Доклады Академии наук СССР
1971. Том 199, № 1

УДК 62-50+519.3 КИБЕРНЕТИКА И ТЕОРИЯ РЕГУЛИРОВАНИЯ

Ю. Н. ЖЕЛНИН

ОБ ОПТИМАЛЬНОМ УПРАВЛЕНИИ ПРИ НЕПОЛНОЙ ИНФОРМАЦИИ

(Представлено академиком А. А. Дородницыным 28 XII 1970)

Известные в настоящее время методы решения вариационных задач оптимального управления позволяют определить управление либо как функцию от начальных условий и аргумента (программное управление), либо как функцию текущих фазовых координат системы и аргумента (синтез управления). Следовательно, при создании оптимальной системы управления необходимо располагать полной информацией о начальном и текущем фазовом состоянии объекта управления. Однако существует обширный класс динамических систем, в которых информация о положении в фазовом пространстве является неполной и ограничена измерительным устройством, которым располагает система. В этом случае возникает задача об отыскании оптимального управления в условиях неполной информации. В настоящей работе предлагается возможный подход к решению этой задачи для нелинейных детерминированных систем, который позволяет свести задачу управления с ограниченной информацией о текущем фазовом состоянии к игровой задаче с ограничениями на фазовые координаты в некотором расширенном пространстве. При этом ограничению на информацию в исходной задаче управления соответствуют ограничения, типа равенств, на фазовые координаты в редуцированной игровой задаче в расширенном пространстве.

1. Пусть поведение динамической системы описывается уравнениями

$$\dot{x} = f(x, u, \xi), \quad y(t) - \varphi(x, u, \xi, w) = 0,$$
$$I = \int_{t_0}^{T} f^0(x, u, \xi)\, dt + F[x(T)], \quad x(t_0) = x^0 \in X^0. \tag{1}$$

Здесь $x(x_1 \ldots x_n)$ — вектор фазовых координат, $u(u_1 \ldots u_r) \in U$ — вектор управлений, $\xi(\xi_1 \ldots \xi_q) \in \Xi$ — вектор возмущений. Качество управления оценивается функционалом I. Информация о текущем фазовом состоянии определяется вектором измерений $y(y_1 \ldots y_p)$, $w(w_1 \ldots w_m) \in W$ — вектор погрешностей измерений. Векторы $x^0 \in X^0$, $\xi \in \Xi$, $w \in W$ считаются неизвестными, однако границы области их возможных значений известны; X^0, Ξ, W, U — ограниченные, замкнутые области возможных значений величин x^0, $\xi(t)$, $w(t)$, $u(t)$. Функции f, φ, f^0, F обладают обычными для этого класса задач свойствами [1-5]. Предполагается, что система уравнений $y(t) - \varphi(x, u, \xi, w) = 0$ не может быть однозначно разрешена относительно x, u, ξ, w (т. е. $p < n + r + q + m$). Следовательно, вектор $y(t)$ представляет «неполную» информацию о текущем фазовом состоянии. Зависимость $u(y(t), t)$ назовем **синтезом управления при «неполной» информации**. Задача заключается в определении оптимального синтеза $u^*(y(t), t)$, минимизирующего функционал I.

2. Предварительно рассмотрим задачу о наблюдении. Пусть на интервале (t_0, T) известен вектор $y(t)$. Требуется определить функционал I при условии, что x^0, ξ, w неизвестны, а $u(t)$ известно. Поскольку $y(t)$ неоднозначно определяет $x(t)$, $\xi(t)$, $w(t)$, то существует некоторое множество значений x^0 и функций $\hat{\xi}(t)$, $\hat{w}(t)$ и, следовательно, множество значений

