ИЗВЕСТИЯ РАН. ТЕОРИЯ И СИСТЕМЫ УПРАВЛЕНИЯ, 2020, № 4, с. 18–27

УСТОЙЧИВОСТЬ

УДК 517.977

ИГРОВОЕ УПРАВЛЕНИЕ СЛУЧАЙНОЙ СКАЧКООБРАЗНОЙ СТРУКТУРОЙ ОБЪЕКТА В ЧИСТЫХ СТРАТЕГИЯХ[1]

© 2020 г. В. А. Болдинов[a],*, В. А. Бухалёв[a], А. А. Скрынников[a]

[a] Московский научно-исследовательский телевизионный ин-т, МАИ (национальный исследовательский ун-т), ФГУП "ГосНИИАС", Москва, Россия

**e-mail: victorboldinov@mail.ru*

Поступила в редакцию 26.11.2019 г.
После доработки 10.02.2020 г.
Принята к публикации 30.03.2020 г.

Рассматривается задача оптимального управления случайной скачкообразной структурой объекта в условиях противодействия. Смена состояний структуры объекта наблюдается противоборствующими сторонами с помощью индикаторов, работающих с ошибками. Критерием оптимальности управлений является некоторый функционал состояния объекта, который один из противников стремится минимизировать, а другой – максимизировать. Игроки управляют структурой объекта в чистых стратегиях, применяя конечное число возможных стратегий. Оптимальные управления находятся в классе детерминированных зависимостей от результатов наблюдений, предшествующих текущему моменту. Приводится пример решения задачи оптимизации управления структурой объекта с двумя состояниями методами теории систем со случайной скачкообразной структурой в игровой постановке.

DOI: 10.31857/S0002338820040022

Введение. В настоящей статье приводится динамическая стохастическая система со случайной скачкообразной структурой (ССС) [1–6], имеющей конечное число возможных состояний. Переходы из одного состояния в другое происходят в случайные моменты времени и управляются двумя противоборствующими сторонами (военными противниками, экономическими или политическими конкурентами), преследующими строго противоположные интересы [1, 3]. При этом каждый из противников располагает конечным числом возможных стратегий (управлений) и руководствуется некоторым своим априорным представлением об управляемом объекте и информацией, которую он получает от своего индикатора структуры, регистрирующего с ошибками текущее состояние структуры объекта.

Ставится задача построения алгоритмов управления противников ("игроков"), состоящая в нахождении оптимальных управлений с обратной связью по состоянию объекта в каждый момент времени k в классе детерминированных зависимостей от показаний индикаторов структуры – на отрезке времени от начального момента до текущего k.

Для решения задачи используется теория стохастического динамического программирования на основе метода динамического программирования Беллмана, байесовская обработка информации и марковские математические модели [1, 3, 7–12]. Применение этих методов позволяет построить алгоритмы, сочетающие точность решения с простотой реализации. Их достоинствами являются: обратная связь управлений с состоянием объекта, комплексирование априорной и апостериорной информации о состоянии объекта и рекуррентная форма алгоритмов, не требующая запоминания всей совокупности наблюдений на отрезке времени, который предшествует текущему моменту. Это особенно важно, например, для реализации в системах управления, навигации и наведения летательных аппаратов при существующих ограничениях по памяти в бортовых цифровых вычислительных машинах [13].

1. Постановка задачи. Дано: рассматривается объект ССС, управляемый двумя игроками, которые преследуют строго противоположные интересы. Структура s_k описывается марковской це-

[1] Работа выполнена при финансовой поддержке РФФИ (проект № 19-08-00502а).

пью с конечным числом возможных состояний $s_k = \overline{1, n^{(s)}}$, где k – текущий момент времени: $k = \overline{0, n}$.

Информация, которой располагают игроки о вероятностях переходов из состояния s_k в состояние s_{k+1}, неодинакова:

$$q_k^A(s_{k+1}|s_k, \Theta_k, \vartheta_k) \quad \text{и} \quad q_k^B(s_{k+1}|s_k, \Theta_k, \vartheta_k), \tag{1.1}$$

где Θ_k, ϑ_k – управления игроков A и B, имеющие конечное число возможных стратегий: $\Theta_k = \overline{1, n^{\Theta}}$, $\vartheta_k = \overline{1, n^{\vartheta}}$.

Состояние структуры регистрируется индикаторами с ошибками. Измерения состояния структуры описываются условно-марковскими цепями с конечным числом возможных состояний $r_k = \overline{1, n^r}$ и $\rho_k = \overline{1, n^{\rho}}$. Условно-марковские цепи заданы условными вероятностями переходов из r_k в r_{k+1} и из ρ_k в ρ_{k+1} при фиксированных s_{k+1}, Θ_k, ϑ_k:

$$\pi_{k+1}^A(r_{k+1}|r_k, s_{k+1}, \Theta_k, \vartheta_k) \quad \text{и} \quad \pi_{k+1}^B(\rho_{k+1}|\rho_k, s_{k+1}, \Theta_k, \vartheta_k). \tag{1.2}$$

Зависимость $\pi_{k+1}^A(\cdot)$ от Θ_k означает, что игрок A может управлять как структурой объекта, так и характеристикой индикатора структуры. Зависимость $\pi_{k+1}^A(\cdot)$ от ϑ_k означает, что игрок B может управлять не только структурой объекта s_k, но и осуществлять информационное противодействие игроку A. Аналогичный смысл имеет зависимость $\pi_{k+1}^B(\cdot)$ от ϑ_k и Θ_k.

