Math-Net.Ru

Общероссийский математический портал

Ю. Хинц, Н. Яп, Алгоритмы для оптимального управления стохастических систем переключающегося типа, *Теория вероятн. и ее примен.*, 2015, том 60, выпуск 4, 770–800

DOI: <https://doi.org/10.4213/tvp5034>

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением
<http://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 54.92.164.108

26 апреля 2023 г., 12:41:22

© 2015 г.

ХИНЦ Ю.*, ЯП Н.*

**АЛГОРИТМЫ ДЛЯ ОПТИМАЛЬНОГО УПРАВЛЕНИЯ
СТОХАСТИЧЕСКИХ СИСТЕМ
ПЕРЕКЛЮЧАЮЩЕГОСЯ ТИПА**

(Перевод Муравлева А. Н.)

Для задач оптимального управления переключающегося типа с линейной динамикой состояния представлен новый алгоритмический метод построения решения, который требует минимальных предположений, связанных с выпуклостью, демонстрируя при этом отличную вычислительную эффективность, в том числе, на задачах большой размерности. Кроме того, разработаны прямые двойственные процедуры для оценки расстояния от построенных приближенных решений до оптимальных.

Ключевые слова и фразы: оптимальное управление, американские опционы, потраекторное управление, дуальность.

1. Введение. Основная трудность при принятии решений в условиях неопределенности — понять, каким образом следует обновлять оценки и решения, чтобы достичь оптимальности на определенном временном промежутке. Подобные вопросы часто изучаются в рамках *марковской теории принятия решений*, которую можно рассматривать как *оптимальное стохастическое управление* в дискретном времени.

Теоретические основы марковской теории принятия решений в настоящее время хорошо изучены, строгий математический анализ доступен в форме учебных пособий (см. [2], [4], [12] и [23]).

Однако, несмотря на богатый набор теоретических средств, имеющих в настоящее время, вопросы практического применения вызывают неизменный интерес. В этом контексте, *приближенное динамическое программирование* (см. [22]) появилось в результате попыток предоставить одновременно практически осуществимые эвристики и возникающие из теории идеи, объясняющие, почему они хорошо проявляют себя на практике.

*University of Technology Sydney; e-mail: juri.hinz@uts.edu.au

¹⁾ Данное исследование частично поддержано Австралийским советом по исследованиям через программу Проектов научных открытий (проект № DP13010335).

При управлении большой системой, практический подход к работе с пространством состояний большой размерности состоит в том, чтобы первым делом построить конечную дискретизацию пространства. С другой стороны, можно задаться вопросом аппроксимации функций на данном пространстве. В этом смысле, функциональные методы предлагают приближать функцию цены на пространстве состояний. Одним из таких методов является подход, основанный на методе наименьших квадратов и методе Монте-Карло, который предлагает аппроксимацию с помощью соответствующим образом параметризованного набора базисных функций. Так как параметры в нем вычисляются путем выполнения последовательных регрессий, этот метод относится к семейству регрессионных.

Вслед за [7], [28], [29], работа [18] стала источником последующих исследований, направленных на теоретическое обоснование предложенного подхода. Вопросы сходимости были рассмотрены в [8] и позднее обобщены в [27], [9] и [10], распространения на случай с возможностью многократного исполнения были рассмотрены в [6], а также изучались в [3], где была подчеркнута связь между теорией статистического обучения и теорией эмпирических процессов. Для обзора применения методов Монте-Карло в финансовой инженерии мы отсылаем заинтересованного читателя к книге Глаззермана [13], где можно также найти и список литературы. Помимо финансовых приложений методы, основанные на аппроксимации функций, также используются для того, чтобы уловить локальное поведение функции цены. Так появились продвинутые регрессионные методы, например, ядерные методы [20], [21], локальная полиномиальная регрессия [11], нейронные сети [5].

Одно из главных преимуществ методов, основанных на регрессии, заключается в том, что они сводят вычисления к простым линейным алгебраическим операциям в пространствах малой размерности. Однако из-за последовательных итераций, необходимых для реализации динамического программирования, вычисляемые решения часто нестабильны. В данной работе предложены некоторые численные аспекты нового функционального метода, который использует предположения выпуклости для получения стабильных и быстрых решений в терминах алгебраических операций над матрицами. Хотя при нашем подходе мы ограниваемся задачами довольно специфической структуры, данный класс охватывает большое количество важных практических приложений.

Цель настоящей работы — убрать ряд довольно ограничительных предположений, сделанных в [15], что позволит использовать нашу методологию для широкого класса задач. Кроме того, разработана техника измерения качества численных решений с точки зрения оценки их расстояния до оптимальных, это реализовано путем вычисления так называемых доверительных интервалов (или *границ*). Представлен новый и

численно эффективный алгоритм оценивания функции цены как сверху, так и снизу.

Предлагаемая статья организована следующим образом. После введения общих понятий в разд. 2, в разд. 3 вводится понятие *выпуклой системы переключающегося типа* и обсуждаются решения для стохастических задач из этого класса. В разд. 4 проводится обзор и анализ численной схемы из [15], которая дает быстрые и стабильные решения для выпуклых задач переключающегося типа. Раздел 5 представляет собой первый шаг к ослаблению требования выпуклости. Замечательное обобщение получено в разд. 6, где предложен метод, позволяющий обойти любые требования выпуклости, в то же время обеспечивая значительное улучшение в скорости вычислений. В разд. 7 представлен еще один ключевой результат настоящей работы, где мы предлагаем адаптацию подхода [25] получения рекурсивных схем построения оценок для верхних границ приближенных решений. Раздел 8 содержит два численных примера.

2. Марковская теория принятия решений. Мы начнем с обзора классических результатов, касающихся общего подхода к работе с задачами марковской теории принятия решений на конечном горизонте. При изложении мы будем следовать гл. 2 из [2], адаптируя ее под наши цели. Рассмотрим систему с конечным временным горизонтом $0, \dots, T$, состояние которой изменяется в измеримом пространстве $(E, \mathcal{E})$ и находится под воздействием элементов из множества A возможных управлений. Для каждого $a \in A$ мы предполагаем, что $K_t^a(x, dx')$ является стохастическим переходным ядром на $(E, \mathcal{E})$. Рассмотрим фиксированную последовательность $(X_t)_{t=0}^T$ случайных величин, которые можно рассматривать как координатные проекции, действующие на прямом произведении $E^{\{0, \dots, T\}}$ копий $(E, \mathcal{E})$. Отображение $\pi_t: E \rightarrow A$, согласно которому осуществляется выбор управления в момент времени t , называется *решающим правилом*. Последовательность решающих правил $\pi = (\pi_t)_{t=0}^{T-1}$ называется *стратегией*. Для каждой начальной точки $x_0 \in E$ и каждой стратегии $\pi = (\pi_t)_{t=0}^{T-1}$ существует вероятностная мера $\mathbf{P}^{x_0, \pi}$, для которой $\mathbf{P}^{x_0, \pi}(X_0 = x_0) = 1$ и соотношение

$$\mathbf{P}^{x_0, \pi}(X_{t+1} \in B \mid X_0, \dots, X_t) = K_t^{\pi_t(X_t)}(X_t, B) \quad (2.1)$$

имеет место для всех измеримых множеств $B \in \mathcal{E}$ и $t = 0, \dots, T - 1$. То есть, если система находится в состоянии X_t в момент времени t , управление $a = \pi_t(X_t)$ используется для того, чтобы выбрать переходную вероятность $K_t^{a=\pi_t(X_t)}(X_t, \cdot)$, которая задает случайный переход из состояния X_t в X_{t+1} с распределением $K_t^{\pi_t(X_t)}(X_t, \cdot)$. Для удобства записи, если выбрано управление $a \in A$, будем использовать обозначение $\mathcal{K}_t^a$ для одношагового переходного оператора, соответствующего пере-

ходному ядру K_t^a . Другими словами, для каждого управления $a \in A$ действие оператора $\mathcal{K}_t^a$ на функцию φ задается соотношением

$$(\mathcal{K}_t^a \varphi)(x) = \int_E \varphi(x') K_t^a(x, dx'), \quad x \in E, \quad (2.2)$$

если интегралы в правой части существуют.

В каждый момент времени t нам задана *функция выигрыша* (reward function) $r_t: E \times A \rightarrow \mathbf{R}$ на шаге t , где $r_t(x, a)$ представляет собой выигрыш при выборе управления $a \in A$ для системы, находящейся в состоянии $x \in E$ в момент времени t . В конце временного горизонта в момент T предполагается, что никакое управление не может быть выбрано. В этой ситуации, если система находится в состоянии x , выигрышем является *терминальная цена* (scrap value) $r_T(x)$, которая описывается заранее выбранной *терминальной функцией* $r_T: E \rightarrow \mathbf{R}$. Для заданной начальной точки x_0 наша цель состоит в максимизации ожидаемого общего выигрыша на конечном горизонте. Другими словами, требуется найти такую $\pi^* = (\pi_t^*)_{t=0}^{T-1}$, что

$$\pi^* = \operatorname{argmax}_{\pi \in \mathcal{A}} \mathbf{E}^{x_0, \pi} \left(\sum_{t=0}^{T-1} r_t(X_t, \pi_t(X_t)) + r_T(X_T) \right), \quad (2.3)$$

где $\mathcal{A}$ — набор всех стратегий и $\mathbf{E}^{x_0, \pi}$ обозначает математическое ожидание по мере управляемой марковской цепи, определенной в (2.1). При некоторых дополнительных предположениях задача (2.3) является хорошо поставленной (см. [2, с. 199]).

Вычисление оптимальной стратегии происходит по следующей схеме. Для $t = 0, \dots, T-1$ определим *оператор Беллмана*

$$\mathcal{T}_t v(x) = \sup_{a \in A} (r_t(x, a) + \mathcal{K}_t^a v(x)), \quad x \in E, \quad (2.4)$$

область определения которого состоит из измеримых функций $v: E \rightarrow \mathbf{R}$, для которых интегралы $\mathcal{K}_t^a v$ существуют для всех $a \in A$. Далее, рассмотрим *рекуррентные соотношения Беллмана*:

$$v_T = r_T, \quad v_t = \mathcal{T}_t v_{t+1} \quad \text{для } t = T-1, \dots, 0. \quad (2.5)$$

При соответствующих предположениях у рекуррентных соотношений Беллмана существует решение, которое задает так называемые *функции цены* (value functions) $(v_t^*)_{t=0}^T$ и определяет оптимальную стратегию π^* согласно

$$\pi_t^*(x) = \operatorname{argmax}_{a \in A} (r_t(x, a) + \mathcal{K}_t^a v_{t+1}^*(x)), \quad x \in E,$$

для всех $t = 0, \dots, T-1$.

3. Выпуклые системы переключающегося типа. В оставшейся части работы рассмотрим марковские задачи принятия решений, структура которых удовлетворяет специальным дополнительным предположениям, при которых решение рекуррентных соотношений Беллмана существует, что позволяет нам сосредоточить свое внимание на вопросе построения численных аппроксимаций решения. В литературе можно найти большое количество разнообразных достаточных условий, гарантирующих, что эти предположения выполнены в конкретных примерах, возникающих на практике (см. [2]).

