

Киселев Ю.Н.
(кафедра оптимального управления)
**МЕТОД ДИНАМИЧЕСКОГО ПРОГРАММИРОВАНИЯ В НЕПРЕРЫВНЫХ
УПРАВЛЯЕМЫХ СИСТЕМАХ**

Лекция 1

1 Метод динамического программирования в задаче быстрого действия

1.1 Предварительные замечания

Метод *динамического программирования* [2] разработан Р.Беллманом (1920-1984) для оптимального управления процессами, имеющими более общий характер, чем процессы, описываемые обыкновенными дифференциальными уравнениями. Для управляемых процессов с уравнениями движения в форме обыкновенных дифференциальных уравнений известны необходимые условия оптимальности — *принцип максимума Понтрягина* [1]; эти условия, как известно, формулируются в терминах обыкновенных дифференциальных уравнений. В методе динамического программирования условия оптимальности выражаются в терминах дифференциального уравнения в частных производных — уравнения Беллмана. Мы рассмотрим сначала метод динамического программирования применительно к задаче быстрого действия. Для обоснования метода динамического программирования требуются определенные предположения о гладкости так называемой функции Беллмана $V(x)$, которая вводится ниже. Эти предположения не вытекают из постановки задачи и в ряде простых примеров не выполняются, что приводит к ограниченности области применимости метода динамического программирования в непрерывных задачах оптимального управления, см. [1], §9.

В классическом *втором методе Ляпунова*, см. [3], для исследования устойчивости систем обыкновенных дифференциальных уравнений центральную роль играет операция *дифференцирования функции V в силу системы*. Эта операция, по существу, является также и основным техническим инструментом динамического программирования для управляемых систем. В этом смысле теория динамического программирования может рассматриваться как определенное развитие теории второго метода Ляпунова применительно к управляемым системам с оптимизируемым функционалом. Поэтому функция Беллмана обозначается буквой V , традиционно применяемой для функции Ляпунова. Вычислительные схемы динамического программирования в дискретных системах описаны в [4].

Замечание 1.1. Дифференциальное уравнение Беллмана является неклассическим дифференциальным уравнением в частных производных, содержащим операцию взятия максимума. Экстремальные операции ухудшают свойства гладкости функций, см. [5]. Это качественное утверждение иллюстрируется следующими двумя примерами.

Пример 1.1. Рассмотрим набор из N гладких функций $\phi_1(x), \dots, \phi_N(x)$ и положим

$$\Phi(x) = \max\{\phi_1(x), \dots, \phi_N(x)\} = \max_{i=1, \dots, N} \phi_i(x).$$

Функция $\Phi(x)$, как правило, оказывается негладкой. Например, при $n = 2, \phi_1(x) = x, \phi_2(x) = -x$ имеем $\Phi(x) = |x|$; последняя функция *недифференцируема* в точке $x = 0$.

Пример 1.2. Рассмотрим гладкую функцию двух переменных $f(x, y)$ и положим

$$F(x) = \max_{y \in Y} f(x, y),$$

где множество Y замкнуто и ограничено. Функция $F(x)$, как правило, недифференцируема. Например, опорная функция выпуклого компакта U

$$c(\psi) = \max_{u \in U} (\psi, u)$$

может оказаться недифференцируемой, хотя линейная функция $f(\psi, u) = (\psi, u)$, к которой применяется операция максимизации, является гладкой. В частности, если множество $U = S_1(0)$ — шар радиуса единица, то его опорная функция $c(\psi) = \|\psi\|$ *недифференцируема* в точке $\psi = 0$; если $U = \{|u_1| \leq 1, |u_2| \leq 1\}$ — квадрат на плоскости, то его опорная функция $c(\psi_1, \psi_2) = |\psi_1| + |\psi_2|$ *недифференцируема* на множестве $\{\psi_1 = 0\} \cup \{\psi_2 = 0\}$.

Перечислим основные идеи метода динамического программирования:

- *погружение исходной оптимизационной задачи в семейство аналогичных задач, введение функции Беллмана;*
- *формулировка и обоснование принципа оптимальности Беллмана для динамических систем, развивающихся во времени;*
- *вывод уравнения Беллмана для функции Беллмана, исследование связи гладких решений уравнения Беллмана с решением исходной оптимизационной задачи.*

Обратимся сначала к краткому описанию *идеи погружения*. Рассмотрим оптимизационную задачу

$$f(x) \rightarrow \min_{x \in X}. \quad (p)$$

Вместе с задачей (p) рассматривается семейство оптимизационных задач

$$f(x; \alpha) \rightarrow \min_{x \in X(\alpha)}, \quad (p_\alpha),$$

зависящее от некоторого параметра $\alpha \in A$, так что

$$f(x; \alpha)|_{\alpha=\alpha_0} = f(x), \quad X(\alpha)|_{\alpha=\alpha_0} = X.$$

Таким образом, задача (p_α) при $\alpha = \alpha_0$ превращается в задачу (p) . Задача (p) включена (погружена) в семейство задач (p_α) . Предполагая разрешимость задачи (p_α) при всех $\alpha \in A$, введем обозначение

$$T(\alpha) = \min_{x \in X(\alpha)} f(x; \alpha)$$

для оптимального значения в задаче (p_α) . Вместе с функцией $T(\alpha)$ традиционно рассматривается функция

$$V(\alpha) = -T(\alpha),$$

которая называется *функцией Беллмана* и играет центральную роль в методе динамического программирования.

