УДК 519.2

*Д. И. Кудрявцев*

# Управление марковским процессом в задаче с ограничениями

В настоящей работе рассматривается условная задача динамического программирования с затухающим действием в случае любого конечного числа ограничений и конечных множеств состояний и управлений. Безусловная задача хорошо известна.

Условную задачу с одним ограничением рассмотрел Фрид [1]. К сожалению, подход к задаче, использованный в его работе, остался незамеченным из-за большого числа чисто технических приемов. В этой работе, являющейся продолжением [1], ставится цель не только обобщить задачу на случай произвольного конечного числа ограничений, но и выделить основную идею, не затемняя ее второстепенными выкладками. Поэтому рассматриваются лишь конечные множества состояний и управлений, хотя данный подход позволяет получать аналогичные результаты и в более общих случаях (см., например, [1]).

Подход к решению этой задачи, как и в [1], основан на принципе Лагранжа, что позволяет получить необходимые и достаточные условия оптимальности стратегии, а также выделить определенный класс стратегий, в котором существует хотя бы одна оптимальная. Основной результат этой работы заключается в следующем: существует оптимальная стратегия, являющаяся взвесью не более чем $m+1$-й стационарной стратегии, где $m$ — число ограничений.

Пусть $S$ и $A$ — непустые конечные множества. Пусть $f_0, f_1, \ldots, f_m$ — ограниченные функции на $S \times A$, а $q$ ставит в соответствие каждой паре $(s, a) \in S \times A$ вероятностную меру $q(\cdot \mid s, a)$, определенную на множестве состояний $S$, в соответствии с которой происходит переход из состояния $s_i$ в $s_{i+1}$. Положим $\mathcal{H}_n = S_1 \times A_1 \times \ldots \times S_n$, где $S_i \equiv S$, $A_i \equiv A$ для любого $i = 1, \ldots, n$. Стратегией $\pi$ будем называть последовательность функций $\{\pi^n(\cdot \mid h_n)\}$, $n = 1, 2, \ldots$, $h_n \in \mathcal{H}_n$, каждая из которых является переходной функцией из $\mathcal{H}_n$ в $A$. Класс всех возможных стратегий обозначим через $\Pi$. Введем понятие взвешенной стратегии.

Определение. *$\alpha^n$-взвесью стратегий $\pi_1, \ldots, \pi_n$ с весом $\alpha^n = (\alpha_1, \ldots, \alpha_n)$, где $\sum_{i=1}^{n} \alpha_i = 1$, $\alpha_i \geqslant 0$, назовем стратегию $\pi(\alpha^n, \pi_1, \ldots, \pi_n)$, которую определим как последовательность переходных функций $\{\pi^l(\alpha^n, \pi_1, \ldots, \pi_n)\}$, $l = 1, 2, \ldots$, каждая из которых* $\pi^l(\alpha^n, \pi_1, \ldots, \pi_n) = \sum_{i=1}^{n} \alpha_i \pi_i^l(\cdot \mid h_l)$.

Таким образом, $\Pi$ — выпуклое множество.

Для любого $n \geqslant 2$ и $s_1 \in S$ распределение $q$ и стратегия $\pi$ определяют меру на $\mathcal{H}_n$. В силу теоремы Ионеску — Тулча [2] при фиксированном $s_1$ распределение $q$ и стратегия $\pi$ определяют меру $P_{s_1}^{\pi}$ на пространстве траекторий $S_1 \times A_1 \times S_2 \times A_2 \times \ldots$, которое обозначим $\Omega$. Математическое ожидание по этой мере будем обозначать $M_{s_1}^{\pi}$. В дальнейшем, полагая $s_1$ всегда фиксированным, будем опускать этот индекс.

Замечание 1. Пусть $F$ — ограниченная функция на пространстве траекторий $\Omega$ и $n$ стратегий $\pi_1, \ldots, \pi_n$, тогда $M^{\lambda}F = \sum_{i=1}^{n} \alpha_i M^{\pi_i} F$, где $\lambda = \pi(\alpha^n, \pi_1, \ldots, \pi_n)$.