Рассмотрим марковскую модель принятия решений, эволюция состояний которой описывается одной дискретной и одной непрерывной компонентами. Более конкретно, предположим, что пространство состояний $E = P \times \mathbf{R}^d$ является прямым произведением конечного пространства P и евклидова пространства $\mathbf{R}^d$. Мы предполагаем, что первая компонента $p \in P$ изменяется детерминистически под действием управлений из конечного множества A с помощью функции

$$\alpha: P \times A \rightarrow P, \quad (p, a) \rightarrow \alpha(p, a),$$

где $\alpha(p, a) \in P$ является новым значением дискретной компоненты состояния системы, если предыдущее значение ее дискретной компоненты было p , а в качестве управления было выбрано $a \in A$. Кроме того, мы предполагаем, что непрерывная компонента состояния представляет собой неуправляемый марковский процесс $(Z_t)_{t=0}^T$ на $\mathbf{R}^d$, эволюция которого задается случайными линейными преобразованиями

$$Z_{t+1} = W_{t+1} Z_t$$

с выбранными заранее независимыми и интегрируемыми матрицами возмущений $(W_t)_{t=1}^T$. Предположим, наконец, что функции выигрыша

$$r_t(p, z, a), \quad t = 0, \dots, T-1, \quad p \in P, \quad a \in A,$$

и терминальные функции

$$r_T(p, z), \quad p \in P,$$

при фиксированном p являются выпуклыми и липшицевыми по непрерывной компоненте пространства состояний z на $\mathbf{R}^d$. В такой постановке, переходные операторы задаются соотношениями

$$\mathcal{K}_t^a v(p, z) = \mathbf{E}(v(\alpha(p, a), W_{t+1} z)), \quad t = 0, \dots, T-1, \quad a \in A, \quad (3.1)$$

а операторы Беллмана принимают вид

$$\mathcal{J}_t v(p, z) = \sup_{a \in A} \left(r_t(p, z, a) + \mathbf{E}(v(\alpha(p, a), W_{t+1} z)) \right) \quad (3.2)$$

для всех $p \in P$, $z \in \mathbf{R}^d$ и $t = 0, \dots, T-1$. В дальнейшем марковские задачи принятия решений, удовлетворяющие данным предположениям, будем называть *выпуклыми системами переключающегося типа*.

4. Алгоритмические решения. Для рассматриваемых систем обратная индукция, описанная в (2.5) дает решение рассматриваемой задачи управления (см. [15]). Однако, глядя на оператор Беллмана (3.2), мы видим, что при решении рекуррентных соотношений Беллмана возникает ряд проблем, наиболее острая из которых заключается в том, что нам требуется построить поточечное решение для каждого $z \in \mathbf{R}^d$. В [15] был предложен метод построения решения в «функциональной» форме. Далее мы приводим подробный обзор полученных в [15] результатов и обоснование предложенного подхода.

Первым делом, аппроксимировав математическое ожидание в операторе Беллмана (3.2) с помощью конечной суммы, получаем *модифицированный* оператор Беллмана $\mathcal{T}_t^n$, который действует на заданную функцию цены в соответствии с

$$\mathcal{T}_t^n v(p, z) = \max_{a \in A} \left(r_t(p, z, a) + \sum_{k=1}^n \nu_{t+1}(k) v(\alpha(p, a), W_{t+1}(k)z) \right), \quad (4.1)$$

где $(W_{t+1}(k))_{k=1}^n$ обозначают набор реализаций возмущений с соответствующими вероятностными весами $(\nu_{t+1}(k))_{k=1}^n$. Заменяя настоящий оператор Беллмана (3.2) в обратной индукции (2.5) ее модифицированным аналогом (4.1), получаем модифицированную обратную индукцию

$$v_T^n = r_T, \quad v_t^n = \mathcal{T}_t^n v_{t+1}^n \quad \text{для } t = T-1, \dots, 0. \quad (4.2)$$

Хотя интегрирование теперь заменено на конечную сумму, отыскание $(v_t^n)_{t=0}^T$ является по-прежнему алгоритмически неразрешимой задачей, поскольку вычисление должно производиться в каждой точке $z \in \mathbf{R}^d$. В этой ситуации важным является следующее наблюдение: поскольку терминальная функция $r_T(p, z)$ и функции выигрыша $r_t(p, z, a)$, $t = 0, \dots, T-1$, используемые в (4.1) и (4.2), являются выпуклыми по непрерывной компоненте, построенные путем обратной индукции функции цены $(v_t^n)_{t=0}^T$ должны быть также выпуклыми по этой компоненте.

Далее мы предлагаем приближать данные функции $(v_t^n)_{t=0}^T$ с помощью максимума по конечному набору их субградиентов. Прежде чем приступить к объяснению преимущества подобных кусочно-линейных аппроксимаций, нам необходимо определить несколько понятий.

Будем называть счетное подмножество $G \subset \mathbf{R}^d$ *сеткой*. Для сетки G *субградиентная оболочка* $\mathcal{S}_G f$ выпуклой функции f определяется как максимум из субградиентов $\nabla_g f$ функции f по множеству точек сетки $g \in G$:

$$\mathcal{S}_G f = \vee_{g \in G} \nabla_g f.$$

По заданному семейству $\{(W_t(k))_{t=1}^T: k = 1, \dots, n\}$ траекторий возмущений, количество которых увеличивается с ростом $n \in \mathbf{N}$, и семейству сеток $(G^m)_{m \in \mathbf{N}}$, плотность которых также увеличивается с ростом

$m \in \mathbf{N}$, определим для каждой пары $n, m \in \mathbf{N}$ дважды модифицированные операторы Беллмана $\mathcal{T}_t^{m,n}$ для $t = 0, \dots, T-1$:

$$(\mathcal{T}_t^{m,n}v)(p, \cdot) = \mathcal{S}_{G^m} \max_{a \in A} \left(r_t(p, \cdot, a) + \sum_{k=1}^n \nu_{t+1}(k) v(\alpha(p, a), W_{t+1}(k) \cdot) \right).$$

Используя эти операторы, дважды модифицированные функции цены $(v_t^{m,n})_{t=0}^T$ получаются с помощью обратной индукции, которая начиная с

$$v_T^{m,n}(p, \cdot) = \mathcal{S}_{G^m} r_T(p, \cdot), \quad p \in P, \quad (4.3)$$

задает рекуррентным образом функции

$$v_t^{m,n} = \mathcal{T}_t^{m,n} v_{t+1}^{m,n}, \quad t = T-1, \dots, 0. \quad (4.4)$$

Очевидно, данная схема включает в себя два параметра аппроксимации $n \in \mathbf{N}$ и $m \in \mathbf{N}$, которые соответствуют выборке возмущений $(W_{t+1}(k))_{k=1}^n$ с весами $(\nu_{t+1}(k))_{k=1}^n \subset \mathbf{R}_+$ и стягивающимися сетками $(G^m)_{m \in \mathbf{N}}$. При соответствующих предположениях данная схема имеет отличные свойства сходимости (см. [15]). Однако, далее мы сосредоточимся исключительно на ее алгоритмическом аспекте.

Поскольку функции цены $(v_t^{m,n})_{t=0}^T$, вычисляемые согласно (4.3) и (4.4), являются кусочно-линейными и выпуклыми (по непрерывной компоненте), далее мы займемся представлением функций этого класса в терминах матриц, чтобы переписать алгоритм обратной индукции (4.3) и (4.4) в компактной матричной форме.

Матрица L с d столбцами называется *матрицей-представителем* кусочно-линейной выпуклой функции $l: \mathbf{R}^d \rightarrow \mathbf{R}$, если имеет место соотношение $l(z) = \max(Lz)$ для всех $z \in \mathbf{R}^d$. Мы будем использовать обозначение $l \sim L$, если кусочно-линейная выпуклая функция l обладает матрицей-представителем L .

Как оказывается, построение субградиентной оболочки можно описать непосредственно в терминах матриц-представителей. А именно, если l обладает матрицей-представителем L , то ее субградиентной оболочке $\mathcal{S}_G$ на сетке $G = \{g^1, \dots, g^m\}$ соответствует матрица-представитель $\Upsilon_G[L]$, где действие оператора Υ_G состоит в перестановке строк матрицы и определяется как

$$\Upsilon_G[L]_{i,\cdot} = L_{\arg\max(Lg^i)} \quad \text{для } i = 1, \dots, m.$$

Суть данной операции в том, что когда Υ_G применяется к матрице L с d столбцами, результатом $\Upsilon_G[L]$ перестановки строк является $(m \times d)$ -матрица, i -я строка которой есть строка L , на которой достигается максимум в i -й точке сетки g^i . Таким образом, связь между субградиентной

оболочкой функции и ее матрицей-представителем действительно задается в терминах оператора перестановки строк Υ_G :

$$l \sim L \implies \mathcal{S}_G l \sim \Upsilon_G[L].$$

Подобное соответствие имеет место для суммирования кусочно-линейных выпуклых функций с последующим взятием субградиентной оболочки. А именно, этому действию соответствует поэлементная сумма их матриц-представителей после применения оператора перестановки строк к каждой из матриц:

$$l \sim L, \quad f \sim F \implies \mathcal{S}_G(f + l) \sim \Upsilon_G[L] + \Upsilon_G[F].$$

Аналогичным образом, максимизация кусочно-линейных выпуклых функций с последующим взятием субградиентной оболочки реализуется на матричном уровне путем склеивания строк матриц-представителей с их последующей перестановкой:

$$l \sim L, \quad f \sim F \implies \mathcal{S}_G(l \vee f) \sim \Upsilon_G[L \sqcup F].$$

Оператор склеивания по строкам $L \sqcup F$ выполняет здесь конкатенацию двух матриц L и F . Заметим, что строки матрицы L , представляющей субградиентную оболочку $\mathcal{S}_G f$ выпуклой кусочно-линейной функции f всегда могут быть расположены таким образом, что $L = \Upsilon_G(L)$. Мы говорим, что субградиентный представитель L задан в *нормальной форме*, если $L = \Upsilon_G(L)$. В дальнейшем мы будем предполагать, что матрицы-представители заданы в нормальной форме.

Для последующего использования введем также эквивалентную, но алгоритмически более удобную процедуру максимизации на уровне матриц-представителей. Пусть задана сетка $G = \{g^1, \dots, g^m\} \subset \mathbf{R}^d$ и $(m \times d)$ -матрицы $F(a)$, $a \in A$, заданные в нормальной форме, т.е. $F(a) = \Upsilon_G(F(a))$ для всех $a \in A$. Используем обозначение

$$F := \bigsqcup_{a \in A} F(a)$$

для $(m \times d)$ -матрицы F , i -й строкой которой

$$F_i = F_{i,}(a(i)), \quad i = 1, \dots, m,$$

является i -я строка той матрицы $F(a(i))$, на которой достигается максимум в i -й точке сетки g^i , т.е.

$$a(i) = \operatorname{argmax}\{F_{i,}(a)g^i: a \in A\}, \quad i = 1, \dots, m.$$

Эта максимизация используется для получения субградиентной оболочки максимумов по семейству f^a , $a \in A$, кусочно-линейных выпуклых

функций в терминах матриц-представителей их субградиентных оболочек (заданных в нормальной форме):

$$\mathcal{S}_G f^a \sim F(a), \quad a \in A \quad \Rightarrow \quad \mathcal{S}_G \left(\bigvee_{a \in A} f^a \right) \sim \bigsqcup_{a \in A} F(a).$$

Наконец, подчеркнем, что определение субградиентной оболочки для композиции функции с линейным отображением соответствует простому матричному умножению с последующей операцией перестановки строк. Другими словами, для каждой $(d \times d)$ -матрицы W

$$l \sim L \quad \Rightarrow \quad \mathcal{S}_G(l(W \cdot)) \sim \Upsilon_G[LW].$$

Поскольку дважды модифицированная обратная индукция состоит из максимизации, суммирования и композиций с линейными отображениями, применяемыми к кусочно-линейным выпуклым функциям, ее можно переписать в терминах операций над матрицами. Приведем соответствующий алгоритм

Предварительные вычисления. Для заданной сетки $G^m = \{g^1, \dots, g^m\}$ определены операторы перестановки строк $\Upsilon = \Upsilon_{G^m}$ и максимизации по строкам $\sqcup_{a \in A}$. Строим выборку из распределения $(W_t(k))_{k=1}^n$ для каждого возмущения W_t с соответствующими весами $(\nu_t(k))_{k=1}^n$ для $t = 1, \dots, T$. По заданным функциям выигрыша $(r_t)_{t=0}^{T-1}$ и терминальному значению r_T строим нормальную форму матриц-представителей для их субградиентных оболочек:

$$\mathcal{S}_{G^m} r_t(p, \cdot, a) \sim R_t(p, a), \quad \mathcal{S}_{G^m} r_T(p, \cdot) \sim R_T(p)$$

для $t = 0, \dots, T-1$, $p \in P$ и $a \in A$. Пусть $V_t(p)$ для $t = 0, \dots, T$, $p \in P$ обозначают матриц-представителей соответствующих функций цены

$$v_t^{n,m}(p, \cdot) \sim V_t(p) \quad \text{для } t = 0, \dots, T, \quad p \in P,$$

которые получаются в результате следующей матричной обратной индукции.