Принцип оптимальности — вторая фундаментальная идея динамического программирования для систем, развивающихся во времени — может быть кратко выражена следующим образом: любой кусок оптимальной траектории оптимален. В каждой конкретной задаче или в классе задач принцип оптимальности Беллмана подлежит обоснованию. Опираясь на принцип оптимальности для функции $V(\alpha)$, выводят некоторое специальное уравнение — уравнение Беллмана, составляющее

основу метода динамического программирования. При определенных, довольно жестких, предположениях метод динамического программирования позволяет описать необходимые и достаточные условия оптимальности.

Применение метода динамического программирования мы рассмотрим сейчас в задаче быстродействия.

1.2 Постановка задачи быстродействия. Погружение. Проверка принципа оптимальности. Формулировка Предположений 1.1, 1.2 о функции Беллмана

Рассмотрим задачу быстродействия

$$\begin{cases} \dot{x} = f(x, u), & x \in \mathbb{R}^n, & u \in U, \\ x(t_0) = x_0, & x(t_1) = x_1, \\ T \equiv t_1 - t_0 \rightarrow \min, \end{cases} \quad (P)$$

в которой допустимыми управлениями являются кусочно-непрерывные функции времени t со значениями в области управления U , $f(x, u)$ — n -мерная векторная функция с координатами $f^1(x, u), \dots, f^n(x, u)$, причем функции $f^i(x, u)$, $i = 1, \dots, n$, и их частные производные $\frac{\partial f^i(x, u)}{\partial x^j}$, $i, j = 1, \dots, n$, определены и непрерывны на прямом произведении $\mathbb{R}^n \times U$.

При построении семейства задач (P_α) в качестве параметра погружения α выбирается начальное состояние x_0 управляемого объекта. Задача изучается при определенных условиях (**Предположения 1.1, 1.2**).

Предположение 1.1. При любом начальном состоянии $x_0 \in \mathbb{R}^n$ существует оптимальное управление, которое осуществляет перевод объекта из x_0 в конечную точку x_1 со временем быстродействия $T(x_0)$, причем

$$T(x_1) = 0, \quad T(x_0) > 0 \quad \text{при} \quad x_0 \neq x_1.$$

Таким образом, функция Беллмана

$$V(x) = -T(x)$$

определена при всех $x \in \mathbb{R}^n$, причем

$$V(x_1) = 0, \quad V(x) < 0, \quad x \neq x_1.$$

Предположение 1.2. Функция $T(x)$ непрерывна при всех $x \in \mathbb{R}^n$ и имеет непрерывный градиент $T'(x)$ при $x \in \mathbb{R}^n \setminus \{x_1\}$. Таким образом, функция Беллмана $V(x)$ непрерывна в $\mathbb{R}^n$ и имеет непрерывный градиент $V'(x)$ в $\mathbb{R}^n \setminus \{x_1\}$.

Замечание 1.2. Предположение о гладкости функции Беллмана может не выполняться. Например, в задаче быстродействия для тележки [1] функция оптимального времени $T(x)$ недифференцируема на линии переключения AOB , [1], §5, Пример 1.

Для задачи быстродействия (P) легко выполнить обоснование принципа оптимальности. Покажем, что любой "хвост" оптимальной траектории оптимален. Пусть

$$x(t), \quad t_0 \leq t \leq t_1; \quad x(t_0) = x_0, \quad x(t_1) = x_1, \quad (1)$$

— оптимальная траектория, отвечающая оптимальному управлению

$$u(t), \quad t_0 \leq t \leq t_1; \quad u(t) \in U. \quad (2)$$

Возьмем любое число $\tau \in (t_0, t_1)$ и проверим, что траектория

$$x(t), \tau \leq t \leq t_1,$$

("хвост"траектории (1)) оптимальна по быстродействию из точки $x(\tau)$ в момент времени τ в точку x_1 в момент времени t_1 . Если допустить противное, то существуют допустимое управление

$$\hat{u}(t), \tau \leq t \leq \hat{t}_1,$$

и отвечающая ему траектория

$$\hat{x}(t), \tau \leq t \leq \hat{t}_1,$$

такие, что

$$\hat{x}(t)|_{t=\tau} = x(\tau), \quad \hat{x}(t)|_{t=\hat{t}_1} = x_1, \quad \hat{t}_1 < t_1.$$

