величины θ_t^i (5.75), используя оценку (5.106) и лемму 5.7:

$$\theta_t^i = \sum_{s=1}^N c_t^{is} +$$

$$+ \frac{1}{\Delta n_t} \sum_{\tau=n_t}^{n_{t+1}-1} \left[\mathsf{M}\{\chi(z_\tau = z(i)) \,|\, \mathscr{F}_t\} - \sum_{s=1}^N c_t^{is} \right] \geqslant N\varepsilon_t -$$

$$- \frac{r}{\Delta n_t} \sum_{k=0}^{[\Delta n_t/r]} \left| \frac{1}{r} \sum_{l=0}^{r-1} \left[p_i(d_{n_t}, z_{n_t}, n_t + rk + l) - \sum_{s=1}^N c_t^{is} \right] \right| +$$

$$+ O\left(\frac{1}{\Delta n_t}\right) \geqslant N\varepsilon_t - \frac{r}{\Delta n_t} \sum_{k=0}^{\infty} \left[1 - C\varepsilon_t^{r(K-r)}\right]^{\frac{n_t + rk + 1 - r}{r(K-r)} - 1} + O\left(\frac{1}{\Delta n_t}\right) =$$

$$= \varepsilon_t \left[N + O\left((\Delta n_t)^{-1} \varepsilon_t^{-1-r(K-r)} e^{-C n_t \varepsilon_t^{r(K-r)}} \right) + \right.$$

$$\left. + O\left((\Delta n_t \varepsilon_t)^{-1} \right) \right] \geqslant \frac{N}{2} \varepsilon_t,$$

где $t \geqslant T$, а C и T — достаточно велики. Таким образом, с вероятностью 1

$$\bar{S}_{n+1}^{il} \geqslant \frac{N}{2n} \sum_{t=T}^{t(n)} \varepsilon_t^2 \Delta n_t.$$

Далее доказательство леммы 5.6 полностью повторяется после (5.79) с использованием последнего неравенства. Лемма доказана. ▲

Следствие. *Если последовательности* $\{\varepsilon_k\}$ *и* $\{n_k\}$ *таковы, что*

$$\varepsilon_k \sim k^{-\theta}, \quad n_k \sim k^{\varkappa}, \quad \varkappa > \max\left\{1+\theta,\ \theta\frac{K^2}{4}\right\}, \quad 0 \leqslant \theta < \frac{1}{4}, \tag{5.120}$$

то для всех $i=\overline{1, K}$ *и* $l=\overline{1, N}$ *с вероятностью* 1 *справедливо соотношение*

$$\varliminf_{n\to\infty} \frac{s_n^{il}}{n^{1-2\theta/\varkappa}} > 0. \tag{5.121}$$

Доказательство этого результата проводится аналогично доказательству следствия из леммы 5.6.

Теорема 5.5. *Пусть выполнены предположения* П1‴ — П3‴, *а параметры алгоритма* (5.61) — (5.64) $\{\gamma_k\}$, $\{\varepsilon_k\}$, $\{n_k\}$ *удовлетворяют условиям*

$$0 < \varepsilon_k \xrightarrow[k\to\infty]{} 0, \quad \gamma_k > 0, \quad \varepsilon_k^{1+K^2/4}\Delta n_k \xrightarrow[k\to\infty]{} \infty,$$

$$\lim_{k\to\infty} \gamma_k^{-1}\left[n_k^{-1}\Delta n_k + \Delta n_k\left(\sum_{t=1}^{k} \varepsilon_t^2 \Delta n_t\right)^{-1}\right] = 0,$$

$$\gamma_{k+1}^{-1}\Delta n_{k+1} \geqslant \gamma_k^{-1}\Delta n_k, \quad \sum_{k=1}^{\infty}\left[\gamma_k\varepsilon_k^{-2}(\Delta n_k)\, n_k^{-1} + \left(\frac{\Delta n_k}{n_k}\right)^2\right] < \infty$$

и существуют последовательность положительных чисел $\{h_k\}$ *и* $\delta > 0$, *такие, что*

$$\sum_{k=1}^{\infty} h_k^2 < \infty, \quad \varlimsup_{k\to\infty}\left[h_k\varepsilon_k^{-\delta} + \Delta n_k\left(h_k\sum_{t=1}^{k}\varepsilon_t^2\Delta n_t\right)^{-1}\right] < \infty.$$

Тогда при любых начальных $c_1 \in \hat{C}_{\varepsilon_1}$ *и* $z_1 \in Z$ *для произвольной управляемой связной марковской цепи из класса* $M(K, N)$, *функционирующей в соответствии с алгоритмом* (5.61) — (5.64), *с вероятностью* 1 *выполняется целевое условие* (5.47), *т. е. предельные средние потери* Φ (5.8) *с вероятностью* 1 *достигают в асимптотике минимально возможного значения* Φ^* (5.48).

Доказательство. Выберем произвольную точку $\tilde{c} \in C$ и для каждого $t = 1, 2, \ldots$ рассмотрим

$$\tilde{\tilde{c}}_t^{il} \triangleq a_t\tilde{c}^{il} + b_t^i, \quad i=\overline{1, K}, \quad l=\overline{1, N},$$

где

$$a_t \triangleq 1 - 2N\varepsilon_t \left(\min_i \bar{b}^i\right)^{-1}, \quad b_t^i \triangleq 2\bar{q}^i \varepsilon_t \left(\min_i \bar{b}^i\right)^{-1},$$

а стохастический вектор $\bar{b} = (\bar{b}^1, \ldots, \bar{b}^K)^{\mathrm{T}}$ однозначно определяется из системы уравнений

$$\bar{b}^j = \frac{1}{N} \sum_{i=1}^{K} \bar{b}^i \sum_{l=1}^{N} \pi_{ij}^l, \quad \sum_{j=1}^{K} \bar{b}^j = 1.$$

Решение этой системы уравнений существует, единственно и $\min_i \bar{b}^i > 0$, поскольку оно соответствует решению системы уравнений (5.33) при $d^{il} = N^{-1}$ (см. теорему 5.2). Отметим, что $\approx{c}_t \in \mathbf{C}_{2\varepsilon t}$ (5.55).

Рассмотрим также

$$\tilde{c}_t \triangleq \hat{\pi}_{\varepsilon_t}\left\{\approx{c}_t\right\}, \quad t = 1, 2, \ldots$$

Поскольку $\mathbf{C}_{2\varepsilon_t} \subset \mathbf{C}_{\varepsilon_t}$ и выполнены условия леммы 5.8, то существует конечный с вероятностью 1 момент времени t_0, такой, что для всех $t > t_0$

$$\left\|\tilde{c}_t - \approx{c}_t\right\| \leqslant \mathrm{const} \left\|\hat{\pi}_{n_t} - \pi\right\| = o\left(\left(\sum_{s=1}^{t} \varepsilon_s^2 \Delta n_s\right)^{-\rho/2}\right) \quad \forall \rho \in [0, 1)$$

и

$$\left\|\tilde{c}_{t+1} - \tilde{c}_t\right\| \leqslant \mathrm{const} \left\|\hat{\pi}_{n_{t+1}} - \hat{\pi}_{n_t}\right\| \leqslant O^*\left(\Delta n_t \left(\sum_{s=1}^{t} \varepsilon_s^2 \Delta n_t\right)^{-1}\right).$$

Далее, повторяем доказательство теоремы 5.3 до соотношения (5.90) включительно. Поскольку (5.92) в условиях настоящей теоремы может не выполняться, обоснуем (5.91). С учетом (5.106) и условий данной теоремы будем иметь

$$\sum_{\tau=1}^{\infty} \tau^{-2} \sum_{\alpha,\beta} \mathsf{M}\left\{\left(c_{t(\tau)}^{\alpha\beta} - \tilde{c}_{t(\tau)}^{\alpha\beta}\right) \frac{\xi_\tau^2 \chi\left(z_\tau = z(\alpha),\ x_\tau = x(\beta)\right)}{\left(c_{t(\tau)}^{\alpha\beta}\right)^2} \middle| \mathscr{F}_\tau\right\} \leqslant$$

$$\leqslant \mathrm{const} \sum_{\tau=1}^{\infty} \tau^{-2} \sum_{\alpha,\beta} \left(c_{t(\tau)}^{\alpha\beta} \sum_{s=1}^{N} c_{t(\tau)}^{\alpha\beta}\right)^{-1} \leqslant$$

$$\leqslant \mathrm{const} \sum_{\tau=1}^{\infty} \tau^{-2} \varepsilon_{t(\tau)}^{-2} \leqslant \mathrm{const} \sum_{t=1}^{\infty} \varepsilon_t^{-2} \left(\frac{1}{n_t - 1} - \frac{1}{n_{t+1} - 1}\right) < \infty.$$

Последний ряд сходится, так как $\sum_{k=1}^{\infty} \gamma_k \varepsilon_k^{-2} \Delta n_k n_k^{-1} < \infty$ и $n_{k+1}\gamma_k \xrightarrow[k\to\infty]{} \infty$, что и доказывает (5.91). При этом в (5.91)

$$\frac{1}{n}\sum_{\tau=1}^{n} \mathrm{M}\{\theta_\tau \mid \mathscr{F}_\tau\} =$$

$$= \frac{1}{n}\sum_{\tau=1}^{n}\sum_{\alpha,\beta} v_{\alpha\beta}\chi(z_\tau = z(\alpha))\, d_{n'(\tau)}^{\alpha\beta} -$$

$$- \sum_{\alpha,\beta} v_{\alpha\beta}\tilde{c}^{\alpha\beta} - r_{1n} + O^*\left(\frac{1}{n}\sum_{k=1}^{t(n)} \Delta n_k \varepsilon_k\right),$$

где

$$r_{1n} = r'_{1n} + r''_{1n},$$

$$r'_{1n} \triangleq \sum_{\alpha,\beta} v_{\alpha\beta}\frac{1}{n}\sum_{\tau=1}^{n}\left(\tilde{c}_{t(\tau)}^{\alpha\beta} - \tilde{\tilde{c}}_{t(\tau)}^{\alpha\beta}\right)\chi(z_\tau = z(\alpha))\left(\sum_{s=1}^{N} c_{t(\tau)}^{\alpha s}\right)^{-1},$$

$$r''_{1n} \triangleq \sum_{\alpha,\beta} v_{\alpha\beta}\frac{1}{n}\sum_{\tau=1}^{n}\tilde{\tilde{c}}_{t(\tau)}^{\alpha\beta}\left[\chi(z_t = z(\alpha))\left(\sum_{s=1}^{N} c_{t(\tau)}^{\alpha s}\right)^{-1} - 1\right].$$

Имея в виду эти новые обозначения для r'_{1n}, r''_{1n} покажем, что в условиях данной теоремы правая часть неравенства (5.95) в пределе стремится к нулю; при этом будем опускать те места доказательства, которые в точности повторяют доказательство теоремы 5.3.

Из (5.99) и леммы 5.6 для сходимости к нулю r'_{1n_k} достаточно, чтобы при некотором $\rho \in (0, 1)$

$$\frac{1}{n_k}\sum_{t=1}^{k}\frac{\Delta n_t}{\varepsilon_t}\left(\sum_{s=1}^{t}\varepsilon_s^2 \Delta n_s\right)^{-\rho/2} \xrightarrow[k\to\infty]{} 0.$$

По лемме П.7 в условиях данной теоремы это имеет место, поскольку

$$\varepsilon_t^{-2/\rho}\left(\sum_{s=1}^{t}\varepsilon_s^2\Delta n_s\right)^{-1} = O\left(\varepsilon_t^{-2/\rho}\Delta n_t^{-1} h_t\right) =$$

$$= O\left(h_t \varepsilon_t^{-2/\rho+1+K^2/4}\,\Delta n_t^{-1}\varepsilon_t^{-1-K^2/4}\right) =$$

$$= \left(\Delta n_t \varepsilon_t^{1+K^2/4}\right)^{-1} O\left(h_t \varepsilon_t^{-\delta}\right) \xrightarrow[t\to\infty]{} 0,$$

$$\delta \triangleq 2(\rho^{-1} - 1).$$

Покажем, что $r''_{1n_k} \to 0$ при $k \to \infty$. Аналогично (5.100), в силу леммы П.13 справедливо соотношение

$$\frac{1}{n_{k+1}-1} \sum_{\tau=1}^{n_{k+1}-1} \tilde{\tilde{c}}_{t(\tau)}^{\alpha\beta} \chi(z_\tau = z(\alpha)) \left(\sum_{s=1}^{N} c_{t(\tau)}^{\alpha s} \right)^{-1} =$$

$$= \frac{1}{n_{k+1}-1} \sum_{t=1}^{k} \left(\sum_{s=1}^{N} c_t^{\alpha s} \right)^{-1} \tilde{\tilde{c}}_t^{\alpha\beta} \sum_{\tau=n_t}^{n_{t+1}-1} \chi(z_\tau = z(\alpha)) \overset{\text{п.н.}}{=}$$

$$\overset{\text{п.н.}}{=} \frac{1}{n_{k+1}-1} \sum_{t=1}^{k} \left(\sum_{s=1}^{N} c_t^{\alpha s} \right)^{-1} \tilde{\tilde{c}}_i^{\alpha\beta} \sum_{\tau=n_t}^{n_{t+1}-1} \mathsf{P}\{z_\tau = z(\alpha) \mid \mathcal{F}_{n_t}\} + o(1),$$

поскольку сходятся ряды

$$\sum_{\tau=1}^{\infty} \tau^{-2} \mathsf{M} \left\{ \left[\tilde{\tilde{c}}_{t(\tau)}^{\alpha\beta} \left(\left(\sum_{s=1}^{N} c_{t(\tau)}^{\alpha s} \right)^{-1} \chi(z_\tau = z(\alpha)) - \right.\right.\right.$$

$$\left.\left.\left. - \mathsf{P}\{z_\tau = z(\alpha) \mid \mathcal{F}_{n'(\tau)}\} \right) \right]^2 \middle| \mathcal{F}_{n'(\tau)} \right\} \leqslant$$

$$\leqslant \sum_{k=1}^{\infty} \sum_{\tau=n_k}^{n_{k+1}-1} \tau^{-2} \mathsf{M} \left\{ \left[\chi(z_\tau = z(\alpha)) \left(\sum_{s=1}^{N} c_{t(\tau)}^{\alpha s} \right)^{-1} - \right.\right.$$

$$\left.\left. - \mathsf{P}\{z_\tau = z(\alpha) \mid \mathcal{F}_{n'(\tau)}\} \right]^2 \middle| \mathcal{F}_{n'(\tau)} \right\} \overset{\text{п.н.}}{\leqslant}$$

$$\overset{\text{п. н.}}{\leqslant} C(\omega) \sum_{k=1}^{\infty} \varepsilon_k^{-2} \sum_{\tau=n_k}^{n_{k+1}-1} \tau^{-2} \leqslant C(\omega) \sum_{k=1}^{\infty} \varepsilon_k^{-2} \left(\frac{1}{n_k - 1} - \frac{1}{n_{k+1}-1} \right) < \infty.$$

Отсюда и из неравенства (5.108) леммы П.7 получим

$$r''_{1n_k} \leqslant \frac{C_1}{n_{k+1}-1} \sum_{t=1}^{k} \varepsilon_t^{-1} \sum_{\alpha=1}^{K} \sum_{\tau=0}^{[\Delta n_t/r]} \left(\left| \sum_{l=0}^{r-1} \mathsf{P}\{z_{n_t+l+\tau r} = \right.\right.$$

$$\left.\left. = z(\alpha) \mid \mathcal{F}_{n_t}\} - \sum_{s=1}^{N} c_t^{\alpha s} \right| + r \right) + o(1) \leqslant$$

$$\overset{\text{п.н.}}{\leqslant} \frac{C_1}{n_{k+1}-1} \sum_{t=1}^{k} \varepsilon_t^{-1} \left(1 + \sum_{\tau=0}^{[\Delta n_t/r]} \left[1 - C\varepsilon_t^{r(K-r)} \right]^{\frac{n_t - r + 1 + \tau r}{r(K-r)}} \right) \leqslant$$

$$\leqslant \frac{C_1}{n_{k+1}-1} \sum_{t=1}^{k} \varepsilon_t^{-1-r(K-r)} + o(1) \xrightarrow[k\to\infty]{} 0$$

в условиях данной теоремы. Итак, $r_{1n_k} \xrightarrow[k\to\infty]{} 0$.

Величины r_{2n_k} стремятся к нулю при тех же условиях, что и в теореме 5.3.

Рассмотрим r_{3n_k}. Используя полученную выше оценку для $\|\tilde{c}_t - \tilde{\tilde{c}}_t\|$, приходим к неравенству

$$r_{3n_k} \leqslant \frac{K_2}{n_{k+1}-1} \sum_{\tau=1}^{n_{k+1}-1} \theta'_\tau + \frac{K_2 C(\omega)}{n_{k+1}-1} \sum_{t=1}^{k} \theta''_t,$$

где

$$\theta'_\tau \triangleq \gamma_{t(\tau)} \xi_\tau^2 \left(c_{t(\tau)}^{\alpha\beta}\right)^{-2},$$

$$\theta''_t \triangleq (\Delta n_t)^2 \left(\gamma_t \sum_{s=1}^{t} \varepsilon_s^2 \Delta n_s\right)^{-1}.$$

Среднее арифметическое неотрицательных величин θ'_τ стремится к нулю с вероятностью 1 в силу леммы П.8, поскольку в условиях данной теоремы сходится ряд

$$\sum_{\tau=1}^{\infty} \tau^{-1} \mathrm{M}\left\{\theta'_\tau \mid \mathscr{F}_{n'(\tau)}\right\} \leqslant C(\omega) \sum_{\tau=1}^{\infty} \tau^{-1} \gamma_{t(\tau)} \varepsilon_{t(\tau)}^{-2} \leqslant$$

$$\leqslant C(\omega) \sum_{k=1}^{\infty} \gamma_k \varepsilon_k^{-2} \sum_{\tau=n_k}^{n_{k+1}-1} \tau^{-1} \leqslant C(\omega) \sum_{k=1}^{\infty} \gamma_k \varepsilon_k^{-2} (\Delta n_k)\, n_k^{-1} < \infty.$$

Вторoе слагаемое в r_{3n_k} стремится к нулю по лемме П.7. Таким образом, $r_{3n_k} \to 0$ при $k \to \infty$.

Асимптотическая эквивалентность Φ_n и $\overline{\Phi}_n$, по аналогии с теоремой 5.3, также имеет место в силу леммы П.13, (5.106) и условий данной теоремы, поскольку с вероятностью 1 сходится ряд

$$\sum_{\tau=1}^{\infty} \tau^{-2} \mathrm{M}\left\{\xi_\tau^2 \mid \mathscr{F}_{n'(\tau)}\right\} \overset{\text{п.н.}}{\leqslant} \mathrm{const} \sum_{\tau=1}^{\infty} \tau^{-2} \varepsilon_{n'(\tau)}^{-2} \leqslant$$

$$\leqslant \mathrm{const}^1 \sum_{k=1}^{\infty} \varepsilon_k^{-2} \frac{\Delta n_k}{n_k n_{k+1}} < \infty.$$

Следовательно, в силу произвольности $\tilde{c} \in \mathrm{C}$ и соотношений (5.48), (5.54) и (5.57)

$$\overline{\lim_{n\to\infty}}\, \Phi_n = \overline{\lim_{k\to\infty}}\, \Phi_{n_k} \leqslant \min_{\tilde{c} \in \mathrm{C}} \sum_{\alpha,\beta} v_{\alpha\beta} \tilde{c}^{\alpha\beta} = \Phi^*.$$

Теорема доказана. ▲

Условия этой теоремы выполняются в классе последовательностей $\{\gamma_k\}$, $\{\varepsilon_k\}$, $\{n_k\}$ вида

$$\begin{gathered}\gamma_k = \gamma k^{-\nu},\ \varepsilon_k = \varepsilon k^{-\theta},\ n_k = [k^{\varkappa}],\\ 0 < 2\theta < \nu < 1 - 2\theta,\ \varkappa > 1 + \theta(1 + K^2/4).\end{gathered} \tag{5.122}$$

На рис. 11 показаны допустимые области значений (ν, θ) и $(\varkappa, \theta)$, которые гарантируют работоспособность алгоритма адаптивного управления (5.61)—(5.64): одинарная штриховка соответствует регулярным цепям,

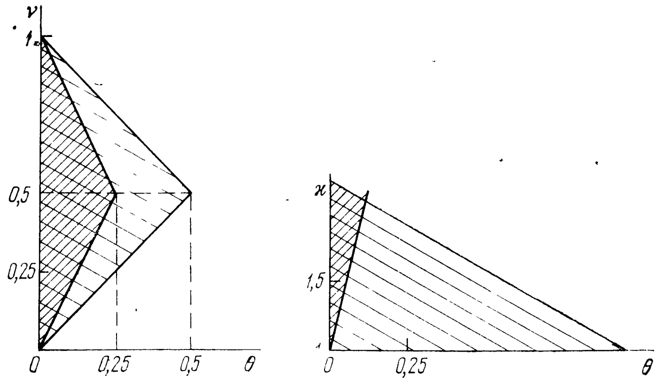

Рис. 11

а двойная — связным. Более жесткие требования (5.122) к параметрам алгоритма в случае связных цепей по сравнению с условиями (5.105) являются как бы «платой» за возможность управлять более широким классом процессов.

Оценим теперь скорость достижения равенства в (5.47) при адаптивном управлении связной марковской цепью с помощью алгоритма (5.61)—(5.64).