Инициализация. Начинаем с задания матриц

$$V_T(p) = R_T(p), \quad \text{для всех } p \in P.$$

Цикл. Для $t = T-1, \dots, 0$ вычисляем для $p \in P$

$$V_t(p) = \bigsqcup_{a \in A} \left(R_t(p, a) + \sum_{k=1}^n \nu_{t+1}(k) \Upsilon \left[V_{t+1}(\alpha(p, a)) W_{t+1}(k) \right] \right). \quad (4.5)$$

5. Обобщение на невыпуклый случай. В данном разделе мы продемонстрируем, что приведенный выше алгоритм может быть также адаптирован для случая, когда функции выигрыша и терминальная функция не являются выпуклыми, но могут быть представлены в виде разности двух выпуклых функций. Более конкретно, предположим, что для всех $t = 0, \dots, T-1$, и $p \in P$, $a \in A$ имеет место

$$r_t(p, \cdot, a) = \check{r}_t(p, \cdot, a) - \hat{r}_t(p, \cdot, a) \quad (5.1)$$

и

$$r_T(p, \cdot) = \check{r}_T(p, \cdot) - \hat{r}_T(p, \cdot) \quad (5.2)$$

с выпуклыми функциями $\check{r}_t(p, \cdot, a)$, $\hat{r}_t(p, \cdot, a)$, $\check{r}_T(p, \cdot)$ и $\hat{r}_T(p, \cdot)$. Идея состоит в том, чтобы разделить обратную индукцию на параллельные процедуры, каждая из которых имеет дело с выпуклыми функциями. Предположим, что на шаге t , функция цены v_{t+1} может быть представлена в виде разности $v_{t+1} = \check{v}_{t+1} - \hat{v}_{t+1}$ выпуклых функций $\check{v}_{t+1}(p, \cdot)$ и $\hat{v}_{t+1}(p, \cdot)$ для $p \in P$. Тогда оператор Беллмана можно переписать в виде

$$\begin{aligned} \mathcal{T}_t v(p, z) &= \sup_{a \in A} (r_t(p, z, a) + \mathcal{K}_t^a v_{t+1}(p, z)) \\ &= \sup_{a \in A} ([\check{r}_t(p, z, a) + \mathcal{K}_t^a \check{v}_{t+1}(p, z)] - [\hat{r}_t(p, z, a) + \mathcal{K}_t^a \hat{v}_{t+1}(p, z)]). \end{aligned}$$

Отсюда видно, что максимизации по $a \in A$ подлежат разности выпуклых функций. Однако непосредственное применение оператора максимизации по строкам оказывается несовместимым с этим разложением. Следовательно требуется найти способ выразить максимум разностей выпуклых функций в виде разности двух выпуклых функций. Здесь помогает следующее простое наблюдение.

Рассмотрим для каждого $a \in A$ разность $\check{f}_a - \hat{f}_a$ двух выпуклых функций $\check{f}_a$ и $\hat{f}_a$, и пусть $\hat{f} := \sum_{a \in A} \hat{f}_a$. Тогда для каждого $a \in A$ функции $\check{f}_a - \hat{f}_a + \hat{f}$ и $\hat{f}$ являются выпуклыми и дают желаемое разложение

$$\max_{a \in A} (\check{f}_a - \hat{f}_a) = \max_{a \in A} (\check{f}_a - \hat{f}_a + \hat{f}) - \hat{f}. \quad (5.3)$$

Следуя данной схеме, мы предлагаем следующий алгоритм.

Предварительные вычисления. Представить функции выигрыша $(r_t)_{t=0}^{T-1}$ и терминальную функцию r_T в виде разности выпуклых функций, как в (5.1) и (5.2), и построить их матрицы-представителей (в нормальной форме)

$$\begin{aligned} \mathcal{S}_{G^m} \check{r}_t(p, \cdot, a) &\sim \check{R}_t(p, a), & \mathcal{S}_{G^m} \hat{r}_t(p, \cdot, a) &\sim \hat{R}_t(p, a), \\ \mathcal{S}_{G^m} \check{r}_T(p, \cdot) &\sim \check{R}_T(p), & \mathcal{S}_{G^m} \hat{r}_T(p, \cdot) &\sim \hat{R}_T(p) \end{aligned} \quad (5.4)$$

для всех $t = 0, \dots, T-1$, $p \in P$ и $a \in A$. Пусть $(v_t^{n,m})_{t=0}^T$ обозначают приближенные функции цены, для которых

$$v_t^{n,m} = \check{v}_t^{n,m} - \hat{v}_t^{n,m}, \quad (5.5)$$

где $\check{v}_t^{n,m}(p, \cdot)$ и $\hat{v}_t^{n,m}(p, \cdot)$ являются кусочно-линейными выпуклыми функциями с матрицами-представителями

$$\check{v}_t^{n,m}(p, \cdot) \sim \check{V}_t(p) \quad \text{и} \quad \hat{v}_t^{n,m}(p, \cdot) \sim \hat{V}_t(p) \quad (5.6)$$

для $t = 0, \dots, T$, $p \in P$.

И н и ц и а л и з а ц и я. Начинаем с задания матриц

$$\check{V}_T(p) = \check{R}_T(p) \quad \text{и} \quad \hat{V}_T(p) = \hat{R}_T(p) \quad \text{для всех } p \in P.$$

Ц и к л. Для $t = T-1, \dots, 1$ вычисляем сначала

$$\begin{aligned} \check{\Psi}_t(p, a) &= \check{R}_t(p, a) + \sum_{k=1}^n \nu_{t+1}(k) \Upsilon \left[\check{V}_{t+1}(\alpha(p, a)) \cdot W_{t+1}(k) \right], \\ \hat{\Psi}_t(p, a) &= \hat{R}_t(p, a) + \sum_{k=1}^n \nu_{t+1}(k) \Upsilon \left[\hat{V}_{t+1}(\alpha(p, a)) \cdot W_{t+1}(k) \right], \end{aligned} \quad (5.7)$$

а затем

$$\begin{aligned} \hat{V}_t(p) &= \sum_{a \in A} \hat{\Psi}_t(p, a), \\ \check{V}_t(p) &= \bigsqcup_{a \in A} (\check{\Psi}_t(p, a) - \hat{\Psi}_t(p, a) + \hat{V}_t(p)) \end{aligned} \quad (5.8)$$

для всех $p \in P$.

6. Эффективная аппроксимация. Хотя численные эксперименты показывают стабильные и надежные результаты, возникает ощущение, что вычислительная производительность страдает из-за того, что большая часть времени тратится на необходимые оператору Υ перестановки в матрицах. Из (4.5) видно, что для вычисления

$$\sum_{k=1}^n \nu_{t+1}(k) \Upsilon \left[V_{t+1}(\alpha(p, a)) W_{t+1}(k) \right] \quad (6.1)$$

на каждом шаге алгоритма, оператор перестановки строк применяется n раз, при умножении на каждую из матриц возмущений. Эта задача становится все более трудоемкой при растущих значениях размера выборки возмущений n , особенно для матриц больших размерностей. Прежде чем продолжить, для простоты обозначений опустим временной индекс $t+1$ в (6.1). Итак, сосредоточимся на двух основных источниках вычислительных затрат, возникающих при подсчете данного выражения, а именно

$$\begin{aligned} &\text{перестановке строк } \Upsilon[VW(k)] \text{ в} \\ &\text{больших матрицах } VW(k) \end{aligned} \quad (6.2)$$

и

$$\begin{aligned} &\text{суммированию матриц } \Upsilon[VW(k)] \text{ по} \\ &\text{большому диапазону индексов } k = 1, \dots, n. \end{aligned} \quad (6.3)$$

Оставшаяся часть этого раздела будет поделена на две части. В разд. 6.1 мы предложим метод аппроксимации (6.1), устраняющий одновременно обе проблемы. Уменьшение вычислительных затрат делает возможным получить *приближенные* решения для больших сеток и размеров выборок распределения. Кроме того, мы увидим, что в отличие от (6.1), данная аппроксимация не требует, чтобы $V = V_{t+1}$ была выпуклой. В разд. 6.2 мы строим аппроксимацию первого порядка, которая дает эффективный способ вычисления функций без необходимости раскладывать их на выпуклые составляющие, и оказывается приемлемой для наших целей. Объединяя эту аппроксимацию с методом из разд. 6.1, мы получаем эффективный алгоритм, в котором больше не обременены требованиями выпуклости.

6.1. Оценивание условных математических ожиданий. Ключевым моментом является то, что процедуру, описанную в (6.2), можно приблизить, заменив операцию перестановки строк умножением на подходящим образом выбранную матрицу. Другими словами, идея состоит в том, чтобы для $k = 1, \dots, n$

$$\begin{aligned} &\text{построить матрицу } Y(k), \text{ такую, что} \\ &Y(k)VW(k) \text{ приближает } \Upsilon[VW(k)]. \end{aligned} \quad (6.4)$$

Прежде, чем обосновывать эту аппроксимацию, давайте сначала посмотрим, как она может быть использована для решения вычислительной задачи (6.3). При наличии (6.4) имеет место следующая аппроксимация для (6.1):

$$\sum_{k=1}^n \nu(k) \Upsilon[VW(k)] \approx \sum_{k=1}^n \nu(k) Y(k) VW(k), \quad (6.5)$$

которая в свою очередь требует эффективного вычисления сумм матриц. В примерах, возникающих на практике, размер выборки распределения n и размер сетки m (количество строк в V) обычно на несколько порядков больше размерности d матриц возмущений $W(k)$. Например, для достижения приемлемого уровня сходимости в типичных приложениях, размер выборки n и размер сетки m должны быть выбраны в диапазоне нескольких тысяч, в то время как размерность пространства состояний d оказывается порядка нескольких десятков. Это наблюдение показывает, что значительное снижение вычислительных затрат может быть достигнуто с помощью аддитивного разложения реализаций возмущений. Предположим, что матрица возмущений W представляется в

виде линейной комбинации

$$W = \overline{W} + \sum_{j=1}^J \varepsilon_j E(j) \quad (6.6)$$

с неслучайными матрицами $\overline{W}$ и $(E(j))_{j=1}^J$, и случайными коэффициентами $(\varepsilon_j)_{j=1}^J$. С помощью этого разложения, каждая реализация $W(k)$ матрицы возмущения W переписывается в виде

$$W(k) = \overline{W} + \sum_{j=1}^J \varepsilon_j(k) E(j), \quad k = 1, \dots, n. \quad (6.7)$$

Используя это представление, перепишем правую часть (6.5) в виде

$$\begin{aligned} \sum_{k=1}^n \nu(k) Y(k) V W(k) &= \left(\sum_{k=1}^n \nu(k) Y(k) \right) V \overline{W} \\ &+ \sum_{j=1}^J \left(\sum_{k=1}^n \nu(k) \varepsilon_j(k) Y(k) \right) V E(j), \end{aligned}$$

где мы поменяли порядок суммирования. Если вычислить матрицы

$$D_0 = \sum_{k=1}^n \nu(k) Y(k), \quad D_j = \sum_{k=1}^n \nu(k) \varepsilon_j(k) Y(k), \quad j = 1, \dots, J, \quad (6.8)$$

заранее, то подсчет (6.5) существенно упростится:

$$\sum_{k=1}^n \nu(k) Y(k) V W(k) = D_0 V \overline{W} + \sum_{j=1}^J D_j V E(j), \quad (6.9)$$

поскольку в этом случае остается произвести уже небольшое количество сложений и умножений матриц. Оператор, выполняющий данное эффективное вычисление условного математического ожидания, будем обозначать в дальнейшем через $\mathcal{E}$, т.е.