Рандомизированной марковской стратегией будем называть такую стратегию, в которой $\pi^n$ зависит только от состояния $s_n$, в котором система находится в текущий момент времени. Марковская стратегия — последовательность $\{g_i\}$, $i = 1, 2, \ldots$, где каждая $g_i$ — измеримая функция, отображающая $S$ в $A$, и $g_n(s)$ — управление, которое выбираем на $n$-м шагу, если $n$-е состояние — $s$. Стационарная стратегия — марковская стратегия, в которой $g_n = g$ для всех $n$, где $g$ — некоторая измеримая функция, отображающая $S$ в $A$. Понятно, что класс всех стационарных стратегий в нашем случае ($S$ и $A$ конечны) конечен.

Рассмотрим $\beta \in [0, 1)$. Определим функции $F_i^\beta$, $i = 0, \ldots, m$, на пространстве траекторий $\Omega$: $F_i^\beta = \sum_{j=1}^{\infty} \beta^{j-1} f_i(s_j, a_j)$. Далее будем опускать индекс $\beta$, полагая его везде фиксированным. Заметим, что для любого фиксированного $\beta \in [0, 1)$ функции $F_i$ ограничены.

Зафиксируем $K = (K_1, \ldots, K_m)$, где $K_i < \infty$ для любого $i = 1, \ldots, m$. Будем говорить, что стратегия $\pi$ принадлежит классу $\mathcal{D}$, если $M^\pi F_i \leqslant K_i$ для любого $i = 1, \ldots, m$. Предположим, что клас $\mathcal{D}$ непуст, и назовем ценой игры в точке $s_1$ число $v = \sup_{\pi \in \mathcal{D}} M^\pi F_0$. Стратегию $\pi_0 \in \mathcal{D}$, на которой достигается $v$ (если она существует), назовем оптимальной.

Поставим перед собой задачу исследовать существование оптимальной стратегии, а также выяснить, к какому классу стратегий могла бы принадлежать $\pi_0$. Существенно, что в отличие от безусловной задачи, где доказывалось существование оптимальной стратегии сразу во всех точках, в условной задаче может не существовать стратегии, оптимальной даже в двух точках (см. пример 2 из [1]).

Если зафиксировать функцию $F$ на $\Omega$, то функция $M^\pi F$ как функция от стратегии $\pi$ является линейной и ограниченной, если ограничена функция $F$. Как уже было сказано, множество всех стратегий $\Pi$ — выпуклое множество. Поэтому поставленная задача, которую формально можно представить в виде

$$M^\pi F_0 \to \sup, \tag{1}$$

$$M^\pi F_i - K_i \leqslant 0, \quad i = 1, \ldots, m, \tag{2}$$

$$\pi \in \Pi, \tag{3}$$

является выпуклой экстремальной задачей оптимального управления.

Введем функцию Лагранжа $\mathcal{L}(\pi, \lambda, \lambda_0) = \lambda_0 M^\pi F_0 - \sum_{i=1}^{m} \lambda_i (M^\pi F_i - K_i)$, где $\lambda = (\lambda_1, \ldots, \lambda_m)$. Согласно теореме Куна — Таккера [3] справедливо:

1) если $\pi_0$ — оптимальная стратегия, то существует вектор $(\hat\lambda_0, \hat\lambda_1, \ldots, \hat\lambda_m) \neq 0$ такой, что выполняются условия

$$\hat\lambda_i \geqslant 0, \quad i = 0, \ldots, m, \tag{4}$$

$$\hat\lambda_i (M^{\pi_0} F_i - K_i) = 0, \quad i = 1, \ldots, m, \tag{5}$$

$$\mathcal{L}(\pi_0, \hat\lambda, \hat\lambda_0) = \sup_{\pi \in \Pi} \mathcal{L}(\pi, \hat\lambda, \hat\lambda_0); \tag{6}$$

2) если $\hat\lambda_0 > 0$ и $\pi_0 \in D$, то условия (4)—(6) являются достаточными для того, чтобы стратегия $\pi_0$ была оптимальной;

3) для того чтобы $\hat\lambda_0 \neq 0$ достаточно, чтобы нашлась стратегия $\bar\pi \in \Pi$ такая, что выполняются условия: $M^{\bar\pi} F_i - K_i < 0$, $i = 1, \ldots, m$.