Теорема 5.6. *Пусть в условиях теоремы* 5.5 *последовательности* $\{\gamma_k\}$, $\{\varepsilon_k\}$, $\{n_k\}$ *удовлетворяют соотношениям* (5.122). *Тогда при любых* $c_1 \in \widehat{C}_{\varepsilon_1}$ *и* $z_1 \in Z$ *для произвольной связной управляемой марковской цепи из класса* $M(K, N)$, *функционирующей в соответствии с алгоритмом* (5.61)—(5.64), *с вероятностью* 1 *при любом* $\delta > 0$

$$\begin{gathered}\Phi_n - \Phi^* \leqslant o\left(n^{\delta - \varphi_1}\right) + O^*\left(n^{-\varphi_2}\right),\\ \varphi_1 = \varkappa^{-1} \min\left\{\frac{1}{2} - \theta,\ \nu - 2\theta\right\},\\ \varphi_2 = \varkappa^{-1} \min\{\varkappa - 1 - \theta(1 + K^2/4),\ 1 - \nu - 2\theta,\ \theta\},\end{gathered} \tag{5.123}$$

т. е. скорость достижения равенства в (5.47) *по порядку либо сколь угодно близка к* $n^{-\varphi_1}$ *при* $\varphi_1 \leqslant \varphi_2$, *либо равна* $n^{-\varphi_2}$ *при* $\varphi_2 < \varphi_1$.

Доказательство проводится по той же схеме, что и для теоремы 5.4, с использованием соотношений, полученных при доказательстве теоремы 5.5.

Аналогично (5.95), имеем

$$\overline{\Phi}_{n_{k+1}-1} - \widetilde{V}(\tilde{c}) \leqslant \tilde{r}_{1n_k} + r_{2n_k} + r_{3n_k} + r_{4n_k} + O^*(\varepsilon_k),$$

где $\tilde{r}_{1n_k} = (r'_{1n} + r''_{2n})\big|_{n=n_{k+1}-1}$, величины r'_{1n} и r''_{2n} определены в теореме 5.5, а r_{2n_k}, r_{3n_k} и r_{4n_k} такие же, как и в теоремах 5.3, 5.4.

С учетом специфики рассматриваемого случая (см. (5.106) — (5.108)), получим следующие оценки порядка скорости сходимости к нулю последовательностей, введенных при доказательстве теорем 5.3—5.5 и фигурирующих в правой части неравенства (5.95):

$$r_{4n_k} \overset{\text{п.н.}}{=} o\left(n_k^{\delta - \frac{\varkappa-2\theta}{2\varkappa}}\right),$$

$$r'_{1n} \overset{\text{п.н.}}{=} O^*\left(n^{\delta - \frac{\varkappa-2\theta}{2\varkappa}}\right),$$

$$r''_{n_k} \overset{\text{п.н.}}{\leqslant} o\left(n_k^{\delta-(1/2-\theta)/\varkappa}\right) + O^*\left(n_k^{\frac{1-\varkappa+\theta(1+K^2/4)}{\varkappa}}\right),$$

$$r_{1n_k} = r'_{1n_k} + r''_{1n_k}, \quad |r_{2n_k}| \overset{\text{п.н.}}{\leqslant} O^*\left(n_k^{-\frac{1-\nu}{\varkappa}}\right),$$

$$r_{3n_k} \overset{\text{п.н.}}{\leqslant} o\left(n_k^{\delta-\frac{\nu-2\theta}{\varkappa}}\right) + O^*\left(n_k^{\frac{2\theta+\nu-1}{\varkappa}}\right),$$

где $\delta > 0$ — произвольное сколь угодно малое число.

Таким образом, получаем в силу произвольности $\tilde{c} \in \mathbf{C}$

$$\Phi_n - \Phi^* = \overline{\Phi}_n - \Phi^* + o\left(n^{\delta - \frac{1-2\theta}{2}}\right) =$$

$$= \overline{\Phi}_{n'(n)} - \Phi^* + O\left(n^{-1/\varkappa}\right) + o\left(n^{\delta-\frac{1-2\theta}{2}}\right),$$

откуда с учетом (5.95) вытекает требуемая оценка. Теорема доказана. ▲

Следствие. *Наибольший порядок скорости достижения равенства в* (5.47), *гарантируемый теоремой* 5.6, *равен*

$o(n^{\delta-\varphi})$, $\varphi=(8+K^2/4)^{-1}$, *т. е.*

$$\Phi_n - \Phi^* \overset{\text{п.н.}}{\leqslant} o(n^{\delta-\varphi}) \quad \forall\delta>0$$

и достигается при

$$\nu=\nu^* \triangleq 1/2, \quad \theta=\theta^* \triangleq 1/6, \quad \varkappa=\varkappa^* \triangleq 4/3+K^2/24.$$

Таким образом, параметры ν^*, θ^* и $\varkappa^*$ являются оптимальными в смысле верхней оценки, даваемой теоремой 5.6. Интересно отметить, что число состояний K управляемой марковской цепи влияет только на значение $\varkappa^*$. С ростом K порядок φ скорости достижения равенства в (5.47) убывает, в пределе стремясь к нулю при $K\to\infty$. Наибольшее возможное значение φ реализуется для управляемых связных цепей с двумя состояниями ($K=2$), и при этом $\varkappa^*=3/2$, $\varphi=1/9$; в этом случае порядок скорости почти в два раза меньше, чем для класса регулярных управляемых марковских цепей (см. следствие из теоремы 5.4).

§ 5.6. Обобщение и обсуждение результатов

Рассмотренная в этой главе задача адаптивного управления конечной марковской цепью существенно отличается от задач адаптивного выбора, исследованных в предыдущих главах. Это отличие состоит в том, что ранее рассмотренные постановки задач касались так называемых процессов с независимыми значениями [117, 118], здесь же приходится иметь дело с зависимыми процессами, вложимыми в конечную марковскую цепь. Необходимость исследовать управляемые процессы такого типа исключает возможность использования ранее разработанного математического аппарата при анализе и синтезе алгоритмов адаптивного выбора и требует привлечения другой техники доказательства теорем о достижении цели управления.

В рамках принятого здесь подхода не удается доказать сходимость рандомизированных правил выбора d_n (5.57) к некоторому правилу d, однако обосновывается, что цель управления, состоящая в достижении равенства в соотношении (5.47), как в случае регулярности, так и в случае связности управляемой марковской цепи достигается при адаптивном управлении с помощью алгоритма (5.61)—(5.64). Естественно, что в более общем случае связной цепи требования (5.122) к параметрам этого алгоритма оказываются более жесткими, чем соответствующие требования (5.105) для регулярных цепей. Оказывается, что эти требования к алгоритму (5.61)—(5.64) гарантируют достижение целевого условия (5.36), (5.37) при адаптивном управлении конечными марковскими цепями общего вида.

Теорема 5.7. *Пусть выполнены условия теоремы* 5.6 *и*

$$\lim_{k\to\infty}\left(\sum_{t=1}^{k}\Delta n_t\varepsilon_t^K\right)^{-1}\ln k=0. \tag{5.124}$$

Тогда алгоритм адаптивного управления (5.61)—(5.64) *при любых начальных* $c_1\in\widehat{C}_{\varepsilon_1}$ *и* $z_1\in Z$ *с вероятностью* 1 *обеспечивает выполнение целевого условия* (5.36), (5.37) *для любой управляемой марковской цепи из класса* $M(K, N)$.

Доказательство. Покажем сначала, что в условиях теоремы произвольная управляемая марковская цепь из класса $M(K, N)$ за конечное число тактов с вероятностью 1 окажется в одной из связных компонент $Z_+^{(l)}$, $l=\overline{1, L}$ и более из нее не выйдет.

Пусть l_0 — число элементов класса $Z_+^{(0)}$ несущественных состояний, а $\Pi^{(0)}(d)$ — квадратная матрица переходных вероятностей, соответствующая состояниям класса $Z_+^{(0)}$ при невырожденной стационарной стратегии $\{d\}\in\in\sum_S^+$ (см. лемму 5.1). Будем для удобства считать, что состояниям из $Z_+^{(0)}$ соответствуют номера 1, 2, ..., l_0, т. е. $Z_+^{(0)}=\{z(i)\,|\,i=\overline{1, l_0}\}$ Как отмечалось в § 5.1, эта матрица не является стохастической, причем у нее существует такая строка с номером α, что

$$\sum_{j=1}^{K}\sum_{l=1}^{N}\pi_{\alpha j}^{l}d^{\alpha l}<1$$

для любых $d\in\operatorname{int}D$.

Покажем, что при выполнении условия (5.124)

$$\sum_{k=1}^{\infty}\mathsf{P}\,\{z_{n_k}\in Z_+^{(0)}\}<\infty. \tag{5.125}$$

Поскольку в (5.61)—(5.64) правила d_n выбора управлений «заморожены» на интервалах времени n_k, $n_{k+1}-1$, то в силу предположения ПЗ''' для каждых $i, j=\overline{1, l_0}$

$$\mathsf{P}\{z_{n_{k+1}}=z(j)\,|\,\mathcal{F}_{n_k}, z_{n_k}=z(i)\}=\left[\left(\Pi^{(0)}(d_k)\right)^{\Delta n_k}\right]_{ij},$$

где в правой части этого равенства стоит элемент Δn_k-й степени матрицы $\Pi^{(0)}(d_k)$, стоящий на пересечении i-й строки и j-го столбца.

Обозначим целую часть $[\Delta n_k/l_0]\triangleq m_k$ Для любого $i=\overline{1, l_0}$ получим

$$\sum_{j=1}^{l_0}\mathsf{P}\{z_{n_{k+1}}=z(j)\,|\,\mathcal{F}_{n_k}, z_{n_k}=z(i)\}=$$

$$=\sum_{j=1}^{l_0}\sum_{j_1=1}^{l_0}\cdots\sum_{j_{m_k}=1}^{l_0}[(\Pi^{(0)}(d_k))^{l_0}]_{ij_1}\cdots[(\Pi^{(0)}(d_k))^{l_0}]_{j_{m_k-1}j_{m_k}}\times$$

$$\times[(\Pi^{(0)}(d_k))^{l_0}]_{j_{m_k}j}\leqslant\sum_{j_1=1}^{l_0}[(\Pi^{(0)}(d_k))^{l_0}]_{ij_1}\times$$

$$\times\sum_{j_2=1}^{l_0}[(\Pi^{(0)}(d_k))^{l_0}]_{j_1j_2}\cdots\sum_{j_{m_k}=1}^{l_0}[(\Pi^{(0)}(d_k))^{l_0}]_{j_{m_k-1}j_{m_k}}\leqslant$$

$$\leqslant\left(\max_{i=\overline{1,l_0}}\max_{d\in D_{\varepsilon_k}}\sum_{j=1}^{l_0}[(\Pi^{(0)}(d))^{l_0}]_{ij}\right)^{m_k},$$

где

$$D_\varepsilon\triangleq\left\{\|d^{il}\|\,|\,d^{il}\geqslant\varepsilon,\quad\sum_{l=1}^{N}d^{il}=1\,(i=\overline{1, K},\ l=\overline{1, N})\right\}.$$

Но при любой невырожденной стационарной стратегии $\{d\}$ все строки матрицы $[\Pi^{(0)}(d)]^{l_0}$ не являются стохастическими. В самом деле, однородная марковская цепь с переходной матрицей $\Pi(d)$, стартуя из любого состояния $z\in Z_+^{(0)}$, с положительной вероятностью за l_0-1 тактов пройдет через состояние (множества $Z_+^{(0)}$), которому соответствует нестохастическая строка матрицы $\Pi^{(0)}(d)$; это следует из того, что $Z_+^{(0)}$ — множество несущественных состояний и, следовательно, оно не может содержать в себе связных компонент. Таким образом, при любом достаточно малом $\varepsilon>0$

$$\max_{d\in D_\varepsilon}\sum_{j=1}^{l_0}[(\Pi^{(0)}(d))^{l_0}]_{ij}<1.$$

Эту оценку можно уточнить, если учесть линейность по d матрицы $\Pi^{(0)}(d)$ и вид симплекса D_ε:

$$\max_{d\in D_\varepsilon}\sum_{j=1}^{l_0}\left[\left(\Pi^{(0)}(d)\right)^{l_0}\right]_{ij}\leqslant 1-b\varepsilon^{l_0}$$

для любых $i=\overline{1,\ l_0}$ и достаточно малого $\varepsilon>0$, где число $b>0$.

Таким образом, при достаточно больших k

$$\sum_{j=1}^{l_0}\mathsf{P}\left\{z_{n_{k+1}}=z(j)\,|\,\mathscr{F}_{n_k},\ z_{n_k}=z(i)\right\}\leqslant\left(1-b\varepsilon_k^{l_0}\right)^{m_k}.$$

Отсюда сразу получаем рекуррентное неравенство

$$\mathsf{P}\left\{z_{n_{k+1}}\in Z_+^0\right\}=\mathsf{M}\left\{\sum_{j=1}^{l_0}\mathsf{P}\left\{z_{n_{k+1}}=z(j)\,|\,\mathscr{F}_{n_k}\right\}\right\}=$$

$$=\mathsf{M}\left\{\sum_{i=1}^{l_0}\chi\left(z_{n_k}=z(i)\right)\sum_{j=1}^{l_0}\mathsf{P}\left\{z_{n_{k+1}}=z(j)\,|\,\mathscr{F}_{n_k},\ z_{n_k}=z(i)\right\}\right\}\leqslant$$

$$\leqslant\left(1-b\varepsilon_k^{l_0}\right)^{m_k}\mathsf{P}\left\{z_{n_k}\in Z_+^{(0)}\right\},$$

из которого в силу (5.124) вытекает (5.125). Но тогда по лемме Бореля — Кантелли процесс z_n с вероятностью 1 может оставаться в множестве $Z_+^{(0)}$ лишь конечное число тактов. Перейдя в одну из связных компонент $Z_+^{(l)}$, процесс z_n более из нее не выйдет, а следовательно, для почти всех $\omega\in\Omega_l$ (см. леммы 5.1, 5.2) справедливо утверждение теоремы 5.5, относящееся к связным управляемым марковским цепям: алгоритм (5.61) — (5.64) обеспечивает достижение оптимизационной цели

$$\overline{\lim_{n\to\infty}}\,\Phi_n=\Phi^*(l)\triangleq\inf_{\{d\}\in\Sigma_S^+}\sum_{z(i)\in Z_+^{(l)}}\sum_{k=1}^{N}v_{ik}d^{ik}p_i(d).\qquad(5.126)$$

Таким образом, из (5.126) вытекает, что с вероятностью 1

$$\overline{\lim_{n\to\infty}}\,\Phi_n\leqslant\max_{l=\overline{0,L}}\Phi^*(l)=\Phi^*$$

(где Φ^* определено в (5.37)), т. е. выполняется целевое условие (5.36), (5.37). Теорема доказана. ▲

Замечание. Поскольку попадание процесса z_n в ту или иную связную компоненту $Z_+^{(l)}$ марковской цепи при любой невырожденной стратегии управления является случайным событием, зависящим от реализованных управлений и начального распределения вероятностей состояний, то нельзя в общем случае требовать выполнения с вероятностью 1 какой-либо оптимизационной цели, отличной от (5.36), (5.37) (за исключением случая равнодоходных цепей, у которых $\Phi^*(l)=\Phi^*$, $l=\overline{1, L}$ [118]).

Очевидно, что в классе последовательностей $\{\gamma_k\}$, $\{\varepsilon_k\}$, $\{n_k\}$ вида (5.122) условие (5.124) будет автоматически выполняться, т. е. теорема 5.7 «работает» в широком классе последовательностей $\{\gamma_k\}$, $\{\varepsilon_k\}$, $\{n_k\}$, входящих в алгоритм адаптивного управления (5.61)—(5.64).

Процесс z_n, порождаемый алгоритмом (5.61)—(5.64) в условиях (5.122) при управлении произвольной марковской цепью из класса $M(K, N)$, с вероятностью 1 остается в одной из связных компонент начиная с конечного момента времени. Следовательно, скорость достижения целевого условия (5.36), (5.37) такая, как предписывается теоремой 5.6 и следствием из нее. Таким образом, справедлива следующая теорема.

Теорема 5.8. *Пусть выполнены условия теоремы* 5.6 *и, кроме того, последовательности* $\{\gamma_k\}$, $\{\varepsilon_k\}$, $\{n_k\}$ *выбраны в соответствии с* (5.122). *Тогда для произвольной управляемой марковской цепи из класса* $M(K, N)$, *функционирующей в соответствии с алгоритмом* (5.61)—(5.64), *справедлива оценка* (5.123) *скорости достижения целевого условия* (5.36), (5.37). *При этом максимальная гарантируемая скорость реализуется при*

$$\nu=\nu^*=\frac{1}{2},\quad \theta=\theta^*=\frac{1}{6},\quad \varkappa=\varkappa^*=\frac{K^2+32}{24}$$

и равна $o(n^{\delta-\varphi})$ $\forall\,\delta>0, \varphi=4/(K^2+32)$.

Рассмотренный в данной главе алгоритм адаптивного управления относится к классу алгоритмов стохастической аппроксимации, причем он реализует асимптотически оптимальную в смысле целевого условия (5.36), (5.37) (в классе любых допустимых, в том числе и нестационарных рандомизированных стратегий) стратегию управления конечной марковской цепью в условиях априорной неопределенности. Естественно, что этот алгоритм не является единственным алгоритмом адаптивного управления конечной марковской цепью. Существуют и другие алгоритмы [100, 117, 118, 149], к каждому из которых можно добавить еще и целую серию соответствующих модификаций. Однако большинство из них предполагает текущее оценивание переходных матриц $\|\pi_{ij}^{l}\|$ управляемой цепи Маркова либо других матриц (например, матрицы чувствительности) той же размерности [148, 149]. Интересно было бы получить рекуррентный алгоритм адаптивного управления, не требующий такого дополнительного оценивания. По-видимому, это можно сделать, используя поисковые методы.

Интересно отметить также, что при $K=1$, т. е. в случае, когда задача адаптивного управления марковской цепью вырождается в задачу безусловной минимизации средних потерь (методы решения которой рассматривались во второй главе), алгоритм (5.61) — (5.64) совпадает с алгоритмом (2.47) при $n_k=k$, $k=1, 2, \ldots$

ЗАКЛЮЧЕНИЕ

Рассмотренные здесь задачи адаптивного выбора вариантов представляют собой задачи адаптивного управления случайными процессами с конечным множеством управляющих воздействий. С единых позиций сформулированы задачи безусловного и условного выбора, игровая задача и задача адаптивного управления конечными однородными марковскими цепями. В каждой из них целью управления является оптимизационная цель типа минимизации с вероятностью 1 предельных значений текущих средних потерь. Показано, что эти задачи могут решаться на основе единого подхода, основанного на использовании рандомизированных правил выбора и метода стохастической аппроксимации. Полученные на этом пути алгоритмы имеют рекуррентный характер и допускают сравнительно простую реализацию на цифровых вычислительных машинах при неявных априорных данных о решаемой задаче. Эти алгоритмы представляют собой, по сути, алгоритмы дискретного адаптивного управления в реальном масштабе времени.

Малый объем априорной информации отражается в медленной сходимости этих методов (со степенной скоростью). Исключение составляют лишь задачи с бинарными потерями, которые могут быть решены достаточно быстро (со скоростью геометрической прогрессии), но для этого требуется дополнительная априорная информация о достаточном различии средних потерь за выбор оптимального и неоптимальных вариантов.

Полученные результаты по скорости сходимости рассмотренных алгоритмов имеют характер оценок сверху. Остается, таким образом, открытым вопрос о точности этих оценок. Кроме того, интересно было бы изучить предельные возможности по скорости сходимости всего класса алгоритмов адаптивного выбора вариантов для каждой задачи (т. е. получить оценки снизу скорости достижения цели адаптации) и установить вид алгоритмов, обладающих предельно возможной скоростью.

Важным представляется также получение и исследование рекуррентных алгоритмов типа стохастической аппроксимации, предназначенных для решения задач адаптивного управления частично наблюдаемыми марковскими цепями (когда состояния цепи не доступны наблюдению), задач управления цепями при более сложных целевых условиях (условная минимизация, игры и т. д.), задач управления полумарковскими процессами. Эти вопросы ждут своего решения.

НЕКОТОРЫЕ СВЕДЕНИЯ ИЗ ТЕОРИИ ВЕРОЯТНОСТЕЙ И СЛУЧАЙНЫХ ПРОЦЕССОВ

Пусть $\Omega = \{\omega\}$ — множество *элементарных событий* (исходов) ω. Система $\mathscr{F}$ подмножеств Ω называется *σ-алгеброй*, если выполняются следующие свойства:

1) $\Omega \in \mathscr{F}$;

2) $\bigcup\limits_{n=1}^{\infty} A_n \in \mathscr{F}$, $\bigcap\limits_{n=1}^{\infty} A_n \in \mathscr{F}$ для любых множеств $A_n \in \mathscr{F}$, $n = 1, 2, \ldots$;

3) $\overline{A} \triangleq \{\omega \in \Omega \mid \omega \overline{\in} A\} \in \mathscr{F}$ для любых $A \in \mathscr{F}$.