Тогда допустимое управление

$$u_*(t) = \begin{cases} u(t), & t_0 \leq t \leq \tau, \\ \hat{u}(t), & \tau < t \leq \hat{t}_1, \end{cases}$$

порождает траекторию

$$x_*(t) = \begin{cases} x(t), & t_0 \leq t \leq \tau, \\ \hat{x}(t), & \tau \leq t \leq \hat{t}_1, \end{cases}$$

вдоль которой переход из точки x_0 в точку x_1 происходит быстрее, чем вдоль траектории (1), что противоречит ее оптимальности. Принцип оптимальности обоснован. Заметим, что управление $u_*(t)$ допустимо, так как класс допустимых управлений состоит из кусочно-непрерывных функций и возможный скачек управления в точке τ не выводит управление $u_*(t)$ из класса допустимых управлений.

1.3 Вывод уравнения Беллмана. Необходимые условия оптимальности

Рассмотрим оптимальную пару $(x(t), u(t))$, $t_0 \leq t \leq t_1$. На основе принципа оптимальности можно записать равенство

$$T(x(t)) = \Delta t + T(x(t + \Delta t))$$

при любом $\Delta t > 0$; $t, t + \Delta t \in [t_0, t_1]$. Отсюда вытекает соотношение

$$\frac{V(x(t + \Delta t)) - V(x(t))}{\Delta t} = 1.$$

Переходя в нем к пределу при $\Delta t \rightarrow 0$, получаем

$$\frac{d}{dt}V(x(t)) = 1,$$

или

$$(V'(x(t)), \dot{x}(t)) = 1.$$

При выводе последнего соотношения мы использовали **Предположение 1.2** о гладкости функции Беллмана и считали, что в точке $t \in [t_0, t_1]$ управление $u(t)$ непрерывно, что влечет существование производной $\dot{x}(t)$. Привлекая дифференциальное уравнение управляемого движения, из последнего равенства получаем *необходимое условие оптимальности* пары $(x(t), u(t))$, $t_0 \leq t \leq t_1$:

$$(V'(x(t)), f(x(t), u(t))) = 1, \quad t_0 \leq t < t_1. \quad (3)$$

Момент времени $t = t_1$ здесь исключен, так как в точке x_1 существование градиента $V'(x)$ функции Беллмана не предполагается. Справедливость равенства (3) в точках $\tau_j \in (t_0, t_1)$ скачков управления проверяется предельным переходом $t \rightarrow \tau_j - 0$ (с учетом соглашения $u(\tau_j - 0) = u(\tau_j)$). Соотношение (3) при $t = t_0$ влечет

$$(V'(x_0), f(x_0, u(t_0))) = 1 \quad \forall x_0 \in \mathbb{R}^n \setminus \{x_1\}. \quad (4)$$

Пусть теперь v — произвольная точка области управления U . Рассмотрим движение $y(t)$, $t_0 \leq t \leq t_0 + \Delta t$, $\Delta t > 0$, фазовой точки из начального состояния $x_0 \neq x_1$ под воздействием постоянного управления, равного v . При малых Δt точка $y(t_0 + \Delta t) \neq x_1$. Из точки $y(t_0 + \Delta t)$ дальнейшее движение фазовой точки выполняем оптимальным образом до попадания в конечную точку x_1 . Имеет место неравенство

$$T(x_0) \leq \Delta t + T(y(t_0 + \Delta t)),$$

из которого следует

$$\frac{V(y(t_0 + \Delta t)) - V(x_0)}{\Delta t} \leq 1.$$

Но $y(t_0 + \Delta t) = y(t_0) + \dot{y}(t_0)\Delta t + o(\Delta t) = x_0 + f(x_0, v)\Delta t + o(\Delta t)$, поэтому предельный переход при $\Delta t \rightarrow +0$ в последнем соотношении приводит к неравенству

$$(V'(x_0), f(x_0, v)) \leq 1 \quad \forall v \in U, \forall x_0 \neq x_1. \quad (5)$$

Сравнение соотношений (4) и (5) дает:

$$\max_{v \in U} (V'(x_0), f(x_0, v)) = 1 \quad \forall x_0 \neq x_1. \quad (6)$$

Итак, для функции $V(x)$ получено дифференциальное уравнение Беллмана

$$\begin{cases} \max_{u \in U} (V'(x), f(x, u)) = 1, & x \neq x_1, \\ V(x_1) = 0, \end{cases} \quad (7)$$

к которому мы присоединили граничное условие; напомним, что $V(x) < 0$ при $x \neq x_1$. Таким образом, доказана

Теорема 1.1. *В Предположениях 1,2 функция Беллмана $V(x)$ удовлетворяет условиям (7). Для оптимальной пары $(x(t), u(t))$, $t_0 \leq t \leq t_1$, имеет место равенство (3).*

Уравнение Беллмана (7) можно переписать в форме уравнения

$$\max_{u \in U} \sum_{i=1}^n \frac{\partial V(x)}{\partial x^i} f^i(x, u) = 1$$

с частными производными первого порядка неизвестной функции $V(x)$, причем в его левой части содержится операция взятия максимума. Оказывается, что уравнение Беллмана позволяет также описать и достаточные условия оптимальности в задаче быстрогодействия (P).