Покажем, что найдется вектор $(1, \lambda) \in \mathbb{R}_+^{m+1}$ и стратегия $\pi_0 \in \mathcal{D}$ такие, что выполняются условия (4) — (6). Как обычно $\mathbb{R}_+^{m+1} = \{x \in \mathbb{R}^{m+1}: x_i \geqslant 0, i = 0, \ldots, m\}$. Тем самым докажем существование оптимальной стратегии и укажем класс стратегий, в котором она всегда существует.

Введем функцию $w(\lambda) = \sup_{\pi \in \Pi} \mathcal{L}(\pi, \lambda, 1)$. Так как мы ищем вектор

$(1, \hat{\lambda}) \in \mathbb{R}_{+}^{m+1}$, то условие (6) принимает вид

$$w(\hat{\lambda}) = \sup_{\pi \in \Pi} \mathscr{L}(\pi, \hat{\lambda}, 1) = \mathscr{L}(\pi_0, \hat{\lambda}, 1). \tag{7}$$

Функция $\mathscr{L}(\pi, \lambda, 1)$ линейна по $\lambda$ при любой фиксированной стратегии $\pi$. Поскольку справедлива цепочка равенств

$$F_0 - \sum_{i=1}^{m} \lambda_i (F_i - K_i) = \sum_{j=1}^{\infty} \beta^{j-1} f_0(s_j, a_j) - \sum_{i=1}^{m} \lambda_i \left( \sum_{j=1}^{\infty} \beta^{j-1} f_i(s_j, a_j) - K_i \right) =$$

$$= \sum_{j=1}^{\infty} \beta^{j-1} \left( f_0(s_j, a_j) - \sum_{i=1}^{m} \lambda_i f_i(s_j, a_j) \right) + \sum_{i=1}^{m} \lambda_i K_i,$$

функция $f_0(s, a) - \sum_{i=1}^{m} \lambda_i f_i(s, a)$ ограничена, а сумма $\sum_{i=1}^{m} \lambda_i K_i = \mathrm{const}$ при любом фиксированном $\lambda$, то согласно теореме 7 $(b)$ из [4] точная верхняя грань функции $\mathscr{L}(\pi, \lambda, 1)$ по $\pi$ достигается в любой фиксированной точке $\lambda$ на стационарной стратегии.

Поэтому для исследования функции $w(\lambda)$ достаточно рассматривать функцию $\mathscr{L}(\pi, \lambda, 1)$ только на таких стратегиях. Как было замечено, в случае конечных множеств состояний и управлений стационарных стратегий конечное число, поэтому справедливо:

1) для любого фиксированного $\lambda$ существует стационарная стратегия $\pi(\lambda)$ на которой достигается $\omega(\lambda)$. Заметим, что этой фразой стратегия $\pi(\lambda)$ определяется, вообще говоря, неоднозначно;

2) $w(\lambda)$ как функция от $\lambda$ представляет собой выпуклую вниз кусочно линейную функцию, причем число таких «кусков» не превышает числа стационарных стратегий, а значит конечно.

Рассмотрим в $\mathbb{R}^{m+1}$ график функции $w(\lambda)$, который обозначим $W$. Графиком функции $\mathscr{L}(\pi, \lambda, 1)$ для каждой фиксированной стратегии $\pi$ будет гиперплоскость, которую обозначим $T(\pi)$. Будем рассматривать также гиперплоскости $T(c) = \{(c, \lambda) : \lambda \in \mathbb{R}^m, c = \mathrm{const}\}$.

Для того чтобы лучше понять смысл дальнейших рассуждений, предположим, что имеем оптимальную стратегию $\pi_0$ в задаче (1)—(3) и точку $\hat{\lambda} \in \mathbb{R}_{+}^{m}$ такую, что $\hat{\lambda}$ удовлетворяет (5) и $\hat{\lambda}_i \neq 0$ для всех $i = 1, \ldots, m$. Тогда $M^{\pi_0} F_i - K_i = 0$, $i = 1, \ldots, m$, или, другими словами, гиперплоскость $T(\pi_0)$ совпадает с $T(\hat{c})$, если $\hat{c} = M^{\pi_0} F_0$.