Пара $(\Omega, \mathscr{F})$ образует *измеримое пространство*. Говорят, что на измеримом пространстве $(\Omega, \mathscr{F})$ задана *вероятностная мера* P, если определена функция множеств $\mathsf{P}\{A\}$, $A \in \mathscr{F}$ со значениями в отрезке $[0, 1]$, такая, что для любых попарно непересекающихся множеств $A_1, A_2, \ldots$ из $\mathscr{F}$

$$\mathsf{P}\left\{\bigcup_{n=1}^{\infty} A_n\right\} = \sum_{n=1}^{\infty} \mathsf{P}\{A_n\}$$

и

$$\mathsf{P}\{\Omega\} = 1.$$

Тройка $(\Omega, \mathscr{F}, \mathsf{P})$ образует *вероятностное пространство*. Далее, предполагается, что оно является полным по вероятностной мере P, т. е. σ-алгебра $\mathscr{F}$ содержит все такие множества A, для каждого из которых найдутся множества $A_1, A_2 \in \mathscr{F}$, удовлетворяющие условиям $A_1 \subseteq A \subseteq A_2$, $\mathsf{P}\{A_2 \setminus A_1\} = 0$.

Действительная функция $\xi = \xi(\omega)$, $\omega \in \Omega$ называется *случайной величиной*, определенной на $(\Omega, \mathscr{F})$, если она *$\mathscr{F}$-измерима*, т. е. для любых $x \in (-\infty, \infty)$

$$\{\omega \mid \xi(\omega) \leqslant x\} \in \mathscr{F}.$$

Пусть $\xi_1, \ldots, \xi_k$ — случайные величины, определенные на вероятностном пространстве $(\Omega, \mathscr{F}, \mathsf{P})$. Минимальная σ-алгебра, содержащая для любых $x = (x_1, \ldots, x_k) \in R^k$ события

$$\{\omega \mid \xi_1 \leqslant x_1, \ldots, \xi_k \leqslant x_k\},$$

называется *σ-алгеброй, порожденной случайными величинами* $\xi_1, \ldots, \xi_k$ и обозначается $\sigma(\xi_1, \ldots, \xi_k)$.

Математическим ожиданием $\mathsf{M}\{\xi\}$ по вероятностной мере P случайной величины $\xi = \xi(\omega)$, определенной на $(\Omega, \mathscr{F})$, называется интеграл Лебега

$$\mathsf{M}\{\xi\} \triangleq \int_{\Omega} \xi(\omega)\, \mathsf{P}\{d\omega\}.$$

Случайная величина $\eta = \eta(\omega)$ называется *условным математическим ожиданием* случайной величины $\xi = \xi(\omega)$ относительно σ-алгебры $\mathscr{F}_0 \subseteq \mathscr{F}$ и обозначается $M\{\xi|\mathscr{F}_0\}$, если она $\mathscr{F}_0$-измерима и для всякого множества $A \in \mathscr{F}_0$

$$\int\limits_A \eta(\omega)\,\mathsf{P}\{d\omega\} = \int\limits_A \xi(\omega)\,\mathsf{P}\{d\omega\}$$

(интегралы понимаются в смысле Лебега).

Условное математическое ожидание обладает следующими свойствами: если $\xi = \xi(\omega)$ и $\theta = \theta(\omega)$ — случайные величины, определенные на $(\Omega, \mathscr{F})$, θ — $\mathscr{F}_0$-измерима, $\mathscr{F}_0 \subseteq \mathscr{F}$, то

1) $\mathsf{M}\{\theta|\mathscr{F}_0\} = \theta$,

2) $\mathsf{M}\{\xi\theta|\mathscr{F}_0\} = \theta\mathsf{M}\{\xi|\mathscr{F}_0\}$,

3) $\mathsf{M}\{\mathsf{M}\{\xi|\mathscr{F}_1\}|\mathscr{F}_0\} = \mathsf{M}\{\xi|\mathscr{F}_0\}$, где $\mathscr{F}_0 \subseteq \mathscr{F}_1$.

Здесь и далее равенство двух случайных величин понимается в сильном смысле, т. е. с вероятностью 1 (по вероятностной мере P).

Неравенство Чебышева. *Для любых* $a > 0$ *и* $r > 0$

$$\mathsf{P}\{|\xi| \geqslant a\} \leqslant a^{-r}\mathsf{M}\{|\xi|^r\}.$$

Доказательство вытекает из неравенства

$$\mathsf{M}\{|\xi|^r\} \geqslant \int\limits_{\{\omega:|\xi|\geqslant a\}} |\xi|^r \mathsf{P}\{d\omega\} \geqslant a^r \mathsf{P}\{|\xi| \geqslant a\}. \quad \blacktriangle$$

Неравенство Коши — Буняковского. *Если случайные величины* ξ, η *таковы, что*

$$\mathsf{M}\{\xi^2\} < \infty, \quad \mathsf{M}\{\eta^2\} < \infty,$$

то

$$\mathsf{M}^2\{|\xi\eta|\} \leqslant \mathsf{M}\{\xi^2\}\mathsf{M}\{\eta^2\}.$$

Доказательство. Если $\mathsf{M}\{\xi^2\} = 0$, то $\mathsf{P}\{\xi = 0\} = 1$ и неравенство выполнено. Пусть $\mathsf{M}\{\xi^2\} > 0$ и $\mathsf{M}\{\eta^2\} > 0$, Тогда требуемое неравенство получается усреднением соотношения

$$2|uv| \leqslant u^2 + v^2$$

при

$$u \triangleq \xi/\sqrt{\mathsf{M}\{\xi^2\}}, \quad v \triangleq \eta/\sqrt{\mathsf{M}\{\eta^2\}}. \quad \blacktriangle$$

Неравенство Иенсена. *Пусть* $f = f(x)$, $x \in R^1$ — *выпуклая функция, т. е. для любых* x, $y \in R^1$ *и* $\lambda \in [0, 1]$

$$f(\lambda x + (1-\lambda)y) \leqslant \lambda f(x) + (1-\lambda)f(y),$$

и пусть $\xi = \xi(\omega)$ — *случайная величина, такая, что* $\mathsf{M}\{|\xi|\} < \infty$. *Тогда*

$$\mathsf{M}\{f(\xi)\} \geqslant f(\mathsf{M}\{\xi\}).$$

Доказательство. Для выпуклых функций справедливо следующее утверждение: для любого $x_0 \in R^1$ найдется число $g(x_0)$, такое, что для всех $x \in R^1$

$$f(x) \geqslant f(x_0) + g(x_0)(x - x_0).$$

Усредняя это неравенство при $x = \xi$ и $x_0 = \mathsf{M}\{\xi\}$, получаем требуемое неравенство. ▲

Лемма Фату. *Пусть* η, $\xi_1, \ldots, \xi_n$ — *случайные величины. Тогда*

1) *если* $\mathsf{P}\{\xi_n \geqslant \eta\} = 1$ *для всех* $n = 1, 2, \ldots$ *и* $\mathsf{M}\{\eta\} > -\infty$, *то* $\mathsf{M}\left\{\varliminf_{n\to\infty} \xi_n\right\} \leqslant \varliminf_{n\to\infty} \mathsf{M}\{\xi_n\}$;

2) *если* $\mathsf{P}\{\xi_n \leqslant \eta\} = 1$ *для всех* $n = 1, 2, \ldots$ *и* $\mathsf{M}\{\eta\} < \infty$, *то* $\varlimsup_{n\to\infty} \mathsf{M}\{\xi_n\} \geqslant \mathsf{M}\left\{\varlimsup_{n\to\infty} \xi_n\right\}$;

3) *если* $\mathsf{P}\{|\xi_n| \leqslant \eta\} = 1$ *для всех* $n = 1, 2, \ldots$ *и* $\mathsf{M}\{\eta\} < \infty$, *то* $\mathsf{M}\left\{\varliminf_{n\to\infty} \xi_n\right\} \leqslant \varliminf_{n\to\infty} \mathsf{M}\{\xi_n\} \leqslant \varlimsup_{n\to\infty} \mathsf{M}\{\xi_n\} \leqslant \mathsf{M}\left\{\varlimsup_{n\to\infty} \xi_n\right\}$.

Доказательство случая 3) вытекает из 1) и 2), но неравенство в 2) есть следствие 1). Докажем 1).

Обозначим $\varphi_n \triangleq \inf\limits_{m\geqslant n} \xi_m$. Тогда

$$\varliminf_{n\to\infty} \xi_n = \lim_{n\to\infty} \inf_{m\geqslant n} \xi_m = \lim_{n\to\infty} \varphi_n$$

и, значит, по теореме о монотонной сходимости [147, с. 202]

$$\mathsf{M}\left\{\varliminf_{n\to\infty} \xi_n\right\} = \mathsf{M}\left\{\lim_{n\to\infty} \varphi_n\right\} = \lim_{n\to\infty} \mathsf{M}\{\varphi_n\} =$$

$$= \varliminf_{n\to\infty} \mathsf{M}\{\varphi_n\} \leqslant \varliminf_{n\to\infty} \mathsf{M}\{\xi_n\}. \ ▲$$

Теорема Лебега (о мажорируемой сходимости). *Пусть* η, ξ_1, $\xi_2 \ldots$ — *случайные величины, такие, что*

$$\mathsf{P}\{|\xi_n| \leqslant \eta\} = 1, \quad \mathsf{P}\left\{\lim_{n\to\infty} \xi_n = \xi\right\} = 1, \quad \mathsf{M}\{\eta\} < \infty.$$

Тогда

$$\mathsf{M}\{|\xi|\} < \infty, \quad \lim_{n\to\infty} \mathsf{M}\{\xi_n\} = \mathsf{M}\{\xi\}, \quad \lim_{n\to\infty} \mathsf{M}\{|\xi_n - \xi|\} = 0.$$

Доказательство. Поскольку для почти всех $\omega \in \Omega$

$$\varlimsup_{n\to\infty} \xi_n = \varliminf_{n\to\infty} \xi_n = \lim_{n\to\infty} \xi_n = \xi,$$

то

$$\mathsf{M}\left\{\varlimsup_{n\to\infty} \xi_n\right\} = \mathsf{M}\left\{\varliminf_{n\to\infty} \xi_n\right\} = \mathsf{M}\left\{\lim_{n\to\infty} \xi_n\right\}$$

и в силу леммы Фату (случай 3)) $\lim\limits_{n\to\infty} \mathsf{M}\{\xi_n\} = \mathsf{M}\{\xi\}$. Очевидно, что $\mathsf{P}\{|\xi| \leqslant \eta\} = 1$, поэтому $\mathsf{M}\{|\xi|\} < \infty$. Поскольку $\mathsf{P}\{|\xi_n - \xi| \leqslant 2\eta\} = 1$ то, используя уже доказанный факт, получим

$$\lim_{n\to\infty} \mathsf{M}\{|\xi_n - \xi|\} = \mathsf{M}\{\lim_{n\to\infty} |\xi_n - \xi|\} = 0. \quad ▲$$

Лемма Бореля — Кантелли. *Пусть* $(\Omega, \mathscr{F}, P)$ — *вероятностное пространство, а* $\{A_n\}$ — *некоторая последовательность событий из*

$\mathscr{F}$. Обозначим

$$\chi(A) \triangleq \begin{cases} 1, & \omega \in A, \\ 0, & \omega \overline{\in} A. \end{cases}$$

Тогда

1) *если* $\sum_{n=1}^{\infty} \mathsf{P}\{A_n\} < \infty$, *то* $\mathsf{P}\left\{\sum_{n=1}^{\infty} \chi(A_n) < \infty\right\} = 1$,

2) *если события A_n независимы в совокупности, т. е. для любого $k = 1, 2, \ldots$ и произвольных различных натуральных чисел $n_1, n_2, \ldots, n_k$*

$$\mathsf{P}\left\{\bigcup_{t=1}^{k} A_{n_t}\right\} = \prod_{t=1}^{k} \mathsf{P}\{A_{n_t}\}$$

и если, кроме того,

$$\sum_{n=1}^{\infty} \mathsf{P}\{A_n\} = \infty,$$

то

$$\mathsf{P}\left\{\sum_{n=1}^{\infty} \chi(A_n) = \infty\right\} = 1.$$

Доказательство. Так как

$$\left\{\omega \,\middle|\, \sum_{n=1}^{\infty} \chi(A_n) = \infty\right\} = \left\{\omega \,\middle|\, \bigcap_{n=1}^{\infty} \bigcup_{k \geqslant n} A_k\right\},$$

то по теореме Лебега (о мажорируемой сходимости)

$$\mathsf{P}\left\{\sum_{n=1}^{\infty} \chi(A_n) = \infty\right\} = \mathsf{P}\left\{\bigcap_{n=1}^{\infty} \bigcup_{k \geqslant n} A_k\right\} = \\ = \lim_{n \to \infty} \mathsf{P}\left\{\bigcup_{k \geqslant n} A_k\right\} \leqslant \lim_{n \to \infty} \sum_{k=n}^{\infty} \mathsf{P}\{A_k\} = 0,$$

что и доказывает случай 1). Докажем случай 2). События $\overline{A}_1$, $\overline{A}_2, \ldots$ также независимы в совокупности, а следовательно, по теореме Лебега

$$\mathsf{P}\left\{\bigcap_{k=n}^{\infty} \overline{A}_k\right\} = \prod_{k=n}^{\infty} \mathsf{P}\{\overline{A}_k\} = \prod_{k=n}^{\infty} (1 - \mathsf{P}\{A_k\}) = 0,$$

поскольку в силу неравенства $1 - x \leqslant e^{-x}$

$$\prod_{k=n}^{s} (1 - \mathsf{P}\{A_k\}) \leqslant \exp\left\{-\sum_{k=n}^{s} \mathsf{P}\{A_k\}\right\} \xrightarrow[s \to \infty]{} 0. \quad \blacktriangle$$

Напомним определения некоторых видов сходимости случайных последовательностей, используемых в теории случайных процессов. Последовательность $\{\xi_n\}$ случайных величин, определенных на вероятностном пространстве $(\Omega, \mathscr{F}, \mathsf{P})$, *сходится к случайной величине* ξ

а) *по вероятности* $\left(\xi_n \xrightarrow[n\to\infty]{\mathsf{P}} \xi\right)$, если для любого $\varepsilon > 0$

$$\mathsf{P}\{|\xi_n - \xi| > \varepsilon\} \underset{n\to\infty}{\longrightarrow} 0;$$

б) *с вероятностью 1 или почти наверное* $\left(\xi_n \xrightarrow[n\to\infty]{\text{п.н.}} \xi\right)$, *если*

$$\mathsf{P}\left\{\lim_{n\to\infty} \xi_n = \xi\right\} = 1;$$

в) *в среднеквадратическом* (l. i. m. $\xi_n = \xi$), если

$$\mathsf{M}\{(\xi_n - \xi)^2\} \underset{n\to\infty}{\longrightarrow} 0.$$

Из сходимости с вероятностью 1 и в среднеквадратическом следует сходимость по вероятности. Если имеет место достаточно быстрая сходимость в среднеквадратическом, такая, что $\sum\limits_{n=1}^{\infty} \mathsf{M}\{(\xi_n - - \xi)^2\} < \infty$, то $\xi_n \xrightarrow[n\to\infty]{\text{п.н.}} \xi$.

Пусть последовательность $\{\xi_n\}$ случайных величин, определенных на вероятностном пространстве $(\Omega, \mathscr{F}, \mathsf{P})$, согласована с последовательностью $\{\mathscr{F}_n\}$ неубывающих σ-алгебр, т. е. ξ_n измерима относительно $\mathscr{F}_n$ и $\mathscr{F}_n \subseteq \mathscr{F}_{n+1} \subseteq \mathscr{F}$ для любых $n = 1, 2, \ldots$ Последовательность $\{\xi_n, \mathscr{F}_n\}$ называется

а) *мартингалом*, если для каждого $n = 1, 2, \ldots$ существует $\mathsf{M}\{|\xi_n|\} < \infty$ и

$$\mathsf{M}\{\xi_{n+1} \mid \mathscr{F}_n\} \overset{\text{п.н.}}{=\!=} \xi_n;$$

б) *субмартингалом*, если для каждого $n = 1, 2, \ldots$ существует $\mathsf{M}\{\max\{0, \xi_n\}\} < \infty$ и

$$\mathsf{M}\{\xi_{n+1} \mid \mathscr{F}_n\} \overset{\text{п.н.}}{\geqslant} \xi_n;$$

в) *супермартингалом*, если для каждого $n = 1, 2, \ldots$ существует $\mathsf{M}\{\max\{0, -\xi_n\}\} < \infty$ и

$$\mathsf{M}\{\xi_{n+1} \mid \mathscr{F}_n\} \overset{\text{п.н.}}{\leqslant} \xi_n.$$

Теорема Дуба. *Пусть $\{\xi_n, \mathscr{F}_n\}$ — неотрицательный $\left(\xi_n \overset{\text{п.н.}}{\geqslant} 0\right)$ супермартингал, причем $\sup\limits_n \mathsf{M}\{\xi_n\} < \infty$. Тогда существует неотрицательная случайная величина ξ, такая, что $\mathsf{M}\{\xi\} < \infty$ и $\xi_n \xrightarrow[n\to\infty]{\text{п.н.}} \xi$.*

Доказательство. Выберем произвольные неотрицательные числа a и b, $a < b$ и обозначим $\beta_n(a, b)$ — число пересечений процессом $\{\xi_n\}$ отрезка $[a, b]$ снизу вверх за n тактов. Введем также последовательность $\{\tau_n\}$ моментов «первого выхода» процесса

$\{\xi_n\}$ из отрезка $[a, b]$, т. е.

$$\tau_1 \triangleq \min\{n \mid \xi_n < a, \xi_t \geqslant a \ \forall t = \overline{1, n-1}\},$$
$$\tau_{2k} \triangleq \min\{n \mid > \tau_{2k-1}, \xi_n > b, \xi_t \leqslant b \ \forall t = \overline{\tau_{2k-1}, n-1}\},$$
$$\tau_{2k+1} \triangleq \min\{n \mid n > \tau_{2k}, \xi_n < a, \xi_n \geqslant a \ \forall t = \overline{\tau_{2k}, n-1}\}.$$

Обозначим

$$\chi_n \triangleq \begin{cases} 1, & \text{если} \quad \tau_{2k-1} < n \leqslant \tau_{2k}, \\ 0, & \text{если} \quad \tau_{2k} < n \leqslant \tau_{2k+1}. \end{cases}$$

Тогда

$$\sum_{t=1}^{n-1} \chi_t (\xi_{t+1} - \xi_t) \geqslant$$
$$\geqslant \begin{cases} (b-a)\beta_n(a, b), & \text{если} \quad \tau_{2k} < n \leqslant \tau_{2k+1}, \\ (b-a)\beta_n(a, b) + \xi_n - \xi_{\tau_{2k-1}}, & \text{если} \quad \tau_{2k-1} < n \leqslant \tau_{2k}. \end{cases}$$

Поскольку $\xi_{\tau_{2k-1}} < a$, то

$$(b-a)\beta_n(a, b) \leqslant \sum_{t=1}^{n} \chi_t (\xi_{t+1} - \xi_t) + \max\{0, a - \xi_n\}.$$

Учитывая измеримость χ_n относительно $\mathscr{F}_n$ и свойства условных математических ожиданий, получим

$$\mathsf{M}\{\chi_t(\xi_{t+1} - \xi_t)\} = \mathsf{M}\{\mathsf{M}\{\chi_t(\xi_{t+1} - \xi_t) \mid \mathscr{F}_t\}\} =$$
$$= \mathsf{M}\{\chi_t \mathsf{M}\{\xi_{t+1} - \xi_t \mid \mathscr{F}_t\}\} \leqslant 0.$$

Отсюда следует оценка сверху среднего числа пересечений

$$\mathsf{M}\{\beta_n(a, b)\} \leqslant (b-a)^{-1} \mathsf{M}\{\max\{0, a - \xi_n\}\}.$$

Доказательство самой теоремы проведем от противного: допустим, что

$$\mathsf{P}\left\{\varliminf_{n\to\infty} \xi_n < \varlimsup_{n\to\infty} \xi_n\right\} > 0.$$

Тогда найдутся неотрицательные a и b, $a < b$, такие, что

$$\mathsf{P}\left\{\varliminf_{n\to\infty} \xi_n \leqslant a < b \leqslant \varlimsup_{n\to\infty} \xi_n\right\} > 0,$$

а значит,

$$\mathsf{P}\left\{\lim_{n\to\infty} \beta_n(a, b) = \infty\right\} > 0.$$

Отсюда, используя лемму Фату и полученную выше оценку сверху среднего числа пересечений, приходим к противоречию:

$$\infty > (b-a)^{-1} \sup_n \mathsf{M}\{\max\{0, a - \xi_n\}\} \geqslant$$
$$\geqslant \varliminf_{n\to\infty} \mathsf{M}\{\beta_n(a, b)\} \geqslant \mathsf{M}\left\{\varliminf_{n\to\infty} \beta_n(a, b)\right\} = \infty.$$

Теорема доказана. ▲

ПРИЛОЖЕНИЯ

П. 1. Вспомогательные утверждения

Лемма П.1. *Предположим, что последовательности действительных чисел $\{u_n\}$, $\{w_n\}$ для всех $n \geqslant n_0$ удовлетворяют рекуррентным неравенствам*

$$u_{n+1} \leqslant (1 - cn^{-1}) u_n + dn^{-(p+1)},$$
$$w_{n+1} \geqslant (1 - cn^{-1}) w_n + dn^{-(p+1)},$$

где $c > p > 0$, $d \geqslant 0$. Тогда справедливы оценки

$$\varlimsup_{n\to\infty} n^p u_n \leqslant d/(c-p) \leqslant \varliminf_{n\to\infty} n^p w_n,$$

причем существуют последовательности $\{u_n\}$ и $\{w_n\}$, для которых они достигаются.

