УДК 62-506

АДАПТИВНЫЕ МЕХАНИЗМЫ ФУНКЦИОНИРОВАНИЯ АКТИВНЫХ СИСТЕМ

III. ОБУЧЕНИЕ И УПРАВЛЕНИЕ

ЦЫГАНОВ В. В.

(Москва)

Рассматриваются задачи синтеза адаптивных механизмов функционирования двухуровневых активных систем, в которых информация, получаемая от активных элементов в процессе их функционирования, используется для изменения параметров самого механизма и достижения цели центра. Даны постановки задач синтеза и найдены конструктивные достаточные условия существования обучающихся механизмов, в которых планирование и стимулирование активных элементов совмещается с обучением центра. Рассмотрен подход к решению задачи синтеза дуальных механизмов, в которых дополнительно оптимизируется некоторый критерий качества управления. Получены соответствующие достаточные условия оптимальности.

1. Введение

При адаптивных идентификации и управлении сложными иерархическими системами необходимо учитывать человеческий фактор или активность составляющих их элементов, связанную с наличием у них собственных целей [1—5]. Дальновидный активный элемент (АЭ) системы может предсказывать управляющие воздействия центра и выбирать свои состояния так, чтобы, воздействуя на результаты идентификации и управления, максимизировать свою собственную целевую функцию (см. формулу (1) в работе [4]), зависящую от процедур стимулирования в каждом периоде.

Например, при использовании известных процедур обучения и управления [6—9] текущая информация, получаемая от элементов в процессе их функционирования, используется для изменения параметров механизма и достижения цели центра. Именно, управление устанавливается в зависимости от текущих результатов обучения — рекуррентной оценки параметров активной системы $(\mathbf{a}_t)$: $\mathbf{a}_t=I_t(\mathbf{a}_{t-1}, \mathbf{y}_t)$, $t=0, 1, \ldots, \mathbf{a}_{-1}=\mathbf{a}^0$. Здесь I_t — процедура обучения, $\mathbf{y}_t$ — наблюдаемое центром состояние АЭ в периоде t. Активный элемент, зная процедуру обучения, выбирает свое состояние из множества решений игры [4, 5], обеспечивающих в том или ином смысле максимум его целевой функции. Это состояние, вообще говоря, отличается от того, которое бы имело место в отсутствие активности. Поэтому непосредственное применение центром процедур адаптации и обучения, разработанных применительно к автоматическим системам, вообще говоря, неэффективно. Проблема состоит в построении механизмов функционирования $\Sigma=(I, \pi, Q, f)$, включающих процедуры обучения (I), планирования (π), управления (Q) и стимулирования (f), обеспечивающих эффективную адаптацию в активных системах.

В работах [3–5] рассматривалась задача построения адаптивных механизмов, обеспечивающих идентификацию структуры векторного АЭ, определяемой скалярным параметром — потенциалом p. Для случая детерминированной структуры АЭ было получено условие сильной прогрессивности [3, 4], обеспечивающее максимальную скорость сходимости оценки параметра к потенциалу ($a_t = p$ при любых $t \geq 0$). В [5] получено более слабое достаточное условие для скалярного АЭ.

В гл. 4 [3] рассматривалась задача построения механизма функционирования независимого скалярного АЭ с недетерминированной структурой при адаптивном управлении с эталонной моделью. Были найдены достаточные условия его эффективности, включая ограничения на процедуры идентификации и управления.

В настоящей работе рассматривается задача построения адаптивных механизмов функционирования (АМФ) зависимых векторных АЭ с недетерминированной структурой, определяемой случайным векторным параметром, при различных целях центра. Ниже используются единые обозначения, принятые в [4, 5].

2. Решение игры элементов

Рассмотрим активную систему, содержащую N зависимых АЭ. Как и ранее, будем обозначать состояние i-го АЭ в периоде t через $\mathbf{y}_t^i = (y_{1t}^i, \ldots, y_{mt}^i)$, $i = \overline{1, N}$, а состояние системы — как $y_t = (\mathbf{y}_t^1, \ldots, \mathbf{y}_t^N)$. Предположим, что $m = N$ и центр наблюдает лишь одну компоненту состояния каждого векторного АЭ («выход» АЭ). Обозначим ее для i-го АЭ $\tilde{y}_{it} \equiv y_{it}^i$, $i = \overline{1, N}$. Остальные компоненты его состояния y_{jt}^i можно трактовать как децентрализованные ресурсы, выделяемые i-м АЭ j-му АЭ перед началом периода t, $i, j = \overline{1, N}$, $i \neq j$, $y_{jt}^i \in [0, p_{jt}^i]$, где p_{jt}^i — случайный параметр. Наблюдаемое центром состояние $\tilde{y}_{it}$ ограничено не только его собственным потенциалом — случайным параметром p_{it}^i, но и ресурсами, выделяемыми i-му АЭ остальными элементами

$$(1) \qquad \tilde{y}_{it} \in [0, W_t^i(\mathbf{u}_t^i)], \ \mathbf{u}_t^i = (y_{it}^1, \ldots, y_{it}^{i-1}, p_{it}^i, y_{it}^{i+1}, \ldots, y_{it}^N),$$

причем W_t^i — монотонно возрастающая функция своих аргументов. Тогда множество возможных состояний i-го АЭ $Y_t^i(\mathbf{u}_t^i) = [0, W_t^i(\mathbf{u}_t^i)] \times \prod_{i=1}^{N} [0, p_j^i]$. Легко видеть, что $W_t^i(\mathbf{u}_t^i) \leq z_t^i \equiv W_t^i(\mathbf{p}_t^i)$, где z_t^i — наблюдаемое состояние i-го АЭ в периоде t в отсутствие активности (или кратко — предельное состояние), $\mathbf{p}_t^i = (p_{1t}^i, \ldots, p_{Nt}^i)$ — его потенциал, $i = \overline{1, N}$. Множество возможных значений составного параметра $p_t = (\mathbf{p}_t^1, \ldots, \mathbf{p}_t^N)$ — потенциала системы в периоде t — обозначим через P_t.