$$\mathcal{E}(V) := D_0 V \overline{W} + \sum_{j=1}^J D_j V E(j). \quad (6.10)$$

Перейдем теперь к обоснованию аппроксимации (6.4). Предположим, что сетка $\{g^1, \dots, g^m\}$ представлена матрицей G , i -я строка которой соответствует точке сетки g^i , т.е. G состоит из m строк, таких, что $G_{i,\cdot} = g^i$ для $i = 1, \dots, m$. Пусть $\tilde{L} = \Upsilon[L]$ обозначает результат применения Υ к матрице L . Тогда матрица $\tilde{L}$ характеризуется тем, что

$$\begin{aligned} \tilde{L} = \Upsilon[L] \text{ состоит из } m \text{ строк,} \\ \text{полученных перестановкой строк } L, \end{aligned} \quad (6.11)$$

и

$$\tilde{L}_{i,\cdot} G_{i,\cdot}^\top \geq L_{j,\cdot} G_{i,\cdot}^\top \quad \text{для всех } i, j = 1, \dots, m. \quad (6.12)$$

В соответствии с (6.11) мы делаем вывод о том, что

$$\begin{aligned} \Upsilon[VW(k)] \text{ состоит из } m \text{ строк,} \\ \text{полученных перестановкой строк } VW(k). \end{aligned} \quad (6.13)$$

Поскольку любая перестановка строк может быть реализована с помощью умножения слева на подходящим образом выбранную матрицу, то существует перестановочная матрица $Y_V(k)$ такая, что

$$Y_V(k) VW(k) = \Upsilon[VW(k)]. \quad (6.14)$$

Вычисление $Y_V(k)$ требует больших усилий, поскольку каждая матрица зависит не только от $W(k)$, но также и от соответствующей V . Поэтому мы предлагаем строить для каждой $Y_V(k)$ суррогатную матрицу $Y(k)$, которая зависит только от $W(k)$ и не зависит от V . Поскольку $Y_V(k)$ должны удовлетворять (6.14), заметим по аналогии с (6.12), что должно быть выполнено условие

$$(Y_V(k)V)_{i,\cdot} (W(k)G)_{i,\cdot}^\top \geq V_{j,\cdot} (W(k)G)_{i,\cdot}^\top \quad \text{для } i, j = 1, \dots, m. \quad (6.15)$$

Для каждого $i = 1, \dots, m$ поставим строке $(W(k)G)_{i,\cdot}$ в соответствие ближайшую строку $G_{h_k(i),\cdot}$ в матрице, представляющей исходную сетку:

$$h_k(i) = \operatorname{argmin} \{ j = 1, \dots, m: \|(W(k)G)_{i,\cdot} - G_{j,\cdot}\| \}, \quad i = 1, \dots, m. \quad (6.16)$$

С помощью данной функцией близости $h_k: \{1, \dots, m\} \rightarrow \{1, \dots, m\}$ вместо (6.15) можно рассмотреть условие

$$(Y(k)V)_{i,\cdot} G_{h_k(i),\cdot}^\top \geq V_{j,\cdot} G_{h_k(i),\cdot}^\top \quad \text{для } i, j = 1, \dots, m \quad (6.17)$$

с подходящей перестановочной матрицей $Y(k)$. Хотя (6.17) очевидно не является эквивалентным (6.15), это соотношение дает разумное приближение, когда сетка является достаточно плотной. Определим $Y(k)$ согласно

$$Y(k)_{i,j} = \begin{cases} 1, & \text{если } j = h_k(i), \\ 0 & \text{в других случаях,} \end{cases} \quad (6.18)$$

и заметим, что если V имеет нормальную форму $\Upsilon[V] = V$, то верно следующее утверждение:

$$V_{h_k(i),\cdot} G_{h_k(i),\cdot}^\top \geq V_{j,\cdot} G_{h_k(i),\cdot}^\top \quad \text{для } i, j = 1, \dots, m.$$

То есть, требуемая аппроксимация (6.4) осуществляется с помощью (6.18).

Предварительные вычисления, необходимые для аппроксимации (6.1) (т.е. вычисление $D_0, \dots, D_J$), являются затратными. Выигрыш в производительности может быть достигнут, только если возмущения $(W_t)_{t=1}^T$ одинаково распределены, что дает возможность произвести необходимые вычисления один раз сразу для всех матриц возмущений. В этом случае идеи, представленные в этом разделе, в краткой форме выражаются в следующем алгоритме.

Предварительные вычисления. Строим выборку $(W(k))_{k=1}^n$ из целевого распределения. Для каждого возмущения $W(k)$ находим соответствующую перестановочную матрицу $Y(k)$ как в (6.18), используя функцию близости (6.16). Используем эти матрицы и компоненты описанного в (6.7) разложения каждой $W(k)$ для вычисления матриц из (6.8).

Продолжение. Выполняем алгоритм (5.4)–(5.8), заменив (5.7) на

$$\begin{aligned}\check{\Psi}_t(p, a) &= \check{R}_t(p, a) + \mathcal{E} \left[\check{V}_{t+1}(\alpha(p, a)) \right], \\ \hat{\Psi}_t(p, a) &= \hat{R}_t(p, a) + \mathcal{E} \left[\hat{V}_{t+1}(\alpha(p, a)) \right],\end{aligned}$$

заменив также условные математические ожидания на их эффективно вычисляемый аналог (6.10).

6.2. Прямая схема. До сих пор мы работали с параллельными процедурами на выпуклых функциях. Однако важно отметить, что ни в одной части процедуры подсчета эффективного условного математического ожидания выпуклость целевой функции не требовалась. Учитывая этот факт, представим теперь дальнейшее упрощение предложенного алгоритма, основанное на аппроксимации первого порядка. Для невыпуклых кусочно-гладких функций f ранее была предложена следующая схема: сначала функция $f = \hat{f} - \check{f}$ раскладывается на выпуклые составляющие $\hat{f}$ и $\check{f}$, а затем каждой из них ставится в соответствие ее матрица-представитель, $\hat{F}$ и $\check{F}$, соответственно, заданная в нормальной форме относительно сетки G . Значение $f(z)$ в точке z в этом случае вычисляется согласно

$$f(z) = \max(\hat{F}z) - \max(\check{F}z).$$

Однако, если известна только разность матриц $\hat{F} - \check{F}$, то можно использовать аппроксимацию первого порядка

$$f(z) \approx (\hat{F} - \check{F})_{h(z), \cdot} z,$$

где h — так называемая узловая функция (host function) сетки $G = \{g^1, \dots, g^m\}$ с $m \in \mathbf{N}$,

$$h(z) = \operatorname{argmin}\{\|z - g^i\|: i \in \{1, \dots, m\}\},$$

которая возвращает каждому аргументу $z \in \mathbf{R}^d$ соответствующий номер узла — индекс точки сетки G с наименьшим расстоянием до z . Аппроксимация первого порядка использует непосредственно разность $\hat{F} - \check{F}$ и, в отличие от выпуклого разложения, не требует отдельных вычислений выпуклых и вогнутых компонент. Если при обращении к функциям использовать аппроксимацию первого порядка, то не возникает необходимости следить за выпуклой и вогнутой составляющими отдельно. Это дает значительное упрощение алгоритма, что выражается в следующей прямой схеме.

Предварительные вычисления. Определяем оператор $\mathcal{E}$ как в (6.10) при необходимых для этого предположениях. Определяем для $p \in P$, $a \in A$ и $t = 0, \dots, T-1$ матрицы

$$R_t(p, a) = \hat{R}_t(p, a) - \check{R}_t(p, a), \quad R_T(p) = \hat{R}_T(p) - \check{R}_T(p), \quad (6.19)$$

как в (5.4). Строим приближенные функции цены, их выпуклое разложение и представителей, как в (5.5) и (5.6). Матрицы

$$V_t(p) = \check{V}_t(p) - \hat{V}_t(p) \quad \text{для } t = 0, \dots, T, \quad p \in P,$$

получаются по следующей схеме.

Инициализация. Начинаем с задания матриц

$$V_T(p) = \check{R}_T(p) - \hat{R}_T(p) \quad \text{для всех } p \in P.$$

Цикл. Для $t = T-1, \dots, 0$ вычисляем для $p \in P$

$$V_t(p) = \bigsqcup_{a \in A} \left(R_t(p, a) + \mathcal{E}(V_{t+1}(p, a)) \right). \quad (6.20)$$

Замечание. В отличие от случая выпуклого разложения (5.4)–(5.8), *прямая* схема (6.19)–(6.20) возвращает сразу разность $V_t(p) = \check{V}_t(p) - \hat{V}_t(p)$. Это значит, что доступ к приближенным функциям цены предоставляется с помощью соотношения

$$v_t^{m,n}(p, z) \approx V_t(p)_{h(z), \cdot} z$$

с использованием принимающей функции h сетки G^m . В частности, это дает аппроксимацию оптимальной стратегии $\pi^{m,n} = (\pi_t^{m,n})_{t=0}^{T-1}$ в виде

$$\pi_t^{m,n}(p, z) = \operatorname{argmax}_{a \in A} \left((R_t(p, a) + \mathcal{E}(V_{t+1}(p, a)))_{h(z), \cdot} \cdot z \right) \quad (6.21)$$

для $t = 0, \dots, T-1$, $z \in \mathbf{R}^d$, $p \in P$. Для задания узловой функции h удобно использовать структуры типа деревьев на сетке, которые могут быть построены с использованием иерархических методов кластеризации.

7. Диагностика решения. В разд. 6.1 был получен эвристический метод построения эффективной аппроксимации условного математического ожидания в рекуррентных соотношениях Беллмана. В разд. 6.2 мы увидели, что, объединяя его с аппроксимацией первого порядка, для заданной сетки можно получить аппроксимацию (6.21) оптимальной стратегии в (2.3). Чтобы говорить о расстоянии от этого приближенного решения до оптимального, нужно сначала задать подходящую меру для расстояния.

Предположим, что нам задана произвольная стратегия $\pi = (\pi_t)_{t=0}^{T-1}$. Для нее можно определить соответствующий набор *цен стратегии* (*policy values*) $(v_t^\pi(p, z))_{t=0}^T$, которые задаются рекуррентными соотношениями

$$v_T^\pi(p, z) = r_T(p, z), \quad (7.1)$$

$$v_t^\pi(p, z) = r_t(p, z, \pi_t(p, z)) + \mathbf{E}\left(v_{t+1}^\pi(\alpha(p, \pi_t(p, z)), W_{t+1}z)\right), \quad (7.2)$$

для $t = T - 1, \dots, 0$. Рассмотрим систему переключающегося типа с заданными начальными положением $p_0^\pi = p_0 \in P$ и состоянием $Z_0 = z_0 \in \mathbf{R}^d$. В каждый момент времени t управление и новое положение определяется рекуррентным образом в соответствии со стратегией $\pi = (\pi_t)_{t=0}^{T-1}$ согласно

$$a_t^\pi := \pi_t(p_t^\pi, Z_t), \quad p_{t+1}^\pi := \alpha(p_t^\pi, a_t^\pi), \quad t = 0, \dots, T - 1.$$

Эти значения определяют *работу стратегии* (*policy run*) $\mathcal{V}_0^\pi(p_0^\pi, z_0)$, где

$$\mathcal{V}_0^\pi(p_0^\pi, Z_0) = \sum_{s=0}^{T-1} r_s(p_s^\pi, Z_s, a_s^\pi) + r_T(p_T^\pi, Z_T).$$

Согласно определению, $v_0^\pi(p_0, z_0)$ есть математическое ожидание работы стратегии

$$v_0^\pi(p_0, z_0) = \mathbf{E}(\mathcal{V}_0^\pi(p_0, z_0)), \quad p \in P, \quad z \in \mathbf{R}^d.$$

На практике, чтобы оценить это значение, можно использовать метод Монте-Карло. А именно, при наличии последовательности $(\omega_k)_{k \in \mathbf{N}}$ независимых случайных исходов, по усиленному закону больших чисел имеет место соотношение

$$v_0^\pi(p_0, z_0) = \mathbf{E}(\mathcal{V}_0^\pi(p_0, z_0)) = \lim_{k \rightarrow \infty} \frac{1}{k} \sum_{k=1}^K \mathcal{V}_0^\pi(p_0, z_0)(\omega_k). \quad (7.3)$$

Подобная процедура Монте-Карло может быть полезной для оценки эффективности конкретной стратегии π согласно критерию (2.3), однако, она никак не проясняет, насколько $v_0^\pi(p_0, z_0)$ оказывается далеко

от теоретически возможного значения $v_0^{\pi^*}(p_0, z_0)$ для оптимальной стратегии π^* , если бы она была известна.