Доказательство. Без ограничения общности можно считать, что $n_0 \geqslant c$. Введем последовательность $\{y_n\}$ с помощью рекуррентного соотношения

$$y_{n+1} = (1 - cn^{-1}) y_n + dn^{-(p+1)}, \qquad n \geqslant n_0.$$

Покажем, что

$$\lim_{n\to\infty} n^p y_n = d/(c-p).$$

Имеем для всех $n \geqslant n_0$

$$y_{n+1} = y_{n_0} \prod_{k=n_0}^{n} \left(1 - \frac{c}{k}\right) + \sum_{k=n_0}^{n} \frac{d}{k^{p+1}} \prod_{m=k+1}^{n} \left(1 - \frac{c}{m}\right).$$

Введем следующие определения. Пусть $\{\alpha_n\}$ — некоторая последовательность положительных чисел. Будем обозначать

1) $\beta_n \triangleq O(\alpha_n) \Leftrightarrow \beta_n/\alpha_n$ — ограниченная последовательность;

2) $\beta_n \triangleq O^*(\alpha_n) \Leftrightarrow \varlimsup\limits_{n\to\infty} \dfrac{\beta_n}{\alpha_n} = \varliminf\limits_{n\to\infty} \dfrac{\beta_n}{\alpha_n} \in (0, \infty)$;

3) $\beta_n \triangleq o(\alpha_n) \Leftrightarrow \lim\limits_{n\to\infty} \dfrac{\beta_n}{\alpha_n} = 0$.

Так как

$$n^{-p} - (n+1)^{-p} = n^{-p} - n^{-p}(1 + n^{-1})^{-p} =$$
$$= n^{-p} - n^{-p}[1 - pn^{-1} + O(n^{-2})] = pn^{-(p+1)} + O(n^{-(p+2)}),$$

то

$$\frac{d}{n^{p+1}} = \frac{d}{c-p}\left[\frac{1}{(n+1)^p} - \left(1-\frac{c}{n}\right)\frac{1}{n^p} + O\left(n^{-(p+2)}\right)\right].$$

Таким образом, после подстановки и некоторых преобразований, получим

$$y_{n+1} = y_{n_0}\prod_{k=n_0}^{n}\left(1-\frac{c}{k}\right) + \frac{d}{c-p}\left[\frac{1}{(n+1)^p} - \frac{1}{n_0^p}\prod_{m=n_0}^{n}\left(1-\frac{c}{m}\right) + \right.$$
$$\left. + \sum_{k=n_0}^{n} O\left(k^{-(p+2)}\right)\prod_{m=k+1}^{n}\left(1-\frac{c}{m}\right)\right].$$

Поскольку

$$\prod_{k=n_0}^{n}\left(1-\frac{c}{k}\right) \leqslant \exp\left\{-\sum_{k=n_0}^{n}\frac{c}{k}\right\} = O^*(1)\exp\{-c\ln n\} = O^*\left(n^{-c}\right)$$

и

$$\sum_{k=n_0}^{n} O\left(k^{-(p+2)}\right)\prod_{m=k+1}^{n}\left(1-\frac{c}{m}\right) =$$
$$= \prod_{m=n_0}^{n}\left(1-\frac{c}{m}\right)\sum_{k=n_0}^{n} O\left(k^{-2}\right)k^{-p}\left[\prod_{s=n_0}^{k}\left(1-\frac{c}{s}\right)\right]^{-1} =$$
$$= O\left(n^{-c}\right)\sum_{k=n_0}^{n} O\left(k^{-2}\right)k^{c-p} =$$
$$= \frac{O(1)}{n^{p+\varepsilon}}\sum_{k=n_0}^{n} O\left(k^{-(2-\varepsilon)}\right)\frac{k^{c-p-\varepsilon}}{n^{c-p-\varepsilon}} = o\left(n^{-p}\right),$$

где $\varepsilon \in (0,\ c-p)$. Учитывая это, получим

$$y_{n+1} = \frac{d}{c-p}(n+1)^{-p} + o\left(n^{-p}\right).$$

Отсюда следует справедливость данной леммы, поскольку $y_n \geqslant u_n$ при $y_{n_0} = u_{n_0}$, а при $y_{n_0} = w_{n_0}$ для всех $n \geqslant n_0$ выполняется неравенство $y_n \leqslant w_n$. Лемма доказана. ▲

Замечание. Оценки, установленные в этой лемме, совпадают с оценками, даваемыми известными леммами Чжуна [13, 142]. Здесь устанавливается также достижимость этих оценок.

Следствие [13]. *Пусть* $\{u_n\}$ — *неотрицательная числовая последовательность, такая, что*

$$u_{n+1} = \left(1-\frac{c_n}{n}\right)u_n + \frac{d_n}{n^{p+1}},\quad n \geqslant n_0,$$

где $\{c_n\}$ и $\{d_n\}$ — последовательности действительных чисел, удовлетворяющих условиям:

$$\lim_{n\to\infty} c_n \triangleq c > p > 0, \quad \lim_{n\to\infty} d_n = d > 0.$$

Тогда

$$\lim_{n\to\infty} n^p u_n = d/(c-p).$$

Доказательство. Для любого $\varepsilon > 0$ найдется $n_1 > n_0$, такое, что $|c_n - c| \leqslant \varepsilon$ и $|d_n - d| \leqslant \varepsilon$ для всех $n \geqslant n_1$, а следовательно,

$$u_{n+1} \leqslant \left(1 - \frac{c-\varepsilon}{n}\right) u_n + \frac{d+\varepsilon}{n^{p+1}},$$

$$u_{n+1} \geqslant \left(1 - \frac{c+\varepsilon}{n}\right) u_n + \frac{d-\varepsilon}{n^{p+1}}.$$

В силу леммы П.1 отсюда получаем

$$\frac{d-\varepsilon}{c+\varepsilon-p} \leqslant \varliminf_{n\to\infty} n^p u_n \leqslant \varlimsup_{n\to\infty} n^p u_n \leqslant \frac{d+\varepsilon}{c-\varepsilon-p}.$$

Предельным переходом при $\varepsilon \to +0$ получаем требуемый результат. Следствие доказано. ▲

Лемма П.2 [69]. *Пусть последовательность неотрицательных чисел u_n для всех $n \geqslant n_0$ удовлетворяет рекуррентному неравенству*

$$u_{n+1} \leqslant (1-\nu_n) u_n + \beta_n + \delta_n u_n^r,$$

в котором $r \in (0, 1)$, а последовательности $\{\nu_n\}$, $\{\beta_n\}$, $\{\delta_n\}$ для некоторых $a \geqslant 0$, $c > 0$, $d \geqslant 0$, $\alpha \in (0, 1)$, $t > \alpha$, $s > \alpha$ удовлетворяют условиям

$$\varliminf_{n\to\infty} n^\alpha \nu_n \geqslant c, \quad \varlimsup_{n\to\infty} n^t \beta_n \leqslant d, \quad \varlimsup_{n\to\infty} n^s \delta_n \leqslant a.$$

Тогда

а) *если $s > r\alpha + (1-r)t$ и при $\alpha = 1$ $c > t-1$, то*

$$\varlimsup_{n\to\infty} n^{t-\alpha} u_n \leqslant d/\hat{c}(\alpha), \quad \hat{c}(\alpha) = c - (t-1)\chi(\alpha = 1);$$

б) *если $s = r\alpha + (1-r)t$ и при $\alpha = 1$ $c > t-1$, то*

$$\varlimsup_{n\to\infty} n^{t-\alpha} u_n \leqslant f,$$

где $f > 0$ — корень уравнения $\hat{c}(\alpha) f = d + af^r$; при $r = 1/2$ решением этого уравнения является

$$f = \left(\frac{a + \sqrt{a^2 + 4\hat{c}(\alpha) d}}{2\hat{c}(\alpha)}\right)^2;$$

в) *если* $s < r\alpha + (1-r)t$ *и при* $\alpha = 1$ $c > (s-\alpha)/(1-r)$, *то*

$$\varlimsup_{n\to\infty} n^{\frac{s-\alpha}{1-r}} u_n \leqslant \left(\frac{a}{c - \frac{s-\alpha}{1-r}\chi(\alpha=1)} \right)^{\frac{1}{1-r}}.$$

Доказательство. Используя неравенство

$$x^r \leqslant (1-r)x_0^r + \frac{r}{x_0^{1-r}}x \quad \forall x, x_0 > 0$$

для $x = u_n$ и $x_0 = fn^{-\rho}$, $f > 0, \rho \triangleq \min\{t-\alpha, (s-\alpha)/(1-r)\}$, получим для всех $n \geqslant n_0$

$$u_{n+1} \leqslant \left(1 - \frac{c+o(1)}{n^\alpha} + \frac{r(a+o(1))}{f^{1-r}n^{s-(1-r)\rho}}\right)u_n + \\ + \frac{d+o(1)}{n^t} + \frac{(1-r)(a+o(1))f^r}{n^{s+r\rho}}.$$

Теперь, применяя для каждого из трех случаев а), б), в) лемму П.1 (при $\alpha = 1$) и Чжуна (при $\alpha < 1$) [13, с. 227], получим неравенства вида

$$\varlimsup_{n\to\infty} n^\rho u_n \leqslant u(f),$$

где $u(f)$ — величина, зависящая, вообще говоря, от f. Беря инфинум $u(f)$ по тем значениям $f > 0$, для которых справедливо последнее неравенство, получим результат леммы. Лемма доказана. ▲

Лемма П.3. *Пусть* $a \in [0, 1)$, $\beta \in (0, 1)$. *Тогда справедливы неравенства*:

а) $$\prod_{k=1}^{\infty}(1 + a\beta^k) \geqslant \exp\left\{\frac{a\beta}{1-\beta}\left(1 - \frac{a\beta}{2(1+\beta)}\right)\right\},$$

б) $$\prod_{k=1}^{\infty}(1 - a\beta^k) \geqslant (1-a)^{-\ln^{-1}\beta},$$

в) $$\prod_{k=1}^{\infty}(1 - a\beta^{k-1})^k \geqslant (1-a)^{1-\ln^{-1}\beta+\ln^{-2}\beta}.$$

Доказательство:

а) пользуясь неравенством

$$1 + z \geqslant \exp\{z - z^2/2\} \quad \forall z \geqslant 0,$$

получаем

$$\prod_{k=1}^{\infty}(1 + a\beta^k) \geqslant \exp\left\{\sum_{k=1}^{\infty}(a\beta^k - a^2\beta^{2k}/2)\right\},$$

откуда следует требуемое неравенство;

б) из неравенства

$$1 - z \geqslant \exp\{z/(z-1)\} \quad \forall z \in [0, 1)$$

получаем

$$\prod_{k=1}^{\infty} (1 - a\beta^k) \geqslant \exp\left\{\sum_{k=1}^{\infty} a\beta^k/(a\beta^k - 1)\right\} \geqslant$$

$$\geqslant \exp\left\{\int_0^{\infty} \frac{a\beta^x}{a\beta^x - 1}\, dx\right\} = (1-a)^{-\ln^{-1}\beta};$$

в) используя неравенство из случая б) доказательства при $z \triangleq a\beta^{k-1}$, получаем

$$\prod_{k=1}^{\infty} (1 - a\beta^{k-1})^k = (1-a)\exp\left\{\sum_{k=2}^{\infty} k\ln(1 - a\beta^{k-1})\right\} \geqslant$$

$$\geqslant (1-a)\exp\left\{-\sum_{k=2}^{\infty} k\frac{a\beta^{k-1}}{1 - a\beta^{k-1}}\right\} \geqslant (1-a)\exp\left\{-\int_1^{\infty} x\frac{a\beta^{x-1}}{1 - a\beta^{x-1}}dx\right\} =$$

$$= (1-a)\exp\left\{-\int_0^{\infty} \frac{a\beta^x}{1 - a\beta^x}\, dx - \int_0^{\infty} x\frac{a\beta^x}{1 - a\beta^x}\, dx\right\}.$$

Интегрируя по частям и пользуясь вторично неравенством из случая б) доказательства при $z \triangleq a\beta^x$, получаем

$$\int_{-\infty}^{\infty} x\frac{a\beta^x}{1 - a\beta^x}\, dx = \ln^{-1}\beta\int_0^{\infty} \ln(1 - a\beta^x)\, dx \leqslant \ln^{-1}\beta\int_0^{\infty} \frac{a\beta^x}{a\beta^x - 1}\, dx.$$

Таким образом,

$$\prod_{k=1}^{\infty} (1 - a\beta^{k-1})^k \geqslant (1-a)\exp\left\{-(1 - \ln^{-1}\beta)\int_0^{\infty} \frac{a\beta^x}{1 - a\beta^x}\, dx\right\} =$$

$$= (1-a)\exp\left\{-(1 - \ln^{-1}\beta)\frac{\ln(1-a)}{\ln\beta}\right\} = (1-a)^{1 - \ln^{-1}\beta + \ln^{-2}\beta}.$$

Лемма доказана. ▲

Лемма П.4. *Для последовательностей* $\{\gamma_n\}$ *вида* $\gamma_n = \gamma(n+a)^{-1}$, $n = 1, 2, \ldots$ *справедливы соотношения:*

а) *при* $\gamma \in (0, 1)$ *и* $a > \gamma$

$$\left(\frac{1+a}{n + a - \gamma + 1}\right)^{\gamma} \geqslant \prod_{k=1}^{n} (1 - \gamma_k) \geqslant \left(\frac{a-\gamma}{n+a}\right)^{\gamma},$$

б) *при* $\gamma = 1$ *и* $a > 0$

$$\prod_{k=1}^{n} (1 - \gamma_k) = \frac{a}{n+a}.$$

Доказательство:
а) справедливы соотношения:

$$\exp\left\{\int_1^{n+1} \ln\left(1 - \frac{\gamma}{x+a}\right) dx\right\} \geqslant \prod_{k=1}^{n} (1 - \gamma_k) =$$

$$= \exp\left\{\sum_{k=1}^{n} \ln\left(1 - \frac{\gamma}{k+a}\right)\right\} \geqslant \exp\left\{\int_0^n \ln\left(1 - \frac{\gamma}{x+a}\right) dx\right\},$$

$$\int \ln\left(1 - \frac{\gamma}{x+a}\right) dx =$$

$$= (x + a - \gamma) \ln (x + a - \gamma) - (x + a) \ln (x + a) + C;$$

требуемые оценки вытекают из этих соотношений и выпуклости функции $x \ln x$ $(x > 0)$, для которой выполнено неравенство

$$(x + \Delta x) \ln (x + \Delta x) - x \ln x \geqslant \Delta x (1 + \ln x);$$

б) при $\gamma = 1$ выполняется рекуррентное соотношение

$$\frac{\gamma_n}{1 - \gamma_n} = \gamma_{n-1} \qquad (n \geqslant 2),$$

откуда следует равенство

$$\prod_{k=1}^{n} (1 - \gamma_k) = (1 - \gamma_1) \prod_{k=2}^{n} \frac{\gamma_k}{\gamma_{k-1}} = (1 - \gamma_1) \frac{\gamma_n}{\gamma_1}.$$

Лемма доказана. ▲

Лемма П.5. *Пусть последовательность неотрицательных чисел* u_n *для всех* $n \geqslant n_0$ *удовлетворяет рекуррентному неравенству*

$$u_{n+1} \leqslant (1 - \nu_n) u_n + \beta_n + \delta_n u_n^r,$$

в котором $\nu_n \in (0, 1]$, $r \in (0, 1)$, $\beta_n \geqslant 0$, $\delta_n \geqslant 0$,

$$\sum_{n=n_0}^{\infty} \nu_n = \infty, \quad \overline{\lim_{n\to\infty}}\, \beta_n/\nu_n \triangleq b, \quad \overline{\lim_{n\to\infty}}\, \delta_n/\nu_n \triangleq d.$$

Тогда

$$\overline{\lim_{n\to\infty}}\, u_n \leqslant \min_{c>r} u(c),$$

где

$$u(c) \triangleq \left(1 - \frac{r}{c}\right)^{-1} \left[b + (1 - r)\, c^{\frac{r}{1-r}} d^{\frac{1}{1-r}}\right].$$

Доказательство. Используя неравенство

$$x^r \leqslant (1-r)\, x_0^r + r x_0^{r-1} x \qquad \forall x,\, x_0 > 0$$

для $x = u_n$, $x_0 = (c\delta_n/\nu_n)^{\frac{1}{1-r}}$, $c > r$, при всех $n \geqslant n_0$ получим

$$u_{n+1} \leqslant [1 - \nu_n (1 - r/c)]\, u_n + \beta_n + (1-r)\,(c\delta_n/\nu_n)^{\frac{r}{1-r}} \delta_n,$$

или

$$u_{n+1} \leqslant (1 - \tilde{\nu}_n)\, u_n + \tilde{\beta}_n \qquad (n \geqslant n_0),$$

где

$$\tilde{\nu}_n \triangleq \nu_n (1 - r/c), \qquad \tilde{\beta}_n \triangleq \beta_n + (1-r)\,\delta_n\,(c\delta_n/\nu_n)^{\frac{r}{1-r}}.$$

Докажем теперь, что в условиях леммы

$$\overline{\lim_{n\to\infty}}\, u_n \leqslant \overline{\lim_{n\to\infty}}\, \tilde{\beta}_n/\tilde{\nu}_n \triangleq u(c).$$

Для любого $\varepsilon > 0$ найдется номер $n_1 \geqslant n_0$, для которого

$$\sup_{t \geqslant n_1} \tilde{\beta}_t/\tilde{\nu}_t \leqslant u(c) + \varepsilon,$$

поэтому при $n \geqslant n_1$

$$u_{n+1} \leqslant u_{n_1} \prod_{t=n_1}^{n} (1 - \tilde{\nu}_t) + \sum_{t=n_1}^{n} \tilde{\beta}_t \prod_{s=t+1}^{n} (1 - \tilde{\nu}_s) \leqslant$$

$$\leqslant u_{n_1} \prod_{t=n_1}^{n} (1 - \tilde{\nu}_t) + (u(c) + \varepsilon) \sum_{t=n_1}^{n} \tilde{\nu}_t \prod_{s=t+1}^{n} (1 - \tilde{\nu}_s).$$

Воспользовавшись тождеством

$$\prod_{t=n_1}^{n} (1 - \tilde{\nu}_t) + \sum_{t=n_1}^{n} \tilde{\nu}_t \prod_{s=t+1}^{n} (1 - \tilde{\nu}_s) = 1,$$

получим

$$u_{n+1} \leqslant u(c) + \varepsilon + (u_{n_1} - u(c) - \varepsilon) \prod_{t=n_1}^{n} (1 - \tilde{\nu}_t) \xrightarrow[n\to\infty]{} u(c) + \varepsilon.$$

В силу произвольности $c > r$ и $\varepsilon > 0$. Лемма доказана. ▲

Лемма П.6 [71]. *Пусть числовые последовательности* $\{\gamma_n\}$, $\{g_n\}$ *таковы, что* $0 < \gamma_{n+1} \leqslant \gamma_n$ $(n = 1, 2, \ldots)$ *и*

$$G \triangleq \sup_n \left| \sum_{t=1}^{n} g_t \right| < \infty.$$

Тогда справедлива оценка

$$\left| \sum_{t=1}^{n} g_t \gamma_t^{-1} \right| \leqslant 2G\gamma_n^{-1}, \quad n = 1, 2, \ldots$$

и, следовательно,

$$\lim_{n\to\infty} \frac{1}{h_n} \sum_{t=1}^{n} \frac{g_t}{\gamma_t} = 0,$$

если последовательность $\{h_n\}$ *положительных чисел удовлетворяет условию*

$$\lim_{n\to\infty} h_n \gamma_n = \infty.$$

Доказательство. Воспользуемся формулой суммирования по частям:

$$\sum_{t=1}^{n} \alpha_t \beta_t = \alpha_n \sum_{t=1}^{n} \beta_t - \sum_{t=1}^{n} (\alpha_t - \alpha_{t-1}) \sum_{\tau=1}^{t-1} \beta_\tau, \quad \alpha_0 = 0.$$

Положим $\alpha_t = 1/\gamma_t$, $\beta_t = g_t$ $(t = 1, 2, \ldots)$. Тогда при каждом $n = 1, 2, \ldots$ получим (считая $\gamma_0 = \infty$ и учитывая условия леммы)

$$\left| \sum_{t=1}^{n} \frac{g_t}{\gamma_t} \right| = \left| \frac{1}{\gamma_t} \sum_{t=1}^{n} g_t - \sum_{t=1}^{n} \left(\frac{1}{\gamma_t} - \frac{1}{\gamma_{t-1}} \right) \sum_{\tau=1}^{t-1} g_\tau \right| \leqslant$$

$$\leqslant \frac{1}{\gamma_n} \sup_n \left| \sum_{t=1}^{n} g_t \right| + \sum_{t=1}^{n} \left(\frac{1}{\gamma_t} - \frac{1}{\gamma_{t-1}} \right) \sup_m \left| \sum_{\tau=1}^{m} g_\tau \right| =$$

$$= \left(\sup_n \left| \sum_{t=1}^{n} g_t \right| \right) \frac{2}{\gamma_n},$$

откуда и следует требуемый результат. Лемма доказана. ▲

Лемма П.7 (Тёплиц [147]). *Пусть* $\{a_n\}$ — *последовательность неотрицательных чисел, такая, что для всех* $n \geqslant n_0$

$$0 < b_n \triangleq \sum_{t=1}^{n} a_t \xrightarrow[n\to\infty]{} \infty,$$

а $\{x_n\}$ — *последовательность чисел, сходящаяся к* x^*, *т. е.* $x_n \xrightarrow[n\to\infty]{} x^*$. *Тогда*

$$b_n^{-1} \sum_{t=1}^{n} a_t x_t \xrightarrow[n\to\infty]{} x^*.$$

Доказательство. Выберем $\varepsilon > 0$ и номер n_0, такой, что $|x_n - x^*| \leqslant \varepsilon$ для всех $n \geqslant n_0$. Тогда

$$\left| b_n^{-1} \sum_{t=1}^{n} a_t x_t - x^* \right| \leqslant b_n^{-1} \sum_{t=1}^{n} a_t \left| x_t - x^* \right| \leqslant$$

$$\leqslant b_n^{-1} \sum_{t=1}^{n_0-1} a_t \left| x_t - x^* \right| + b_n^{-1} \sum_{t=n_0}^{n} a_t \varepsilon \leqslant$$

$$\leqslant b_n^{-1} n_0 \max_{t=\overline{1,n_0}} a_t \left| x_t - x^* \right| + \varepsilon,$$

откуда из произвольности $\varepsilon > 0$ и того, что $b_n \xrightarrow[n\to\infty]{} \infty$, следует утверждение леммы. Лемма доказана. ▲

Следствие. *Если* $x_n \xrightarrow[n\to\infty]{} x^*$, *то* $n^{-1} \sum_{t=1}^{n} x_t \xrightarrow[n\to\infty]{} x^*$.