Так как интересы игроков строго противоположны, то показатели качества (эффективности) игры для обоих аналогичны:

$$\begin{aligned} J^A(\Theta_{\overline{0,n-1}}, \vartheta_{\overline{0,n-1}}, r_{\overline{0,n-1}}) &\triangleq \sum_{k=1}^{n} \mathrm{M}[W_k(s_k, \Theta_{k-1}, \vartheta_{k-1}) \mid r_{\overline{0,k-1}}] = \\ &= \sum_{k=1}^{n} \sum_{s_k} \sum_{\Theta_{k-1}} \sum_{\vartheta_{k-1}} W_k(s_k, \Theta_{k-1}, \vartheta_{k-1}) \mathrm{P}[s_k, \Theta_{k-1}, \vartheta_{k-1} \mid r_{\overline{0,k-1}}] = \\ &= \sum_{k=1}^{n} \sum_{s_k} W_k(s_k, \Theta_{k-1}, \vartheta_{k-1}) \mathrm{P}[s_k \mid r_{\overline{0,k-1}}], \end{aligned} \tag{1.3}$$

$$\begin{aligned} J^B(\Theta_{\overline{0,n-1}}, \vartheta_{\overline{0,n-1}}, \rho_{\overline{0,n-1}}) &\triangleq \sum_{k=1}^{n} \mathrm{M}[W_k(s_k, \Theta_{k-1}, \vartheta_{k-1}) \mid \rho_{\overline{0,k-1}}] = \\ &= \sum_{k=1}^{n} \sum_{s_k} W_k(s_k, \Theta_{k-1}, \vartheta_{k-1}) \mathrm{P}[s_k \mid \rho_{\overline{0,k-1}}], \end{aligned} \tag{1.4}$$

где $W_k(\cdot)$ – текущая функция потерь; $\mathrm{M}[\cdot]$, $\mathrm{P}[\cdot]$, $\triangleq$ – символы соответственно математического ожидания, вероятности и равенства по определению. При этом полагается, что управления Θ_{k-1}, ϑ_{k-1}, как было сказано во Введении, детерминированно зависят от наблюдений $r_{\overline{0,k-1}}$, $\rho_{\overline{0,k-1}}$.

Поскольку рассматривается задача игрового управления, в которой показатели эффективности игроков (1.3), (1.4) различны, так как основываются на различной информации – r_k и ρ_k, то для того, чтобы подчеркнуть это обстоятельство, в качестве показателя выбраны суммы условных математических ожиданий текущей функции потерь $W_k(s_k, \Theta_{k-1}, \vartheta_{k-1})$ при фиксированных наблюдениях $r_{\overline{0,k-1}}$, $\rho_{\overline{0,k-1}}$ соответственно.

Как следует из (1.3), (1.4), критерии оптимальности J^{A^*}, J^{B^*} определяются выражениями

$$J^{A^*}(r_{\overline{0,n-1}}) \triangleq \min_{\Theta_{\overline{0,n-1}}} \max_{\vartheta_{\overline{0,n-1}}} J^A(\Theta_{\overline{0,n-1}}, \vartheta_{\overline{0,n-1}}, r_{\overline{0,n-1}}), \tag{1.5}$$

$$J^{B^*}(\rho_{\overline{0,n-1}}) = \max_{\vartheta_{\overline{0,n-1}}} \min_{\Theta_{\overline{0,n-1}}} J^B(\Theta_{\overline{0,n-1}}, \vartheta_{\overline{0,n-1}}, \rho_{\overline{0,n-1}}), \tag{1.6}$$

т.е. игрок A выбирает оптимальную стратегию на отрезке $[0, n-1]$, добиваясь минимума показателя качества $J^A(\cdot)$ и предполагая, что его противник будет придерживаться стратегии, максимизирующей этот показатель. Противоположным образом действует игрок B, который максимизирует показатель $J^B(\cdot)$ в расчете на стратегию игрока A, минимизирующую этот показатель.

Априорные сведения о начальных значениях вероятностей состояний структуры, которыми располагают игроки, различны: $p_0^A(s_0)$ и $p_0^B(s_0)$.

Требуется найти: оптимальные управления $\Theta_k^*(r_{\overline{0,k}}, \rho_{\overline{0,k}})$, $\vartheta_k^*\left(r_{\overline{0,k}}, \rho_{\overline{0,k}}\right)$ с обратной связью по состоянию объекта в классе детерминированных зависимостей от наблюдений $r_{\overline{0,k}}$, $\rho_{\overline{0,k}}$.

2. Алгоритм игрока *A*. 2.1. Регулятор структуры. Найдем уравнения регулятора структуры (блока управления), связывающие оптимальное управление с вероятностью состояния структуры.

С учетом специфики поставленной задачи применим подход, разработанный Р. Беллманом и известный как метод динамического программирования [7]. Его обобщения и модификации широко используются для синтеза оптимальных управлений с обратной связью в стохастических системах [8–11].

Обозначим

$$J_k^A(\Theta_{\overline{k-1,n}}, \vartheta_{\overline{k-1,n}}, r_{\overline{0,k-1}}) \triangleq \sum_{i=k}^{n} W_i^A(\Theta_{i-1}, \vartheta_{i-1}, r_{\overline{0,k-1}}), \tag{2.1}$$

$$W_i^A(\cdot) \triangleq \mathrm{M}[W_i(s_i, \Theta_{i-1}, \vartheta_{i-1}) \mid r_{\overline{0,k-1}}] = \sum_{s_i} W_i(s_i, \Theta_{i-1}, \vartheta_{i-1})\mathrm{P}[s_i \mid r_{\overline{0,k-1}}], \tag{2.2}$$

где $J_k^A(\cdot)$ – функция оставшихся потерь – показатель качества на отрезке $[k, n]$, в отличие от показателя качества $J^A(\cdot)$ на отрезке $[1, n]$, определяемого формулой (1.3).