1.4 Уравнение Беллмана. Достаточные условия оптимальности

Теорема 1.2. *Пусть функция $V(x)$*

1) определена и непрерывна при всех $x \in \mathbb{R}^n$, $V(x_1) = 0, V(x) < 0, x \neq x_1$,

2) имеет непрерывный градиент $V'(x)$ при $x \in \mathbb{R}^n \setminus \{x_1\}$.

Тогда, если допустимый процесс

$$(x(t), u(t)), t_0 \leq t \leq t_1, \quad x(t_0) = x_0, x(t_1) = x_1, \quad x(t) \neq x_1, t_0 \leq t < t_1,$$

удовлетворяет условию

$$(V'(x(t)), f(x(t), u(t))) = 1, \quad t_0 < t < t_1, \quad (9)$$

то этот процесс оптимален по быстродействию в задаче (P), а время быстродействия $T(x_0) \equiv t_1 - t_0 = -V(x_0)$.

Доказательство. Покажем сначала, что для времени перехода при условии (9) имеет место равенство

$$t_1 - t_0 = -V(x_0).$$

Находим:

$$\begin{aligned} -V(x_0) &= 0 - V(x_0) = V(x_1) - V(x_0) = V(x(t))|_{t=t_0}^{t=t_1} = \int_{t_0}^{t_1} \frac{d}{dt} V(x(t)) dt = \\ &= \int_{t_0}^{t_1} (V'(x(t)), \dot{x}(t)) dt = \int_{t_0}^{t_1} (V'(x(t)), f(x(t), u(t))) dt = \{(9)\} \\ &= \int_{t_0}^{t_1} 1 dt = t_1 - t_0. \end{aligned}$$

Любой другой допустимый процесс

$$\hat{x}(t), \hat{u}(t), t_0 \leq t \leq \hat{t}_1, \quad \hat{x}(t_0) = x_0, \hat{x}(\hat{t}_1) = x_1, \hat{x}(t) \neq x_1, t_0 \leq t < \hat{t}_1,$$

имеет время перехода

$$\hat{t}_1 - t_0 \geq t_1 - t_0.$$

Действительно,

$$\begin{aligned} t_1 - t_0 &= -V(x_0) = V(x_1) - V(x_0) = V(\hat{x}(t))|_{t=t_0}^{t=\hat{t}_1} = \int_{t_0}^{\hat{t}_1} \frac{d}{dt} V(\hat{x}(t)) dt = \\ &= \int_{t_0}^{\hat{t}_1} (V'(\hat{x}(t)), \dot{\hat{x}}(t)) dt = \int_{t_0}^{\hat{t}_1} (V'(\hat{x}(t)), f(\hat{x}(t), \hat{u}(t))) dt \leq \\ &\leq \int_{t_0}^{\hat{t}_1} \max_{v \in U} (V'(\hat{x}(t)), f(\hat{x}(t), v)) dt = \\ &= \{p\} = \int_{t_0}^{\hat{t}_1} 1 dt = \hat{t}_1 - t_0. \end{aligned}$$

Теорема 1.2 доказана.

1.5 Схема решения задачи быстродействия методом динамического программирования

Приведем схему решения задачи быстродействия (P) методом динамического программирования, сопроводив ее некоторыми комментариями.

1.5.1 а) *Ищется гладкое решение $V(x)$ уравнения Беллмана*

$$\max_{u \in U} (V'(x), f(x, u)) = 1, \quad x \neq x_1,$$

с граничным условием $V(x_1) = 0$, причем, естественно, $V(x) < 0, x \neq x_1$.

Априори существование такой функции $V(x)$ неизвестно. Даже при известном факте существования такого решения вопрос о конструктивном его нахождении остается открытым. Предположим, что это решение $V(x)$ с непрерывным градиентом $V'(x)$, $x \neq x_1$, найдено. Тогда оптимальное время перехода из начального состояния x в конечное состояние x_1 определяется равенством $T(x) = -V(x)$.

1.5.2 б) *Ищется максимизирующая точка $u(x)$ из условия максимума:*

$$\max_{u \in U} (V'(x), f(x, u)) = (V'(x), f(x, u(x))), \quad x \neq x_1.$$

Предположим, что максимизирующая точка (максимизатор)

$$u(x) = \operatorname{argmax}_{u \in U} (V'(x), f(x, u)), \quad x \neq x_1,$$

является единственной при каждом x . Функция $u(x)$, $x \neq x_1$, является оптимальным по быстродействию управлением в форме синтеза (обратной связи). Обычно при решении задач пункты а) и б) этой схемы не разделяются и выполняются параллельно.