Этот результат вытекает из леммы П.7, если в ней принять $a_n = 1$ $(n = 1, 2, \ldots)$.

Лемма П.8 (Кронекер [147]). *Пусть* $\{b_n\}$ — *последовательность положительных чисел, такая, что*

$$b_n \xrightarrow[n\to\infty]{} \infty, \quad b_{n+1} \geqslant b_n \quad \forall n \geqslant 1,$$

а $\{x_n\}$ *последовательность чисел такая, что сходится ряд* $\sum_{n=1}^{\infty} x_n$. *Тогда*

$$b_n^{-1} \sum_{t=1}^{n} b_t x_t \xrightarrow[n\to\infty]{} 0.$$

Доказательство. Суммируя по частям, получим

$$b_n^{-1} \sum_{t=1}^{n} b_t x_t = S_t + b_n^{-1} \sum_{t=0}^{n-1} (b_t - b_{t+1}) S_t = S_t - b_n^{-1} \sum_{t=1}^{n-1} a_t S_t,$$

где $S_t \triangleq \sum_{m=1}^{t} x_m$, $S_0 = b_0 = 0$, $a_t = b_{t+1} - b_t \geqslant 0$. Поскольку $S_n \xrightarrow[n\to\infty]{} \sum_{t=1}^{\infty} x_t \triangleq S$,

$$b_n^{-1} \sum_{t=1}^{n-1} a_t S_t \xrightarrow[n\to\infty]{} S,$$

а значит,

$$S_n - b_n^{-1} \sum_{t=1}^{n-1} a_t S_t \xrightarrow[n\to\infty]{} 0.$$

Лемма доказана. ▲

Лемма П.9 (Роббинс, Сигмунд [104]). *Пусть* $\{\mathcal{F}_n\}$ — *поток σ-алгебр, а* x_n, α_n, β_n, ξ_n — $\mathcal{F}_n$-*измеримые неотрицательные случайные величины, причем при всех* $n = 1, 2, \ldots$ *существуют* $\mathsf{M}\{x_{n+1} | \mathcal{F}_n\}$ *и с вероятностью* 1 *удовлетворяют неравенству*

$$\mathsf{M}\{x_{n+1} \mid \mathcal{F}_n\} \leqslant x_n (1 + \alpha_n) + \beta_n - \xi_n.$$

Тогда для почти всех $\omega \in \Omega_0$, *где*

$$\Omega_0 \triangleq \left\{\omega \in \Omega \,\middle|\, \sum_{n=1}^{\infty} \alpha_n < \infty, \sum_{n=1}^{\infty} \beta_n < \infty\right\},$$

существует конечный предел $\lim_{n\to\infty} x_n = x^*(\omega)$ *и сходится ряд*

$$\sum_{n=1}^{\infty} \xi_n < \infty.$$

Доказательство. Введем величины

$$\tilde{x}_n \triangleq x_n \prod_{t=1}^{n-1} (1+\alpha_t)^{-1}, \quad \tilde{\beta}_n \triangleq \beta_n \prod_{t=1}^{n} (1+\alpha_t)^{-1},$$

$$\tilde{\xi}_n \triangleq \xi_n \prod_{t=1}^{n} (1+\alpha_t)^{-1}.$$

Тогда

$$\mathsf{M}\{\tilde{x}_{n+1} \mid \mathcal{F}_n\} = \mathsf{M}\{x_{n+1} \mid \mathcal{F}_n\} \prod_{t=1}^{n} (1+\alpha_t)^{-1} \leqslant \tilde{x}_n + \tilde{\beta}_n - \tilde{\xi}_n.$$

Определим

$$u_n \triangleq \tilde{x}_n - \sum_{t=1}^{n-1} (\tilde{\beta}_t - \tilde{\xi}_t),$$

для которого верно неравенство

$$\mathsf{M}\{u_{n+1} \mid \mathcal{F}_n\} \overset{\text{п.н.}}{\leqslant} u_n,$$

и случайный момент

$$\tau = \tau(a) \triangleq \inf \left\{ n \,\middle|\, \sum_{t=1}^{n} \tilde{\beta}_t \geqslant a \right\}, \quad a = \mathrm{const} > 0.$$

Рассмотрим «остановленный» процесс:

$$u_{\tau\wedge n} \triangleq u_\tau \chi(\tau < n) + u_n \chi(\tau \geqslant n),$$

который ограничен снизу, так как

$$u_{\tau\wedge n} \geqslant -\sum_{t=1}^{(\tau\wedge n)-1} \tilde{\beta}_t \geqslant -a$$

$$(\tau \wedge n = \tau\chi(\tau < n) + n\chi(\tau \geqslant n)).$$

Поскольку

$$\mathsf{M}\{u_{\tau\wedge(n+1)} \mid \mathcal{F}_n\} = u_\tau \chi(\tau \leqslant n) + \chi(\tau > n)\, \mathsf{M}\{u_{n+1} \mid \mathcal{F}_n\} \leqslant$$
$$\leqslant u_\tau \chi(\tau \leqslant n) + u_n \chi(\tau > n) = u_{\tau\wedge n},$$

то $(u_{\tau\wedge n}, \mathcal{F}_n)$ — ограниченный снизу супермартингал, а значит, существует конечный предел

$$\lim_{n\to\infty} u_{\tau\wedge n} \overset{\text{п.н.}}{=} u^*(\omega).$$

Так как $\beta_n \geqslant 0$ и для почти всех $\omega \in \Omega_0$ найдется a, при котором $\tau = \tau(a) = \infty$, то $u_{\tau\wedge n} = u_{\infty\wedge n} = u_n$, а следовательно,

$u_n \xrightarrow[n\to\infty]{\text{п.н.}} u^*(\omega)$ на множестве Ω_0, или, иначе,

$$\tilde{x}_n + \sum_{t=1}^{n-1} \tilde{\xi}_t \xrightarrow[n\to\infty]{\text{п.н.}} u^*(\omega) + \sum_{t=1}^{\infty} \tilde{\beta}_t.$$

Но так как $S_n \triangleq \sum_{t=1}^{n} \tilde{\xi}_t$ — монотонная, ограниченная сверху последовательность для почти всех $\omega \in \Omega_0$, то она имеет предел S_∞ на Ω_0, а значит, и $\tilde{x}_n$ имеет предел $\tilde{x}_\infty(\omega)$. А раз так, то

$$x_n \xrightarrow[n\to\infty]{} \tilde{x}_\infty(\omega) \prod_{t=1}^{\infty} (1+\alpha_t) \triangleq x^*(\omega)$$

и

$$\sum_{t=1}^{n-1} \xi_t = \sum_{t=1}^{n-1} \tilde{\xi}_t \prod_{m=1}^{t} (1+\alpha_m) \leqslant S_\infty \prod_{t=1}^{\infty} (1+\alpha_t) < \infty$$

для почти всех $\omega \in \Omega_0$. Лемма доказана. ▲

Лемма П.10. *Пусть последовательности неотрицательных случайных величин* $\{\eta_n\}$ *и* $\{\theta_n\}$ *согласованы с неубывающей последовательностью σ-алгебр* $\{\mathcal{F}_n\}$ *(т. е.* η_n *и* θ_n *измеримы относительно* $\mathcal{F}_n$*),* $\mathsf{M}\{\eta_1\} < \infty$ *и, кроме того,*

а) $$\sum_{t=1}^{\infty} \mathsf{M}\{\theta_t\} < \infty,$$

б) $$\mathsf{M}\{\eta_{n+1} \mid \mathcal{F}_n\} \overset{\text{п.н.}}{\leqslant} \eta_n + \theta_n, \quad n = 1, 2, \ldots$$

Тогда с вероятностью 1 *существует конечный предел*

$$\lim_{n\to\infty} \eta_n \overset{\text{п.н.}}{=} \eta.$$

Доказательство. Из условия а) и леммы Фату с вероятностью 1 имеем $\sum_{n=1}^{\infty} \theta_n < \infty$. Утверждение данной леммы вытекает из леммы П.9, если в ней положить $\alpha_n = 0$, $\beta_n = \theta_n$, $\xi_n = 0$. Лемма доказана. ▲

Лемма П.11. *Пусть в условиях леммы* П.10 *вместо* б) *последовательность* $\{\eta_n\}$ *удовлетворяет соотношению*

$$\mathsf{M}\{\eta_{n+1} \mid \mathcal{F}_n\} \overset{\text{п.н.}}{\leqslant} (1 - \lambda_{n+1} + \nu_{n+1})\,\eta_{n+1} + \theta_n, \quad n = 1, 2, \ldots,$$

где $\{\lambda_n\}$, $\{\nu_n\}$ *— последовательности неотрицательных чисел, для которых*

$$\sum_{n=1}^{\infty} \lambda_n = \infty, \quad \sum_{n=1}^{\infty} \nu_n < \infty.$$

Тогда

$$\lim_{n\to\infty} \eta_n \overset{\text{п.н.}}{=} 0.$$

Доказательство. Из леммы П.9 и сходимости ряда $\sum_{n=1}^{\infty} \theta_n \overset{\text{п.н.}}{<} \infty$ вытекает, что с вероятностью 1 существует предел $\lim_{n\to\infty} \eta_n = \eta^*(\omega)$ и, кроме того,

$$\sum_{n=1}^{\infty} \lambda_{n+1}\eta_n \overset{\text{п.н.}}{<} \infty.$$

Но из того, что $\sum_{n=1}^{\infty} \lambda_n = \infty$, следует существование подпоследовательности $\eta_{n_k} \xrightarrow[k\to\infty]{\text{п.н.}} 0$ и, следовательно, $\eta^*(\omega) \overset{\text{п.н.}}{=} 0$. Лемма доказана. ▲

Лемма П.12. *Пусть случайные последовательности* $\{\zeta_n\}$ *и* $\{\chi_n\}$ *таковы, что*

1) $\mathsf{M}\{\zeta_n\} \triangleq m, \quad \mathsf{M}\{\zeta_n^2\} \leqslant K < \infty, \quad n = 1, 2, \ldots;$

2) $\{\chi_n\}$ — *последовательность нулей и единиц;*

3) ζ_{n+1} *не зависит от совокупности* $\{\zeta_1, \ldots, \zeta_n, \chi_1, \ldots, \chi_{n+1}\}$ *при всех* $n = 1, 2, \ldots$

Тогда для почти всех

$$\omega \in A \triangleq \left\{\omega \,\middle|\, \sum_{n=1}^{\infty} \chi_n = \infty\right\}$$

$$\lim_{n\to\infty} \sum_{t=1}^{n} \chi_t\zeta_t \Big/ \sum_{t=1}^{n} \chi_t = m.$$

Доказательство. Определим λ_n и y_n следующим образом: если $\sum_{t=1}^{n} \chi_t > 0$, то

$$\lambda_n \triangleq \chi_n \Big/ \sum_{t=1}^{n} \chi_t, \quad y_t \triangleq \sum_{t=1}^{n} \chi_t\zeta_t \Big/ \sum_{t=1}^{n} \chi_t,$$

если $\sum_{t=1}^{n} \chi_t = 0$, то

$$\lambda_n \triangleq y_n = 0.$$

Введем σ-алгебру $\mathcal{F}_n$, порожденную совокупностью $\{\zeta_1, \ldots, \zeta_n, \chi_1, \ldots, \chi_{n+1}\}$ и положим $y_0 = 0$. Тогда

$$y_{n+1} = (1 - \lambda_{n+1}) y_n + \lambda_{n+1}\zeta_{n+1}.$$

Отсюда

$$\eta_{n+1} \triangleq (y_{n+1} - m)^2 = (1 - \lambda_{n+1})^2 \eta_n + \lambda_{n+1}^2 (\zeta_{n+1} - m)^2 + \\ + 2\lambda_{n+1}(1 - \lambda_{n+1})(y_n - m)(\zeta_{n+1} - m),$$

$$\mathsf{M}\{\eta_{n+1}|\mathcal{F}_n\} \leqslant (1 - \lambda_{n+1})\eta_n + (K - m^2)\lambda_{n+1}^2.$$

Так как $\lambda_n \in [0, 1]$ и

$$\sum_{t=1}^{\infty} \lambda_t^2 \leqslant \sum_{k=1}^{\infty} k^{-2} < \infty,$$

то последовательности $\{\eta_n\}$, $\{\lambda_n\}$ и $\{\mathscr{F}_n\}$ удовлетворяют условиям леммы П.9 и, следовательно, с вероятностью 1 существует конечный предел $\lim\limits_{n\to\infty} \eta_n = \eta(\omega)$. Из рекуррентности неравенства для $\{\eta_n\}$ и леммы Фату, по аналогии с доказательством леммы П.11, получаем с вероятностью 1

$$\sum_{n=1}^{\infty} \lambda_{n+1}\eta_n < \infty.$$

Для всех $\omega \in A$

$$\sum_{n=1}^{\infty} \lambda_n = \sum_{k=1}^{\infty} k^{-1} = \infty$$

и поэтому $\eta(\omega) = 0$. Лемма доказана. ▲

Лемма П.13. *Пусть $\{v_n\}$ — последовательность случайных величин, согласованных с неубывающей последовательностью σ-алгебр $\{\mathscr{F}_n\}$, т. е. $\mathscr{F}_n \subseteq \mathscr{F}_{n+1}$ и v_n измерима относительно $\mathscr{F}_n$ для любых $n = 1, 2, \ldots$ Если существуют $\mathsf{M}\{v_n \mid \mathscr{F}_{n-1}\}$ $(n=1, 2, \ldots)$ и*

$$\sum_{t=1}^{\infty} a_t^{-2}\mathsf{M}\left\{(v_t - \mathsf{M}\{v_t \mid \mathscr{F}_{t-1}\})^2 \mid \mathscr{F}_{t-1}\right\} \overset{\text{п.н.}}{<} \infty,$$

где $\{a_n\}$ — монотонно неубывающая неограниченная последовательность положительных чисел, то с вероятностью 1

$$\lim_{n\to\infty}\left(\frac{1}{a_n}\sum_{t=1}^{n} v_t - \frac{1}{a_n}\sum_{t=1}^{n}\mathsf{M}\{v_t \mid \mathscr{F}_{t-1}\}\right) = 0.$$

Доказательство. По лемме П.8 случайные величины

$$\frac{1}{a_n}\sum_{t=1}^{n}(v_t - \mathsf{M}\{v_t \mid \mathscr{F}_{t-1}\})$$

стремятся к нулю при $n \to \infty$ на множестве тех ω, где имеют предел величины

$$S_n \triangleq \sum_{t=1}^{n} a_t^{-1}(v_t - \mathsf{M}\{v_t \mid \mathscr{F}_{t-1}\}).$$

Но в условиях данной леммы этот предел существует с вероятностью 1 по лемме П.9, поскольку

$$\mathsf{M}\{S_n^2 \mid \mathscr{F}_{n-1}\} \overset{\text{п.н.}}{=} S_{n-1}^2 + a_n^{-2}\mathsf{M}\{(v_n - \mathsf{M}\{v_n \mid \mathscr{F}_{n-1}\})^2 \mid \mathscr{F}_{n-1}\}.$$

Лемма доказана. ▲

Следствие. *Если $\nu_n = \chi_n$, $n = 1, 2, \ldots,$ — случайные величины, принимающие значения 0, 1 и согласованные с неубывающей последовательностью σ-алгебр $\{\mathscr{F}_n\}$, то с вероятностью 1*

$$\lim_{n\to\infty}\left[\frac{1}{n}\sum_{t=1}^{n}\chi_t - \frac{1}{n}\sum_{t=1}^{n}\mathsf{P}\{\chi_t = 1 \mid \mathscr{F}_{t-1}\}\right] = 0.$$

Лемма П.13А. *Пусть $\{\nu_n\}$ — последовательность случайных величин, согласованных с неубывающей последовательностью σ-алгебр $\{\mathscr{F}_n\}$, т. е. $\mathscr{F}_n \subseteq \mathscr{F}_{n+1}$ и ν_n измерима относительно $\mathscr{F}_n$ для любых $n = 1, 2, \ldots$ Если существуют $\mathsf{M}\{\nu_n \mid \mathscr{F}_{n-1}\}$ $(n = 1, 2, \ldots)$ и при этом*

$$\sum_{t=1}^{\infty} t^{-2}\eta_t \mathsf{M}\{(\nu - \mathsf{M}\{\nu_t \mid \mathscr{F}_{t-1}\})^2 \mid \mathscr{F}_{t-1}\} \overset{\text{п.н.}}{<} \infty,$$

где $\{\eta_t\}$ — последовательность положительных чисел, удовлетворяющих условию

$$\lim_{n\to\infty}\left(\frac{\eta_n}{\eta_{n-1}} - 1\right)n = \lambda < 2,$$

то с вероятностью 1

$$\eta_n S_n^2 \xrightarrow[n\to\infty]{} 0, \quad S_n \triangleq \frac{1}{n}\sum_{t=1}^{n}(\nu_t - \mathsf{M}\{\nu_t \mid \mathscr{F}_{t-1}\}),$$

или, иначе,

$$S_n \overset{\text{п.н.}}{=} o\left(\eta_n^{-1/2}\right).$$

Доказательство. Имеем при больших n

$$\mathsf{M}\{S_n^2 \mid \mathscr{F}_{n-1}\} \overset{\text{п.н.}}{=} S_{n-1}^2\left(1 - \frac{1}{n}\right)^2 +$$
$$+ n^{-2}\mathsf{M}\{(\nu_n - \mathsf{M}\{\nu_n \mid \mathscr{F}_{n-1}\})^2 \mid \mathscr{F}_{n-1}\} = S_{n-1}^2\left(1 - \frac{2 + o(1)}{n}\right) +$$
$$+ n^{-2}\mathsf{M}\{(\nu_n - \mathsf{M}\{\nu_n \mid \mathscr{F}_{n-1}\})^2 \mid \mathscr{F}_{n-1}\},$$

откуда следует, что для $W_n \triangleq \eta_n S_n^2$ справедливо соотношение

$$\mathsf{M}\{W_n \mid \mathscr{F}_{n-1}\} \overset{\text{п.н.}}{=} W_{n-1}\left(1 - \frac{2 - \lambda + o(1)}{n}\right) +$$
$$+ \eta_n n^{-2}\mathsf{M}\{(\nu_n - \mathsf{M}\{\nu_n \mid \mathscr{F}_{n-1}\})^2 \mid \mathscr{F}_{n-1}\},$$

из которого в силу условий данной леммы и леммы П.9 вытекает, что $W_n \xrightarrow[n\to\infty]{\text{н.н.}} 0$. Лемма доказана. ▲

Следствие. *Если в условиях леммы* П.13А $\eta_n = O^*(n^\varkappa)$, *то* $\lambda = \varkappa$ *и, следовательно, при* $\varkappa < 2$

$$S_n \overset{\text{п.н.}}{=} o\left(n^{-\varkappa/2}\right).$$

Замечание. Если $\eta_n \equiv 1$, то лемма П.13А переходит в лемму П.13.

Лемма П.13Б. *Пусть* $\{\mathcal{F}_n\}$ — *поток* σ-*алгебр*, $\{u_n\}$ — *последовательность неотрицательных случайных величин, причем для всех* $n = 1, 2, \ldots$ *существуют* $\mathsf{M}\{u_n | \mathcal{F}_{n-1}\}$, u_n *измеримы относительно* $\mathcal{F}_n$ *и с вероятностью* 1 *справедливы неравенства*

$$\mathsf{M}\{u_{n+1} | \mathcal{F}_n\} \leqslant (1 - \alpha_n) u_n + \beta_n,$$

где $\{\alpha_n\}$, $\{\beta_n\}$ — *числовые последовательности, для которых*

$$\alpha_n \in (0, 1], \quad \beta_n \geqslant 0, \quad \nu_n > 0 \qquad (n = 1, 2, \ldots);$$

$$\sum_{n=1}^{\infty} \alpha_n = \infty, \quad \sum_{n=1}^{\infty} \beta_n \nu_{n+1} < \infty,$$

$$\lim_{n\to\infty} \frac{\nu_{n+1} - \nu_n}{\alpha_n \nu_n} \triangleq \mu < 1.$$

Тогда $u_n = o\left(\nu_n^{-1}\right)$ *с вероятностью* 1.