Рассматривается следующий порядок функционирования системы в периоде t. Вначале центр, исходя из принятого им АМФ и основываясь на предыдущих наблюдениях $\tilde{\mathbf{y}}_\tau = (\tilde{y}_{1\tau}, \ldots, \tilde{y}_{N\tau})'$ [1], $\tau < t$, производит оценку параметров активной системы a_{t-1} и назначает вектора планов (нормативов стимулирования) $\mathbf{x}_t$ и управлений (централизованных ресурсов) $\mathbf{r}_t$. Затем после реализации случайного параметра $\mathbf{p}_t^i$ i-й АЭ выбирает состояния y_{jt}^i, $i \neq j$, $1 \leq j \leq N$ и сообщает об этом остальным АЭ. Затем он же

---

[1] Штрих обозначает транспонирование.

выбирает свое наблюдаемое состояние $\tilde{y}_{it}$, уже зная фактическую реализацию $W_t^i(\mathbf{u}_t^i)$, и получает поощрения $\varphi_t^i=f_t^i(\mathbf{x}_t, \tilde{\mathbf{y}}_t)$.

Определим теперь решение игры элементов. В соответствии с [5] элементы решают задачи оптимизации собственных целевых функций с прогнозами потенциала, состояния и механизма. При этом i-й АЭ использует свои собственные прогнозы множества возможных значений потенциала системы $P_{i\tau}(P_{i\tau}\subset P_\tau)$, а также прогнозы множеств возможных выборов k-го АЭ $Y_{i\tau}^k(\mathbf{u}_\tau^k)$, $k=\overline{1,N}$ $(p_\tau\in P_{i\tau},\ Y_{i\tau}^k(\mathbf{u}_\tau^k)\subset Y_\tau^k(\mathbf{u}_\tau^k))$, $\tau=\overline{t+1, t+T}$, $i=\overline{1,N}$. На основе этих прогнозов i-й АЭ определяет свой гарантированный выигрыш при состоянии системы y_t:

$$(2)\qquad \hat{w}_t^i(x_t, y_t)=\min_{p_\tau\in P_{i\tau},\ \tau=\overline{t+1,\, t+T}}\ \min_{\mathbf{y}_\tau^k\in Y_{i\tau}^k,\ k=\overline{1,N}} w_t^i(\varphi_t^i, \ldots, \varphi_{t+T}^i).$$

В качестве решения игры элементов рассматривается равновесие Нэша. Именно, множество решений игры как состояний системы, максимизирующих гарантированный результат каждого из элементов, имеет вид

$$(3)\qquad R_t(\Sigma, p_t)=\{y_t^*\,|\,\hat{w}_t^i(x_t, y_t^*)\geqslant\hat{w}_t^i(x_t, y_t),\ \mathbf{y}_t^k\in Y_t^k(\mathbf{u}_t^k),\ i,\ k=\overline{1,N}\}.$$

Как и в работах [3, 5], предполагается, что каждый АЭ экстраполирует действующий АМФ на весь период его дальновидности (см. (3) в [5]): $\mathbf{a}_\tau=I_t(\mathbf{a}_{\tau-1}, y_\tau)$, $x_{\tau+1}=\pi_{t+1}(\mathbf{a}_\tau)$, $r_\tau=Q_t(\mathbf{a}_\tau)$, $\varphi_\tau^i=f_t^i(x_\tau, \tilde{y}_\tau)$, $\tau=\overline{t, t+T}$. Кроме того, АЭ благожелателен по отношению к центру: если $R_t(\Sigma, p_t)\ni\ni v_t=(\mathbf{v}_t^1, \ldots, \mathbf{v}_t^N)$, $\mathbf{v}_t^i=(p_{1t}^i, \ldots, p_{i-1t}^i, z_t^i, p_{i+1t}^i, \ldots, p_{Nt}^i)$, то $y_t^*=v_t$. Содержательно это означает, что АЭ не занижают свои показатели, если им это не выгодно. Вектор $\mathbf{v}_t^i$ есть состояние i-го элемента в отсутствие активности. Перейдем теперь к исследованию задач синтеза адаптивных механизмов при различных предположениях относительно целей центра.

3. Постановка задачи синтеза обучающегося механизма

В обучающемся механизме функционирования активной системы центр совмещает обучение [6—8] с планированием и стимулированием активных элементов. Предположим, что векторы $\mathbf{p}_t^i$ являются функциями наблюдаемого центром входного воздействия r_t, возмущаемыми некоторой случайной ненаблюдаемой помехой ξ_t. Тогда предельное состояние АЭ z_t^i также является случайной функцией $\mathbf{r}_t$. Предполагая, что все элементы функционируют в стационарном режиме (или режиме нормальной работы) [8], рассмотрим настраиваемую модель ограничений i-го элемента в следующем виде:

$$(4)\qquad \hat{z}_t^i=a_{t-1}^i q^i(\mathbf{r}_t),\ \hat{\mathbf{z}}_t=(\hat{z}_t^1, \ldots, \hat{z}_t^N)',\ \mathbf{a}_t=(a_t^1, \ldots, a_t^N)',$$