1.5.3 в) *Составляется замкнутая система*

$$\dot{x} = f(x, u(x)) \equiv F(x), \quad x \neq x_1.$$

При любом начальном состоянии $x_0 \neq x_1$ решение $x(t, x_0)$, $x(t_0, x_0) = x_0$, замкнутой системы обладает свойством

$$x(t, x_0) \rightarrow x_1 \quad \text{при} \quad t \rightarrow t_1 - 0, \quad \text{где} \quad t_1 = t_0 + T(x_0) = t_0 - V(x_0),$$

и является оптимальной по быстродействию траекторией, соединяющей точку x_0 с точкой x_1 . Оптимальное управление в программной форме для задачи (P) имеет вид

$$u(t) = u(x(t, x_0)),$$

где $u(x)$ — оптимальное управление в форме синтеза (обратная связь, регулятор).

1.5.4 Основные достоинства и недостатки метода динамического программирования

Достоинства метода динамического программирования:

- уравнение Беллмана дает необходимые и достаточные условия оптимальности (в отличие от принципа максимума, который является только необходимым условием оптимальности);
- оптимальное управление $u(x)$ имеет форму синтеза (обратной связи).

Недостатки метода динамического программирования:

- жесткие предположения о гладкости функции Беллмана при обосновании подхода (имеются простые примеры, в которых эти предположения не выполняются, при этом метод динамического программирования нельзя считать обоснованным, тогда как принцип максимума в этих примерах успешно работает и полностью обоснован);
- аналитические трудности при построении решения $V(x)$ уравнения Беллмана (метод динамического программирования требует решения неклассического дифференциального уравнения с частными производными, тогда как в принципе максимума привлекаются только обыкновенные дифференциальные уравнения). Общие рецепты для построения функции $V(x)$ не формулируются.

Некоторые примеры применения изложенной схемы рассмотрены ниже.

1.6 Примеры

1.6.1 Пример 1.3. Задача быстрогодействия в начало координат для простого движения с областью управления, имеющей форму шара $U = S_1(0)$:

$$\dot{x} = u, \quad u \in U; \quad x(0) = x_0, \quad x(T) = 0; \quad T \rightarrow \min.$$

Очевидно, что оптимальное управление в форме синтеза имеет вид

$$u(x) = -\frac{x}{\|x\|}, \quad x \neq 0,$$

а оптимальное время движения из начального состояния x в начало координат определяется равенством

$$T(x) = \|x\|.$$

Функция Беллмана в этой задаче имеет вид

$$V(x) = -\|x\|.$$

Проиллюстрируем порядок применения метода динамического программирования в этом примере. Уравнение Беллмана

$$\max_{u \in U} (V'(x), u) = 1 \quad (8)$$

можно записать в виде

$$\|V'(x)\| = 1.$$

Оно имеет очевидные решения

$$V(x) = \pm \|x\| + C.$$

Произвольная постоянная C определяется с помощью граничного условия $V(0) = 0$: получаем $C = 0$; условие $V(x) < 0, x \neq 0$, дает:

$$V(x) = -\|x\|.$$

Эта функция имеет градиент

$$V'(x) = -\frac{x}{\|x\|}, \quad x \neq 0; \quad \|V'(x)\| = 1.$$

Максимизирующая точка $u(x)$ в (8)

$$u(x) = \operatorname{argmax}_{u \in U} (V'(x), u) = \frac{V'(x)}{\|V'(x)\|} = -\frac{x}{\|x\|}, \quad x \neq 0,$$

определяет оптимальный синтез для рассматриваемого примера. Замкнутая система принимает вид

$$\dot{x} = -\frac{x}{\|x\|}, \quad x \neq 0.$$

Правая часть этой системы определена при всех $x \neq 0$; точка $x = 0$ является особой точкой этого уравнения. Задача Коши для этого уравнения при начальном условии $x|_{t=0} = x_0$ имеет решение

$$x(t, x_0) = \frac{x_0}{\|x_0\|} (\|x_0\| - t), \quad 0 \leq t < T(x_0) \equiv \|x_0\|, \quad (9)$$

$$x(t, x_0) \rightarrow 0 \quad \text{при} \quad t \rightarrow \|x_0\| - 0. \quad (10)$$

Решение (9) определено и при $t = \|x_0\|$, когда оно обращается в нуль, однако его подстановка в замкнутую систему в этот момент времени невозможна. Поэтому мы вполне удовлетворимся предельным соотношением (10). Оптимальная программа постоянна:

$$u(t) = -\frac{x_0}{\|x_0\|}, \quad 0 \leq t \leq \|x_0\|.$$

Упражнение 1.1. Уравнение Беллмана в полярных координатах. В двумерном случае уравнение Беллмана в рассмотренном примере 1.3 можно записать в виде

$$V_{x_1}'^2(x) + V_{x_2}'^2(x) = 1, \quad x = (x_1, x_2) \neq 0,$$

при дополнительных условиях $V(0) = 0, \quad V(x) < 0, x \neq 0$. Решение последней задачи было предъявлено в форме $V(x) = -\|x\|$; при $n = 2$ в подробной записи имеем:

$$V(x) = -\sqrt{x_1^2 + x_2^2}.$$

Рассмотрим прямой метод получения этого решения, переходя в уравнении к полярным координатам r, ϕ : $x_1 = r \cos \phi, \quad x_2 = r \sin \phi$.