Представим $J_k^A(\cdot)$ в виде суммы двух слагаемых: $W_k^A(\cdot)$ и оставшейся части суммы из (2.1). Тогда на основании (2.1), (2.2) получаем

$$\begin{aligned} J_k^A(\Theta_{\overline{k-1,n-1}}, \vartheta_{\overline{k-1,n-1}}, r_{\overline{0,k-1}}) &= W_k^A(\Theta_{k-1}, \vartheta_{k-1}, r_{\overline{0,k-1}}) + \\ &+ \sum_{i=k+1}^{n} \mathrm{M}[W_i(s_i, \Theta_{i-1}, \vartheta_{i-1}) \mid r_{\overline{0,k-1}}] = \\ &= W_k^A(\Theta_{k-1}, \vartheta_{k-1}, r_{\overline{0,k-1}}) + J_{k+1}^A(\Theta_{\overline{k,n-1}}, \vartheta_{\overline{k,n-1}}, r_{\overline{0,k-1}}) = \\ &= W_k^A(\Theta_{k-1}, \vartheta_{k-1}, r_{\overline{0,k-1}}) + \sum_{r_k} J_{k+1}^A(\Theta_{\overline{k,n-1}}, \vartheta_{\overline{k,n-1}}, r_{\overline{0,k}})\mathrm{P}[r_k \mid r_{\overline{0,k-1}}]. \end{aligned} \tag{2.3}$$

Обозначив

$$J_k^{A^*}(r_{\overline{0,k-1}}) \triangleq \min_{\Theta_{\overline{k-1,n-1}}} \max_{\vartheta_{\overline{k-1,n-1}}} J_k^A(\Theta_{\overline{k-1,n-1}}, \vartheta_{\overline{k-1,n-1}}, r_{\overline{0,k-1}})$$

и применяя операцию $\min\limits_{\Theta_{\overline{k-1,n-1}}} \max\limits_{\vartheta_{\overline{k-1,n-1}}}$ к обеим частям равенства (2.3), получаем

$$\begin{aligned} J_k^{A^*}(r_{\overline{0,k-1}}) &= \min_{\Theta_{\overline{k-1,n-1}}} \max_{\vartheta_{\overline{k-1,n-1}}} \times \\ \times \left[W_k^A(\Theta_{k-1}, \vartheta_{k-1}, r_{\overline{0,k-1}}) + \sum_{r_k} J_{k+1}^A(\Theta_{\overline{k,n-1}}, \vartheta_{\overline{k,n-1}}, r_k, r_{\overline{0,k-1}})\mathrm{P}\left[r_k \mid r_{\overline{0,k-1}}\right] \right] &= \\ = \min_{\Theta_{k-1}} \max_{\vartheta_{k-1}} \min_{\Theta_{\overline{k,n-1}}} \max_{\vartheta_{\overline{k,n-1}}} \times & \\ \times \left[W_k^A(\Theta_{k-1}, \vartheta_{k-1}, r_{\overline{0,k-1}}) + \sum_{r_k} J_{k+1}^A(\Theta_{\overline{k,n-1}}, \vartheta_{\overline{k,n-1}}, r_k, r_{\overline{0,k-1}})\mathrm{P}\left[r_k \mid r_{\overline{0,k-1}}\right] \right]. & \end{aligned} \tag{2.4}$$

Формула (2.4) отражает традиционный способ, разработанный Р. Беллманом [7] и используемый всеми авторами, применяющими теорию динамического программирования для оптимизации управления стохастическими системами, например, Р.А. Ховард [8], А.А. Фельдбаум [9], М. Аоки [10], А.Е. Брайсон и Хо Ю Ши [11].

Так как $W_k^A(\cdot)$ от $\Theta_{\overline{k,n-1}}$, $\vartheta_{\overline{k,n-1}}$ не зависит, а

$$\min_{\Theta_{\overline{k,n-1}}} \max_{\vartheta_{\overline{k,n-1}}} \sum_{r_k} J_{k+1}^A(\Theta_{\overline{k,n-1}}, \vartheta_{\overline{k,n-1}}, r_k, r_{\overline{0,k-1}}) \mathrm{P}[r_k \mid r_{\overline{0,k-1}}] =$$
$$= \sum_{r_k} J_{k+1}^{A^*}(r_k, r_{\overline{0,k-1}}) \mathrm{P}[r_k \mid r_{\overline{0,k-1}}],$$

то из (2.4) следует рекуррентное уравнение для $J_k^{A^*}(\cdot)$:

$$J_k^{A^*} = \min_{\Theta_{k-1}} \max_{\vartheta_{k-1}} \left[W_k^A(\Theta_{k-1}, \vartheta_{k-1}) + \sum_{r_k} J_{k+1}^{A^*}(\Theta_{k-1}, \vartheta_{k-1}, r_k) \sigma_k^A(r_k) \right] =$$
$$= \min_{\Theta_{k-1}} \max_{\vartheta_{k-1}} [W_k^A(\Theta_{k-1}, \vartheta_{k-1}) + \tilde{J}_{k+1}^{A^*}(\Theta_{k-1}, \vartheta_{k-1})], \tag{2.5}$$
$$k = n, n-1, \ldots, 1; \quad \tilde{J}_{n+1}^{A^*} \equiv 0,$$

где

$$\sigma_k^A(r_k) \triangleq \mathrm{P}[r_k \mid r_{\overline{0,k-1}}];$$
$$\tilde{J}_{k+1}^{A^*}(\Theta_{k-1}, \vartheta_{k-1}) \triangleq \sum_{r_k} J_{k+1}^{A^*}(\Theta_{k-1}, \vartheta_{k-1}, r_k) \sigma_k^A(r_k). \tag{2.6}$$

Аргумент $r_{\overline{0,k-1}}$ у всех функций здесь и далее опущен для простоты записи. Его наличие обозначено символами "$\hat{}$", "$\tilde{}$" и пр.

Вероятность $\sigma_k^A(r_k)$ находится по формуле полной вероятности

$$\sigma_k^A(r_k) = \sum_{s_k} \pi_k^A(r_k \mid r_{k-1}, s_k, \Theta_{k-1}, \vartheta_{k-1}) \tilde{p}_k^A(s_k), \tag{2.7}$$

а $W_k^A(\cdot)$, как следует из (2.2), – по формуле

$$W_k^A(\Theta_{k-1}, \vartheta_{k-1}) = \sum_{s_k} W_k(s_k, \Theta_{k-1}, \vartheta_{k-1}) \tilde{p}_k^A(s_k), \tag{2.8}$$

где $\tilde{p}_k^A(s_k) \triangleq \mathrm{P}[s_k \mid r_{\overline{0,k-1}}]$ – вероятность состояния структуры s_k, прогнозируемая на один шаг дискретности вперед и определяемая по формуле полной вероятности

$$\tilde{p}_k^A(s_k) = \sum_{s_{k-1}} q_{k-1}^A(s_k \mid s_{k-1}, \Theta_{k-1}, \vartheta_{k-1}) \hat{p}_{k-1}^A(s_{k-1}), \tag{2.9}$$

где $\hat{p}_{k-1}^A(s_{k-1}) \triangleq \mathrm{P}[s_{k-1} \mid r_{\overline{0,k-1}}]$ – апостериорная вероятность состояния структуры s_{k-1}.