$I(\hat{x}^0, u, \hat{\xi}, \hat{w}) = I$, которые удовлетворяют дифференциальному уравнению, описывающему поведение системы, и уравнению измерительного устройства

$$\begin{gathered}\dot{\hat{x}} = f(\hat{x}, u, \hat{\xi}), \quad y(t) - \varphi(\hat{x}, u, \hat{\xi}, \hat{w}) = 0, \\ \hat{I} = \int_{t_0}^{T} f^0(\hat{x}, u, \xi)\, dt + F[\hat{x}(T)], \quad \hat{x}(t_0) = \hat{x}^0 \in X^0.\end{gathered} \tag{2}$$

Очевидно, что искомой величиной I является одно из значений $\hat{I}$. Поскольку нельзя установить однозначно, какое из этих значений совпадает с действительным, то в качестве оценки искомой величины выберем такое $\hat{I} = \hat{I}^*$, которое обеспечивает минимум максимальной возможной ошибки оценки. Пусть $\rho(I - \hat{I})$ — некоторая функция, характеризующая величину отклонения оценки $\hat{I}$ от действительного значения I. Тогда оптимальная оценка $\hat{I}^* = I(\hat{x}^0, u, \hat{\xi}, \hat{w})$ определится из того, что $\hat{x}^0$, $\hat{\xi}(t)$, $\hat{w}(t)$, удовлетворяющие (2), минимизируют, а x^0, $\xi(t)$, $w(t)$, удовлетворяющие (1), максимизируют функционал $\rho(I - \hat{I})$. Это условие может неоднозначно определять вектор оценки $\hat{x}(t)$: $\hat{x}^0$, $\hat{\xi}$, $\hat{w}$. Тогда для однозначного определения его могут быть сформулированы дополнительные условия. Условия задачи в этом случае имеют вид

$$\begin{gathered}\dot{x} = f(x, u, \xi), \quad y(t) - \varphi(x, u, \xi, w) = 0, \\ \dot{\hat{x}} = f(\hat{x}, u, \hat{\xi}), y(t) - \varphi(\hat{x}, u, \hat{\xi}, \hat{w}) = 0, \\ \rho^* = \min_{\hat{x}^0, \hat{\xi}, \hat{w}} \max_{x^0, \xi, w} \rho(I - \hat{I}).\end{gathered} \tag{3}$$

Сформулированная задача является игровой с ограничениями типа равенств на фазовые координаты x, $\hat{x}$ и управления ξ, $\hat{\xi}$, w, $\hat{w}$. Роль ограничений выполняют уравнения измерительного устройства. Не останавливаясь на методах решения таких задач в общем случае, укажем, что известные результаты теории дифференциальных игр позволяют при определенных допущениях установить необходимые условия оптимальности как в чистых ([1-4]), так и в смешанных стратегиях ([5, 7]). Приведем необходимые условия оптимальности в чистых стратегиях в области «регулярности» функционала ([2, 6]) (для простоты принято $f^0 = 0$, φ — скаляр).

$$\begin{gathered}\mathscr{H} = \max_{(\hat{\xi} \in \Xi,\, \hat{w} \in W,\, y - \hat{\varphi} = 0)} \min_{(\xi \in \Xi,\, w \in W,\, y - \varphi = 0)} \psi \cdot f + \hat{\psi} \cdot \hat{f}, \\ \dot{\psi} = -\frac{\partial \mathscr{H}}{\partial x} + \lambda \frac{\partial \varphi}{\partial x}, \quad \psi(T) = -\frac{\partial \rho}{\partial x}, \quad \dot{\hat{\psi}} = -\frac{\partial \mathscr{H}}{\partial \hat{x}} + \hat{\lambda} \frac{\partial \varphi}{\partial \hat{x}}, \\ \hat{\psi}(T) = -\frac{\partial \rho}{\partial \hat{x}}.\end{gathered} \tag{4}$$