Имеем:

$$\begin{aligned} V_r' &= V_{x_1}' \cdot \frac{\partial x_1}{\partial r} + V_{x_2}' \cdot \frac{\partial x_2}{\partial r} = V_{x_1}' \cos \phi + V_{x_2}' \sin \phi, \\ V_\phi' &= V_{x_1}' \cdot \frac{\partial x_1}{\partial \phi} + V_{x_2}' \cdot \frac{\partial x_2}{\partial \phi} = r(-V_{x_1}' \sin \phi + V_{x_2}' \cos \phi), \end{aligned}$$

или

$$\begin{aligned} V'_{x_1} \cos \phi + V'_{x_2} \sin \phi &= V'_r, \\ - V'_{x_1} \sin \phi + V'_{x_2} \cos \phi &= \frac{1}{r} V'_\phi. \end{aligned}$$

Почленное возведение в квадрат и сложение двух последних равенств дает: $V'^2_{x_1} + V'^2_{x_2} = V'^2_r + \frac{1}{r^2} V'^2_\phi$. Поэтому уравнение Беллмана в полярных координатах принимает вид

$$V'^2_r + \frac{1}{r^2} V'^2_\phi = 1.$$

Ищем решение последнего уравнения в виде функции одной (радиальной) переменной $r : V = R(r)$; при этом уравнение принимает вид $R'^2(r) = 1$, или $R'(r) = \pm 1$, откуда

$$R(r) = \pm r + C.$$

Условие $V|_{x=0} = 0 \iff R(r)|_{r=0} = 0$ дает $C = 0$, а неравенство $V(x) < 0, x \neq 0$, позволяет выбрать правильный знак перед r и получить требуемый результат $V = -r$, или, в декартовых координатах, $V(x) = -\sqrt{x_1^2 + x_2^2}$.

Упражнение 1.2. Рассмотреть задачу предыдущего упражнения в трехмерном случае и в случае размерности вектора x , равной $n > 3$.

1.6.2 Пример 1.4. Задача быстродействия для линейной системы с кососимметрической матрицей A и областью управления U , имеющей форму шара $S_R(0)$ радиуса R :

$$\begin{cases} \dot{x} = Ax + u, & u \in U \equiv S_R(0), R > 0; \quad A^* = -A; \\ x(0) = x_0, \quad x(T) = 0; & T \rightarrow \min. \end{cases} \quad (11)$$

Условие кососимметричности $A^* = -A$ матрицы A влечет тождество

$$(Ax, x) = 0 \quad \forall x. \quad (12)$$

Действительно, $(Ax, x) = (x, A^*x) = (x, -Ax) = -(Ax, x)$, откуда следует равенство (12). Уравнение Беллмана для задачи (11)

$$\max_{u \in U} (V'(x), Ax + u) = 1 \quad (13)$$

можно переписать в виде

$$(V'(x), Ax) + R\|V'(x)\| = 1. \quad (14)$$

Легко проверяется, что функция

$$V(x) = -\frac{\|x\|}{R},$$

с градиентом

$$V'(x) = -\frac{1}{R} \frac{x}{\|x\|}, \quad \|V'(x)\| = \frac{1}{R}, \quad x \neq 0,$$

удовлетворяет уравнению Беллмана (14) и неравенству $V(x) < 0$ при $x \neq 0$, а также граничному условию $V(0) = 0$. Оптимальное управление $u(x)$ находим как максимизирующую точку в (13):

$$u(x) = \arg \max_{\|u\| \leq R} (V'(x), u) = R \frac{V'(x)}{\|V'(x)\|} = -R \frac{x}{\|x\|}.$$

Замкнутая система имеет вид:

$$\dot{x} = Ax - R \frac{x}{\|x\|}, \quad x \neq 0.$$

При $A = 0, R = 1$ пример 1.4 превращается в пример 1.3.