Доказательство. Определим $\tilde{u}_n = u_n \nu_n$. Тогда

$$\mathsf{M}\{\tilde{u}_{n+1} | \mathcal{F}_n\} \leqslant \tilde{u}_n (1 - \alpha_n) \nu_{n+1} \nu_n^{-1} + \nu_{n+1} \beta_n =$$
$$= \tilde{u}_n \left[1 - \alpha_n \left(1 - \mu + o(1)\right)\right] + \nu_{n+1} \beta_n,$$

откуда в силу леммы П.9 вытекает, что с вероятностью 1 $\tilde{u}_n \xrightarrow[n\to\infty]{} 0$, или $u_n = o\left(\nu_n^{-1}\right)$. Лемма доказана. ▲

П.2. Оператор проектирования π_ε^N на ε-симплекс

Пусть для любого $q \in R^N$ вектор $\pi_\varepsilon^N\{q\}$, принадлежащий ε-симплексу

$$S_\varepsilon^N = \left\{ p = (p_1, \ldots, p_N) \,\middle|\, \sum_{i=1}^{N} p_i = 1,\ p_i \geqslant \varepsilon \ \left(i = \overline{1; N}\right) \right\}$$

определяется условием

$$\left\| \pi_\varepsilon^N\{q\} - q \right\| = \min_{p \in S_\varepsilon^N} \| p - q \|, \tag{П2.1}$$

Для любого $q \in R^N$ вектор $\pi_\varepsilon^N\{q\}$ существует и единствен и $q = \pi_\varepsilon^N\{q\}$ тогда и только тогда, когда $q \in S_\varepsilon^N$. Обозначим π_0^N оператор π_ε^N при $\varepsilon = 0$.

Лемма П.14. *Для любых* $q \in R^N$ *и* $\varepsilon \in [0, 1/N)$ *справедливо равенство*

$$\pi_\varepsilon^N\{q\} = (1-\varepsilon N)\,\pi_0^N\left\{\frac{q-\varepsilon e^N}{1-\varepsilon N}\right\} + \varepsilon e^N,$$

где $e^N = (1, \ldots, 1) \in R^N$.

Доказательство. Любой вектор $\tilde{p} \in S_0^N$ может быть представлен в виде

$$\tilde{p} = \frac{p-\varepsilon e^N}{1-\varepsilon N},$$

где $p \in S_\varepsilon^N$. Поэтому для всех $q \in R^N$ и $p \in S_\varepsilon^N$

$$\left\| q - \left[(1-\varepsilon N)\,\pi_0^N\left\{\frac{q-\varepsilon e^N}{1-\varepsilon N}\right\} + \varepsilon e^N\right]\right\| = (1-\varepsilon N)\left\|\frac{q-\varepsilon e^N}{1-\varepsilon N} - \right.$$

$$\left. - \pi_0^N\left\{\frac{q-\varepsilon e^N}{1-\varepsilon N}\right\}\right\| \leqslant (1-\varepsilon N)\left\|\frac{q-\varepsilon e^N}{1-\varepsilon N} - \tilde{p}\right\| = \|q-p\|.$$

Лемма доказана. ▲

Лемма П.15. *Для любых* $q \in R^N$

$$\pi_0^N\{q\} = \pi_0^N\{\tilde{\pi}^N\{q\}\},$$

где $\tilde{\pi}^N$ — *оператор проектирования на гиперплоскость* $\tilde{S}^N \triangleq \triangleq \left\{q \mid q \in R^N,\ \sum_{i=1}^N q_i = 1\right\}$, *действующий по правилу*

$$\left\|\tilde{\pi}_0^N\{q\} - q\right\| = \min_{p \in \tilde{S}^N} \|p - q\|.$$

Доказательство следует из равенства

$$\|q-p\|^2 = \left\|q - \tilde{\pi}_0^N\{q\}\right\|^2 + \left\|\tilde{\pi}_0^N\{q\} - p\right\|^2,$$

справедливого при всех $q \in R^N$ и $p \in \tilde{S}^N$. Лемма доказана. ▲

Лемма П.16. *Для любых* $q \in R^N$

$$\tilde{\pi}^N\{q\} = q + N^{-1}(1 - \langle q, e^N\rangle)e^N.$$

Доказательство очевидно. ▲

Лемма П.17. *Пусть* $q \in \tilde{S}^N$ *и* $q_{i_\nu} < 0$ $(\nu = \overline{1, k})$, $q_i \geqslant 0$ $(i \neq i_\nu,$ $\nu = \overline{1, k})$. *Тогда* $\left(\pi_0^N\{q\}\right)_{i_\nu} = 0$, $\nu = \overline{1, k}$.

Доказательство. Обозначим $y = (y_1, \ldots, y_N) = \pi_0^N\{q\}$ и предположим противное, т. е. $\sum_{\nu=1}^k y_{i_\nu} > 0$. Тогда вектор $z = (z_1, \ldots$ $\ldots, z_N)$, у которого $z_{i_\nu} = 0$ и $z_i = y_i + (N-k)^{-1}\sum_{\nu=1}^k y_{i_\nu}\ (\nu = \overline{1, k},$

$i \neq i_\nu$) принадлежит симплексу S^N. Кроме того,

$$\|q - z\|^2 = \|q - y\|^2 - \sum_{\nu=1}^{k} y_{i_\nu}^2 + 2 \sum_{\nu=1}^{k} q_{i_\nu} y_{i_\nu} - \\ - (N-k)^{-1} \left(\sum_{\nu=1}^{k} y_{i_\nu} \right)^2 + 2 (N-k)^{-1} \left(\sum_{\nu=1}^{k} y_{i_\nu} \right) \left(\sum_{\nu=1}^{k} q_{i_\nu} \right) < \\ < \|q - y\|^2,$$

что противоречит определению π_0^N. Лемма доказана. ▲

Из леммы П.17 следует, что операция проектирования π_ε^N на ε-симплекс S_ε^N сводится к комбинации простого линейного преобразования и операции проектирования π_0^N на полный симплекс S^N. В свою очередь, оператор π_0^N, как следует из лемм П.15 — П.17, может быть реализован следующим алгоритмом:

шаг 0: $k = 0$, $q^0 = q$;

шаг i: проектируем вектор q^{i-1} на $\tilde{S}^{N-k}$ (лемма П.15); если $\tilde{\pi}^{N-k}\{q^{i-1}\} \in S^{N-k}$, то конец; если нет, то зануляем все отрицательные компоненты вектора $\tilde{\pi}^{N-k}\{q^{i-1}\}$ (пусть их k штук) и рассматриваем в качестве q^i укороченный вектор размерности $N-k$; переходим к шагу $i+1$.

Очевидно, при любом $q \in R^N$ этот алгоритм реализует оператор π_0^N за конечное число шагов, меньшее N. Он проще и экономичнее алгоритма, описанного в [90], так как не требует определения наименьшей отрицательной компоненты.

Ниже приводится подпрограмма на языке фортран-4 для ЭВМ, реализующая описанный способ проектирования на ε-симплекс S_ε^N.

```
      SUBROUTINE PROEK (Q, N, E, SQ1, RN, MON)
      DIMENSION Q(1), MON(1)
      R = E*RN
      AR = 1.—R
      AK = RN
      DO 1 I = 1, N
      MON (I) = 0
    1 Q(I) = (Q(I) — E)/AR
      SQ1 = SQ1/AR
   10 DO 2 I = 1, N
      IF (MON(I).EQ. 1) GO TO 2
      Q(I) = Q(I) + SQ1/AK
    2 CONTINUE
      DO 4 I = 1, N
      IF (MON(I). EQ. 1) GO TO 4
      IF (Q(I). GE.0.) GO TO 4
      SQ1 = Q(I)
      Q(I) = 0
      MON(I) = 1
      AK = AK — 1.
```

```
      GO TO 10
    4 CONTINUE
      DO 3 I = 1, N
    3 Q(I) = E + AR*Q(I)
      RETURN
      END
```

Входными переменными в подпрограмме PROEK являются
Q — вектор (Q_1, ..., Q_N), проектируемый на ε-симплекс;
N — размерность вектора Q;
E — величина $\varepsilon \in [0, N^{-1})$;

SQ1 — величина, равная $1 - \sum_{i=1}^{N} Q_i$;

RN — вспомогательная переменная (типа REAL), равная по величине N,

MON — рабочий массив (типа INTEGER) размерности N.

П.3. Подпрограммы для ЭВМ на языке фортран-4, реализующие алгоритмы адаптивного выбора вариантов

Каждая из приводимых ниже подпрограмм при одном обращении реализует очередной шаг соответствующего алгоритма, т. е. осуществляет рандомизированный выбор варианта и вычисляет «новые» значения рекуррентно пересчитываемых условных вероятностей выбора и, возможно, других величин. Работа с этими подпрограммами требует использования также подпрограмм-функций, которые вычисляют значения потерь, соответствующих выбранному варианту. Эти подпрограммы-функции должны, естественно, прилагаться пользователем. Исключение из сказанного выше составляет подпрограмма STRAF, реализующая алгоритм (3.26), построенный по методу штрафных функций: при одном обращении к этой подпрограмме осуществляется два акта выбора вариантов, в результате чего производится рекуррентный пересчет соответствующих векторов. Отметим также, что приводимые подпрограммы содержат обращения к другим подпрограммам, а именно: к PROEK (см. П.2) и PROJCT, AUTOM и RANDU (см. ниже).

1. Программа NARSHA.

Назначение: реализует один шаг алгоритма Нарендры — Шапиро (2.15). Использует подпрограммы AUTOM и POTERI.

Обращение:

CALL NARSHA (P, N, N1, GAM, CSI)

Описание параметров:

P — одномерный массив (P(1), ..., P(N)); элемент P(i) равен при обращении к NARSHA значению условной вероятности $p_n(i)$ выбора варианта $x(i)$ на n-м шаге, а в результате обращения — условной вероятности $p_{n+1}(i)$ выбора $x(i)$ на следующем шаге алгоритма (2.15); должны выполняться условия нормировки:

$$\sum_{i=1}^{N} p(i) = 1, \quad p(i) > 0 \quad \forall i = \overline{1, N};$$

N — число возможных вариантов, или размер массива P;
N1 = N — 1;
GAM — длина шага γ_n в (2.15);
CSI — потери ξ_n в (2.15), переменная типа INTEGER.

Примечание: используемая подпрограмма-функция POTERI имеет тип INTEGER и вычисляет реализацию потерь ξ_n, принимающих значения 0 или 1; содержит один целочисленный параметр NV, равный номеру выбранного варианта от 1 до N.

```
  SUBROUTINE NARSHA (P, N, N1, GAM, CSI)
  DIMENSION P(1)
  INTEGER M/1/, CSI, POTERI
  CALL AUTOM (M, P, N, N1, NV)
  CSI = POTERI (NV)
  IF (CSI.EQ.1) RETURN
  DO 1 I = 1, N
  IF (I.EQ.NV) GO TO 2
  P(I) = P(I) — GAM*P(I)
  GO TO 1
2 P(I) = P(I) + GAM*(1.—P(I))
1 CONTINUE
  RETURN
  END
```

2. Подпрограмма LUCE.

Назначение: реализует один шаг алгоритма Льюса (2.22). Обращается к подпрограммам AUTOM, POTERI.

Обращение:

```
CALL LUCE (P, N, N1, GAM, A, CSI)
```

Описание параметров:

P, N, N1, GAM, CSI — имеют тот же смысл в отношении алгоритма (2.22), что и соответствующие параметры подпрограммы NARSHA, A — параметр a в (2.22).

Примечание: используемая подпрограмма-функция POTERI имеет тот же смысл и содержит тот же параметр, что и для подпрограммы NARSHA.

```
  SUBROUTINE LUCE (P, N, N1, GAM, A, CSI)
  DIMENSION P(1)
  INTEGER M/1/, CSI, POTERI
  CALL AUTOM (M, P, N, N1, NV)
  CSI = POTERI (NV)
  IF (CSI.EQ.1) GO TO 3
  X = GAM/(A + (1.—A)*P(NV))
  GO TO 4
3 X = —GAM/(1.—(1.—A)*P(NV))
4 DO 1 I = 1, N
  IF (I.EQ.NV) GO TO 2
  P(I) = P(I) — P(NV)*P(I)*X
  GO TO 1
2 P(I) = P(I) + P(NV)*(1.— P(I))*X
1 CONTINUE
  RETURN
  END
```

3. Подпрограмма VARVOR.

Назначение: реализует один шаг алгоритма Варшавского — Воронцовой (2.43). Использует подпрограммы AUTOM, POTERI.

Обращение:

CALL VARVOR (P, N, N1, GAM, CSI)

Описание параметров: все параметры имеют тот же смысл в отношении алгоритма Варшавского — Воронцовой, что и соответствующие параметры подпрограммы NARSHA.

Примечание: см. примечание к подпрограмме LUCE.

```
  SUBROUTINE VARVOR (P, N, N1, GAM, CSI)
  DIMENSION P(1)
  INTEGER M/1/, CSI, POTERI
  CALL AUTOM (M, P, N, N1, NV)
  CSI = POTERI (NV)
  IF (CSI.EQ.1) GO TO 3
  X = GAM
  GO TO 4
3 X = —GAM
4 DO 1 I = 1, N
  IF (I.EQ.NV) GO TO 2
  P(I) = P(I) — P(NV)*X*P(I)
  GO TO 1
2 P(I) = P(I) + P(NV)*(1.— P(I))*X
1 CONTINUE
  RETURN
  END
```

4. Подпрограмма BUMOST.

Назначение: реализует один шаг алгоритма Буша — Мостеллера (2.44). Обращается к подпрограммам AUTOM, POTERI.

Обращение:

CALL BUMOST (P, N, N1, GAM, CSI)

Описание параметров: все параметры имеют тот же смысл в отношении алгоритма (2.44), что и соответствующие параметры подпрограммы NARSHA.

Примечание: см. примечание к подпрограмме LUCE.

```
  SUBROUTINE BUMOST (P, N, N1, GAM, CSI)
  DIMENSION P(1)
  INTEGER M/1/, CSI, POTERI
  CALL AUTOM (M, P, N, N1, NV)
  CSI = POTERI (NV)
  X = CSI
  X = X/N1
  DO 1 I = 1, N
  IF (I.EQ.NV) GO TO 2
  P(I) = P(I) + GAM*(P(I) — X)
  GO TO 1
2 P(I) = P(I) + GAM*(1.— P(I) — CSI)
1 CONTINUE
  RETURN
  END
```

5. Подпрограмма STAPPR.

Назначение: реализует один шаг проекционного алгоритма стохастической аппроксимации (2.47). Использует подпрограммы AUTOM, POTERI, PROEK.

Обращение:

CALL STAPPR (P, N, N1, RN, GAM, EPS, DEL, CSI, MON)

Описание параметров:

P, N, N1, CSI имеют тот же смысл в отношении алгоритма (2.47), что и соответствующие параметры подпрограммы NARSHA;

RN = N;

EPS — параметр ε_{n+1} в (2.47);

DEL — параметр Δ в (2.47);

CSI — переменная типа REAL;

MON — целочисленный рабочий массив размера N.

Примечание: используемая подпрограмма-функция POTERI имеет тип REAL и вычисляет реализацию потерь ξ_n, принимающую любое действительное значение; содержит один целочисленный параметр NV, равный номеру выбранного варианта от 1 до N.

```
SUBROUTINE STAPPR (P, N, N1, RN, GAM, EPS, DEL, CSI, MON)
DIMENSION P(1), MON(1)
INTEGER M/1/
CALL AUTOM (M, P, N, N1, NV)
CSI = POTERI (NV)
S = (CSI — DEL)*GAM/P(NV)
P(NV) = P(NV) — S
CALL PROEK (P, N, EPS, S, RN, MON)
RETURN
END
```

6. Подпрограмма POISK.

Назначение: реализует один шаг алгоритма случайного поиска (2.70). Обращается к подпрограммам AUTOM, POTERI, RANDU, PROEK.

Обращение:

```
CALL POISK (P, PP, N, N1, RN, GAM, EPS, EPS1, CSI, Q,
* MON)
```

Описание параметров:

P, N, N1, RN, GAM, CSI, MON имеют тот же смысл в отношении алгоритма (2.70), что и соответствующие параметры подпрограммы STAPPR;

EPS, EPS1 — параметры ε_n, ε_{n+1} алгоритма (2.70) соответственно;

Q — рабочий массив (Q(1), ..., Q(N)) размера N.

Примечание: см. примечание к подпрограмме STAPPR.

```
SUBROUTINE POISK (P, PP, N, N1, RN, GAM, EPS, EPS1, CSI,
* Q, MON)
DIMENSION P(1), PP(1), Q(1), MON(1)
INTEGER M/1/
CALL AUTOM (M, P, N, N1, NV)
CSI = POTERI (NV)
X = CSI*GAM
```

```
   SQ = 0.
   T = -1./RN
   R = (1.-T)/2.
   T = T/2.
   DO 11 I = 1, N
   CALL RANDU (M, J, Q(I))
   M = J
 1 Q(I) = 2.*Q(I) - 1.
   DO 11 I = 1, N
   S = 0.
   DO 13 J = 1, N
   IF (J.EQ.I) GO TO 12
   S = S + Q(J)*T
   GO TO 11
12 S = S + R*Q(J)
13 CONTINUE
   P(I) = PP(I) + EPS*S
11 PP(I) = PP(I) - Q(I)*X
   CALL PROEK (PP, N, EPS1, SQ, RN, MON)
   RETURN
   END
```

7. Подпрограмма IERAR.

Назначение: реализует один шаг двухуровневых иерархических алгоритмов (2.83) без передачи информации и (2.98) с передачей информации нижнему уровню. Обращается к подпрограммам AUTOM, POTERI, PROEK.

Обращение: CALL IERAR (IALG, P, N, N1, RN, PP, NN, NN1,
* RNN, MON, G1, E1, D1, G2, E2, D2, Q, CSI)

Описание параметров:

IALG — переменная, значение которой указывает, какой алгоритм реализуется:

IALG = 1 — алгоритм без передачи информации;

IALG = 2 — алгоритм с передачей информации нижнему уровню; если при обращении к IERAR переменная IALG имеет значение, отличное от 1 или 2, то на печать выдается текст «STOP IN SUBR. IERAR: IALG = » с указанием значения переменной IALG, а затем выполняется оператор STOP;

P — одномерный массив (P(1), ..., P(N)); элемент P(i) равен при обращении к IERAR значению условной вероятности $p_n(i)$ выбора i-й группы вариантов на n-м шаге, а в результате обращения — условной вероятности $p_{n+1}(i)$ выбора на следующем шаге одного из алгоритмов (2.83) или (2.98);

N — параметр N_1 в (2.76), или размер массива P;

N1 = N — 1;

RN = N;

PP — двумерный массив, описываемый в предложении DIMENSION как PP (N, NN); если первый индекс имеет значение I = i ($i = \overline{1, N_1}$), то вектор (PP(i, 1), ..., PP(i, NN)) равен при обращении к IERAR значению p_n^i, а после обращения — значению p_{n+1}^i в (2.83) или (2.98);

NN — параметр N_2 в (2.76), размерность векторов (PP(i, 1), ..., PP(i, NN));

NN1 = NN — 1;
RNN = NN;

G1, E1, D1, G2, E2 — параметры соответственно γ_n', ε_{n+1}', Δ, γ_n'', ε_{n+1}'' в (2.83) или (2.98);

D2 — одномерный массив размерности N, i-я компонента которого D2(i) равна значению параметра Δ_i в (2.83) или (2.98);

Q — рабочий одномерный массив размера N;

CSI — потери ξ_n в (2.83) или (2.98).

Примечание: используемая подпрограмма-функция POTERI имеет тип REAL и вычисляет реализацию потерь ξ_n, принимающую любое действительное значение; содержит два целочисленных параметра, определяющих выбранный вариант: первый параметр NV задает номер группы вариантов, среди которых осуществляется выбор (на втором этапе), принимает значения от 1 до N, а второй параметр NVV задает номер варианта в этой группе, принимает значения от 1 до NN.

```
      SUBROUTINE IERAR (IALG, P, N, N1, RN, PP, NN, NN1, RNN
     *, MON, G1, E1, D1, G2, E2, D2, Q, CSI)
      DIMENSION P(1), PP(1, 1), Q(1), D2(1), MON(1)
      INTEGER M/1/
      CALL AUTOM (M, P, N, N1, NV)
      DO 10 I = 1, NN
   10 Q(I) = PP(NV, I)
      CALL AUTOM (M, Q, NN, NN1, NVV)
      CSI = POTERI (NV, NVV)
      S = G2*(CSI — D2(NV))/Q(NVV)
      IF (IALG.EQ.1) GO TO 8
      S = S/P(NV)
      IF(IALG. EQ.2) GO TO 8
      PRINT 90, IALG
   90 FORMAT (/2X,' STOP IN SUBR. IERAR : IALG = ', I7//)
      STOP
    8 Q(NVV) = Q(NVV) — S
      CALL PROEK (Q, NN, E2, S, RNN, MON)
      DO 11 I = 1, NN
   11 PP(NV, I) = Q(I)
      S = G1*(CSI — D1)/P(NV)
      P(NV) = P(NV) — S
      CALL PROEK (P, N, E1, S, RN, MON)
      RETURN
      END
```

8. Подпрограмма STRAF.