Пара минимаксных управлений, согласно (2.5), (2.8), (2.9), определяется формулой

$$(\Theta_{k-1}^*, \vartheta_{k-1}^A) = \arg \min_{\Theta_{k-1}} \max_{\vartheta_{k-1}} [W_k^A(\hat{p}_{k-1}^A(s_{k-1}), \Theta_{k-1}, \vartheta_{k-1}) + \tilde{J}_{k+1}^{A^*}(\hat{p}_{k-1}^A(s_{k-1}), \Theta_{k-1}, \vartheta_{k-1})], \tag{2.10}$$

где Θ_{k-1}^* – оптимальное управление игрока A, а ϑ_{k-1}^A – предполагаемое игроком A оптимальное управление игрока B, основанные на показаниях индикатора структуры $r_{\overline{0,k-1}}$, принадлежащего игроку A.

Рекуррентные уравнения (2.5)–(2.10) описывают алгоритм регулятора структуры игрока A. Выходными сигналами регулятора являются управления Θ_k^*, ϑ_k^A, входным сигналом – апостериорная вероятность $\hat{p}_k^A(s_k)$, которая определяется алгоритмом классификатора структуры (в блоке обработки информации).

2.2. Классификатор структуры. Апостериорная вероятность состояния структуры $\hat{p}^A_k(s_k)$, согласно формуле Байеса, обобщенной на класс ССС [1, 2], и формуле полной вероятности, определяется рекуррентными уравнениями

$$\hat{p}^A_{k+1}(s_{k+1}) = \frac{\pi^A_{k+1}(r_{k+1} \mid r_k, s_{k+1}, \Theta^*_k, \vartheta^A_k)\tilde{p}^A_{k+1}(s_{k+1})}{\sum\limits_{s_{k+1}} \pi^A_{k+1}(r_{k+1} \mid r_k, s_{k+1}, \Theta^*_k, \vartheta^A_k)\tilde{p}^A_{k+1}(s_{k+1})}, \tag{2.11}$$

$$\tilde{p}^A_{k+1}(s_{k+1}) = \sum_{s_k} q^A_k(s_{k+1} \mid s_k, \Theta^*_k, \vartheta^A_k)\hat{p}^A_k(s_k), \tag{2.12}$$

$$s_k = \overline{1, n^{(s)}}; \quad k = 0,1, \cdots, n-1; \quad \tilde{p}^A_0(s_0) = p^A_0(s_0).$$

В целом, оптимальный *минимаксный* информационно-управляющий алгоритм игрока A описывается замкнутой системой рекуррентных уравнений (2.5)–(2.12), в которой уравнения регулятора (2.5)–(2.10) решаются в “обратном времени” ($k = n, n-1, \ldots, 1$) при “начальных” условиях $\tilde{J}^{A^*}_{n+1} \equiv 0$, а уравнения классификатора (2.11)–(2.12) – в “прямом времени” ($k = 0,1, \ldots, n-1$) при начальных условиях $\tilde{p}^A_0 = p^A_0(s_0)$.

3. Алгоритм игрока B**.** Аналогичный информационно-управляющий *максиминный* алгоритм игрока B описывается уравнениями (2.5)–(2.12), в которых производятся следующие замены:

$$\min_{\Theta_{k-1}} \max_{\vartheta_{k-1}} [\cdot]^A \to \max_{\vartheta_{k-1}} \min_{\Theta_{k-1}} [\cdot]^B;$$

индекс “A” $\to$ индекс “B”; $r_k \to \rho_k$; $\Theta^*_k \to \Theta^B_k$; $\vartheta^A_k \to \vartheta^*_k$, где ϑ^*_k и Θ^B_k – максиминные управления: ϑ^*_k – оптимальное управление игрока B, а Θ^B_k – предполагаемое игроком B оптимальное управление его противника A, основанные на показаниях индикатора структуры $\rho_{\overline{0,k}}$, принадлежащего игроку B.

3.1. Регулятор структуры. С учетом выполненных преобразований алгоритм регулятора структуры игрока B будет иметь вид:

$$J^{B^*}_k = \max_{\vartheta_{k-1}} \min_{\Theta_{k-1}} [W^B_k(\Theta_{k-1}, \vartheta_{k-1}) + \tilde{J}^{B^*}_{k+1}(\Theta_{k-1}, \vartheta_{k-1})], \tag{3.1}$$

$$\tilde{J}^{B^*}_{k+1}(\Theta_{k-1}, \vartheta_{k-1}) \triangleq \sum_{\rho_k} J^{B^*}_{k+1}(\Theta_{k-1}, \vartheta_{k-1}, \rho_k)\sigma^B_k(\rho_k), \tag{3.2}$$

$$\sigma^B_k(\rho_k) = \sum_{s_k} \pi^B_k(\rho_k \mid \rho_{k-1}, s_k, \Theta_{k-1}, \vartheta_{k-1})\tilde{p}^B_k(s_k), \tag{3.3}$$

$$W^B_k(\Theta_{k-1}, \vartheta_{k-1}) = \sum_{s_k} W_k(s_k, \Theta_{k-1}, \vartheta_{k-1})\tilde{p}^B_k(s_k), \tag{3.4}$$

$$\tilde{p}^B_k(s_k) = \sum_{s_{k-1}} q^B_{k-1}(s_k \mid s_{k-1}, \Theta_{k-1}, \vartheta_{k-1})\hat{p}^B_{k-1}(s_{k-1}), \tag{3.5}$$

$$(\vartheta^*_{k-1}, \Theta^B_{k-1}) = \arg\max_{\vartheta_{k-1}} \min_{\Theta_{k-1}} [W^B_k(\cdot) + \tilde{J}^{B^*}_{k+1}(\cdot)], \tag{3.6}$$

$$k = n, n-1, \cdots, 1; \quad \tilde{J}^{B^*}_{n+1} \equiv 0,$$

где $W^B_k(\cdot) \triangleq W^B_k(\hat{p}^B_{k-1}(s_{k-1}), \Theta_{k-1}, \vartheta_{k-1})$, $\tilde{J}^{B^*}_{k+1}(\cdot) \triangleq \tilde{J}^{B^*}_{k+1}(\hat{p}^B_{k-1}(s_{k-1}), \Theta_{k-1}, \vartheta_{k-1})$.