В оставшейся части этого раздела, мы предлагаем разумное решение этого вопроса в терминах *метода диагностики*. Для заданной начальной точки (p_0, z_0) мы объясним, каким образом может быть оценен зазор

$$[v_0^{\pi}(p_0, z_0), v_0^{\pi^*}(p_0, z_0)] \quad (7.4)$$

между заданной стратегией π и оптимальной π^* . Наша методика основана на использовании конечной выборки $\{\omega_1, \dots, \omega_K\}$ реализаций траекторий и использует встроенную технику уменьшения дисперсии для построения *узких доверительных границ* для верхних и нижних оценок интервала (7.4).

Сначала сконцентрируемся на получении верхней границы. Рассмотрим последовательность $\varphi = (\varphi_t)_{t=1}^T$ случайных отображений

$$\varphi_t: P \times \mathbf{R}^d \times A \times \Omega \rightarrow \mathbf{R}, \quad (p, z, a, \omega) \mapsto \varphi_t(p, z, a)(\omega), \quad (7.5)$$

которые для $t = 1, \dots, T$ удовлетворяют соотношениям

$$\mathbf{E}(\varphi_t(p, z, a)) = 0, \quad p \in P, \quad z \in \mathbf{R}^d, \quad a \in A, \quad (7.6)$$

и таковы, что σ -алгебры

$$\sigma(\varphi_t(p, z, a), W_t; a \in A, p \in P, z \in \mathbf{R}^d), \quad t = 1, \dots, T, \quad (7.7)$$

независимы. Для заданных отображений $\varphi = (\varphi_t)_{t=1}^T$ рассмотрим случайные функции $(\bar{v}_t^\varphi)_{t=0}^T$

$$\bar{v}_t^\varphi: P \times \mathbf{R}^d \times \Omega \rightarrow \mathbf{R}, \quad t = 0, \dots, T,$$

определяемые рекуррентным образом для $t = T, \dots, 1$ согласно

$$\bar{v}_T^\varphi(p, z) = r_T(p, z), \quad (7.8)$$

$$\bar{v}_t^\varphi(p, z) = \max_{a \in A} \left(r_t(p, z, a) + \varphi_{t+1}(p, z, a) + \bar{v}_{t+1}^\varphi(\alpha(p, a), W_{t+1}z) \right). \quad (7.9)$$

Для $(\bar{v}_t^\varphi)_{t=0}^T$ имеет место следующая теорема.

Теорема 1. (i) Для произвольной стратегии $\pi = (\pi_t)_{t=0}^{T-1}$ цены $(v_t^\pi)_{t=0}^T$ мажорируются сверху с помощью

$$v_t^\pi(p, z) \leq \mathbf{E}(\bar{v}_t^\varphi(p, z)) \quad \text{для всех } t = 0, \dots, T, \quad p \in P, \quad z \in \mathbf{R}^d. \quad (7.10)$$

(ii) По заданным значениям цен $(v_t^{\pi^*})_{t=0}^T$ оптимальной стратегии $\pi^* = (\pi_t^*)_{t=0}^{T-1}$ определим $(\varphi_t^*)_{t=1}^T$ согласно

$$\varphi_{t+1}^*(p, z, a) = \mathbf{E}(v_{t+1}^{\pi^*}(\alpha(p, a), W_{t+1}z)) - v_{t+1}^{\pi^*}(\alpha(p, a), W_{t+1}z) \quad (7.11)$$

для всех $p \in P$, $z \in \mathbf{R}^d$, $a \in A$ и $t = 0, \dots, T-1$. Тогда отображения $(\varphi_t^*)_{t=1}^T$ удовлетворяют (7.5)–(7.7), а соотношение (7.10) для них переходит в равенство

$$v_t^{\pi^*}(p, z) = \bar{v}_t^{\varphi^*}(p, z) \quad \text{для всех } t = 0, \dots, T, \quad p \in P, \quad z \in \mathbf{R}^d. \quad (7.12)$$

Д о к а з а т е л ь с т в о. (i) Цена $(v_t^{\pi})_{t=0}^T$ стратегии $\pi = (\pi_t)_{t=0}^{T-1}$ удовлетворяет рекуррентным соотношениям (7.2). С учетом (7.6), получаем

$$\begin{aligned} v_t^{\pi}(p, z) &= \mathbf{E}(r_t(p, z, \pi_t(p, z)) + \varphi_{t+1}(p, z, \pi_t(p, z))) \\ &\quad + \mathbf{E}(v_{t+1}^{\pi}(\alpha(p, \pi_t(p, z)), W_{t+1}z)). \end{aligned} \quad (7.13)$$

Докажем утверждение (7.10) по индукции. Для $t = T$ (7.10) имеет место с равенством в силу базисов обратных индукций в (7.1) и (7.8):

$$v_T^{\pi} = r_T = \bar{v}_T^{\varphi}. \quad (7.14)$$

Используя предположение индукции

$$v_{t+1}^{\pi}(p, z) \leq \mathbf{E}(\bar{v}_{t+1}^{\varphi}(p, z)) \quad \text{для всех } p \in P, \quad z \in \mathbf{R}^d,$$

и (7.7), получаем

$$v_{t+1}^{\pi}(\alpha(p, \pi_t(p, z)), W_{t+1}z) \leq \mathbf{E}(\bar{v}_{t+1}^{\varphi}(\alpha(p, \pi_t(p, z)), W_{t+1}z) | W_{t+1})$$

для всех $p \in P$, $z \in \mathbf{R}^d$. Отсюда, в (7.13) имеет место оценка

$$\begin{aligned} v_t^{\pi}(p, z) &\leq \mathbf{E}(r_t(p, z, \pi_t(p, z)) + \varphi_{t+1}(p, z, \pi_t(p, z))) \\ &\quad + \mathbf{E}(\mathbf{E}(\bar{v}_{t+1}^{\varphi}(\alpha(p, \pi_t(p, z)), W_{t+1}z) | W_{t+1})), \end{aligned}$$

из которой следует необходимое утверждение

$$\begin{aligned} v_t^{\pi}(p, z) &= \mathbf{E}(r_t(p, z, \pi_t(p, z)) + \varphi_{t+1}(p, z, \pi_t(p, z)) \\ &\quad + \bar{v}_{t+1}^{\varphi}(\alpha(p, \pi_t(p, z)), W_{t+1}z)) \\ &\leq \mathbf{E}\left(\max_{a \in A} [r_t(p, z, a) + \varphi_{t+1}(p, z, a) + \bar{v}_{t+1}^{\varphi}(\alpha(p, a), W_{t+1}z)]\right) \\ &\leq \mathbf{E}(\bar{v}_t^{\varphi}(p, z)), \end{aligned}$$

где последнее неравенство справедливо в силу рекуррентных соотношений (7.9).

(ii) Пусть π^* является оптимальной стратегией, и определим $\varphi^* = (\varphi_t^*)_{t=1}^T$ согласно (7.11). Данные функции удовлетворяют предположению (7.6), независимость (7.7) также имеет место, поскольку в

$\varphi_{t+1}^*(p, z, a)$ случайная компонента $v_{t+1}^{\pi^*}(\alpha(p, a), W_{t+1}z)$ является функцией от W_{t+1} для $t = 0, \dots, T-1$. Докажем утверждение (7.12) по индукции, базис которой справедлив в силу (7.8). Из предположения индукции $v_{t+1}^{\pi^*} = \bar{v}_{t+1}^{\varphi^*}$ следует, что для всех $p \in P$, $z \in \mathbf{R}^d$ и $a \in A$ имеет место

$$\begin{aligned} r_t(p, z, a) + \varphi_{t+1}^*(p, z, a) + \bar{v}_{t+1}^{\varphi^*}(\alpha(p, a), W_{t+1}z) \\ = r_t(p, z, a) + \varphi_{t+1}^*(p, z, a) + v_{t+1}^{\pi^*}(\alpha(p, a), W_{t+1}z). \end{aligned} \quad (7.15)$$

Подставляя сюда (7.11), получаем, что для всех $p \in P$, $z \in \mathbf{R}^d$ и $a \in A$ справедливо

$$\begin{aligned} r_t(p, z, a) + \varphi_{t+1}^*(p, z, a) + v_{t+1}^{\pi^*}(\alpha(p, a), W_{t+1}z) \\ = r_t(p, z, a) + \mathbf{E}(v_{t+1}^{\pi^*}(\alpha(p, a), W_{t+1}z)) \\ - v_{t+1}^{\pi^*}(\alpha(p, a), W_{t+1}z) + v_{t+1}^{\pi^*}(\alpha(p, a), W_{t+1}z) \\ = r_t(p, z, a) + \mathbf{E}(v_{t+1}^{\pi^*}(\alpha(p, a), W_{t+1}z)). \end{aligned} \quad (7.16)$$

Для доказательства требуемого утверждения (7.12) остается заменить в рекуррентных соотношениях (7.9) выражение (7.15) на (7.16):

$$\begin{aligned} \bar{v}_t^{\varphi^*}(p, z) &= \max_{a \in A} \left(r_t(p, z, a) + \varphi_{t+1}^*(p, z, a) + \bar{v}_{t+1}^{\varphi^*}(\alpha(p, a), W_{t+1}z) \right) \\ &= \max_{a \in A} \left(r_t(p, z, a) + \mathbf{E}(v_{t+1}^{\pi^*}(\alpha(p, a), W_{t+1}z)) \right) \\ &= v_t^{\pi^*}(p, z), \quad p \in P, \quad z \in \mathbf{R}^d, \quad a \in A. \end{aligned}$$

Теорема 1 доказана.

Остановимся на практическом применении данной техники. Предположим, что мы хотим оценить расстояние между оптимальной стратегией π^* и приближенной $\tilde{\pi}$, полученной с помощью численных процедур, описанных ранее. В соответствии с утверждением (i) теоремы 1, произвольная последовательность $(\varphi_t)_{t=1}^T$, удовлетворяющая (7.6) и (7.7), дает верхнюю границу:

$$v_0^{\tilde{\pi}}(p, z) \leq v_0^{\pi^*}(p, z) \leq \mathbf{E}(\bar{v}_0^{\varphi}(p, z)), \quad p \in P, \quad z \in \mathbf{R}^d. \quad (7.17)$$

Обратите внимание, что математическое ожидание $\mathbf{E}(\bar{v}_0^{\varphi}(p, z))$ будет оцениваться с помощью метода Монте-Карло. Таким образом, получаем следующую процедуру.

Оценка для верхней границы.

1) Для заданной системы переключающегося типа реализуем $(\varphi_t)_{t=1}^T$, удовлетворяющие (7.5), (7.6) и (7.7).

2) Выбираем число испытаний $K \in \mathbf{N}$ в методе Монте-Карло и строим для $k = 1, \dots, K$ независимые реализации возмущений $(W_t(\omega_k))_{t=1}^T$.

3) Начиная с $z_0^k := z_0 \in \mathbf{R}^d$, определяем для $k = 1, \dots, K$ рекуррентным образом траектории $(z_t^k)_{t=0}^T$: $z_{t+1}^k = W_{t+1}(\omega_k)z_t^k$, $t = 0, \dots, T-1$, а также реализации $\varphi_{t+1}(p, z_t^k, a)(\omega_k)$, $t = 0, \dots, T-1$, $k = 1, \dots, K$.

4) Для каждого $k = 1, \dots, K$: при $t = T$ задаем $\bar{v}_T^\varphi(p, z_T^k) = r_T(p, z_T^k)$ для всех $p \in P$, для $t = T-1, \dots, 0$ вычисляем выражения

$$\bar{v}_t^\varphi(p, z_t^k) = \max_{a \in A} \left(r_t(p, z_t^k, a) + \varphi_{t+1}(p, z_t^k, a)(\omega_k) + \bar{v}_{t+1}^\varphi(\alpha(p, a), z_{t+1}^k) \right).$$

Сохраняем значения цены как $\bar{v}_0^\varphi(p, z_0^k)$ для $k = 1, \dots, K$.