Пусть теперь часть координат точки $\hat{\lambda}$ равны нулю. Тогда для тех номеров $i$, при которых $\hat{\lambda}_i \neq 0$, из условия (5) следует строгое равенство $M^{\pi_0} F_i = K_i$, а для остальных $i$ выполняются неравенства $M^{\pi_0} F_i - K_i \leqslant 0$, так как $\pi_0 \in \mathscr{D}$. Но тогда $w(\lambda) \geqslant \mathscr{L}(\pi_0, \lambda, 1) = M^{\pi_0} F_0 - \sum_{i=1}^{m} \lambda_i (M^{\pi_0} F_i - K_i) \geqslant$ $\geqslant M^{\pi_0} F_0$ для любого $\lambda \in \mathbb{R}_{+}^{m}$. Положим $\hat{c} = M^{\pi_0} F_0$, тогда $w(\hat{\lambda}) = \mathscr{L}(\pi_0, \hat{\lambda}, 1) = M^{\pi_0} F_0 = \hat{c}$, т. е. $w(\hat{\lambda}) = \hat{c}$ и $w(\lambda) \geqslant \hat{c}$ для любого $\lambda \in \mathbb{R}_{+}^{m}$. Другими словами, график $W$ лежит не ниже гиперплоскости $T(\hat{c})$ при $\lambda \in \mathbb{R}_{+}^{m}$ и у них существует общая точка $(\hat{c}, \hat{\lambda})$.

Будем искать точку $\hat{\lambda}$, надвигая снизу гиперплоскость $T(c)$ на график $W$, рассматриваемый только в $\mathbb{R}_{+}^{m}$. Возникает вопрос, всегда ли можно это сделать? В лемме 1 покажем, что $\hat{c}$ и точка «первого касания» $\lambda^* \in \mathbb{R}_{+}^{m}$ существуют всегда, если $\mathscr{D}$ непусто, а в леммах 2 и 3 выясним, что точку $\lambda^*$ можно взять за искомую $\hat{\lambda}$ и укажем оптимальную стратегию $\pi_0$.

Лемма 1. *Если множество $\mathscr{D}$ непусто, то существует константа $\hat{c}$ и точка* $\lambda^* \in \mathbb{R}_{+}^{m}$ такие, что $\hat{c} = w(\lambda^*) = \inf_{\lambda \in \mathbb{R}_{+}^{m}} w(\lambda) > -\infty$.

Доказательство. Так как множество $\mathcal{D}$ непусто, то существует стратегия $\pi$ такая, что $M^\pi F_i - K_i \leqslant 0$, $i = 1, \ldots, m$. Положим $c = M^\pi F_0$, тогда $c = M^\pi F_0 \leqslant M^\pi F_0 - \sum_{i=1}^{m} \lambda_i (M^\pi F_i - K_i) \leqslant w(\lambda)$ при любом $\lambda \in \mathbb{R}_+^m$, т. е. функция $w(\lambda)$ ограничена снизу на $\mathbb{R}_+^m$. Поскольку она кусочно линейная и таких «кусочков» конечное число, существует точка $\lambda^*$ такая, что $\hat{c} = w(\lambda^*) = \inf_{\lambda \in \mathbb{R}_+^m} w(\lambda) > -\infty$. Лемма доказана.

Замечание 2. Так как $w(\lambda)$ — выпуклая функция, то существуют односторонние частные производные. Нетрудно заметить, что в точке $\lambda^*$, если $\lambda_i^* \neq 0$, должно выполняться неравенство $w'_{\lambda_i^-}(\lambda^*) \leqslant 0 \leqslant w'_{\lambda_i^+}(\lambda^*)$.

Рассмотрим все стратегии $\pi(\lambda^*)$ и пронумеруем их в произвольном порядке. Обозначим $\pi_i(\lambda^*)$ через $\pi^*(i)$. Их конечное число, которое обозначим через $k$. Заметим, что любой стратегии $\pi = \pi(\alpha^n, \pi_1, \ldots, \pi_n)$ соответствует функция $\mathcal{L}(\pi, \lambda, 1) = \sum_{i=1}^{n} \alpha_i \mathcal{L}(\pi_i, \lambda, 1)$, т. е. являющаяся линейной комбинацией функций $\mathcal{L}(\pi_i, \lambda, 1)$ с теми же весами $\alpha_i$.