3.2. Классификатор структуры. Алгоритм классификатора структуры игрока B описывается уравнениями:

$$\hat{p}^B_{k+1}(s_{k+1}) = \frac{\pi^B_{k+1}(\rho_{k+1} \mid \rho_k, s_{k+1}, \Theta^B_k, \vartheta^*_k)\tilde{p}^B_{k+1}(s_{k+1})}{\sum\limits_{s_{k+1}} \pi^B_{k+1}(\rho_{k+1} \mid \rho_k, s_{k+1}, \Theta^B_k, \vartheta^*_k)\tilde{p}^B_{k+1}(s_{k+1})}, \tag{3.7}$$

$$\tilde{p}^B_{k+1}(s_{k+1}) = \sum_{s_k} q^B_k(s_{k+1} \mid s_k, \Theta^B_k, \vartheta^*_k)\hat{p}^B_k(s_k), \tag{3.8}$$

$$k = 0,1,\cdots,n-1; \quad \tilde{p}^B_0(s_0) = p^B_0(s_0); \quad s_k = \overline{1,n^{(s)}}.$$

4. Пример. Рассмотрим задачу оптимизации управления ССС объекта с двумя состояниями как частный случай общей постановки задачи из разд. 1.

Дано: структура объекта, индикаторы структуры и критерии оптимальности описываются следующими выражениями:

1) для игрока A:

$$q^A_k(s_{k+1} \mid s_k) = q_k(s_{k+1} \mid s_k), \quad s_k = 1,2, \tag{4.1}$$

$$q_k(2 \mid 1) \stackrel{\Delta}{=} q_k = q_{\min}, q_{\max},$$

$$q_k(1 \mid 2) \stackrel{\Delta}{=} g_k = g_{\min}, g_{\max},$$

$$\tilde{p}^A_0(1) = p_0(1), \quad \tilde{p}^A_0(2) = 1 - p_0(1),$$

$$\pi^A_{k+1}(r_{k+1} \mid r_k, s_{k+1}), \quad r_k = 1,2, \tag{4.2}$$

$$J^{A^*}_k = \min_{q_{\overline{0,n-1}}} \max_{g_{\overline{0,n-1}}} \sum_{k=1}^{n} \mathrm{M}[W_k(s_k, q_{k-1}, g_{k-1}) \mid r_{\overline{0,k-1}}]; \tag{4.3}$$

2) для игрока B:

$$q^B_k(s_{k+1} \mid s_k) = q_k(s_{k+1} \mid s_k), \quad \tilde{p}^B_0(s_0) = \tilde{p}^A_0(s_0), \tag{4.4}$$

$$\pi^B_{k+1}(\rho_{k+1} \mid \rho_k, s_{k+1}), \quad \rho_k = 1,2, \tag{4.5}$$

$$J^{B^*}_k = \max_{g_{\overline{0,n-1}}} \min_{q_{\overline{0,n-1}}} \sum_{k=1}^{n} \mathrm{M}[W_k(s_k, q_{k-1}, g_{k-1}) \mid \rho_{\overline{0,k-1}}], \tag{4.6}$$

где $W_k(\cdot)$ – текущая функция потерь:

$$W_k(s_k, q_{k-1}, g_{k-1}) = \delta(s_k,1) + \lambda q_{k-1} - \mu g_{k-1}, \tag{4.7}$$

$$\delta(s_k,1) = \begin{cases} 1 & \text{при } s_k = 1, \\ 0 & \text{при } s_k = 2, \end{cases}$$

где $\delta(s_k,1)$ – символ Кронекера; $\lambda = \mathrm{const}$, $\mu = \mathrm{const}$, $\lambda \in (0,1)$, $\mu \in (0,1)$.

Вероятностью перехода q_k управляет игрок A, а вероятностью g_k – игрок B.

Согласно (2.2), (2.8), (4.7),

$$W^A_k = \tilde{p}^A_k(1) + \lambda q_{k-1} - \mu g_{k-1}, \tag{4.8}$$

откуда следует содержательный смысл критериев оптимальности (4.3), (4.6): игрок A минимизирует вероятность первого состояния структуры, ограничивая свои усилия по переводу структуры из первого состояния во второе и предполагая, что противник будет максимизировать эту вероятность; игрок B максимизирует ту же самую вероятность, предполагая, что игрок A будет ее минимизировать. При этом игрок B также старается ограничить свои усилия по переводу структуры из второго состояния в первое (так как $\max(-\mu g_{k-1}) = \min(\mu g_{k-1})$). Весовые коэффициенты λ и μ характеризуют приоритетность соответствующих частных показателей в общем показателе качества.

Как видно из (4.1), каждый из игроков располагает двумя возможными режимами управления: “экономный” – $q_{\min}$, $g_{\min}$ и “энергичный” – $q_{\max}$, $g_{\max}$.

Требуется найти: оптимальные алгоритмы управления игроков в виде детерминированных зависимостей от показаний их индикаторов структуры $r_{\overline{0,k}}$ и $\rho_{\overline{0,k}}$.

Решение. Синтез оптимальных информационно-управляющих алгоритмов игроков A и B осуществляется методом игрового оптимального управления системами ССС.

Информационно-управляющий алгоритм игрока A.

Алгоритм управления игрока A состоит из регулятора и классификатора структуры.