1.6.3 Пример 1.5. Задача быстродействия для систем с инвариантной нормой:

$$\begin{cases} \dot{x} = f(x) + u, & x \in \mathbb{R}^n; u \in U = S_R(0) \equiv \{\|u\| \leq R\}, \\ x(0) = x_0, \quad x(T) = 0, & T \rightarrow \min. \end{cases} \quad (15)$$

Предполагается, что гладкая нелинейная функция $f(x)$ удовлетворяет условию

$$(f(x), x) = 0 \quad \forall x \in \mathbb{R}^n. \quad (16)$$

Неуправляемая система (15) (при $u=0$)

$$\dot{x} = f(x)$$

при условии (16) обладает первым интегралом $\|x\| = \text{const}$, т.е. любое ее решение сохраняет норму. Обратим внимание на то, что форма области управления $U = \{\|u\| \leq R\}$ согласуется с первым интегралом. Имеются интересные прикладные задачи, которые приводят к оптимизационным задачам вида (15).

Уравнение Беллмана для задачи (15)

$$\max_{u \in U} (V'(x), f(x) + u) = 1$$

переписывается в форме

$$(V'(x), f(x)) + R\|V'(x)\| = 1. \quad (17)$$

Функция

$$V(x) = -\frac{\|x\|}{R},$$

непрерывная при всех x , с непрерывным градиентом

$$V'(x) = -\frac{x}{R\|x\|}, \quad x \neq 0,$$

удовлетворяет уравнению Беллмана (17), граничному условию $V(0) = 0$ и неравенству $V(x) < 0$ при $x \neq 0$. Действительно, имеем:

$$(V'(x), f(x)) = -\frac{1}{R} \left(\frac{x}{\|x\|}, f(x) \right) = -\frac{(x, f(x))}{R\|x\|} = 0 \quad \text{в силу (16),}$$

$$\|V'(x)\| = \frac{1}{R}, \quad x \neq 0.$$

Оптимальное управление (синтез) имеет вид

$$u(x) = R \frac{V'(x)}{\|V'(x)\|} = -R \frac{x}{\|x\|}.$$

Замкнутая система

$$\dot{x} = f(x) - R \frac{x}{\|x\|}, \quad x \neq 0, \quad (18)$$

обладает свойством: при любом начальном состоянии $x_0 \neq 0$ решение $x(t, x_0)$, $x(t_0, x_0) = x_0$, замкнутой системы удовлетворяет условию

$$x(t, x_0) \rightarrow 0 \quad \text{при} \quad t \rightarrow T(x_0) - 0,$$

где

$$T(x_0) = -V(x_0).$$

Функция оптимального времени

$$T(x) = -V(x).$$

Ее производная в силу системы (15)

$$\dot{T}(x)_{(15)} = (T'(x), f(x) + u)$$

удовлетворяет условию

$$-\dot{T}(x)_{(15)} \leq -\dot{T}(x)_{(18)}, \quad (19)$$

так как

$$\begin{aligned} -\dot{T}(x)_{(15)} &= (-T'(x), f(x) + u) = (V'(x), f(x) + u) = (V'(x), u) \leq \\ &\leq \max_{\|u\| \leq R} (V'(x), u) = \max_{u \in U} (V'(x), f(x) + u) = 1 = -\dot{T}(x)_{(18)}. \end{aligned}$$

Неравенство (19) означает, что взятая со знаком минус производная функции оптимального времени $T(x)$ в силу системы управляемого движения (15) принимает максимальное значение на оптимальном управлении $u = u(x) \in U$.

Пример 1.5 при $f(x) = Ax$, $A^* = -A$, превращается в пример 1.4.

1.6.4 Пример 1.6. Задача быстродействия для систем с интегральным инвариантом:

$$\begin{cases} \dot{x} = f(x) + u, & x \in \mathbb{R}^n; \quad u \in U, \\ x(0) = x_0, \quad x(T) = 0, & T \rightarrow \min. \end{cases} \quad (20)$$

Предполагается, что область управления U есть гладкий выпуклый компакт, $0 \in \text{int } U$, определяемый своей опорной функцией

$$c(\psi) = \max_{u \in U} (u, \psi), \quad \psi \in \mathbb{R}^n.$$

Выпуклый компакт U может быть также описан своей дистанционной функцией $d(x)$:

$$U = \{x \in \mathbb{R}^n : d(x) \leq 1\}.$$

Функция $d(x)$ обладает свойствами опорной функции, причем $d(0) = 0$, $d(x) > 0$ при $x \neq 0$. Дистанционная функция является опорной функцией поляры

$$\tilde{U} \equiv \{\psi \in \mathbb{R}^n : c(\psi) \leq 1\}$$

выпуклого компакта U :

$$d(x) = c(\tilde{U}, x) \equiv \max_{\psi \in U} (\psi, x), \quad x \in \mathbb{R}^n.$$

Будем предполагать существование градиентов $c'(\psi)$, $\psi \neq 0$, и $d'(x)$, $x \neq 0$, опорной функции $c(\psi)$ и дистанционной функции $d(x)$. Имеют место легко проверяемые тождества

$$c(d'(x)) = 1 \quad \forall x \neq 0, \quad (21)$$

$$d(c'(\psi)) = 1 \quad \forall \psi \neq 0. \quad (22)$$

Упражнение 1.1. Доказать соотношения (21), (22).