Лемма 2. *Если найдется вектор* $\hat{\alpha} = (\hat{\alpha}_1, \ldots, \hat{\alpha}_k)$ *такой, что* $\sum_{i=1}^{k} \hat{\alpha}_i = 1$ *и* $\hat{\alpha}_i \geqslant 0$, $i = 1, \ldots, k$, *и при этом будет выполнено неравенство* $\sum_{i=1}^{k} \hat{\alpha}_i \mathcal{L}(\pi^*(i), \lambda, 1) \geqslant \hat{c}$ *для любого* $\lambda \in \mathbb{R}_+^m$, *где* $\hat{c} = \inf_{\lambda \in \mathbb{R}_+^m} w(\lambda)$, *то стратегия* $\pi_0 = \pi(\hat{\alpha}, \pi^*(1), \ldots, \pi^*(k))$ — *оптимальная.*

Доказательство. Проверим достаточные условия оптимальности стратегии.

Условие (4) и условие $\lambda_0^* > 0$ выполнены, поскольку рассматриваем вектор $(1, \lambda^*)$, $\lambda^* \in \mathbb{R}_+^m$.

Условие (6) эквивалентно условию (7), которое выполняется, так как
$\mathcal{L}(\pi_0, \lambda^*, 1) = \sum_{i=1}^{k} \hat{\alpha}_i \mathcal{L}(\pi^*(i), \lambda^*, 1) = \sum_{i=1}^{k} \hat{\alpha}_i w(\lambda^*) = w(\lambda^*) = \sup_{\pi \in \Pi} \mathcal{L}(\pi, \lambda^*, 1)$.

Проверим условие (5). Пусть сначала найдется номер $i$ такой, что $\lambda_i^*$ строго больше нуля. Рассмотрим точки $\lambda_1 = (\lambda_1^*, \ldots, \lambda_{i-1}^*, \lambda_i^* - \lambda_0, \lambda_{i+1}^*, \ldots, \lambda_m^*)$, $\lambda_2 = (\lambda_1^*, \ldots, \lambda_{i-1}^*, \lambda_i^* + \lambda_0, \lambda_{i+1}^*, \ldots, \lambda_m^*)$, где $\lambda_0 > 0$, $\lambda_i^* - \lambda_0 > 0$. Тогда $\hat{c} \leqslant \sum_{i=1}^{k} \hat{\alpha}_i \mathcal{L}(\pi^*(i), \lambda_1, 1) = \mathcal{L}(\pi_0, \lambda_1, 1) = \mathcal{L}(\pi_0, \lambda^*, 1) + \lambda_0 (M^{\pi_0} F_i - K_i) = \hat{c} + \lambda_0 (M^{\pi_0} F_i - K_i)$, т. е.

$$0 \leqslant \lambda_0 (M^{\pi_0} F_i - K_i), \tag{8}$$

$$\hat{c} \leqslant \sum_{i=1}^{k} \hat{\alpha}_i \mathcal{L}(\pi^*(i), \lambda_2, 1) = \mathcal{L}(\pi_0, \lambda_2, 1) = \mathcal{L}(\pi_0, \lambda^*, 1) - \lambda_0 (M^{\pi_0} F_i - K_i) = \hat{c} - \lambda_0 (M^{\pi_0} F_i - K_i),$$

т. е.

$$0 \geqslant \lambda_0 (M^{\pi_0} F_i - K_i). \tag{9}$$

Из (8) и (9) заключаем, что $M^{\pi_0} F_i - K_i = 0$ для любого номера $i$ такого, что $\lambda_i^* > 0$. Если $\lambda_i^* = 0$, то условие (5) выполняется для этого номера $i$ очевидным образом.

Теперь проверим, что $\pi_0 \in \mathcal{D}$.

Для этого осталось проверить, что $M^{\pi_0}F_i - K_i \leqslant 0$ для всех номеров $i$ таких, что $\lambda_i^* = 0$. Но это также следует из неравенства (9). Лемма доказана.