где $\hat{z}_t^i$ — оценка предельного состояния i-го АЭ в периоде t, q^i — некоторая функция наблюдаемого центром входного воздействия (внешних ресурсов) r_t, выбираемая центром из более общих соображений; a_t^i — настраиваемый параметр, $i=\overline{1,N}$. Пусть при настройке параметра $\mathbf{a}_t$ центр наблюдает предельное состояние $\mathbf{z}_t=(z_t^1, \ldots, z_t^N)'$, а его целью является минимизация средних потерь $J_t(a)=M_\xi\{\Phi_t|_{\mathbf{a}_{t-1}=\mathbf{a}}\}$, где функция потерь $\Phi_t=\Phi(\varepsilon_t)$ представляет собой функцию невязки $\varepsilon_t=\mathbf{z}_t-\hat{\mathbf{z}}_t$, M — оператор математического ожидания, Φ — выпуклая дважды дифференцируемая функция. Соответствующая процедура настройки параметров (или процедура обучения) согласно [6–8] имеет вид

(5) $$\mathbf{a}_t=\mathbf{a}_{t-1}-A\boldsymbol{\gamma}_t\nabla_{\mathbf{a}}\Phi_t\equiv I_t(\mathbf{a}_{t-1},\ \mathbf{z}_t),$$

где A — оператор, преобразующий вектор коэффициентов усиления γ_t в диагональную матрицу. Обозначим через $J_t(\mathbf{a},\ z^t)=\frac{1}{t}\sum_{\tau=0}^{t}\Phi_\tau|_{\mathbf{a}_{\tau-1}=\mathbf{a}}$ эмпирические средние потери, характеризующие качество обучения, $z^t=(\mathbf{z}_0,\ldots,\mathbf{z}_t)$. При этом предполагается, что имеет место сходимость в вероятностном смысле оптимальной выборочной оценки векторного параметра $\mathbf{a}_t$ к оптимальной оценке $\mathbf{a}^*$ в виде

(6) $$\mathbf{a}_t=\operatorname*{argmin}_{\mathbf{a}} J_t(\mathbf{a},\ z^t)\to\mathbf{a}^*=\operatorname*{argmin}_{\mathbf{a}} J_t(\mathbf{a}).$$

Если элементы обладают свойством активности, то они выбирают свои состояния из множества решений игры (3). Поэтому вместо $\mathbf{z}_t$ центр наблюдает $\tilde{\mathbf{y}}_t^*=(y_{1t}^*,\ldots,y_{Nt}^*)$, так что получаемые на основе процедуры (5) и наблюдений $\tilde{\mathbf{y}}_t^*$ оценки параметров

(7) $$\mathbf{a}_t=I_t(\mathbf{a}_{t-1},\ \tilde{\mathbf{y}}_t^*),$$

вообще говоря, не сходятся к оптимальной оценке (6).

Задача синтеза обучающегося механизма функционирования (ОМФ) ставится как задача выбора механизма $\Sigma=(I,\ \pi,\ f)$ с процедурами обучения $I=\{I_t\}$, планирования $\pi=\{\pi_t\}$ и стимулирования $f=\{f_t\}$, такого, что $\mathbf{a}_t=I_t(\mathbf{a}_{t-1},\ \tilde{\mathbf{y}}_t^*)_t\to\mathbf{a}^*$, $y_t^*\in R_t(\Sigma,\ p_t)$.

4. Решение задачи синтеза обучающегося механизма

Рассмотрим механизм $\Sigma=(I,\ \pi,\ f)$, использующий процедуру обучения (7) и следующие процедуры планирования и стимулирования:

(8) $$\mathbf{x}_{t+1}=\mathbf{a}_t,\quad \varphi_t^i=f_t^i(\mathbf{x}_t,\ \tilde{\mathbf{y}}_t).$$

Везде ниже, как и в [5], предполагается, что f_t^i дифференцируема и монотонно возрастает по каждой компоненте вектора $\tilde{\mathbf{y}}_t$ и убывает по компонентам $\mathbf{x}_t$. Следуя [5], обозначим

$$d\xi=\min_{\Omega_t}\frac{\partial}{\partial\xi},\qquad D\xi=\max_{\Omega_t}\frac{\partial}{\partial\xi},\quad \Omega_t=\{\nu=\overline{t,\ t+T};\qquad p_\tau\in P_\tau,$$

$$\tau=\overline{t,\ \nu};\qquad \mathbf{y}_\tau^k\in Y_\tau^k(\mathbf{u}_\tau^k),\ k=\overline{1,\ N}\},\ \underline{\nabla}_{\mathbf{x}}=(dx^1,\ldots,\ dx^N)',$$

$$\mathbf{F}_t^i=-\underline{\nabla}_{\mathbf{x}}'\varphi_\nu^i,\quad \mathbf{G}_{jt}=d\tilde{y}_j\nabla_{\mathbf{x}}\Phi_\nu,\quad H=\nabla_{\mathbf{x}}\nabla_{\mathbf{x}}\Phi_\nu,$$

E — единичная матрица размерности $N\times N$. Тогда справедлива следующая

Теорема 1. Пусть $\Sigma=(I,\ \pi,\ f)$ определяется согласно (7), (8) и удовлетворяет условию $\tilde{D}_{jt}^i(\Sigma)=(d\varphi_t^i w^i)(d\tilde{y}_j f_t^i)-\mathbf{F}_t^i\sum_{\tau=1}^{T}(D\varphi_{\tau+T}^i w_t^i)(E-A\boldsymbol{\gamma}_t H)^{\tau-1}\times$ $\times A\boldsymbol{\gamma}_t\mathbf{G}_{jt}\geqslant 0$ для любых $i,\ j,\ t$.