Регулятор структуры игрока A. На основании (2.9), (4.8) получаем

$$W_k^A = (1 - q_{k-1} - g_{k-1})\hat{p}_{k-1}^A(1) + g_{k-1} + \lambda q_{k-1} - \mu g_{k-1} = \\ = h_{k-1}\hat{p}_{k-1}^A(1) + \lambda q_{k-1} + (1-\mu)g_{k-1}, \tag{4.9}$$

где

$$h_{k-1} \triangleq 1 - q_{k-1} - g_{k-1}. \tag{4.10}$$

Будем искать решение уравнения (2.5) в виде

$$J_{k+1}^{A^*} = \psi_k^A \hat{p}_k^A(1) + m_k^A, \tag{4.11}$$

где ψ_k^A, m_k^A – неопределенные параметры.

Из (2.6), (2.9), (4.10), (4.11) следует

$$\tilde{J}_{k+1}^{A^*} = \psi_k^A \tilde{p}_k^A(1) + m_k^A = \psi_k^A [h_{k-1}\hat{p}_{k-1}(1) + g_{k-1}] + m_k^A. \tag{4.12}$$

Оптимальные значения q_{k-1}, g_{k-1}, согласно (2.10), (4.9)–(4.11), определяются формулой

$$(q_{k-1}^*, \hat{g}_{k-1}^A) = \arg\min_{q_{k-1}} \max_{g_{k-1}} [W_k^A + \tilde{J}_{k+1}^{A^*}] = \tag{4.13}$$

$$= \arg\min_{q_{k-1}} \max_{g_{k-1}} \{h_{k-1}\hat{p}_{k-1}^A(1) + \lambda q_{k-1} + (1-\mu)g_{k-1} + \psi_k^A[h_{k-1}\hat{p}_{k-1}^A(1) + g_{k-1}] + m_k^A =$$

$$= \arg\min_{q_{k-1}} \max_{g_{k-1}} \{(-q_{k-1} - g_{k-1})\hat{p}_{k-1}^A(1) + \lambda q_{k-1} + (1-\mu)g_{k-1} +$$

$$+ \psi_k^A[(-q_{k-1} - g_{k-1})\hat{p}_{k-1}^A(1) + q_{k-1}] = \arg\min_{q_{k-1}} \max_{g_{k-1}} \{[\lambda - (1+\psi_k^A)\hat{p}_{k-1}^A(1)]q_{k-1} +$$

$$+ [(1+\psi_k^A)\hat{p}_{k-1}^A(2) - \mu]g_{k-1}\} =$$

$$= \arg\min_{q_{k-1}} \max_{g_{k-1}} \{[\lambda_{k-1}^A - \hat{p}_{k-1}^A(1)]q_{k-1} + [\hat{p}_{k-1}^A(2) - \mu_{k-1}^A]g_{k-1}\},$$

где

$$\lambda_{k-1}^A \triangleq \lambda(1+\psi_k^A)^{-1}; \quad \mu_{k-1}^A \triangleq \mu(1+\psi_k^A)^{-1}. \tag{4.14}$$

Из (4.13), (4.14) следует

$$q_{k-1}^* = \begin{cases} q_{\min} & \text{при} \quad \hat{p}_{k-1}^A(1) \le \lambda_{k-1}^A, \\ q_{\max} & \text{при} \quad \hat{p}_{k-1}^A(1) > \lambda_{k-1}^A, \end{cases} \tag{4.15}$$

$$\hat{g}_{k-1}^A = \begin{cases} g_{\min} & \text{при} \quad \hat{p}_{k-1}^A(2) \le \mu_{k-1}^A, \\ g_{\max} & \text{при} \quad \hat{p}_{k-1}^A(2) > \mu_{k-1}^A. \end{cases} \tag{4.16}$$

Подставив (4.11)–(4.13) в (2.5), получаем

$$\psi_{k-1}^A \hat{p}_{k-1}^A(1) + m_{k-1} = \tag{4.17}$$

$$= h_{k-1}^A \hat{p}_{k-1}^A(1) + \lambda q_{k-1}^* + (1-\mu)\hat{g}_{k-1}^A + \psi_k^A[h_{k-1}\hat{p}_{k-1}^A(1) + \hat{g}_{k-1}^A] + m_k^A,$$

где

$$h_{k-1}^A \triangleq 1 - q_{k-1}^* - \hat{g}_{k-1}^A. \tag{4.18}$$

Приравнивая коэффициенты при $\hat{p}_{k-1}^A(1)$ в левой и правой частях уравнения (4.17), получаем рекуррентное уравнение для ψ_k:

$$\psi_{k-1}^A = h_{k-1}^A(1 + \psi_k^A),$$

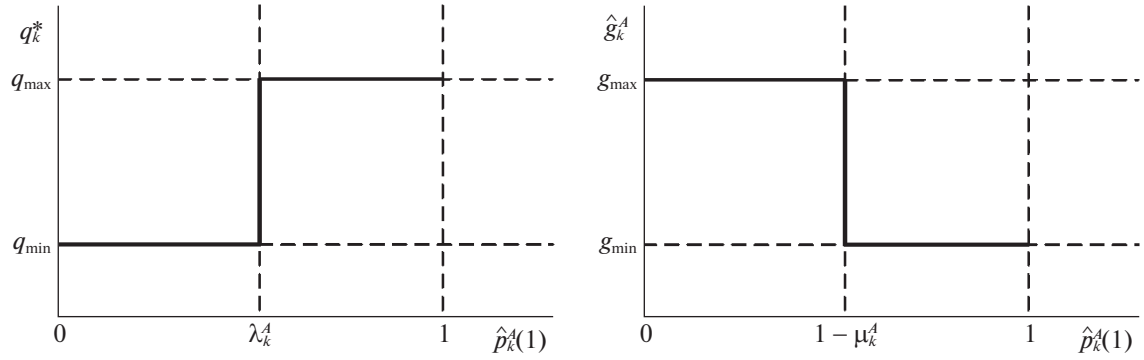

Рис. 1

откуда следует

$$\psi_k^A = h_k^A(1+\psi_{k+1}^A),$$
$$1+\psi_k^A = 1+h_k^A(1+\psi_{k+1}^A),$$

и с учетом (4.14), (4.18) получаем рекуррентное уравнение для $\varepsilon_k^A \triangleq \lambda/\lambda_k^A = \mu/\mu_k^A$:

$$\varepsilon_{k-1}^A = 1+(1-q_k^*-\hat{g}_k^A)\varepsilon_k^A, \quad \varepsilon_n^A = 1, \quad k = n, n-1, \ldots, 1. \tag{4.19}$$

Пороговые значения λ_k^A, μ_k^A вычисляются по формулам

$$\lambda_k^A = \lambda/\varepsilon_k^A; \quad \mu_k^A = \mu/\varepsilon_k^A. \tag{4.20}$$

Учитывая, что $\hat{p}_k^A(2) = 1-\hat{p}_k^A(1)$, алгоритм (4.16) удобно записать в виде

$$\hat{g}_{k-1}^A = \begin{cases} g_{\min} & \text{при} \quad \hat{p}_{k-1}^A(1) \geq 1-\mu_{k-1}^A, \\ g_{\max} & \text{при} \quad \hat{p}_{k-1}^A(1) < 1-\mu_{k-1}^A. \end{cases} \tag{4.21}$$

Алгоритм регулятора игрока A изображен на рис. 1.