Здесь $\mathscr{H}$ — гамильтониан системы, $\psi(\psi_1 \ldots \psi_n)$, $\hat{\psi}(\hat{\psi}_1 \ldots \hat{\psi}_n)$ — векторы вспомогательных переменных, $\lambda(t)$, $\hat{\lambda}(t)$ — неопределенные множители. Необходимо отметить, что условия (4) далеко не всегда определяют решение, так как оно может находиться в области «нерегулярности», для которой условия (4) не определяют решения ([2, 6]). Одним из путей, позволяющих избежать трудности, связанные с наличием областей «нерегулярности», является сведение игровой задачи (3) к двум вариационным задачам:

$$I^+ = \max_{x^0, \xi, w} I, \quad \dot{x} = f(x, u, \xi), \quad y(t) - \varphi(x, u, \xi, w) = 0;$$

$$I^- = \min_{x^0, \xi, w} I, \quad \dot{x} = f(x, u, \xi), \quad y(t) - \varphi(x, u, \xi, w) = 0.$$

После введения дополнительной координаты x_w, удовлетворяющей уравнению $\dot{x}_w = w$ с произвольными краевыми условиями, указанные задачи представляют собой обычные вариационные задачи, методы решения которых хорошо известны. Нетрудно убедиться, что оптимальная оценка $\hat{I}^* = 0{,}5(I^+ + I^-)$.

3. Таким образом, решение задачи о наблюдении (2) позволяет по известному вектору $y(t)$ найти оценку функционала $\hat{I}^*$ с некоторой «гарантированной» точностью. Напомним, что управление $u(t)$ при этом считается заданным и известным. Теперь будем считать управление незаданным и определим его таким образом, чтобы оценка функционала $\hat{I}$ была минимальна, а точность оценки была равна заданной, т. е. $\rho - \bar{\rho} = 0$. Здесь $\bar{\rho}$ — заданное значение. Условия задачи оптимального управления имеют вид

$$\begin{gathered}
\dot{x} = f(x, u, \xi), \quad \dot{\hat{x}} = f(\hat{x}, u, \hat{\xi}), \\
\varphi(x, u, \xi, w) - \varphi(\hat{x}, u, \hat{\xi}, \hat{w}) = 0, \\
\hat{I}^* = \min_{u,\, \hat{x}^0,\, \hat{\xi},\, \hat{w}} \ \max_{x^0, \xi, w} \hat{I} + \lambda^0 \rho (I - \hat{I}).
\end{gathered} \tag{5}$$

Начальные условия определяются заданием начального вектора $y(t_0) = y^0$, множитель λ^0 определится из условия $\rho - \bar{\rho} = 0$. Задача управления (5), как и задача о наблюдении (3), является игровой задачей с ограничениями на фазовые координаты и управления и ее решение принципиально не отличается. Из решения задачи (5) при некотором начальном значении вектора $y(t_0) = y^0$ определим оптимальное управление $u^*(y^0, t)$ (программное управление). Затем, считая вектор y^0 текущим, получим $u^*(y(t), t)$ (синтез оптимального управления).

4. Используя аналогичный подход, сформулируем задачу оптимального управления в дифференциальной игре с двумя игроками, каждый из которых имеет неполную информацию о состоянии игры. Пусть дифференциальная игра описывается уравнениями

$$\begin{gathered}
\dot{x} = f(x, u, v, \xi), \quad x(t_0) = x^0 \in X^0, \\
y_u(t) - \varphi_u(x, u, v, \xi, w_u) = 0, \quad y_v(t) - \varphi_v(x, u, v, \xi, w_v) = 0, \\
I = \int_{t_0}^{T} f^0(x, u, v, \xi)\, dt + F[x(T)].
\end{gathered} \tag{6}$$

Здесь $u(u_1 \ldots u_r) \in U$, $v(v_1 \ldots v_q) \in V$ — векторы управлений игроков P и E, соответственно $\xi(\xi_1 \ldots \xi_p) \in \Xi$ — вектор возмущений, векторы $y_u(y_{u1} \ldots y_{ui})$, $y_v(y_{v1} \ldots y_{vj})$ характеризуют информацию игроков о текущем состоянии игры, $w_u(w_{u1} \ldots w_{uh}) \in W_u$, $w_v(w_{v1} \ldots w_{vl}) \in W_v$ — векторы погрешностей информации. Качество игры оценивается функционалом I. Игрок P минимизирует, игрок E максимизирует этот функционал.