Тогда механизм Σ — обучающийся.

Доказательства теорем 1 и 2 даны в приложении.

Содержательно эта теорема устанавливает условия прогрессивности АМФ [3, 4], при выполнении которых выбирают состояние $y_t^*=v_t$ (т. е. не занижают свои показатели). Это обеспечивает возможность настройки параметров модели ограничений АЭ. Аналогичным образом могут быть получены конструктивные условия прогрессивности и для других итеративных процедур, позволяющие решать задачи анализа и синтеза АМФ, включающих в качестве подсистем типовые процедуры обучения (в том числе идентификации, оценивания, классификации) [6—8, 10].

Рассмотрим, например, задачу оптимального прогрессивного обучения, или, кратко, п-обучения, заключающуюся в выборе векторов γ_t, $t=0, 1, \ldots$, удовлетворяющих условиям теоремы 1 и минимизирующих $J_t(\text{a}, \tilde{y}^t)$. В общем случае это достаточно сложная задача нелинейного программирования, аналитическое решение которой невозможно. Для получения приближенного квазиоптимального решения $(\gamma_t^\text{п})$ предположим, что норма γ_t достаточно мала: $\|\gamma_t\|\ll 1$ [7]. Тогда условия теоремы 1 имеют вид

$$(9)\qquad \tilde{D}_{jt}^i(\Sigma)=\alpha_{jt}^i-(\beta_{jt}^i, \gamma_t)\geqslant 0, \quad \alpha_{jt}^i=(d\varphi_t^i w_t^i)(d\tilde{y}_j f_t^i),$$

$$\beta_{jt}^i=\sum_{\tau=t+1}^{t+T}(D\varphi_{t+T}^i w_t^i)\left[(dx_1 f_t^i)\left(d\tilde{y}_j\frac{\partial\Phi_t}{\partial x_1}\right), \ldots, (dx_N f_t^i)\left(d\tilde{y}_j\frac{\partial\Phi_t}{\partial x_N}\right)\right]$$

Заметим, что $\alpha_{jt}^i\geqslant 0$, $\beta_{jt}^i\geqslant 0$ вследствие естественных предположений о монотонности w_t^i, f_t^i, принятых в [4, 5] и в данной работе. Условия Куна — Таккера для определения $\gamma_t^\text{п}$ имеют вид

$$(10)\qquad \nabla_{\gamma t}J_t(a_t, y^t)+\sum_{i,\, j=1}^{N}\lambda_{jt}^i\beta_{jt}^i=0,$$

$$(11)\qquad (\beta_{jt}^i, \gamma_t)\leqslant\alpha_{jt}^i, \qquad \sum_{i,\, j=1}^{N}[(\beta_{jt}^i, \gamma_t)-\alpha_{jt}^i]\,\lambda_{jt}^i=0,\ \lambda_{jt}^i\geqslant 0.$$

Из (10), используя технику, развитую в § 3.4 [7], нетрудно получить, что $\gamma_t^\text{п}=\gamma_{t\,\text{конт}}-\delta_t$, где $\gamma_{t\,\text{копт}}$ — квазиоптимальный вектор коэффициентов усиления в отсутствие активности [7]:

$$\gamma_{t\,\text{копт}}=[A\nabla_a\Phi_t]^{-1}H_tA\nabla_a\Phi_t \text{ и } \delta_t=[A\nabla_a\Phi_tH_t^{-1}A\nabla_a\Phi_t]^{-1}\times$$
$$\times\sum_{i,\, j=1}^{N}\lambda_{jt}^i\beta_{jt}^i, \qquad H_t=\sum_{\tau=1}^{t}\nabla_a^2\Phi_\tau\,|_{\text{a}_{\tau-1}=\text{a}_{t-1}}.$$

Учитывая, что $A\nabla_\text{a}\Phi_t>0$ при $\beta_{jt}^i>0$, получаем, что $\gamma_t^\text{п}\leqslant\gamma_{t\,\text{копт}}$. В силу (11) при $\lambda_{jt}^i=0$ $\tilde{D}_{jt}^i(\Sigma)=\alpha_{jt}^i-(\beta_{jt}^i, \gamma_t)>0$, т. е. когда условия теоремы 1 заведомо выполняются, имеем $\gamma_t^\text{п}=\gamma_{t\,\text{копт}}$. И, наоборот, при $\lambda_{jt}^i>0$ $\gamma_t^\text{п}<\gamma_{t\,\text{копт}}$, т. е. величины коэффициентов усиления ограничены. Например, в простейшем случае при

$$(12)\qquad m=N=1, \quad \Phi(\varepsilon_t)=\varepsilon_t^2$$

имеем $\gamma_{t\,\text{копт}}=\dfrac{1}{t}\geqslant\gamma_t^\text{п}=\min\left(\dfrac{1}{t}, \dfrac{\alpha_t}{\beta_t}\right)$ (индексы i, $j=1$ опущены).

Содержательно это соответствует ограниченному «шагу» адаптивного планирования (8). Иными словами, при слишком высоких темпах планирования «от достигнутого» теряется заинтересованность АЭ в раскрытии резервов. Это ограничивает возможности использования алгоритмов оптимального обучения в автоматических системах.

Алгоритмы п-обучения могут быть реализованы, если они сходятся. Используя известные достаточные условия сходимости исходного алгоритма (5) [7] и теорему 1, можно получить достаточные условия сходимости алгоритмов п-обучения в виде ограничений на ОМФ (Σ), в частности на процедуру стимулирования (f).