Классификатор структуры игрока A. Апостериорные вероятности $\hat{p}_k^A(1)$, $\hat{p}_k^A(2)$ вычисляются в классификаторе структуры, уравнения которого (2.11), (2.12) принимают вид

$$\hat{p}_{k+1}^A(1) = \left[1+\frac{\pi_{k+1}^A(r_{k+1}\mid r_k,2)\,\tilde{p}_{k+1}^A(2)}{\pi_{k+1}^A(r_{k+1}\mid r_k,1)\,\tilde{p}_{k+1}^A(1)}\right]^{-1},$$
$$\hat{p}_{k+1}^A(2) = 1-\hat{p}_{k+1}^A(1),$$
$$\tilde{p}_{k+1}^A(1) = (1-q_k^*)\hat{p}_k^A(1)+\hat{g}_k^A\hat{p}_k^A(2), \tag{4.22}$$
$$\tilde{p}_k^A(2) = 1-\tilde{p}_k^A(1),$$
$$k = 0,1,2,\ldots,n-1; \quad \tilde{p}_0^A = p_0^A(1).$$

Информационно-управляющий алгоритм игрока B. Алгоритм описывается уравнениями, подобными уравнениям алгоритма игрока A, в которых произведены изменения, соответствующие замене минимакса на максимин:

$$g_{k-1}^* = \begin{cases} g_{\min} & \text{при} \quad \hat{p}_{k-1}^B(1) \geq 1-\mu_{k-1}^B, \\ g_{\max} & \text{при} \quad \hat{p}_{k-1}^B(1) < 1-\mu_{k-1}^B, \end{cases}$$
$$\hat{q}_{k-1}^B = \begin{cases} q_{\min} & \text{при} \quad \hat{p}_{k-1}^B(1) \leq \lambda_{k-1}^B, \\ q_{\max} & \text{при} \quad \hat{p}_{k-1}^B(1) > \lambda_{k-1}^B, \end{cases} \tag{4.23}$$

$$\lambda_k^B = \lambda/\varepsilon_k^B; \quad \mu_k^B = \mu/\varepsilon_k^B; \tag{4.24}$$

$$\varepsilon_{k-1}^B = 1 + (1 - \hat{q}_k^B - g_k^*)\varepsilon_k^B, \quad \varepsilon_n^B = 1, \quad k = n, n-1, \ldots, 1, \tag{4.25}$$

$$\hat{p}_{k+1}^B(1) = \left[1 + \frac{\pi_{k+1}^B(\rho_{k+1} \mid \rho_k, 2)\tilde{p}_{k+1}^B(2)}{\pi_{k+1}^B(\rho_{k+1} \mid \rho_k, 1)\tilde{p}_{k+1}^B(1)}\right]^{-1},$$
$$\hat{p}_{k+1}^B(2) = 1 - \hat{p}_{k+1}^B(1),$$
$$\tilde{p}_{k+1}^B(1) = (1 - \hat{q}_k^B)\hat{p}_k^B(1) + g_k^*\hat{p}_k^B(2), \tag{4.26}$$
$$\tilde{p}_k^B(2) = 1 - \tilde{p}_k^B(1),$$
$$k = 0,1,2,\ldots,n-1; \quad \tilde{p}_0^B = p_0^B(1).$$

Физический смысл полученных оптимальных регуляторов структуры: если вероятность состояния структуры, нежелательного для игрока ($s_k = 1$ – для игрока A и $s_k = 2$ – для игрока B), превышает некоторый порог (λ_k^* – для игрока A и μ_k^* – для игрока B), то включается "энергичный" режим управления ($q_{\max}$ – для игрока A и $g_{\max}$ – для игрока B), повышающий вероятность перехода в желаемое для данного игрока состояние структуры ($s_k = 2$ – для игрока A и $s_k = 1$ – для игрока B).

Если же вероятность нежелательного состояния структуры меньше указанного порога, то включается "экономный" режим управления, при котором вероятность переходов минимальна ($q_{\min}$ – для игрока A и $g_{\min}$ – для игрока B).

Кроме оптимальных управлений q_k^* и g_k^* каждый игрок определяет предполагаемое оптимальное управление своего противника ($\hat{g}_k^A$ или $\hat{q}_k^B$) на основании показаний своего индикатора структуры ($r_{\overline{0,k}}$ или $\rho_{\overline{0,k}}$). Эти оценки необходимы в качестве входных данных для классификаторов структуры.

Приближенно-оптимальные регуляторы структуры. Полученные алгоритмы можно существенно упростить, если в уравнениях (4.19), (4.25) приближенно заменить q_k^*, $\hat{q}_k^B$, g_k^*, $\hat{g}_k^A$ их некоторыми средневзвешенными значениями:

$$q_k^* = \hat{q}_k^B = \tilde{\lambda}_k q_{\min} + (1 - \tilde{\lambda}_k)q_{\max},$$
$$g_k^* = \hat{g}_k^A = \tilde{\mu}_k g_{\min} + (1 - \tilde{\mu}_k)g_{\max}, \tag{4.27}$$

где $\tilde{\lambda}_k = \lambda/\varepsilon_k$, $\tilde{\mu}_k = \mu/\varepsilon_k$.