5) Вычисляем выборочное среднее $K^{-1} \sum_{k=1}^K \bar{v}_0^\varphi(p, z_0^k)$ и его верхнюю доверительную границу, чтобы оценить $v_0^{\pi^*}(p, z_0)$ сверху.

Чтобы получить хорошую оценку сверху, функции $(\varphi_t)_{t=1}^T$ следует выбирать соответствующим образом. Утверждение (ii) теоремы 1 предлагает подходящий выбор. А именно, в гипотетическом случае, когда функции цены оптимальной стратегии $(v_t^{\pi^*})_{t=0}^T$ известны, $(\varphi_t^*)_{t=1}^T$, полученная согласно (7.11), дает точную и неслучайную верхнюю границу. Данная ситуация нереалистична, поскольку на практике оптимальная стратегия π^* неизвестна. Вместо этого мы предлагаем использовать приближенную функцию цены $(v_t)_{t=0}^T$, возвращаемую одним из алгоритмов, описанных в настоящей работе. То есть, следуя (7.11), разумным будет следующий выбор:

$$\varphi_{t+1}(p, z, a) = \mathbf{E} \left(\tilde{v}_{t+1}(\alpha(p, a), W_{t+1}z) \right) - \tilde{v}_{t+1}(\alpha(p, a), W_{t+1}z) \quad (7.18)$$

для $t = 0, \dots, T-1$. Однако, при таком выборе необходимо точное вычисление математического ожидания $\mathbf{E}(\tilde{v}_{t+1}(\alpha(p, a), W_{t+1}z))$, что также невозможно на практике. По этой причине, мы предлагаем модификацию: заменим в (7.18) математическое ожидание на арифметическое среднее по I независимым копиям $(W_{t+1}^{(i)})_{i=1}^I$ возмущения W_{t+1} . То есть, по заданным независимым случайным величинам W_{t+1} и $W_{t+1}^{(i)}$ для $i = 1, \dots, I$ и $t = 0, \dots, T-1$, таким, что распределение $W_{t+1}^{(i)}$ совпадает с распределением W_{t+1} , определим для всех $t = 0, \dots, T-1$, $a \in A$, $p \in P$ и $z \in \mathbf{R}^d$

$$\varphi_{t+1}(p, z, a) = \frac{1}{I} \sum_{i=1}^I \tilde{v}_{t+1}(\alpha(p, a), W_{t+1}^{(i)}z) - \tilde{v}_{t+1}(\alpha(p, a), W_{t+1}z). \quad (7.19)$$

Определенные таким образом $(\varphi_t)_{t=1}^T$ удовлетворяют условиям (7.6) и (7.7), и мы получаем обоснованную и уже вычислимую верхнюю границу.

Перейдем теперь к оцениванию нижней границы интервала (7.4). Для заданной стратегии $\pi = (\pi_t)_{t=0}^{T-1}$, цена $v_0^\pi(p_0, z_0)$ может быть в принципе вычислена как в (7.3) по тестовым запускам стратегии в сериях независимых бэк-тестирующих экспериментов. Однако, как оказывается,

небольшая адаптация метода построения верхней границы дает гораздо лучшие результаты благодаря встроенной технике снижения дисперсии. Аналогично части (ii) предыдущей теоремы, которая показывает, что дисперсия в испытаниях Монте-Карло снижается, если приближенное решение оказывается близко к оптимальному, мы строим рекурсивную процедуру с встроенным контролем отклонения. Идея проста: при наличии почти оптимальной стратегии $\pi = (\pi_t)_{t=0}^{T-1}$ заменить максимизацию в рекуррентных соотношениях (7.8) и (7.9) точным выбором управления в соответствии со стратегией $\pi = (\pi_t)_{t=0}^{T-1}$.

По заданной последовательности $\varphi = (\varphi_t)_{t=1}^T$, удовлетворяющей (7.6) и (7.7), введем случайные функции $(\underline{v}_t^{\pi, \varphi})_{t=0}^T$

$$\underline{v}_t^{\pi, \varphi}: P \times \mathbf{R}^d \times \Omega \rightarrow \mathbf{R}, \quad t = 0, \dots, T,$$

которые определяются рекуррентным образом для $t = T, \dots, 1$,

$$\underline{v}_T^{\pi, \varphi}(p, z) = r_T(p, z), \quad (7.20)$$

$$\begin{aligned} \underline{v}_t^{\pi, \varphi}(p, z) &= r_t(p, z, \pi_t(p, z)) + \varphi_{t+1}(p, z, \pi_t(p, z)) \\ &\quad + \underline{v}_{t+1}^{\pi, \varphi}(\alpha(p, \pi_t(p, z)), W_{t+1}z). \end{aligned} \quad (7.21)$$

Для $(\underline{v}_t^{\pi, \varphi})_{t=0}^T$ имеет место следующая теорема.

Теорема 2. (i) Для заданной последовательности $\varphi = (\varphi_t)_{t=1}^T$ вида (7.5), удовлетворяющей (7.6) и (7.7), и стратегии $\pi = (\pi_t)_{t=0}^{T-1}$ введем $(\underline{v}_t^{\pi, \varphi})_{t=0}^T$ с помощью (7.20), (7.21). Тогда имеет место

$$v_t^{\pi}(p, z) = \mathbf{E}(\underline{v}_t^{\pi, \varphi}(p, z)) \quad \text{для всех } t = 0, \dots, T, \quad p \in P, \quad z \in \mathbf{R}^d. \quad (7.22)$$

(ii) По заданным значениям цен $(v_t^{\pi^*})_{t=0}^T$ оптимальной стратегии $\pi^* = (\pi_t^*)_{t=0}^{T-1}$ определим $(\varphi_t^*)_{t=1}^T$ согласно (7.11). Тогда отображения $(\varphi_t^*)_{t=1}^T$ удовлетворяют (7.5)–(7.7), а соотношение (7.22) для них переходит в детерминистическое равенство:

$$v_t^{\pi^*}(p, z) = \underline{v}_t^{\pi^*, \varphi^*}(p, z), \quad \text{для всех } t = 0, \dots, T, \quad p \in P, \quad z \in \mathbf{R}^d. \quad (7.23)$$

Доказательство. (i) Цена $(v_t^{\pi})_{t=0}^T$ стратегии $\pi = (\pi_t)_{t=0}^{T-1}$ удовлетворяет рекуррентным соотношениям (7.2). С учетом (7.6) получаем

$$\begin{aligned} v_t^{\pi}(p, z) &= \mathbf{E}(r_t(p, z, \pi_t(p, z)) + \varphi_{t+1}(p, z, \pi_t(p, z))) \\ &\quad + \mathbf{E}(\underline{v}_{t+1}^{\pi, \varphi}(\alpha(p, \pi_t(p, z)), W_{t+1}z)). \end{aligned} \quad (7.24)$$

Докажем утверждение (7.22) по индукции. Для $t = T$ утверждение (7.22) справедливо в силу базиса обратных индукций в (7.1) и (7.20)

$$v_T^{\pi} = r_T = \underline{v}_T^{\pi, \varphi}. \quad (7.25)$$

Используя предположение индукции

$$v_{t+1}^\pi(p, z) = \mathbf{E}(\underline{v}_{t+1}^{\pi, \varphi}(p, z)) \quad \text{для всех } p \in P, z \in \mathbf{R}^d,$$

и (7.7), получаем, что

$$v_{t+1}^\pi(\alpha(p, \pi_t(p, z)), W_{t+1}z) = \mathbf{E}(\underline{v}_{t+1}^{\pi, \varphi}(\alpha(p, \pi_t(p, z)), W_{t+1}z) | W_{t+1})$$

для всех $p \in P, z \in \mathbf{R}^d$. Отсюда, получаем в (7.24) равенство

$$\begin{aligned} v_t^\pi(p, z) &= \mathbf{E}(r_t(p, z, \pi_t(p, z)) + \varphi_{t+1}(p, z, \pi_t(p, z))) \\ &\quad + \mathbf{E}(\mathbf{E}(\underline{v}_{t+1}^{\pi, \varphi}(\alpha(p, \pi_t(p, z))W_{t+1}z) | W_{t+1})) \\ &= \mathbf{E}(r_t(p, z, \pi_t(p, z)) + \varphi_{t+1}(p, z, \pi_t(p, z)) \\ &\quad + \underline{v}_{t+1}^{\pi, \varphi}(\alpha(p, \pi_t(p, z))W_{t+1}z)). \end{aligned}$$

Утверждение (7.22) следует из рекуррентных соотношений (7.21).

(ii) Утверждение (7.23) докажем также по индукции, базис которой справедлив в силу (7.20). Из предположения индукции $v_{t+1}^{\pi^*, \varphi^*} = \underline{v}_{t+1}^{\pi^*, \varphi^*}$ следует, что для всех $p \in P, z \in \mathbf{R}^d$ и $a \in A$ имеет место

$$\begin{aligned} r_t(p, z, a) + \varphi_{t+1}^*(p, z, a) + \underline{v}_{t+1}^{\pi^*, \varphi^*}(\alpha(p, a), W_{t+1}z) \\ = r_t(p, z, a) + \varphi_{t+1}^*(p, z, a) + v_{t+1}^{\pi^*, \varphi^*}(\alpha(p, a), W_{t+1}z). \end{aligned} \quad (7.26)$$

Подставляя сюда (7.11), получаем, что для всех $p \in P, z \in \mathbf{R}^d$ и $a \in A$ справедливо

$$\begin{aligned} r_t(p, z, a) + \varphi_{t+1}^*(p, z, a) + v_{t+1}^{\pi^*, \varphi^*}(\alpha(p, a), W_{t+1}z) \\ = r_t(p, z, a) + \mathbf{E}(v_{t+1}^{\pi^*, \varphi^*}(\alpha(p, a), W_{t+1}z)) \\ - v_{t+1}^{\pi^*, \varphi^*}(\alpha(p, a), W_{t+1}z) + v_{t+1}^{\pi^*, \varphi^*}(\alpha(p, a), W_{t+1}z) \\ = r_t(p, z, a) + \mathbf{E}(v_{t+1}^{\pi^*, \varphi^*}(\alpha(p, a), W_{t+1}z)). \end{aligned} \quad (7.27)$$

Для доказательства требуемого утверждения (7.23) остается заменить в рекуррентных соотношениях (7.21) выражение (7.26) на (7.27):

$$\begin{aligned} \underline{v}_t^{\pi^*, \varphi^*}(p, z) &= r_t(p, z, \pi_t^*(p, z)) + \varphi_{t+1}^*(p, z, \pi_t^*(p, z)) \\ &\quad + \underline{v}_{t+1}^{\pi^*, \varphi^*}(\alpha(p, \pi_t^*(p, z)), W_{t+1}z) \\ &= r_t(p, z, \pi_t^*(p, z)) + \mathbf{E}(v_{t+1}^{\pi^*, \varphi^*}(\alpha(p, \pi_t^*(p, z)), W_{t+1}z)) \\ &= v_t^{\pi^*, \varphi^*}(p, z), \quad p \in P, \quad z \in \mathbf{R}^d, \quad a \in A. \end{aligned}$$

Теорема 2 доказана.

Построение оценки для нижней границы основано на уже имеющихся реализациях $(\varphi_t)_{t=1}^T$ из (7.19). Подытожим эту процедуру следующим образом.

Оценка для нижней границы.

1) По заданным функциям цены $(\tilde{v}_t)_{t=0}^T$ и соответствующей стратегии $\tilde{\pi} = (\tilde{\pi}_t)_{t=0}^{T-1}$, выбираем $\varphi = (\varphi_t)_{t=0}^{T-1}$ как в (7.19).

2) Для заданного количества испытаний $K \in \mathbf{N}$ в методе Монте-Карло, строим для $k = 1, \dots, K$ независимые реализации $(W_t(\omega_k))_{t=1}^T$ возмущений.