Например, при выполнении (12) они определяются согласно (9) в виде $\sum\limits_{t=0}^{\infty}\gamma_t^{п}\leqslant\sum\limits_{t=0}^{\infty}(\alpha_t/\beta_t)=\infty$. Если же $\sum\limits_{t=0}^{\infty}(\alpha_t/\beta_t)<\infty$, то процедура (5) при $\gamma_t=\gamma_t^{п}$, вообще говоря, расходится. Это означает, что при слабом стимулировании за рост показателей состояния (α_t мало, см. (9)) или сильных штрафах за невыполнение плана (β_t велико) дальновидному элементу выгодно выбирать предельное состояние лишь при очень малом $\gamma_t^{п}$ (т. е. малом «шаге» адаптивного планирования), но тогда процедура обучения (5) не сходится.

5. Постановка задачи синтеза дуального механизма

До сих пор предполагалось, что размеры выделяемых активным элементам внешних ресурсов $\mathbf{r}_t$ (т. е. управления центра) не зависят от накапливаемой в процессе обучения информации.

При дуальном управлении центр совмещает обучение и управление активной системой [6, 9]. В этом случае оптимальное управление имеет дуальный характер: оптимизирует заданный критерий качества и в то же время способствует накоплению информации о неизвестных параметрах. Решение проблемы дуального управления чрезвычайно сложно даже в случае простых автоматических систем [6, 9]. Для простоты предположим, что $N=1$, и опустим индекс, указывающий номер АЭ ($i=1$). При этом $\tilde{y}_t=y_t$, $\mathbf{u}_t\equiv\mathbf{p}_t$ (см. разд. 2). Заметим, что при дуальном управлении в активной системе элемент должен учитывать влияние выбираемого им состояния (y_t) не только на будущие оценки параметров $\mathbf{a}_\tau$ и планы $\mathbf{x}_\tau$ (как в случае обучающихся механизмов, см. разд. 3), но и на будущие множества его возможных состояний $Y_\tau(\mathbf{p}_\tau)=[0,\ W_\tau(\mathbf{p}_\tau)]$, зависящие от выделяемых ресурсов — управлений центра $\mathbf{r}_\tau$, $\tau>t$. Для исследования этого влияния положим $\mathbf{p}_t=(\mathbf{q}_t',\ \xi_t)'$, где $\mathbf{q}_t=(q_{1t},\ldots,q_{lt})'$ — вектор-функция ресурсов $\mathbf{r}_\tau=(r_{1\tau},\ldots,r_{n\tau})'$, выделенных центром элементу в периодах $\tau=\overline{0,\ t-1}$: $q_{it}=q_i(\mathbf{r}_0,\ldots,\mathbf{r}_{t-1})$, $i=\overline{1,\ l}$; ξ_t — случайная ненаблюдаемая помеха, $\xi_t\in\theta_t$, θ_t — компакт. В свою очередь, при дуальном управлении упомянутые ресурсы предполагаются функциями оценок параметров элемента $\mathbf{a}_\tau$; $r_\tau=Q_\tau(\mathbf{a}_\tau)$.

Решение игры будем искать в виде (3), причем вместо (2) предполагается

$$(13)\qquad \hat{w}_t(\mathbf{x}_t,\ y_t)=\min_{\xi_\tau\in\theta_\tau,\ \tau=\overline{t+1,\ t+T}}\ \max_{y_\tau\in Y_\tau(\mathbf{p}_\tau)} w_t(\varphi_t,\ldots,\varphi_{t+T}).$$

Сравнивая (13) и (2), легко видеть, что в данном случае в роли $Y_{1\tau}^{1}(\mathbf{p}_\tau)$ естественным образом рассматривается множество состояний

АЭ $\operatorname*{Argmax}_{y_\tau \in Y_\tau(\mathbf{p}_\tau)} w_t$, максимизирующих его целевую функцию на $Y_\tau(\mathbf{p}_\tau)$. Содержательно это означает, что при построении прогноза собственной целевой функции $(\hat{w}(\mathbf{x}, y))$ элемент предполагает выбор им оптимальных состояний в будущем в неблагоприятных условиях (минимаксная стратегия учета неопределенности).

Дуальным механизмом будем называть механизм $\Sigma = (I, \pi, Q, f)$, при котором состояние АЭ равно предельному $(y_t = z_t)$ и одновременно экстремален некоторый критерий качества управления. В связи с многообразием возникающих задач синтеза дуальных механизмов целесообразно проиллюстрировать развиваемый подход на одном примере механизмов, использующих фильтр Калмана – Бьюси.

Пусть центр рассматривает модель ограничений АЭ следующего вида:

$$z_t = W(\mathbf{p}_t) = H\mathbf{q}_t + \xi_t, \quad \mathbf{q}_{t+1} = C\mathbf{q}_t + B\mathbf{r}_t,$$

где C, B, H — некоторые матрицы с неотрицательными элементами, ξ_t — последовательность случайных переменных с нулевым средним и известной ковариацией. Предположим, что $\mathbf{q}_\tau^0$ — желательное значение вектора $\mathbf{q}_\tau$ в периоде τ. Минимизируемый центром критерий качества имеет вид

$$J(r) = M_\xi \left\{ \sum_{r=0}^{b} (\mathbf{q}_{\tau+1} - \mathbf{q}_{\tau+1}^0)' K_\tau (\mathbf{q}_{\tau+1} - \mathbf{q}_{\tau+1}^0) + \mathbf{r}_\tau' N_\tau \mathbf{r}_\tau \right\}, \tag{14}$$