Назначение: реализует два шага алгоритма (3.26), построенного по методу штрафных функций. Использует подпрограммы AUTOM, POTERI, PROEK, POTOGR.

Обращение:

```
      CALL STRAF (P, N, N1, RN, GP, EP, UM, D, U, GU, MON
     *     , M, CSI, CS1, CS2)
```

Описание параметров:

P — одномерный массив (P(1), ..., P(N)); элемент P(i) равен при обращении к STRAF значению условной вероятности $p_n(i)$

выбора варианта $x(i)$ на $(n-1)$-м и n-м шагах, а в результате обращения — условной вероятности $p_n(i)$ выбора $x(i)$ на следующих двух шагах алгоритма (3.26);

N — число возможных вариантов, размер массива P;

N1 = N — 1; RN = N;

MON — целочисленный рабочий массив размера N;

GP, EP, UM, D, GU — соответственно параметры γ_n^p, ε_{n+1}, μ_n, δ_n, γ_n^u в алгоритме (3.26);

U — одномерный массив размера M, равный при обращении к STRAF значению вектора $u_n = (u_n(1), \ldots, u_n(m))$, а после обращения — значению u_{n+1} в (3.26);

M — число m ограничений-неравенств в (3.3), размер массива U;

CS1, CS2 — потери соответственно ξ_{n-1}^0, ξ_n^0 в алгоритме (3.26);

CSI, CSJ — одномерные массивы размера M = m, равные значениям векторов $\left(\xi_{n-1}^1, \ldots, \xi_{n-1}^m\right)$, $\left(\xi_n^1, \ldots, \xi_n^m\right)$ соответственно в (3.26).

Примечания: 1) используемая подпрограмма-функция POTERI имеет тот же смысл и содержит тот же параметр, что и для подпрограммы STAPPR; 2) подпрограмма-функция POTOGR имеет два целочисленных параметра: первый указывает номер NV выбранного варианта, а второй — номер I ограничения-неравенства; вычисляет реализацию потерь ξ_{n-1}^i или ξ_n^i, соответствующих выбранному варианту и ограничению-неравенству с номером I = i; 3) величины потерь, вычисляемых подпрограммами-функциями POTERI и POTOGR принимают любые действительные значения.

```
      SUBROUTINE STRAF (P, N, N1, RN, GP, EP, UM, D, U, GU, MON
     *, M, CSI, CSJ, CS1, CS2)
      DIMENSION P(1), U(1), CS1(1), CSJ(1), MON(1)
      INTEGER K/1/
      CALL AUTOM (K, P, N, N1, NV1)
      CS1 = POTERI (NV1)
      DO 1 I = 1, M
    1 CSI(I) = POTOGR (NV1, I)
      CALL AUTOM (K, P, N, N1, NV2)
      CS2 = POTERI (NV2)
      DO 2 I = 1, M
    2 CSJ(I) = POTOGR (NV2, I)
      T1 = UM* CS1 + D
      T2 = UM*CS2 + D
      DO 3 I = 1, M
      T1 = T1 + (CSJ(I) + U(I))*CSI(I)/P(NV1)
    3 T2 = T2 + (CSI(I) + U(I))*CSJ(I)/P(NV2)
      T1 = T1*GP/2.
      T2 = T2*GP/2.
      S = T1 + T2
      P(NV1) = P(NV1) - T1
      P(NV2) = P(NV2) - T2
      CALL PROEK (P, N, EP, S, RN, MON)
      DO 4 I = 1, M
```

```
      U(I) = U(I) - GU*(U(I) + (CS1 + CS2)/2.)
      IF(U(I).LT.0.) U(I) = 0.
    4 CONTINUE
      RETURN
      END
```

9. Подпрограмма LAGRAN.

Назначение: реализует один шаг алгоритма (3.54), построенного по методу множителей Лагранжа. Обращается к подпрограммам AUTOM, POTERI, POTOGR, PROEK.

Обращение:

```
      CALL, LAGRAN (P, N, N1, RN, GAM, EPS, D, DEL, MON
     *, ALAM, DM, M, CSM, CS)
```

Описание параметров:

P, N, N1, RN, MON — имеют тот же смысл в отношении алгоритма (3.54), что и соответствующие параметры подпрограммы STRAF;

GAM, EPS, D, DEL — соответственно параметры γ_n, ε_{n+1}, Δ_0, δ_n в алгоритме (3.54);

ALAM — одномерный массив размера $M = m$, равный при обращении к LAGRAN значению вектора $y_n = (y_n(1), \ldots, y_n(m))$, а после обращения — значению y_{n+1} в алгоритме (3.54);

DM — одномерный массив размера $M = m$, j-й элемент которого $DM(j)$ равен значению параметра Δ_j в алгоритме (3.54);

M — число ограничений-неравенств в (3.3), размер массивов ALAM и DM;

CSM — одномерный массив размера $M = m$, равный значению вектора потерь $\left(\xi_n^1, \ldots, \xi_n^m\right)$ в (3.54);

CS — потери ξ_n^0 в (3.54).

Примечание: используемые подпрограммы-функции POTERI и POTOGR имеют тот же смысл и содержат те же параметры, что и для подпрограммы STRAF.

```
      SUBROUTINE LAGRAN (P, N, N1, RN, GAM, EPS, D, DEL, MON
     *, ALAM, DM, M, CSM, CS)
      DIMENSION P(1), DM(1), ALAM(1), CSM(1), MON(1)
      INTEGER K/1/
      CALL AUTOM (K, P, N, N1, NV)
      CS = POTERI (NV)
      DO 1 I = 1, M
    1 CSM(I) = POTOGR(NV, I)
      S = CS - D
      DO 2 I = 1, M
    2 S = S + ALAM(I)*(CSM(I) - DM(I))
      S = GAM*(S/P(NV) + DEL)
      P(NV) = P(NV) - S
      CALL PROEK (P, N, EPS, S, RN, MON)
      DO 3 I = 1, M
      ALAM(I) = ALAM(I) + GAM*(CSM(I) - DEL*ALAM(I))
      IF (ALAM(I).LT.0.) ALAM(I) = 0.
    3 CONTINUE
      RETURN
      END
```

10. Подпрограмма IGRAN.

Назначение: реализует один шаг алгоритма (4.39) адаптивного выбора вариантов в игре многих лиц. Использует подпрограммы AUTOM, POTERI, PROEK.

Обращение:

CALL IGRAN (P, L, N, N1, RN, G, E, D, CSI, Q, NV, MON, MN)

Описание параметров:

P — двумерный массив, описываемый в предложении DIMENSION как P(L, MN); если первый индекс имеет значение k, то вектор $(P(k, 1), \ldots, P(k, N_k))$ равен при обращении к IGRAN стохастическому вектору p_n^k, а после обращения — стохастическому вектору p_{n+1}^k в (4.39);

L — число игроков l в (4.39);

N — одномерный массив размера L, k-й элемент которого N(k) равен значению параметра N_k;

N1, RN — одномерные массивы размера L; N1(i) = N(i) — 1; RN(i) = N(i);

G, E, D — параметры соответственно γ_n, ε_{n+1}, δ_n в алгоритме (4.39);

CSI — одномерный массив размера L, k-й элемент которого CSI(k) равен реализации потерь k-го игрока ξ_n^k в (4.39);

Q, NV, MON — рабочие массивы, размер которых равен наибольшему элементу массива N.

Примечание: используемая подпрограмма-функция POTERI содержит два параметра: первый — целочисленный массив NV, k-й элемент которого NV(k) задает номер выбранного k-м игроком элемента x_n^k, а второй — номер I игрока, потери которого и вычисляются при текущем обращении к IGRAN. Переменная MN должна быть равна наибольшему элементу массива N.

```
      SUBROUTINE IGRAN (P, L, N, N1, RN, G, E, D, CSI, Q, NV,
     * MON, B)
      DIMENSION P (L, B), N(1), N1(1), RN(1), NV(1), Q(1), CSI(1)
     * , MON(1)
      INTEGER K/1/, B
      DO 7 I = 1, L
      NI = N(I)
      DO 1 J = 1, NI
    1 Q(J) = P(I, J)
      NI1 = N1(I)
      CALL AUTOM (K, Q, N1, NI1, NV(I))
    7 CONTINUE
      DO 2 I = 1, L
      NI = N(I)
      RNI = RN(I)
      CSI(I) = POTERI (NV, I)
      IV = NV(I)
      DO 5 J = 1, NI
    5 Q(J) = P(I, J)
      S = G*(CSI(I)/Q(IV) + D)
      Q(IV) = Q(IV) - S
```

```
      CALL PROEK (Q, NI, E, S, RNI, MON)
      DO 3 J = 1, NI
    3 P(I, J) = Q(J)
    2 CONTINUE
      RETURN
      END
```

11. Подпрограмма UMTSEP.

Назначение: реализует один шаг алгоритма (5.60)—(5.64) адаптивного управления марковской цепью. Обращается к подпрограммам AUTOM, POTERI, PROJCT, TSEP.

Обращение:

```
      CALL UMTSEP (C, K, N, D, GAM, EPS, DNK, CSI, Z, PIT,
     *  S, A, P, N1)
```

Описание параметров:

C — двумерный массив, описываемый в предложении DIMENSION как C(K, N); элемент C(i, l) равен при обращении значению c_k^{il} в (5.61), а в результате — либо c_{k+1}^{il}, если обращение соответствует моменту времени $n = n_{k+1}$, либо по-прежнему c_k^{il}, если $n_k < n < n_{k+1}$;

K — число состояний управляемой марковской цепи;

N — число возможных вариантов (значений управления);

D — двумерный массив, описываемый в предложении DIMENSION как D(K, N); элемент D(i, l) в результате обращения равен условной вероятности d_n^{il} (5.63) выбора варианта с номером l на n-м шаге, если цепь находится в состоянии с номером i;

GAM, EPS — параметры γ_k, ε_{k+1} в (5.61) соответственно;

DNK — переменная типа INTEGER; равна $n_{k+1} - n_k + 1$, если $n_k \leqslant n < n_{k+1}$;

CSI — потери ξ_n в (5.62);

Z — переменная типа INTEGER; при обращении равна номеру состояния z_n, а в результате — номеру z_{n+1} следующего состояния управляемой марковской цепи;

PIT — трехмерный массив, описываемый в предложении DIMENSION как PIT (N, K, K); элемент PIT (l, i, j) равен при обращении значению $(\widehat{\pi}_{ij}^l)_n$, а в результате обращения — значению $(\widehat{\pi}_{ij})_{n+1}^l$ оценки вероятности π_{ij}^l перехода марковской цепи из состояния $z(i)$ в состояние $z(j)$ при управлении $x(l)$, вычисляемой рекуррентным алгоритмом (5.60);

S — целочисленный двумерный массив, описываемый в предложении DIMENSION как S(K, N); элемент S(i, l) равен при обращении значению s_n^{il}, а в результате — значению s_{n+1}^{il} в (5.60); при первом обращении к UMTSEP все элементы должны быть равны нулю;

A — двумерный массив, описываемый в предложении DIMENSION как A(K, N); элемент A(i, l) равен при обращении значению A_n^{il}, а в результате — значению A_{n+1}^{il} в (5.62); при первом обращении все элементы должны быть определены, например, нулями;

P — одномерный рабочий массив (P(1), ..., P(N)) размера N;

N1 = N — 1.

Примечания:

1) используемая подпрограмма-функция POTERI вычисляет реализацию потерь ξ_n в (5.62); содержит два параметра: первый указывает номер состояния марковской цепи, а второй — номер выбранного варианта (управления);

2) подпрограмма TSEP, как и POTERI должна прилагаться пользователем; обращение к ней следующее:

```
CALL TSEP (Z, NV, IZ)
```

Она вычисляет номер последующего состояния марковской цепи, если Z — номер ее текущего состояния, а NV — номер выбранного варианта (управления);

3) подпрограмма PROJCT, обращение к которой имеет вид

```
CALL PROJCT (C, K, N, PIT, EPS)
```

реализует оператор проектирования $\pi^{\widehat{C}}_{\varepsilon_{k+1}}$ матрицы, заданной массивом C, на множество $\widehat{C}_{\varepsilon_{k+1}}$ (5.64); ее параметры K, N, PIT, EPS те же, что и в подпрограмме UMTSEP, а параметр C — двумерный массив (описываемый в предложении DIMENSION как C (K, N)) равен при обращении к PROJCT матрице, которую надо спроектировать, а после обращения — результату проектирования; эта подпрограмма должна быть приложена пользователем; поскольку операция проектирования на множество $\widehat{C}_{\varepsilon_{k+1}}$ (5.64) представляет собой решение задачи квадратичного программирования

$$\sum_{i=1}^{K}\sum_{l=1}^{N}\left(C(i,\,l)-x^{il}\right)^2 \to \min_{\substack{\|x^{il}\| \\ i=\overline{1,\,K} \\ l=\overline{1,\,N}}}$$

$$\sum_{i=1}^{K}\sum_{l=1}^{N} x^{il}=1,$$

$$\sum_{l=1}^{N} x^{il}=\sum_{j=1}^{K}\sum_{l=1}^{N}\mathrm{PIT}\,(l,\,i,\,j)\,x^{jl}\quad\left(i=\overline{1,\,K}\right),$$

$$x^{il}\geqslant \mathrm{EPS}\qquad\left(i=\overline{1,\,K},\ l=\overline{1,\,N}\right),$$

которая, следовательно, может быть решена за конечное число вычислительных операций, то при написании PROJCT естественно использовать какую-либо стандартную программу решения таких задач (см., например, [50], с. 149—153).

```
      SUBROUTINE UMTSEP (C, K, N, D, GAM, EPS, DNK, CSI, Z,
     * PIT, S, A, P, N1)
      DIMENSION C (K, N), D (K, N), PIT (N, K, K), S (K, N),
     * A, (K, N), P(1)
      INTEGER DNK, S, TK/0/, Z, M/1/
      TK = TK + 1
      IF (TK.LT.DNK) GO TO 10
      IF (TK.EQ.DNK) GO TO 11
```

```
101 FORMAT (//2X, 'STOP IN SUBR. UMTSEP: TK > DNK: TK = '
   *, I7, 5X, 'DNK = ', I7/)
    PRINT 101, TK, DNK
    STOP
 11 TK = 1
    DO 1 I = 1, K
    DO 1 J = 1, N
  1 C(I, J) = C(I, J) — GAM*A(I, J)
    CALL PROJCT (C, K, N, PIT, EPS)
    DO 2 I = 1, K
    R = 0.
    DO 3 J = 1, N
  3 R = R + C(I, J)
    DO 2 J = 1, N
  2 D(I, J) = C(I, J)/R
 10 DO 4 J = 1, N
  4 P(J) = D(Z, J)
    CALL AUTOM (M, P, N, N1, NV)
    CSI = POTERI (Z, NV)
    R = 1./TK
    CALL TSEP (Z, NV, IZ)
    S(Z, NV) = S(Z, NV) + 1
    DO 5 I = 1, K
    DO 5 J = 1, N
    A(I, J) = (1.— R)*A(I, J)
    IF (I.EQ.Z.AND.J.EQ.NV) A(I, J) = A(I, J) + CSI*R/C(I, J)
  5 CONTINUE
    R = 1./S(Z, NV)
    DO 6 I = 1, K
    PIT(NV, Z, I) = PIT(NV, Z, I)*(1.— R)
    IF (I.EQ.IZ) PIT(NV, Z, I) = PIT(NV, Z, I) + R
  6 CONTINUE
    Z = IZ
    RETURN
    END
```

12. Подпрограмма AUTOM.

Назначение: вычисление целых случайных чисел между 1 и N > 1 с заданным распределением вероятностей. Во время работы один раз обращается к подпрограмме RANDU.

Обращение:

CALL AUTOM (M, P, N, N1, NC)

Описание параметров:

M — целочисленная переменная, используемая в качестве первого параметра (IX) при обращении к подпрограмме RANDU;

P — одномерный массив P(1), ..., P(N); задает распределение вероятностей целочисленной случайной величины NC, принимающей значение I с вероятностью P(I); должны выполняться условия нормировки:

$$\sum_{i=1}^{N} P(i) = 1, \quad P(i) \geqslant 0 \quad \forall i = \overline{1, N},$$

N — количество возможных значений случайной величины, размер массива P;

N1 = N — 1;

NC — полученное в результате целое случайное число между 1 и N, принимающее значение I с вероятностью P(I).

```
  SUBROUTINE AUTOM (M, P, N, N1, NC)
  DIMENSION P(1)
  NC = N
  CALL RANDU (M, J, X)
  M = J
  A = -X
  DO 1 I = 1, N1
  A = A + P(I)
  IF (A) 1, 2, 2
1 CONTINUE
  RETURN
2 NC = I
  RETURN
  END
```

13. Подпрограмма RANDU ([113]).

Назначение: вычисление равномерно распределенных случайных действительных чисел на отрезке [0, 1] и случайных целых чисел между 0 и 2^{31}. Для каждого обращения в качестве входа служит целое случайное число и образуется новое целое число и вещественное случайное число. Эта программа специально предназначена для системы IBM/360 и образует 2^{29} значений перед началом повторений.

Обращение:

```
CALL RANDU (IX, IY, YFL)
```

Описание параметров:

IX — при первом обращении IX должно содержать нечетное целое положительное число с количеством цифр, не большим девяти. После первого (и каждого последующего) обращения IX должно быть равно значению IY, вычисленному подпрограммой при предыдущем обращении (т. е. после каждого обращения следует предусмотреть оператор присваивания IX = IY);

IY — полученное в результате целое случайное число, требуемое при последующих обращениях к подпрограмме; оно находится между 0 и 2^{31};

YFL — полученное в результате равномерно распределенное случайное число на отрезке [0, 1], представленное в форме с плавающей точкой;

```
  SUBROUTINE RANDU (IX, IY, YFL)
  IY = IX*65539
  IF(IY) 5, 6, 6
5 IY = IY + 2147483647 + 1
6 YFL = YFL*.4656613E-9
  RETURN
  END
```

СПИСОК ЛИТЕРАТУРЫ

1. Айзерман М. А., Браверман Э. М., Розоноэр Л. И. Метод потенциальных функций в теории обучения машин.— М.: Наука, 1970.
2. Алберт А. (Albert A.) Регрессия, псевдоинверсия и рекуррентное оценивание: Пер. с англ.— М.: Наука, 1977.
3. Асо, Кимура (Aso H., Kimura M.) Absolute expediency of learning automata.— Information Sciences, 1979, v. 17, p. 91—112.
4. Афанасьев Г. К., Коновалов Е. В. и др. Организация взаимодействия программ в управляющей машине для коммутационной системы.— В кн.: Автоматы и управление сетями связи, М.: Наука, 1971, с. 26—34.
5. Баба, Саварaги (Baba N., Savaragi Y.) On the learning behavior of stochastic automata under a nonstationary random environment.— IEEE Trans. Systems, Man and Cybernetics, 1975, v. SMC — 5, №. 3, p. 273—275.
6. Бакушинский А. Б., Поляк Б. Т. О решении вариационных неравенств.— ДАН СССР, 1974, т. 219, № 5, с. 1038—1041.
7. Беленький В. З., Волконский В. А. и др. Итеративные методы в теории игр и программировании.— М.: Наука, 1974.
8. Бертсекас (Bertsekas D. P.) Dynamic programming and stochastic control.— N. Y.: Academic, 1976.
9. Боркар, Варайя (Borkar V., Varaiya P.) Identification and adaptive control of Marcov chains.— SIAM J. Control and Optimization, 1982, v. 20, № 4, p. 470—489.
10. Браун (Brown G. W.) Iterative solution of games by fictions play. Activity Analysis of Production and Allocation/Ed. by Koopman T. S.— N. Y.: Wiley, 1951.
11. Бутковский А. Г., Гаврилов Г. А., Красненкер А. С., Позняк А. С., Цыпкин Я. З. Некоторые проблемы технической кибернетики.— В кн.: Кибернетику — на службу коммунизму, М.: Наука, 1977, с. 58—108.
12. Буш, Мостеллер (Bush R. R., Mosteller F.) Stochastic models for learning.— N. Y.: Wiley, 1958.
13. Вазан М. (Wasan M. T.) Стохастическая аппроксимация: Пер. с англ.— М.: Мир, 1972.
14. Варшавский В. И. Коллективное поведение автоматов.— М.: Наука, 1973.
15. Варшавский В. И., Воронцова И. П. О поведении стохастических автоматов с переменной структурой.— АиТ, 1963, т. 24, № 3, с. 353—360.