Рассмотрим функцию $\mathscr{L}(\lambda) = \max_{1 \leqslant i \leqslant k} \mathscr{L}(\pi^*(i), \lambda, 1)$. Очевидно, что $\mathscr{L}(\lambda^*) = \hat{c}$ и $\mathscr{L}(\lambda) \geqslant \hat{c}$ при всех $\lambda \in \mathbb{R}_+^m$.

Лемма 3. *Пусть есть $k$ аффинных функционала $l_i(x)$ на $\mathbb{R}^n$ таких, что при $x_0 \in \mathbb{R}_+^n$ для любого $i = 1, \ldots, k$ $l_i(x_0) = c_0$.*

*Пусть функция $l(x) = \max_{1 \leqslant i \leqslant k} l_i(x) \geqslant c_0$ для любого $x \in \mathbb{R}_+^n$. Тогда:*

1) *существует выпуклая комбинация $\sum_{i=1}^{k} \hat{\alpha}_i l_i(x) \geqslant c_0$ при всех $x \in \mathbb{R}_+^n$;*

2) *отличны от нуля будут не более, чем $n + 1$ координата вектора $\hat{\alpha}$.*

Доказательство. Если докажем, что существует какая-то выпуклая комбинация $l_i(x)$, удовлетворяющая утверждению 1, то утверждение 2 следует немедленно по теореме Каратеодори [3].

Введем индикаторную функцию $\delta B(x) = \begin{cases} +\infty, & x \notin B, \\ 0, & x \in B. \end{cases}$ Тогда условие $l(x) \geqslant c_0$ при любом $x \in \mathbb{R}_+^n$ эквивалентно неравенству $l(x) + \delta\mathbb{R}_+^n(x) \geqslant c_0$. Но тогда точка $x_0$ — точка абсолютного минимума, и по теореме [5, с. 70] $0 \in \partial(l(x_0) + \delta\mathbb{R}_+^n(x_0))$, где символ $\partial$ обозначает субдифференциал соответствующей функции. По теореме Моро — Рокафеллара [3]

$$0 \in \partial(l(x_0) + \delta\mathbb{R}_+^n(x_0)) = \partial l(x_0) + \partial\delta\mathbb{R}_+^n(x_0).$$

Но тогда по теореме Дубовицкого — Милютина о субдифференциале максимума [3] легко получаем утверждение 1 леммы. Лемма доказана.

Таким образом, выбранная точка $(1, \lambda^*)$ действительно удовлетворяет свойствам точки $(1, \hat{\lambda})$.

Замечание 3. Из утверждения 2 леммы 3 следует, что достаточно взвешивать не более $m + 1$ стратегий $\pi^*(i)$. Для них, как и для функционалов $\mathscr{L}_1(\pi^*(i), \lambda, 1)$ (см. доказательство теоремы Каратеодори в [3]), справедливо: если $\pi^*(j) = \pi(\alpha^k, \pi^*(1), \ldots, \pi^*(k))$, то $\alpha_i = \delta_{ij}$, где $\delta_{ij}$ — символ Кронекера.

Итак, доказана следующая теорема.

Теорема. *Пусть множество допустимых стратегий $\mathscr{D}$ непусто, тогда существует оптимальная стратегия $\pi_0$, являющаяся взвесью не более чем $m + 1$ стационарной стратегии $\pi(\hat{\lambda})$, где $\hat{\lambda} \in \mathbb{R}_+$.*


1. *Фрид Е. Б.* Об оптимальных стратегиях в задачах управления с ограничениями // Теория вероятностей и ее применения.— 1972, **17**, № 1.— С. 194—199.
2. *Неве Ж.* Математические основы теории вероятностей.— М. : Мир, 1969.— 310 с.
3. *Блекуэлл Д.* Динамическое программирование в задачах с затухающим действием // Математика: сб. пер.— 1967.— **11**, № 4.— С. 151—160.
4. *Алексеев В. М., Тихомиров В. М., Фомин С. В.* Оптимальное управление.— М. : Наука, 1979.— 429 с.
5. *Алексеев В. М., Галеев Э. М., Тихомиров В. М.* Сборник задач по оптимизации.— М. : Наука, 1984.— 288 с.



Моск. ун-т — Получено 22.05.87