где K_τ, N_τ — квадратные матрицы, $[0, b]$ — интервал функционирования системы. Для получения оценки $\mathbf{a}_t$ вектора $\mathbf{q}_t$ на основе наблюдений z_t используется линейный фильтр Калмана – Бьюси (см., например, п. 7.3 [9]). Оценка записывается в форме

$$\mathbf{a}_{t+1} = A_t \mathbf{a}_t + G_t z_t. \tag{15}$$

Процедура управления имеет вид

$$\mathbf{r}_t = -L_t(\mathbf{a}_t - \mathbf{q}_t^0). \tag{16}$$

Элементы матриц A_t, G_t, L_t неотрицательны и определяются параметрами задачи. Процедуры (15) и (16) обеспечивают минимизацию критерия качества (14) в отсутствие активности элемента, т. е. при $y_t = z_t$. Однако указанная активность может приводить согласно (3), (13) к выбору элементом состояния y_t, отличающегося от предельного (т. е. $y_t < z_t$).

Рассмотрим теперь задачу синтеза дуального механизма функционирования динамической активной системы, использующего алгоритмы оценивания и управления (15), (16), при котором наблюдаемое состояние АЭ равно предельному ($y_t = z_t$, $t = 0, 1, \ldots$) и одновременно минимален критерий качества управления $J(r)$, определяемый согласно (14).

6. Решение задачи синтеза дуального механизма

В качестве вектора плана АЭ рассмотрим оценку $a_t(\mathbf{x}_t = \mathbf{a}_t)$. Тогда согласно (15) уравнения планов имеют вид

$$\mathbf{x}_{t+1} = A_t \mathbf{x}_t + G_t y_t. \tag{17}$$

Введем операторы $d\xi = \min\limits_{\Omega_t} \dfrac{\partial}{\partial \xi}$, $D\xi = \max\limits_{\Omega_t} \dfrac{\partial}{\partial \xi}$,

$$\Omega_t = \{ \nu = \overline{t, t+T}; \quad \mathbf{p}_\tau \in P_\tau, \quad \tau = \overline{t, \nu}; \quad y_\tau \in Y_\tau(\mathbf{p}_\tau) \}$$

и обозначим

$$U_t = dy\varphi_\nu,\quad V_t = Dy\varphi_\nu,\quad \underline{\nabla}_x = (dx_1, \ldots, dx_l)',\quad \mathbf{F}_t = -\underline{\nabla}'_x\varphi_\nu.$$

Будем предполагать, что $d\varphi_\tau w_t \geqslant D\varphi_\nu w_t \geqslant 0$, $t \leqslant \tau < \nu \leqslant t+T$, т. е. значимость стимулов для АЭ в перспективе убывает. Тогда справедлива следующая

Теорема 2. Для того чтобы АМФ $\Sigma = (I, \pi, Q, f)$ с процедурами обучения (15), планирования (17) и управления (16) был дуальным, достаточно

$$\tilde{D}(\Sigma) = U_t - \left[\mathbf{F}_t + \sum_{\tau=2}^{T}\left(\mathbf{F}_t + V_t H \sum_{\nu=1}^{\tau-1} C^{\tau-\nu} B L_t A^{\nu-\tau}\right) A_t^{\tau-1}\right] G_t \geqslant 0. \tag{18}$$

Условия (18) определяют ограничения на процедуру стимулирования (при заданных структуре элемента, процедурах планирования и управления), при которых АМФ прогрессивен и АЭ не занижает свои показатели. Заметим, что таким образом решена задача синтеза оптимального адаптивного механизма, использующего алгоритмы стохастического оптимального управления с обратной связью для линейной системы с известными параметрами, гауссовым шумом и квадратичным критерием качества [9].

Далее, поскольку прогрессивность означает увеличение поощрения АЭ при повышении эффективности его функционирования [1—3], возникает вопрос: как минимизировать затраты на стимулирование в адаптивных механизмах?

Рассмотрим, например, класс прогрессивных АМФ с заданными процедурами обучения (15), планирования (17) и управления (16) и ненулевыми предельными штрафами ($\mathbf{F}_t > 0$): $G = \{\Sigma = (I, \pi, Q, f) \mid \tilde{D}_t(\Sigma) \geqslant 0, \mathbf{F}_t > 0\}$.

Поставим задачу синтеза процедуры f^*, минимизирующей гарантированные затраты на стимулирование в классе АМФ G в форме

$$З_t(\Sigma) = \max_{\xi_t \in \theta_t} f_t(\mathbf{x}_t, z_t) \xrightarrow[\Sigma \in G]{} \min.$$

Здесь учтено, что $y_t = z_t$ в силу теоремы 2.

Соответствующий АМФ $\Sigma^* = (I, \pi, Q, f^*)$ будем называть минимально прогрессивным. Справедливо следующее

Утверждение[2]. Для минимальной прогрессивности АМФ достаточно $\tilde{D}_t(\Sigma^*) = 0$, $t = 0, 1, \ldots$ При этом $f_t^*(\mathbf{x}_t, y_t)$ является линейной функцией своих аргументов.

Доказательство этого утверждения, ввиду его простоты, не приводится. Оно устанавливает минимальные предельные стимулы (или максимальные предельные штрафы) [5], при которых может быть реализован дуальный механизм.

7. Обсуждение результатов

Как уже указывалось, теоремы 1, 2 носят конструктивный характер и позволяют решать задачи анализа и синтеза АМФ, включающих в качестве подсистем типовые процедуры адаптации и обучения [6—9]. При

[2] Аналогичное утверждение справедливо для случая ненулевых предельных стимулов ($U_t > 0$).