По аналогии с задачей управления введем в рассмотрение векторы оценки фазового состояния

$$\hat{x}_u(t) : \hat{x}_u^0, u, \hat{v}_u, \hat{\xi}_u, \hat{w}_u; \quad \hat{x}_v(\hat{t}) : x_v^0, \hat{u}_v, v, \hat{\xi}_v, \hat{w}_v,$$

которым соответствуют оценки функционала $\hat{I}_u$ и $\hat{I}_v$, определяемые каждым из игроков с точностью, характеризуемой функциями ρ_u и ρ_v, и сформулируем исходную игровую задачу (6) следующим образом:

$$\begin{gathered}
\dot{x} = f, \ \dot{\hat{x}}_u = \hat{f}_u, \ \dot{\hat{x}}_v = \hat{f}_v, \ \varphi_u - \hat{\varphi}_u = 0, \ \varphi_v - \hat{\varphi}_v = 0; \\
I_u^* = \min_{\alpha_P} \max_{\beta_E} \hat{I}_u + \lambda_u^0 \rho_u, \ y_u(t_0) = y_u^0; \\
I_v^* = \max_{\alpha_E} \min_{\beta_P} \hat{I}_v + \lambda_v^0 \rho_v, \ y_v(t_0) = y_v^0.
\end{gathered}$$

Здесь символами α_P, β_P, α_E, β_E обозначены величины, которыми могут распоряжаться игроки P и E соответственно. Считается, что каждому из игроков «свое» управление известно, поэтому $\hat{u}_u = u$, $\hat{v}_v = v$.

Сформулированная задача, как и рассмотренная ранее, представляет собой игровую задачу с ограничениями на фазовые координаты и управления. Различные модификации этой задачи могут возникнуть в зависимости от предположений о знании каждым из игроков информации, получаемой противником (знание игроком P функции φ_v и игроком E функции φ_u), и предположений о неизвестном начальном векторе x^0 и возмущений $\xi(t)$, которые, по существу, представляют третьего игрока и могут как содействовать, так и противодействовать каждому из игроков P и E. От этих предположений зависят составляющие величин, входящие в α_P, β_P, α_E, β_E.

5. Изложенный подход позволяет свести к игровой задаче с ограничениями на фазовые координаты, помимо рассмотренных, также задачи с фиксированным правым концом, задачи с ограничениями на фазовые координаты, задачи с запаздывающей информацией и др. Задачи, в которых информация о текущем состоянии определяется дифференциальными соотношениями, легко сводится к рассмотренной.

Отметим, что один из возможных подходов состоит в следующем: продифференцировав соотношение $y(t) - \varphi(x, u, \xi, w) = 0$ и используя уравнение $\dot{x} = f(x, u, \xi)$, получим дифференциальное уравнение, которому удовлетворяет $y(t)$. После этого игровая задача управления может быть сформулирована в пространстве координат вектора $y(t)$. Однако этот подход не упрощает решение, поскольку приводит к ряду дополнительных требований к функциям φ, u, ξ, w.

В заключение автор выражает благодарность А. И. Курьянову за внимание к работе.

Центральный аэрогидродинамический институт
им. Н. Е. Жуковского

Поступило
15 XII 1970

ЦИТИРОВАННАЯ ЛИТЕРАТУРА

[1] М. И. Зеликин, Н. Т. Тынянский, УМН, 136, № 4, 151 (1965). [2] G. Leitman, G. Mon, J. Astronaut. Sci., 14, № 2, 56 (1967). [3] В. М. Гаврилов, Оптимальные процессы в конфликтных ситуациях, М., 1969. [4] А. И. Пропой, Исследование операций, Тр. Вычислит. центра АН СССР, в. 1, 103 (1970). [5] Э. Р. Смольяков, Там же, в. 1, 90 (1970). [6] Р. Айзекс, Дифференциальные игры, М. 1967. [7] Н. Н. Красовский, А. И. Субботин, ДАН, 191, № 2 (1970).