3) Начиная с $z_0^k := z_0 \in \mathbf{R}^d$, определяем для $k = 1, \dots, K$ рекуррентным образом траектории $(z_t^k)_{t=0}^T$:

$$z_{t+1}^k = W_{t+1}(\omega_k)z_t^k, \quad t = 0, \dots, T-1,$$

а также реализации

$$\varphi_{t+1}(p, z_t^k, a)(\omega_k), \quad t = 0, \dots, T-1, \quad k = 1, \dots, K.$$

4) Для каждого $k = 1, \dots, K$: при $t = T$ задаем

$$\underline{v}_T^{\tilde{\pi}, \varphi}(p, z_T^k) = r_T(p, z_T^k) \quad \text{для всех } p \in P$$

для $t = T-1, \dots, 0$ и всех $p \in P$ вычисляем выражения

$$\begin{aligned} \underline{v}_t^{\tilde{\pi}, \varphi}(p, z_t^k) &= r_t(p, z_t^k, \tilde{\pi}_t(p, z_t^k)) + \varphi_{t+1}(p, z_t^k, \tilde{\pi}_t(p, z_t^k))(\omega_k) \\ &\quad + \underline{v}_{t+1}^{\tilde{\pi}, \varphi}(\alpha(p, \tilde{\pi}_t(p, z_t^k)), z_{t+1}^k). \end{aligned} \quad (7.28)$$

Сохраняем значения цены как $\underline{v}_0^{\tilde{\pi}, \varphi}(p, z_0^k)$ для $k = 1, \dots, K$, $p \in P$.

5) Вычисляем выборочное среднее $K^{-1} \sum_{k=1}^K \underline{v}_0^{\tilde{\pi}, \varphi}(p, z_0^k)$ и используем его нижнюю доверительную границу, чтобы оценить $v_0^{\pi^*}(p, z_0)$ для $p \in P$ снизу.

8. Примеры. В литературе оценивание верхней границы для задач об оптимальной остановке в значительной степени опирается на понятие мартингальной двойственности и рассматривалось в [14], [24] и [16]. С вычислительной точки зрения, достижение достаточно точных верхних границ эквивалентно построению «хорошего» мартингала, и в [1] был предложен осуществимый алгоритм для этого. Методы построения верхних оценок были впоследствии распространены на более общий класс задач об оптимальной многократной остановке в [19], [26] и [17]. Наконец, работа [25] обобщает эту технику на более широкий класс задач стохастического управления в дискретном времени. Сочетание методов построения верхних и нижних границ известно как *прямое двойственное Монте-Карло моделирование* (*primal-dual simulation*). В этом разделе мы сравним предложенный нами метод с результатами, полученными с использованием стандартной регрессии на основе метода наименьших квадратов.

Далее мы рассмотрим две конкретные марковские задачи принятия решений, первая из которых является стандартной задачей об оптимальной остановке, а вторая — задачей о многократной остановке. На этих примерах мы продемонстрируем работу предложенного в разд. 6 метода аппроксимации функции цены и описанного в разд. 7 метода диагностики приближенного решения. Задачи об оптимальной остановке являются важным подклассом марковских задач принятия решений (см. гл. 10 и 11 в [2]), для которых построение верхних оценок на основе двойственности хорошо изучено. В качестве иллюстрации нашего подхода, в разд. 8.1 мы получаем границы для важной на практике задачи об оптимальной остановке в дискретном времени — вычислении цены *бермудского* опциона пут (*Bermudan put option*). В разд. 8.2, мы используем предложенные методы для построения численных решений в задаче об оптимальной многократной остановке — расчете цены «свинг» опциона (*swing option*) (см. [6]). Свинг опцион позволяет держателю купить определенное количество базового актива по заранее оговоренной цене более чем один раз до истечения срока действия опциона, однако существует ограничение на максимальное количество раз, которое это можно сделать. Свинг опционы являются преобладающими на товарно-сырьевых рынках, особенно в энергетическом секторе.

Для обоих приложений мы будем рассматривать эволюцию дисконтированной цены актива $(S_t)_{t=0}^T$ в дискретном времени относительно риск-нейтральной меры. Динамика $(S_t)_{t=0}^T$ зависит от типа актива. Для бермудского опциона процесс дисконтированной цены $(S_t)_{t=0}^T$ моделируется в виде мартингала по риск-нейтральной мере. Для «свинг» опциона, мы предполагаем, что процесс цены $(S_t)_{t=0}^T$ моделируется экспонентой процесса Орнштейна–Уленбека, чтобы объяснить свойство возвратности к среднему значению, естественным образом ожидаемое для цен на сырьевые товары.

Логарифм цены $(\tilde{Z}_t)_{t=0}^T$ задает непрерывную компоненту нашей динамики состояний. На практике обычно необходимы дальнейшие преобразования пространства состояний, чтобы динамика стала линейной. Так, в большинстве случаев для описания эволюции непрерывной компоненты состояния необходимо дополнительно рассмотреть константу 1:

$$Z_t = \begin{bmatrix} \tilde{Z}_t \\ 1 \end{bmatrix}, \quad t = 0, \dots, T.$$

В этом представлении состояние системы следует мультипликативной динамике

$$Z_{t+1} = W_{t+1} Z_t, \quad t = 0, \dots, T-1,$$

с независимыми и одинаково распределенными случайными матрицами возмущений $(W_t)_{t=1}^T$, распределения которых отражают модель рассматриваемого процесса.

Выбор сетки является ключевым моментом в алгоритме. Для процессов с многомерным пространством состояний удобным способом ее построения является моделирование и хранение достаточного числа траекторий $(Z_t)_{t=0}^{k_p T}$ нужной длины $k_p T \in \mathbf{N}$. В наших примерах мы использовали количество шагов равное удвоенной длине временного горизонта ($k_p = 2$). Для распределения возмущений аппроксимация дискретными распределениями строится также путем моделирования выборки $(W(k))_{k=1}^n$ независимых реализаций, к которым затем применяются все необходимые шаги из разд. 6 и диагностика методами Монте-Карло. Для вычисления границ мы используем доверительные интервалы, основанные на K сгенерированных траекториях. Более точно, интервалы задаются согласно

$$\left[\underline{\mu} - \Phi^{-1}\left(1 - \frac{x}{2}\right) \frac{\underline{\sigma}}{\sqrt{K}}, \quad \bar{\mu} + \Phi^{-1}\left(1 - \frac{x}{2}\right) \frac{\bar{\sigma}}{\sqrt{K}} \right], \quad (8.1)$$

где $1 - x$ обозначает доверительный уровень, а $(\underline{\mu}, \underline{\sigma})$ и $(\bar{\mu}, \bar{\sigma})$ обозначают выборочные среднее и стандартное отклонение для $(\underline{v}_0^{\tilde{\pi}, \varphi}(p, z_0^k))_{k=1}^K$ и $(\bar{v}_0^{\varphi}(p, z_0^k))_{k=1}^K$ соответственно.

8.1. Бермудский опцион пут. Этот вид опциона позволяет держателю продать базовый актив по заранее оговоренной цене исполнения на дискретном наборе дат исполнения до и включая дату истечения срока опциона. Справедливая цена бермудского опциона пут определяется согласно

$$\sup_{\tau} \mathbf{E} [(K e^{-\rho \tau} - S_{\tau})^+],$$

где τ пробегает множество всех $\{0, \dots, T\}$ -значных моментов останова. Первым делом, представим эту задачу в виде управления системой переключающегося типа. Множество положений $P = \{1, 2\}$ указывает на то, исполняется ли опцион ($p = 1$) или нет ($p = 2$). Множество управлений $A = \{1, 2\}$ представляет, соответственно, выбор между исполнением ($a = 1$) и неисполнением ($a = 2$). Управление α дискретной компонентой пространства состояний задается матрицей

$$(\alpha(p, a))_{p,a=1}^2 \sim \begin{bmatrix} \alpha(1, 1) & \alpha(1, 2) \\ \alpha(2, 1) & \alpha(2, 2) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

и гарантирует, что $p = 1$ является поглощающим состоянием. Непрерывная компонента пространства состояний эволюционирует согласно

$$\tilde{Z}_{t+1} = \tilde{Z}_t + \gamma + \beta \varepsilon_{t=1}^T, \quad \tilde{Z}_0 = \ln S_0, \quad (8.2)$$

где $(\varepsilon)_{t=1}^T$ являются независимыми стандартными нормально распределенными случайными величинами. Мы полагаем параметры равными $\gamma = -\sigma^2 \Delta / 2$ и $\beta = \sigma \sqrt{\Delta}$, где $\Delta > 0$ есть продолжительность времени (в

годах) от момента времени t до $t+1$ и $\sigma > 0$ представляет волатильность процесса, измеренную на годовом масштабе. Для определенного таким образом процесса цены возмущения задаются матрицей

$$W_{t+1} = \begin{bmatrix} 1 & \gamma + \beta \varepsilon_{t+1} \\ 0 & 1 \end{bmatrix}, \quad t = 0, \dots, T-1.$$

Таблица 8.1. Бермудский опцион пут с $K = 40$.

S_0	σ	Время до истечения опциона	доверительный интервал	LSM среднее	LSM отклонение
36	0.2	1	[4.4763, 4.4768]	4.472	.0100
36	0.2	2	[4.8296, 4.8312]	4.821	.0120
36	0.4	1	[7.0989, 7.0992]	7.091	.0200
36	0.4	2	[8.4965, 8.4968]	8.488	.0240
38	0.2	1	[3.2481, 3.2489]	3.244	.0090
38	0.2	2	[3.7355, 3.7370]	3.735	.0110
38	0.4	1	[6.1451, 6.1452]	6.139	.0190
38	0.4	2	[7.6580, 7.6583]	7.669	.0220
40	0.2	1	[2.3119, 2.3129]	2.313	.0090
40	0.2	2	[2.8765, 2.8776]	2.879	.0100
40	0.4	1	[5.3093, 5.3094]	5.308	.0180
40	0.4	2	[6.9075, 6.9077]	6.921	.0220
42	0.2	1	[1.6150, 1.6158]	1.617	.0070
42	0.2	2	[2.2053, 2.2060]	2.206	.0100
42	0.4	1	[4.5797, 4.5798]	4.588	.0170
42	0.4	2	[6.2351, 6.2354]	6.243	.0210
44	0.2	1	[1.1081, 1.1087]	1.118	.0070
44	0.2	2	[1.6836, 1.6843]	1.675	.0090
44	0.4	1	[3.9449, 3.9450]	3.957	.0170
44	0.4	2	[5.6324, 5.6326]	5.622	.0210

Результаты получены с использованием сетки размера $m = 1024$ и возмущений размера $n = 4096$. Диагностика решения проведена для 1024 реализаций траекторий, построен 99% доверительный интервал для значения $x = 0.01$ в (8.1). Для сравнения, в последних двух колонках приведены средние значения и стандартные ошибки из работы [18] (с точностью до трех знаков после запятой), полученные методами Монте-Карло и наименьших квадратов.

Далее, определим на пространстве состояний непрерывной компоненты $\mathbf{R}^2$ функции выигрыша и терминальные функции. Для состояния $(z^{(1)}, z^{(2)}) \in \mathbf{R}^2$ в момент времени $t = 0, \dots, T-1$ управление $a \in \{1, 2\}$ приводит к выигрышу

$$r_t(p, (z^{(1)}, z^{(2)}), a) = (Ke^{-\rho t} - e^{z^{(1)}})^+(p - \alpha(p, a)) \quad (8.3)$$

для всех $p \in P$ и $a \in A$. В последний момент времени T мы полагаем, что опцион пут исполняется автоматически, что дает терминальное значение

$$r_T(p, (z^{(1)}, z^{(2)})) = (Ke^{-\rho T} - e^{z^{(1)}})^+(p - \alpha(p, 1)) \quad (8.4)$$

для всех $p \in P$.