этом можно выделить два типа результатов. Первый связан с тем, что при заданных системах планирования и стимулирования могут быть получены условия прогрессивности типовых процедур обучения [6–8] (см. разд. 3). Второй связан с возможностью построения адаптивных процедур планирования и стимулирования, обеспечивающих прогрессивность механизмов, использующих оптимальные алгоритмы оценивания, идентификации и другие (как это было сделано в разд. 4 на основе теоремы 2). Полученные результаты можно без труда обобщить на случай обучения и управления по нескольким независимым наблюдаемым показателям состояния векторного АЭ ($m>N$), т. е. в случае, если $y_{kt}{}^i \leqslant z_{kt}{}^i = W_{kt}{}^i(\mathbf{p}_t{}^i)$, $i=\overline{1, N}$, $k=1, N+1, \ldots, m$. При этом общее число условий прогрессивности возрастает с N^2 (см. теорему 1) до mN.

В заключение отметим, что в работах [3–5] и в данной работе развивается направление в теории активных систем, связанное с анализом и синтезом адаптивных механизмов функционирования, в которых текущая информация, получаемая от активных элементов во время самого процесса управления, используется для изменения параметров подсистем планирования, управления и стимулирования механизма функционирования с целью достижения определенного, обычно оптимального, состояния активной системы. Тенденции и перспективы развития этого направления рассмотрены в [10].

Полученные результаты по анализу и синтезу прогрессивных адаптивных механизмов занимают центральное место в обосновании принципов и методов совершенствования реальных отраслевых систем планирования и управления [3, 10]. Действительно, последние основаны, как правило, на адаптивных по своей сути процедурах планирования и распределения ресурсов «от достигнутого». Кроме того, эти системы нацелены на использование резервов производства, т. е. должны быть «прогрессивными».

Автор приносит свою искреннюю благодарность В. Н. Буркову за обсуждение результатов работы.

ПРИЛОЖЕНИЕ

Доказательство теоремы 1. Заметим, что если $y_t{}^* = v_t$, то $\tilde{\mathbf{y}}_t{}^* = \mathbf{z}_t$, и из условия (6) следует, что $\mathbf{a}_t \underset{t}{\longrightarrow} \mathbf{a}_t{}^*$. Следовательно, в силу благожелательности АЭ по отношению к центру достаточно доказать, что $R_t(\Sigma, p_t) \ni v_t$.

Заметим, что с учетом (7), (8) процедура планирования имеет вид

$$\text{(П.1)} \qquad \mathbf{x}_t = \mathbf{x}_{t-1} - A\boldsymbol{\gamma}_t \nabla_{\mathbf{x}} \Phi(\tilde{\mathbf{y}}_t - \hat{\mathbf{z}}_t),$$

характерный для рекуррентной процедуры адаптивного планирования (1) [5]. Следовательно, справедлива лемма 2 [5]. Подставляя в (6), (7) [5] выражение (П.1), получаем для $t=0$, с учетом условий теоремы 1, $\hat{D}_{j0}{}^i(\Sigma) = \bar{D}_{j0}{}^i(\Sigma) \geqslant 0$. Но тогда согласно лемме [5] имеем

$$\text{(П.2)} \qquad D^i_{j0} = \min_{p_\tau \in P_\tau,\ \tau=\overline{0, T}} \ \min_{y^k_\tau \in \mathbf{Y}^k_\tau(\mathbf{p}^k_\tau),\ k=\overline{1, N}} \partial w^i_0 / \partial y^i_{j0} \geqslant 0.$$

Учитывая, что $Y_\tau{}^k(\mathbf{u}_\tau{}^k) \subset Y_\tau{}^k(\mathbf{p}_\tau{}^k)$ при $\mathbf{u}_\tau{}^k \leqslant \mathbf{p}_\tau{}^k$, и используя определения (3) и (П.2), нетрудно показать, что $\hat{w}_0{}^i(x_0, y_0)$ – неубывающая функция $y_{j0}{}^i$, $j=\overline{1, N}$. Отсюда в силу $y_{j0}{}^i \leqslant W_0{}^j(\mathbf{u}_0{}^j) \leqslant z_0{}^j$, $W_0{}^j(\cdot) \uparrow y_{j0}{}^k$, $y_{j0}{}^k \leqslant p_{j0}{}^k$, $k \neq j$, $1 \leqslant k \leqslant N$ имеем $\hat{w}_0{}^i(x_0, v_0) \geqslant \hat{w}_0{}^i(x_0, y_0)$, $y_0 = (\mathbf{y}_0{}^1, \ldots, \mathbf{y}_0{}^N)$ $\forall \mathbf{y}_0{}^l \in Y_0{}^l(\mathbf{p}_0{}^l) \subset Y_0{}^l(\mathbf{u}_0{}^l)$.

Но тогда из (3) следует, что $R_0(\Sigma, p_0) \ni v_0$. Аналогично можно показать, что $R_t(\Sigma, p_t) \ni v_t$, $t=1, 2, \ldots$, откуда следует искомое утверждение.