Подставив (4.27) в (4.19), (4.25), получаем рекуррентное уравнение

$$\varepsilon_{k-1} = (1 - q_{\max} - g_{\max})\varepsilon_k + \lambda(q_{\max} - q_{\min}) + \mu(g_{\max} - g_{\min}) + 1, \tag{4.28}$$
$$\varepsilon_n = 1; \quad k = n, n-1, \ldots, 1.$$

В *установившемся* режиме из (4.28) следует

$$\varepsilon_{k-1} = \varepsilon_k = \varepsilon = \frac{1 + \lambda(q_{\max} - q_{\min}) + \mu(g_{\max} - g_{\min})}{q_{\max} + g_{\max}}, \tag{4.29}$$

$$\tilde{\lambda} = \lambda/\varepsilon; \quad \tilde{\mu} = \mu/\varepsilon. \tag{4.30}$$

Как видно из (4.28)–(4.30), приближенные пороговые значения $\tilde{\lambda}_k$, $\tilde{\mu}_k$ вычисляются на основании только априорных данных и не зависят от показаний индикаторов структуры, что значительно упрощает практическую реализацию алгоритмов.

Заключение. Каждый из игровых информационно-управляющих алгоритмов противоборствующих сторон состоит из двух взаимосвязанных блоков: регулятора структуры и классификатора структуры (рис. 2).

Рассмотренная задача представляет собой игру с неполной информацией и ненулевой суммой и не имеет седловой точки вследствие различной информированности игроков о результатах игры. Прежде всего это объясняется различными показаниями индикаторов структуры $r_k \neq \rho_k$,

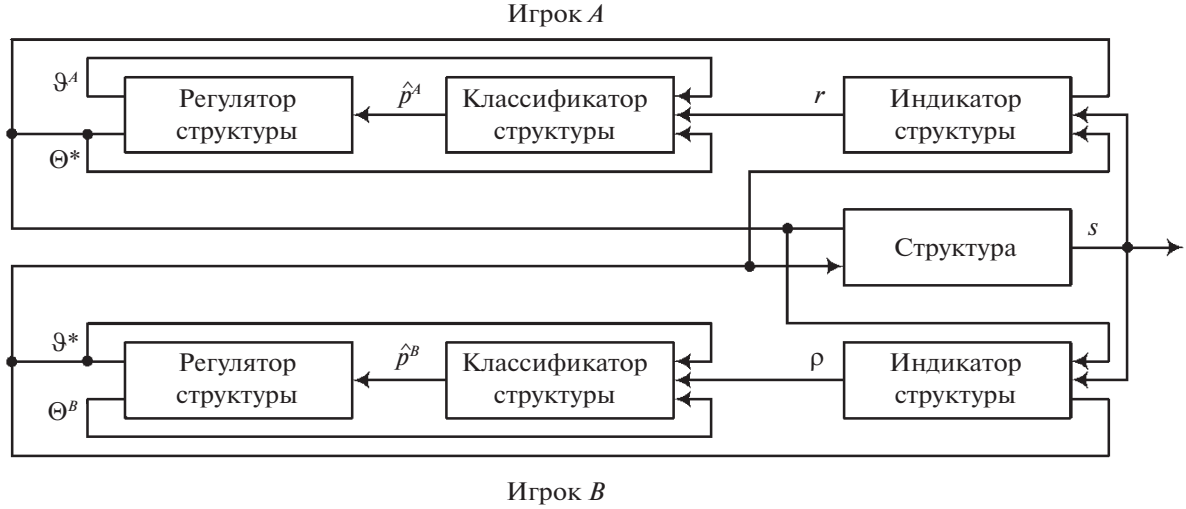

Рис. 2. Игровое управление случайной скачкообразной структурой в чистых стратегиях

откуда следует $\hat{p}_k^A(s_k) \neq \hat{p}_k^B(s_k)$, $\Theta_k^* \neq \Theta_k^B$, $\vartheta_k^* \neq \vartheta_k^A$. Поэтому седловая точка игры отсутствует даже в том случае, когда противники располагают одинаковой априорной информацией ($q^A(\cdot) = q^B(\cdot)$, $\tilde{p}_0^A(s_0) = \tilde{p}_0^B(s_0)$), а текущая функция потерь $W_k(\cdot)$ сепарабельна относительно управлений игроков q_k, g_k.

СПИСОК ЛИТЕРАТУРЫ

1. *Бухалёв В.А.* Распознавание, оценивание и управление в системах со случайной скачкообразной структурой. М.: Физматлит, 1996. 287 с.
2. *Бухалёв В.А.* Оптимальное сглаживание в системах со случайной скачкообразной структурой. М.: Физматлит, 2013. 188 с.
3. *Бухалёв В.А., Скрынников А.А., Болдинов В.А.* Алгоритмическая помехозащита беспилотных летательных аппаратов. М.: Физматлит, 2018. 192 с.
4. *Артемьев В.М.* Теория динамических систем со случайными изменениями структуры. Минск: Высш. шк., 1979. 160 с.
5. *Пакшин П.В.* Дискретные системы со случайными параметрами и структурой. М.: Наука, 1994. 304 с.
6. *Скляревич А.Н.* Линейные системы с возможными нарушениями. М.: Наука, 1975. 352 с.
7. *Беллман Р.* Динамическое программирование. М.: Изд-во иностр. лит., 1960. 400 с.
8. *Ховард Р.А.* Динамическое программирование и марковские процессы. М.: Сов. радио, 1964.
9. *Фельдбаум А.А.* Основы теории оптимальных автоматических систем. М.: Наука, 1966. 623 с.
10. *Аоки М.* Оптимизация стохастических систем. М.: Наука, 1971. 424 с.
11. *Брайсон А.Е., Хо Ю Ши.* Прикладная теория оптимального управления. М.: Мир, 1972. 544 с.
12. *Стратонович Р.Л.* Условные процессы Маркова // Теория вероятностей и ее применения. 1960. Т. 5. Вып. 2. С. 172–195.
13. *Себряков Г.Г., Красильщиков М.Н.* Управление и наведение беспилотных маневренных летательных аппаратов на основе современных информационных технологий. М.: Физматлит, 2003. 280 с.