16. Варшавский В. И., Воронцова И. П. Стохастические автоматы с переменной структурой.— В кн.: Теория конечных и вероятностных автоматов. Труды ИФАК, М.: Наука, 1965, с. 301—309.
17. Висванатан, Нарендра (Viswanathan R., Narendra K. S.) Games of stochastic automata.— IEEE Transactions on systems, man and cybernetics, 1974, v. 4, p. 131—135.
18. Воронцова И. П. Алгоритмы изменения переходных вероятностей стохастических автоматов.— ППИ, 1965, т. 1, вып. 3, с. 122—126.
19. Гаркави Н. Г. Некоторые вопросы теории производительности бульдозеров.— Строительное и дорожное машиностроение, 1960, № 10, с. 16—19.
20. Гёссель М., Сраговичь В. Г. Адаптивное управление марковскими цепями с доходами.— ДАН СССР, 1980, т. 254, № 3, с. 523—526.
21. Гольштейн Е. Г. Обобщенный градиентный метод отыскания седловых точек.— Экономика и мат. методы, 1972, т. VIII, вып. 4, с. 563—579.
22. Гудвин, Син (Goodwin G. C., Sin K. S.) Adaptive filtering prediction and control.— Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1984.
23. Гурвич Е. Т. Метод асимптотического исследования игр автоматов.— АиТ, 1975, № 2, с. 80—94.
24. Денардо (Denardo E. V.) A marcovian decision problem.— In: Mathematical Programming/Eds. T. C. Hu and S. M. Robinson, N. Y.: Academic, 1973.
25. Деревицкий Д. П., Фрадков А. Л. Прикладная теория дискретных адаптивных систем управления.— М.: Наука, 1981.
26. Дерман (Derman C.) Finite state marcovian decision processes.— N. Y.: Academic, 1970.
27. Доши, Шрив (Doshi B., Shreve S. E.) Strong consistency of a modified maximum likelihood estimators for controlled Marcov chains.— Adv. Appl. Prob., 1980, v. 12, p. 313.
28. Дуб Дж. (Doob J.) Вероятностные процессы.— М.: Иностр. лит., 1956.
29. Дынкин Е. Б. Марковские процессы.— М.: Физматгиз, 1963.
30. Дынкин Е. Б., Юшкевич А. А. Управляемые марковские процессы и их применения.— М.: Наука, 1975.
31. Ермольев Ю. М. Методы стохастического программирования.— М.: Наука, 1976.
32. Ершов А. А., Позняк А. С. Об играх обучающихся стохастических автоматов.— В кн.: Адаптивные системы. Вопросы кибернетики, М.: Научный совет по комплексной проблеме «кибернетика» АН СССР, 1974, с. 133—139.
33. Зангвилл У. И. (Zangwill W. I.) Нелинейное программирование: Пер. с англ.— М.: Сов. радио, 1973.
34. Исбелл (Isbell J. R.) On a problem of Robbins.— Annals of Mathematical Statistics, 1959, v. 30, p. 606—610.
35. Исследования по теории самонастраивающихся систем/Под ред. Сраговича В. Г.— М.: ВЦ АН СССР, 1971.
36. Кавер, Хеллман (Cover T. M., Hellman M. E.) The two-armed bandit problem with time invariant finite memory.—

IEEE Trans. on Information Theory, 1970, v. 16, № 2, p. 185—195.
37. Каплинский А. И., Красненкер А. С., Цыпкин Я. З. Рандомизация и сглаживание в задачах и алгоритмах адаптации.— АиТ, 1974, № 6, с. 47—57.
38. Карманов В. Г. Математическое программирование.— М.: Наука, 1975.
39. Катковник В. Я. Линейные оценки и стохастические задачи оптимизации.— М.: Наука, 1976.
40. Кашнер, Тхатхачар, Лакшмиварахан (Kushner H. J., Thathachar M. A. L., Lakshmivarahan S.) Two state automaton — a counter example. Dec. 1970. Appeared in «A note on linear reinforcement scheme for variable structure stochastic automata» by Viswanathan R. and Narendra K. S.— IEEE Trans. on Systems, Man and Cybernetics, 1972, v. 2, p. 292—294.
41. Кемени Дж., Снелл Дж. (Kemeny J. G., Snell J. L.) Конечные цепи Маркова: Пер. с англ.— М.: Наука, 1970.
42. Кифер, Вольфовиц (Kiefer E., Wolfowitz J.) Stochastic estimation of the maximum or a regression function.— Annals of the Mathematical Statistics, 23 (1952), p. 462—466.
43. Коновалов М. Г. Об адаптивном управлении некоторыми классами марковских цепей.— ДАН СССР, 1977, т. 233, № 5, с. 780—783.
44. Коновалов М. Г. Адаптивное управление периодическими процессами с независимыми значениями.— Изв. АН СССР, Техническая кибернетика, 1979, № 1, с. 138—144.
45. Корпелевич Г. М. Об одном итерационном методе решения матричных игр.— В кн.: Проблемы оптимального планирования. Часть I, М.: ЦЭМИ АН СССР, 1973, с. 90—100.
46. Кринский В. И. Асимптотически оптимальный автомат с экспоненциальной скоростью сходимости.— Биофизика, 1964, т. 9, вып. 4, с. 484—487.
47. Крылов В. Ю. Об одном стохастическом автомате асимптотически оптимальном в случайной среде.— АиТ, 1963, т. 24, № 9, с. 1226—1228.
48. Кумар, Беккер (Kumar P. R., Becker A.) A new family of optimal adaptive controllers for Marcov chains.— IEEE Trans. on Automatic Control, 1982, v. AC-27, Feb., p. 137—145.
49. Кумар, Лин (Kumar P. R., Lin W.) Optimal Adaptive controllers for unknown Marcov chains.— IEEE Trans. on Automatic Control, 1982, v. AC-27, № 4, p. 765—774.
50. Кюнци, Шах, Зендер (Künzi H. P., Tzschach H. G., Zehnder C. A.) Numerical methods of mathematical optimization with Algol and Fortran programs.— N. Y. and Lnd.: Academic Press, 1968.
51. Лазарев В. Г., Саввин Г. Г. Сети связи, управление и коммутация.— М.: Связь, 1973.
52. Лазарев В. Г., Соловьев А. В. Адаптивный способ диспетчеризации программ управляющей машины узла коммутации сети связи.— В кн.: Сети связи и дискретные устройства управления, М.: Наука, 1976, с. 3—11.
53. Лакшмиварахан (Lakshmivarahan S.) Learning algorithms theory and applications.— N. Y.; Heidelberg; Berlin; Springer — Verlag, 1981.

54. Лакшмиварахан, Тхатхачар (Lakshmivarahan S., Thathachar M. A. L.) Absolutely expedient learing algorithms for stochastic automata.— IEEE Trans. on Systems, Man and Cybernetics, 1973, v. 3, p. 281—286.
55. Лакшмиварахан, Тхатхачар (Lakshmivarahan S., Thathachar M. A. L.) Absolute expediency of Q- and S-model learning algorithms.— IEEE Trans. on Systems, Man and Cybernetics, 1976, v. 6, p. 222—226.
56. Лакшмиварахан, Тхатхачар (Lakshmivarahan S., Thathachar M. A. L.) Bounds of the probability of convergence of learning automata.— IEEE Trans. on Systems, Man and Cybernetics, 1976, v. 6, p. 756—763.
57. Липцер Р. Ш., Ширяев А. Н. Статистика случайных процессов.— М.: Наука, 1970.
58. Лоэв М. (Loeve M.) Теория вероятностей: Пер. с англ.— М.: Иностр. лит., 1962.
59. Льюс (Luce R. D.) Individual choice behavior.— N. Y.: Willey, 1959.
60. Льюс Р., Райфа Х. (Luce R. D., Raiffa H.) Игры и решения: Пер. с англ.— М.: Иностр. лит., 1961.
61. Любчик Л. М., Позняк А. С. Обучающиеся автоматы в задачах управления стохастическими объектами.— АиТ, 1974, № 5, с. 95—109.
62. Майн Х., Осаки С. (Mine H., Osaki S.) Марковские процессы принятия решений: Пер. с англ.— М.: Наука, 1977.
63. Майстровский Г. Д. О градиентных методах отыскания седловых точек.— Экономика и мат. методы, 1976, т. XII, вып. 5, с. 917—929.
64. Мандл (Mandl P.) Estimation and control in Marcov chains.— Adv. Appl. Prob., 1974, v. 6, p. 40—60.
65. Месарович М., Мако Д., Такахара И. (Mesarovič M., Macko D., Takahara Y.) Теория иерархических многоуровневых систем: Пер. с англ.— М.: Мир, 1973.
66. Назин А. В. Адаптивные автоматные алгоритмы решения задачи дискретной стохастической оптимизации.— В кн.: Материалы 1-й школы-семинара по структурной адаптации, Воронеж, ВПИ, 1977, с. 77—81.
67. Назин А. В. Об игровом подходе к решению одной задачи стохастического программирования.— АиТ, 1978, № 2, с. 72—82.
68. Назин А. В. О классе адаптивных автоматных алгоритмов, минимизирующих в асимптотике средний эмпирический штраф.— В кн.: Адаптация в сложных системах управления, Воронеж: ВПИ, 1978, с. 13—16.
69. Назин А. В. О скорости сходимости и выборе параметров автоматного алгоритма.— АиТ, 1982, № 7, с. 70—80.
70. Назин А. В. О повышении эффективности автоматных алгоритмов адаптивного выбора вариантов.— В кн.: Адаптация и обучение в системах управления и принятия решений, Новосибирск: Наука, 1982, с. 40—46.
71. Назин А. В. Игровая задача адаптивного выбора вариантов и алгоритм ее решения.— АиТ, 1983, № 11, с. 70—75.
72. Назин А. В., Позняк А. С. Адаптивные линейные последовательностные машины.— АиТ, 1975, № 12, с. 114—126.

73. Назин А. В., Позняк А. С. О стохастической игре двух автоматов с нулевой суммой.— АиТ, 1977, № 1, с. 53—61.
74. Назин А. В., Позняк А. С. Матричные игры N лиц в условиях неопределенности.— Экономика и мат. методы, 1978, т. XIV, вып. 5, с. 958—968.
75. Назин А. В., Позняк А. С. О сходимости автоматного алгоритма Нарендры — Шапиро.— АиТ, 1983, № 2, с. 171—174.
76. Назин А. В., Позняк А. С. О сходимости автоматных алгоритмов Льюса и Варшавского — Воронцовой.— АиТ, 1983, № 9, с. 165—169.
77. Назин А. В., Соловьев А. В. Исследование адаптивного алгоритма диспетчеризации программ электронной управляющей машины узла коммутации.— В кн.: Управление на сетях и узлах связи, М.: Наука, 1979, с. 83—87.
78. Нарендра, Тхатхачар (Narendra K. S., Thathachar M. A. L.) Learning automata — a servey.— IEEE Trans. on Systems, Man and Cybernetics. 1974, v. 4, p. 323—334.
79. Невельсон М. Б., Хасьминский Р. З. Стохастическая аппроксимация и рекуррентное оценивание.— М.: Наука, 1972.
80. Нейман Дж. фон, Моргенштерн О. (Neumann J. von, Morgenshtern O.) Теория игр и экономическое поведение: Пер. с англ.— М.: Наука, 1970.
81. Никайдо Х., Исода К. (Nikaido H., Isoda K.) Заметка о бескоалиционных выпуклых играх: Пер. с англ.— В кн.: Бесконечные антагонистические игры, М.: Физматгиз, 1963.
82. Норман (Norman M. F.) Marcov processes and learning models.— N. Y.: Academic Press, 1972.
83. Нэш (Nash J. E.) Equillibrium points in n-person games.— Proc. Nat. Acad. Sci. of USA, 1950, v. 36.
84. Нэш Дж. (Nash J. E.) Бескоалиционные игры: Пер. с англ.— В кн.: Матричные игры, М.: Физматгиз, 1961, с. 205—221.
85. Оуэн Г. (Owen G.) Теория игр: Пер. с англ.— М.: Мир, 1971.
86. Партхасаратхи Т., Рагхаван Т. (Parthasarathy T., Raghavan T. E. S.) Некоторые вопросы теории игр двух лиц: Пер. с англ.— М.: Мир, 1974.
87. Питерсон У. (Peterson W.) Коды, исправляющие ошибки: Пер. с англ.— М.: Мир, 1964.
88. Питерсон У., Уэлдон Э. (Peterson W., Weldon E.) Коды, исправляющие ошибки: Пер. с англ.— М.: Мир, 1976.
89. Позняк А. С. Обучающиеся автоматы в задачах стохастического программирования.— АиТ, 1973, № 10, с. 84—96.
90. Позняк А. С. Исследование сходимости алгоритмов функционирования обучающихся стохастических автоматов.— АиТ, 1975, № 1, с. 88—103.
91. Поляк Б. Т. Введение в оптимизацию.— М.: Наука, 1983.
92. Поляк Б. Т., Третьяков Н. В. Метод штрафных оценок для задач на условный экстремум.— ЖВМ и МФ, 1973, т. 13, № 1, с. 34—46.
93. Поляк Б. Т., Цыпкин Я. З. Псевдоградиентные алгоритмы адаптации и обучения.— АиТ, 1973, № 3, с. 45—68.
94. Пономарев В. А. Об одной конструкции конечного автомата, асимптотически оптимального в стационарной среде.— Биофизика, 1964, т. 9, вып. 1, с. 104—110.
95. Поспелов Д. А. Игры и автоматы.— М.; Л.: Энергия, 1966.

96. Поспелов Д. А. Вероятностные автоматы.— М.: Энергия, 1970.
97. Пресман Э. Л., Сонин И. М. Последовательное управление.— М.: Наука, 1982.
98. Прохоров Ю. В., Розанов Ю. А. Теория вероятностей.— М.: Наука, 1973.
99. Радченко А. Н. О связи теории корректирующих кодов с проблемами самообучения и надежности.— В кн.: Кибернетику — на службу коммунизму, т. 3, М.; Л.: Энергия, 1966, с. 87—114.
100. Риордан (Riordan J. S.) An adaptive automaton controller for discrete time Marcov processes.— Automatica, 1969, v. 5, p. 721—730.
101. Роббинс (Robbins H.)Some aspects of the sequential design of experiments.— Bulletin of American Mathematical Society, 1952, v. 58, № 5, p. 527—535.
102. Роббинс (Robbins H.) A sequential decision problem with finite memory.— Proc. National Academy of Science, 1956, v. 42, № 3, p. 920—923.
103. Робинс, Монро (Robbins H., Monro S.) A stochastic approximation method.— Ann. Math. Statist., 1951, v. 22, № 1, p. 400—407.
104. Роббинс, Сигмунд (Robbins H., Siegmund D.) A convergence theorem for non negative almost supermartingales and some application. Optimizing methods in statistics. Edited by Jagadish S. Rustagi. Proceedings of a symposium held at the Center for Tomorrow the Ohio State University, June 14—16, 1971, p. 233—257.
105. Робинсон Дж. (Robinson J.) Итеративный метод решения игр: Пер. с англ.— В кн.: Матричные игры, М.: Физматгиз, 1961.
106. Розанов Ю. А. Случайные процессы (краткий курс).— М.: Наука, 1971.
107. Розен (Rosen J. B.) Existance and uniqueness of equillibrium points in concave n-person games.— Econometrica, 1965, v. 33, № 3, p. 520—534.
108. Росс (Ross S. M.) Applied probability models with optimization applications.— San Francisco: Holden Day, 1970.
109. Рыжков Ю. И. Управление запасами.— М.: Наука, 1969.
110. Саварагн, Баба (Savaragi Y., Baba N.) Two ε-optimal nonlinear reinforcement schemes for stochastic automata.— IEEE Trans. on Systems, Man and Cybernetics, 1974, January, p. 126—130.
111. Самюэлс (Samuels S. M.) Randomized rules for one two-armed bandit problem with finite memory.— Annals of Mathematical Statistics, 1968, v. 39, № 6, p. 2103—2107.
112. Сарымсаков Т. А. Основы теории процессов Маркова.— М.: Гостехиздат, 1954.
113. Сборник научных программ на Фортране. Вып. 1. Статистика: Пер. с англ.— М.: Статистика, 1974.
114. Смит, Пайк (Smith C. V., Pyke R.) The Robbins — Isbell two-armed bandit problem with finite memory.— Annals of Mathematical Statistics, 1965, v. 36, p. 1375—1386.
115. Соловьев А. В. Некоторые пути повышения эффективности работы управляющей машины узла коммутации.— В кн.:

Построение устройств управления сетями связи, М.: Наука, 1977, с. 21—26.

116. Срагович В. Г. Автоматы с многозначным входом и их поведение в случайных средах.— В кн.: Исследования по теории самонастраивающихся систем. М.: ВЦ АН СССР, 1971, с. 19—65.
117. Срагович В. Г. Теория адаптивных систем.— М.: Наука, 1976.
118. Срагович В. Г. Адаптивное управление.— М.: Наука, 1981.
119. Срагович В. Г., Флеров Ю. А. Построение класса оптимальных автоматов.— ДАН СССР, т. 159, № 6, 1964, с. 1236—1237.
120. Срагович В. Г., Флеров Ю. А. Об одном классе стохастических автоматов.— Изв. АН СССР, Техническая кибернетика, 1965, № 2, с. 66—73.
121. Стратонович Р. Л. Принципы адаптивного приема.— М.: Сов. радио, 1973.
122. Тхатхачар, Лакшмиварахан (Thathachar M. A. L., Lakshmivarahan S.) Application of learning algorithms to hypothesis testing problems.— In: Proc. IEEE Conference on Decision and Control. San Diego: 1973, p. 194—198.
123. Тихонов А. Н. Об устойчивости задач оптимизации функционалов.— ЖВМ и МФ, 1966, т. 6, № 4, с. 631—634.
124. Флеров Ю. А. О некоторых классах многовходовых автоматов.— В кн.: Исследования по теории самонастраивающихся систем, М.: ВЦ АН СССР, 1971, с. 96—110.
125. Флеров Ю. А. Обзор исследований по адаптивным системам в Вычислительном центре АН СССР.— В кн.: Вопросы кибернетики. Адаптивные системы, М.: Научный совет по комплексной проблеме «Кибернетика» АН СССР, 1974, с. 38—45.
126. Фомин В. Н., Фрадков А. Л., Якубович В. А. Адаптивное управление динамическими объектами.— М.: Наука, 1981.
127. Харкевич А. А. Борьба с помехами.— М.: Наука, 1965.
128. Хедли Дж., Уайтин Т. (Hadley G., Whitin T. M.) Анализ систем управления запасами: Пер. с англ.— М.: Наука, 1969.
129. Херкенрат, Калин, Лакшмиварахан (Herkenrath U., Kalin D., Lakshmivarahan S.) On a general class of absorbing-barrier learning algorithms.— Information Sciences, 1981, v. 24, p. 255—263.
130. Херкенрат, Теодореску (Herkenrat U., Theodorescu R.) On a stochastic approximation procedure applied to the bandit problem.— Elektronische Informationsverarbeitung und Kybernetic (EIK), 1979, v. 15, № 5/6, p. 301—307.
131. Ховард Р. (Howard R. A.) Динамическое программирование и марковские процессы: Пер. с англ.— М.: Сов. радио, 1964.
132. Цетлин М. Л. О поведении конечных автоматов в случайных средах.— АиТ, 1961, т. 22, № 10, с. 1345—1354.
133. Цетлин М. Л. Исследования по теории автоматов и моделированию биологических систем.— М.: Наука, 1969.
134. Цыпкин Я. З. Адаптация и обучение в автоматических системах.— М.: Наука, 1968.
135. Цыпкин Я. З. Основы теории обучающихся систем.— М.: Наука, 1970.

136. Цыпкин Я. З. Адаптивные методы выбора решений в условиях неопределенности.— АиТ, 1976, № 4, с. 78—91.
137. Цыпкин Я. З., Позняк А. С. Обучающиеся конечные автоматы.— Техническая кибернетика, 1972, № 3, с. 127—140.
138. Цыпкин, Позняк (Tsypkin Ya. Z., Poznyak A. S.) Learning automata.— Journal of Cybernetics and Information Science. Special Issue on Learning Automata, 1977, ASC, v. 1, № 2, THRU 4, p. 128—160.
139. Цыпкин Я. З., Позняк А. С. Рекуррентные алгоритмы оптимизации в условиях неопределенности.— В кн.: Итоги науки и техники. Сер. Техническая кибернетика, М.: ВИНИТИ, 1983, т. 16, с. 3—70.
140. Чандрасекаран, Лакшманан (Chandrasekaran B., Lakshmanan K. B.) Multiple hypothesis testing with finite memory.— Journal of Cybernetics and Information Science. Special Issue on Learning Automata, 1977, v. 1, № 2, THRU 4, p. 128—160.
141. Чандрасекаран, Шен (Chandrasekaran B., Shen D. W. C.) Stochastic automata games.— IEEE Trans. Syst. Sci. Cybern., 1969, v. SSC-5, p. 145—149.
142. Чжун (Chung K. L.) On stochastic approximation method.— Ann. Math. Stat., 1954, v. 25, № 3, p. 463—483.
143. Шапиро, Нарендра (Shapiro I. J., Narendra K. S.) Use of stochastic automata for parameter self optimization with multimodal performance criteria.— IEEE Trans. SMC-5, 1969, v. 5, № 4, p. 352—361.
144. Шахгильдян В. В., Лохвицкий М. С. Методы адаптивного приема сигналов.— М.: Связь, 1974.
145. Шеннон, Уивер (Shannon C. E., Weaver W.) Mathematical theory of communication.— Urbana: Illinois press Univ., 1949.
146. Шеннон (Shannon C.) Математическая теория связи: Пер. с англ.— В кн.: Работы по теории информации и кибернетике, М.: Иностр. лит., 1963, с. 243—332.
147. Ширяев А. Н. Вероятность.— М.: Наука, 1980.
148. Эль-Фаттах (El-Fattah Y. M.) Recursive algorithms for adaptive control of finite Marcov chains.— IEEE Trans. on System, Man and Cybernetics, 1981, v. SMC-11, № 2, p. 135—144.
149. Эль-Фаттах (El-Fattah Y. M.) Gradient approach for recursive estimation and control in finite Markov chains.— Adv. Appl. Prob., 1981, v. 13, № 4, p. 778—803.
150. Эрроу К. Дж., Гурвиц Л., Удзава Х. (Arrow K. J., Hurwics L., Uzawa H.) Исследования по линейному и нелинейному программированию: Пер. с англ.— М.: Иностр. лит., 1962.