Доказательство теоремы 2. Достаточно показать, что $y_t{}^* = z_t$, $t=0, 1, \ldots$ Для доказательства используем рассуждения по индукции. Обозначим $c_t(y_t, \ldots, y_{t+T}) \equiv w_t(\varphi_t, \ldots, \varphi_{t+T})$ и зафиксируем некоторую произвольную последовательность значений случайной помехи $\xi^t=\{\xi_t, \ldots, \xi_{t+T}\}$. При $\tau=t+T$ имеем $\partial c_t/\partial y_{t+T} \geqslant \geqslant (d\varphi_{t+T} w_t) U_t \geqslant 0$ согласно условию (18) (напомним, что в силу сделанных предположений все члены кроме первого в $\bar{D}_t(\Sigma)$ неположительны). Поэтому $c_t(y_t, \ldots \ldots, y_{t+T-1}, W_{t+T}(\mathbf{p}_{t+T})) \geqslant c_t(y_t, \ldots, y_{t+T-1}, y_{t+T})$ при любых $y_{t+1} \in Y_{t+T}(\mathbf{p}_{t+T}) = =[0, W_{t+T}(\mathbf{p}_{t+T})]$. Предположим теперь, что для некоторого ν, $t<\nu\leqslant t+T$ $c_t(y_1, \ldots \ldots y_\nu, W_{\nu+1}(\mathbf{p}_{\nu+1}), \ldots) \geqslant c_t(y_1, \ldots, y_\nu, y_{\nu+1}, W_{\nu+2}(\mathbf{p}_{\nu+2}), \ldots)$, $y_{\nu+1} \leqslant W_{\nu+1}(\mathbf{p}_{\nu+1})$, и покажем, что при $y_\nu \leqslant W_\nu(\mathbf{p}_\nu)$ имеет место

$$c_t(y_1, \ldots, W_\nu(\mathbf{p}_\nu), W_{\nu+1}(\mathbf{p}_{\nu+1}), \ldots) \geqslant c_t(y_1, \ldots, y_\nu, W_{\nu+1}(\mathbf{p}_{\nu+1}), \ldots). \tag{П.3}$$

Действительно, при этом $\partial c_t/\partial y_\nu \geqslant (d\varphi_\nu w_t)(dy\, \varphi_\nu) - (D\varphi_{\nu+1} w_t)\mathbf{F}_t G_t - \sum_{\tau=\nu+2}^{T} (D\varphi_\tau w_t) \times$

$$\times \left[\mathbf{F}_t + V_t H \sum_{\mu=1}^{\tau-1} c^{\tau-\mu} B L_t A_t^{\mu-\tau}\right] A^{\tau-1} G_t \geqslant (d\varphi_\nu w_t)\bar{D}_t(\Sigma) \geqslant 0.$$

Первое неравенство получено с помощью техники, развитой при доказательстве леммы [5]. При этом использовано свойство неположительности всех слагаемых (кроме первого) в правой его части. Второе неравенство справедливо в силу $d\varphi_\nu w_t \geqslant D\varphi_\tau w_t \geqslant 0$, $\tau \geqslant \mu$, третье – по условию теоремы. Следовательно, имеет место (П.3) при $t\leqslant\nu\leqslant t+T$. Отсюда $c_t(W_t(\mathbf{p}_t), \ldots, W_{t+T}(\mathbf{p}_{t+T})) \geqslant c_t(y_t, \ldots, y_{t+T})$ $\forall y_\tau \in Y_\tau(\mathbf{p}_\tau)$, $\tau=\overline{t, t+T}$ при любой последовательности ξ^t. Но тогда согласно (3), (13) $R_t(\Sigma, \mathbf{p}_t) \ni W_t(\mathbf{p}_t) = z_t$, и в силу принципа благожелательности $y_t^* = z_t$. Учитывая, что при этом управление (16) минимизирует (14), получаем, что рассматриваемый механизм по определению является дуальным, что и требовалось доказать.

ЛИТЕРАТУРА

1. *Бурков В. Н.* Математические основы теории активных систем. М.: Наука, 1977.
2. *Бурков В. Н., Кондратьев В. В.* Механизмы функционирования организационных систем. М.: Наука, 1981.
3. *Бурков В. Н., Кондратьев В. В., Цыганов В. В., Черкашин А. М.* Теория активных систем и совершенствование хозяйственного механизма. М.: Наука, 1984.
4. *Бурков В. Н., Цыганов В. В.* Адаптивные механизмы функционирования активных систем. I // АиТ. 1985. № 9. С. 87–94.
5. *Цыганов В. В.* Адаптивные механизмы функционирования активных систем. II // АиТ. 1985, № 10. С. 116–123.
6. *Цыпкин Я. З.* Адаптация и обучение в автоматических системах. М.: Наука, 1968.
7. *Цыпкин Я. З.* Основы теории обучающихся систем. М.: Наука, 1970.
8. *Цыпкин Я. З.* Основы информационной теории идентификации. М.: Наука, 1984.
9. *Саридис Дж.* Самоорганизующиеся стохастические системы управления. М.: Наука. 1980.
10. *Бурков В. Н., Цыганов В. В.* Тенденции и перспективы построения адаптивных механизмов функционирования активных систем // Измерения, контроль, автоматизация. 1985. № 4. С. 53–60.

Поступила в редакцию
16.I.1986

ADAPTIVE FUNCTIONING MECHANISMS OF ACTIVE SYSTEMS.

III. LEARNING AND CONTROL

TSYGANOV V. V.

In designing adaptive functioning mechanisms for two-level active systems the information submitted by the active elements is employed to change the parameters of the mechanism itself and is conductive to the goals of the Center. Design problems are stated and constructive sufficient conditions found for the learning mechanisms where the planning and encouragment of the active elements proceed simultaneously with the Center acquiring new skills. An approach is discussed to design of dual mechanisms where a management performance criterion is also optimized. The associated sufficient optimality conditions are obtained.