Отметим, что функции выигрыша, как и терминальные, не являются выпуклыми по непрерывной компоненте $z = (z^{(1)}, z^{(2)})$. Поэтому, согласно предложенной методике, мы представляем их в виде разности двух выпуклых функций

$$r_t(p, \cdot, a) = \check{r}_t(p, \cdot, a) - \hat{r}_t(p, \cdot, a), \quad (8.5)$$

$$r_T(p, \cdot) = \check{r}_T(p, \cdot) - \hat{r}_T(p, \cdot), \quad (8.6)$$

заданных как

$$\begin{aligned} \check{r}_t(p, (z^{(1)}, z^{(2)}), a) &= (e^{z^{(1)}} - Ke^{-\rho t})^+(p - \alpha(p, a)), \\ \hat{r}_t(p, (z^{(1)}, z^{(2)}), a) &= (e^{z^{(1)}} - Ke^{-\rho t})(p - \alpha(p, a)), \\ \check{r}_T(p, (z^{(1)}, z^{(2)})) &= (e^{z^{(1)}} - Ke^{-\rho T})^+(p - \alpha(p, 1)), \\ \hat{r}_T(p, (z^{(1)}, z^{(2)})) &= (e^{z^{(1)}} - Ke^{-\rho T})(p - \alpha(p, 1)) \end{aligned}$$

для всех $p \in P, a \in A$ и $(z^{(1)}, z^{(2)}) \in \mathbf{R}^2$.

Для полученных результатов проведено сравнение с известными в литературе мало смещенными оценками для опционов с безрисковой ставкой 0.06 и ценой исполнения равной 40. Результаты приведены в таблице 8.1 для различных комбинаций значений начальной цены, волатильности и времени до истечения опциона.

8.2. «Свинг» опцион. Мы рассматриваем специальный случай «свинг» опциона, удовлетворяющего условию единичного периода рефракции (*unit-time refraction period*), которое ограничивает держателя актива исполнением только одного контракта в любой момент времени. Если, как и ранее, $(S_t)_{t=0}^T$ обозначает дисконтированную цену базового актива, справедливая цена «свинг» опциона с N датами исполнения определяется согласно

$$\sup_{0 \leq \tau_1 < \dots < \tau_N \leq T} \mathbf{E} \left[\sum_{k=1}^N (S_{\tau_k} - Ke^{-\rho \tau_k})^+ \right],$$

где супремум берется по всем моментам остановки $\tau_1, \dots, \tau_N$ со значениями в $\{0, \dots, T\}$. Чтобы представить эту задачу в виде управления системой переключающегося типа, используем множество положений $P = \{1, \dots, N+1\}$, указывающих на то, сколько прав исполнения осталось, т.е. $p \in P$ означает, что осталось еще $p-1$ право исполнения. Множество управлений $A = \{1, 2\}$ такое же, как и в случае бермудского опциона пут, а матрица управления α имеет вид

$$(\alpha(p, a))_{p,a} \sim \begin{bmatrix} \alpha(1, 1) & \alpha(1, 2) \\ \alpha(2, 1) & \alpha(2, 2) \\ \alpha(3, 1) & \alpha(3, 2) \\ \dots & \dots \\ \alpha(N+1, 1) & \alpha(N+1, 2) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 2 & 3 \\ \dots & \dots \\ N & N+1 \end{bmatrix}.$$

Дисконтированная цена моделируется экспонентой процесса со свойством возвратности к своему среднему, с параметром возвратности $\kappa \in [0, 1)$, долгосрочным средним $\mu > 0$ и волатильностью $\sigma > 0$. Поэтому предполагается, что логарифм дисконтированной цены эволюционирует согласно

$$\tilde{Z}_{t+1} = (1 - \kappa)(\tilde{Z}_t - \mu) + \mu + \sigma \varepsilon_{t+1}, \quad \tilde{Z}_0 = \ln S_0 \quad (8.7)$$

с матрицами возмущений

$$W_{t+1} = \begin{bmatrix} (1 - \kappa) & \kappa\mu + \sigma\varepsilon_{t+1} \\ 0 & 1 \end{bmatrix}, \quad t = 0, \dots, T - 1.$$

Функции цены и терминальное значение задаются с помощью

$$r_t(p, (z^{(1)}, z^{(2)}), a) = (e^{z^{(1)}} - Ke^{-\rho t})^+(p - \alpha(p, a)), \quad t = 0, \dots, T - 1, \quad (8.8)$$

и

$$r_T(p, (z^{(1)}, z^{(2)})) = (e^{z^{(1)}} - Ke^{-\rho T})^+(p - \alpha(p, 1)) \quad (8.9)$$

соответственно для всех $p \in P$ и $a \in A$.

В таблице 8.2 проведено сравнение наших результатов с построенными границами из работы [19], в которой предполагалось, что базовый актив следует динамике (8.7) с параметрами

$$\rho = 0, \quad \kappa = 0.9, \quad \mu = 0, \quad \sigma = 0.5, \quad S_0 = 1, \quad K = 0 \quad \text{и} \quad T = 1000.$$

Таблица 8.2. Численные результаты для «свинг» опциона.

Положения системы ($N + 1$)	CSS	МН
	доверительный интервал	доверительный интервал
2	[4.737, 4.761]	[4.773, 4.794]
3	[9.005, 9.031]	[9.016, 9.091]
4	[13.001, 13.026]	[12.959, 13.100]
5	[16.805, 16.830]	[16.773, 16.906]
6	[20.465, 20.491]	[20.439, 20.580]
11	[37.339, 37.363]	[37.305, 37.540]
16	[52.694, 52.718]	[52.670, 53.009]
21	[67.070, 67.095]	[67.050, 67.525]
31	[93.811, 93.835]	[93.662, 94.519]
41	[118.639, 118.663]	[118.353, 119.625]
51	[142.059, 142.084]	[141.703, 143.360]
61	[164.368, 164.392]	[163.960, 166.037]
71	[185.757, 185.781]	[185.335, 187.729]
81	[206.362, 206.386]	[205.844, 208.702]
91	[226.284, 226.308]	[225.676, 228.985]
101	[245.601, 245.625]	[244.910, 248.651]

Результаты получены с использованием сетки размера $m = 1024$ и возмущений размера $n = 1024$. Диагностика решения проведена для 1024 реализаций траекторий, построен 99% доверительный интервал для значения $x = 0.01$. Колонка МН соответствует результатам из [19].

9. Заключение. В настоящей работе был представлен новый класс алгоритмов для решения стохастических задач переключающегося типа, состояние системы в которых следует линейной динамике. Предложенная методика применима к задачам большой размерности, показывает высокую вычислительную эффективность и хорошую точность. Что еще более важно, нами был адаптирован прямой двойственный подход оценивания расстояния от приближенных решений до оптимальных на основе методов Монте-Карло. Это дает надежный инструмент диагностики и оценки качества для постериорного обоснования численной аппроксимации. Авторы полагают, что подобная комбинация эффективных численных схем с последующей диагностической проверкой может быть очень полезной в практических приложениях, а рассматриваемый подход может помочь в развитии и обосновании дальнейших приближенных методов. В частности, в деталях могут быть изучены естественные обобщения предложенной схемы (например, с кусочно-линейной динамикой вместо линейной). Мы оставляем это многообещающее направление для дальнейших исследований.

СПИСОК ЛИТЕРАТУРЫ

1. Andersen L., Broadie M. A primal-dual simulation algorithm for pricing multidimensional American options. — *Management Sci.*, 2004, v. 50, № 9, p. 1222–1234.
2. Bäuerle N., Rieder U. *Markov Decision Processes with Applications to Finance*. Heidelberg: Springer, 2011, 388 p.
3. Belomestny N., Kolodko A., Schoenmakers J. Regression methods for stochastic control problems and their convergence analysis. — *SIAM J. Control Optim.*, 2010, v. 48, № 5, p. 3562–3588.
4. Bertsekas D. P. *Dynamic Programming and Optimal Control*, 3rd ed., Belmont, MA: Athena Scientific, 2005, 558 p.
5. Bertsekas D. P., Tsitsiklis J. N. *Neuro-Dynamic Programming*. Belmont, MA: Athena Scientific, 1996, 512 p.
6. Carmona R., Touzi N. Optimal multiple stopping and valuation of swing options. — *Mathem. Finance*, 2008, v. 18, № 2, p. 239–268.
7. Carriere J. F. Valuation of the early-exercise price for options using simulations and nonparametric regression. — *Insurance: Mathematics and Economics*, 1996, v. 19, № 1, p. 19–30.
8. Clement E., Lamberton D., Protter P. An analysis of the Longstaff-Schwartz algorithm for American option pricing. — *Finance and Stochastics*, 2002, v. 6, № 4, p. 449–471.
9. Egloff D. Monte Carlo algorithms for optimal stopping and statistical learning. — *Appl. Probab.*, 2005, v. 15, № 2, p. 1396–1432.
10. Egloff D., Kohler M., Todorovic N. A dynamic look-ahead Monte Carlo algorithm. — *Ann. Appl. Probab.*, 2007, v. 17, № 4, p. 1138–1171.
11. Fan J., Gijbels I. *Local Polynomial Modelling and Its Applications*. Boca Raton: Chapman & Hall/CRC, 1996, 360 p.
12. Feinberg E. A., Shwartz A. *Handbook of Markov Decision Processes*. Norwell, MA: Kluwer, 2002, 565 p.
13. Glasserman P. *Monte Carlo Methods in Financial Engineering*. New York: Springer, 2003, 596 p.

14. *Haugh M., Kogan L.* Pricing American options: A duality approach. — *Oper. Res.*, 2004, v. 52, № 2, p. 258–270.
15. *Hinz J.* Optimal stochastic switching under convexity assumptions. — *SIAM J. Control Optim.*, 2014, v. 52, № 1, p. 164–188.
16. *Jamshidian F.* Numeraire-invariant option pricing & American, Bermudan, and trigger stream rollover. July 2004. Version 1.6.
17. *Joshi M.S., Yap N.* The multiplicative dual for multiple-exercise options. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2430558, 2014.
18. *Longstaff F., Schwartz E.* Valuing American options by simulation: a simple least-squares approach. — *Rev. Financial Studies*, 2001, v. 14, № 1, p. 113–147.
19. *Meinshausen N., Hambly B.M.* Monte Carlo methods for the valuation of multiple-exercise options. *Mathem. Finance*, 2004, v. 14, № 4, p. 557–583.
20. *Ormoneit D., Glynn P.* Kernel-based reinforcement learning. — *Machine Learning*, 2002, v. 49, № 2–3, p. 161–178.
21. *Ormoneit D., Glynn P.* Kernel-based reinforcement learning in average-cost problems. — *IEEE Transactions in Automatic Control*, 2002, v. 47, № 10, p. 1624–1636.
22. *Powell W.B.* Approximate Dynamic Programming: Solving the Curses of Dimensionality. New York: Wiley, 2007, 656 p.
23. *Puterman M.L.* Markov Decision Processes: Discrete Stochastic Dynamic Programming. New York: Wiley, 1994, 649 p.
24. *Rogers L.C.G.* Monte Carlo valuation of American options. — *Mathem. Finance*, 2002, v. 12, № 13, p. 271–286.
25. *Rogers L.C.G.* Pathwise stochastic optimal control. — *SIAM J. Control Optimisation*, 2007, v. 46, № 3, p. 1116–1132.
26. *Schoenmakers J.* A pure martingale dual for multiple stopping. — *Finance and Stochastics*, 2012, v. 16, № 2, p. 319–334.
27. *Stentoft L.* Convergence of the least squares Monte Carlo approach to American option valuation. — *Management Science*, 2001, v. 50, № 9, p. 576–611.
28. *Tsitsiklis J.N., Van Roy B.* Regression methods for pricing complex American-style options. — *IEEE Transactions on Neural Networks*, 2001, v. 12, № 4, p. 694–703.
29. *Tsitsiklis J.N., Van Roy B.* Optimal stopping of Markov processes: Hilbert space, theory, approximation algorithms, and an application to pricing high-dimensional financial derivatives. — *IEEE Transactions on Automatic Control*, 1999, v. 44, № 10, p. 1840–1851.

Поступила в редакцию
24.XII.2014

Исправленный вариант
15.VI.2015