Будем изучать задачу быстродействия (20), предполагая выполненным условие

$$(d'(-x), f(x)) = 0 \quad \forall x \neq 0. \quad (23)$$

При условии (23) неуправляемая система (20) (при $u = 0$)

$$\dot{x} = f(x) \quad (24)$$

имеет первый интеграл

$$d(-x) = \text{const}. \quad (25)$$

Рассматриваемая управляемая система называется системой с *интегральным инвариантом*. Любое решение системы (24) с ненулевым начальным условием остается на поверхности, определяемой уравнением (25) с положительной константой. Эта поверхность является границей выпуклого компакта, содержащего нулевую точку в качестве внутренней.

При сделанных предположениях нелинейная задача быстродействия (20) допускает точное решение, которое может быть найдено методом динамического программирования. Действительно, уравнение Беллмана

$$\max_{u \in U} (V'(x), f(x) + u) = 1, \quad x \neq 0,$$

может быть записано в форме

$$(V'(x), f(x)) + c(V'(x)) = 1. \quad (26)$$

Решением уравнения (26) служит функция

$$V(x) = -d(-x); \quad V(0) = 0, \quad V(x) < 0 \quad \text{при} \quad x \neq 0.$$

Действительно, ее градиент $V'(x) = d'(-x)$, и, в силу предположения (23), первый член в уравнении (26) обращается в нуль:

$$(V'(x), f(x)) = (d'(-x), f(x)) = 0,$$

а второй член этого уравнения, как показывает тождество (21), равен единице:

$$c(V'(x)) = c(d'(-x)) = 1 \quad \forall x \neq 0.$$

Оптимальное управление в форме синтеза находим как максимизирующую точку в уравнении Беллмана:

$$u(x) = c'(V'(x)) = c'(d'(-x)). \quad (27)$$

Выражение (27) может быть преобразовано, если воспользоваться тождеством

$$c'(d'(\xi)) = \frac{\xi}{d(\xi)} \quad \forall \xi \neq 0. \quad (28)$$

Заметим, что имеет место также двойственное тождество

$$d'(c'(\chi)) = \frac{\chi}{c(\chi)} \quad \forall \chi \neq 0. \quad (29)$$

Упражнение 1.2. Проверить тождества (28), (29).

Из (27) в силу (28) получаем следующее окончательное представление для оптимального синтеза в задаче (20)

$$u(x) = \frac{-x}{d(-x)}, \quad x \neq 0, \quad (30)$$

с привлечением которого строится замкнутая система

$$\dot{x} = f(x) + \frac{-x}{d(-x)}.$$

Полученное выражение (30) для оптимального управления показывает, что оптимальный управляющий сигнал — вектор $u(x)$ — всегда направлен к началу координат от текущей фазовой точки x , причем его интенсивность (норма) максимально возможная с учетом формы области управления U . Функция оптимального времени $T(x)$ перехода объекта из начального состояния x в начало координат определяется равенством

$$T(x) = d(-x).$$

Таким образом, для нелинейной задачи быстродействия (20) с интегральным инвариантом, в которой форма области управления определенным образом согласована с формой инвариантных поверхностей, получено точное решение. Пример 1.6 включает в себя предыдущие примеры 1.2-1.5.

1.6.5 Пример 1.7. Задача быстродействия для тележки

Рассмотрим задачу быстродействия

$$\begin{cases} \dot{x}_1 = x_2, & \dot{x}_2 = u, & |u| \leq 1, \\ x_1(0) = a, & x_2(0) = b, & x_1(T) = x_2(T) = 0, \quad T \rightarrow \min. \end{cases}$$

В этом примере номера фазовых координат обозначаются нижними индексами. Решение этой задачи хорошо известно. На фазовой плоскости $x_1 x_2$ строится линия переключения AOB , определяемая уравнением

$$AOB: \quad x_1 + \frac{1}{2} \text{sign}(x_2) x_2^2 = 0.$$

Функция оптимального времени $T(x_1, x_2)$ перевода начального состояния (x_1, x_2) в начало координат определяется равенством

$$T(x_1, x_2) = \begin{cases} x_2 + 2\sqrt{x_1 + \frac{1}{2}x_2^2}, & \text{если точка } (x_1, x_2) \text{ выше линии } AOB, \\ -x_2 + 2\sqrt{-x_1 + \frac{1}{2}x_2^2}, & \text{если точка } (x_1, x_2) \text{ ниже линии } AOB. \end{cases}$$

Для точек линии переключения AOB обе формулы дают одинаковый результат. Функция Беллмана $V(x_1, x_2) = -T(x_1, x_2)$ отличается от функции оптимального времени только знаком.

Легко проверяется, что функция $T(x_1, x_2)$ непрерывна на всей фазовой плоскости, но в каждой точке линии переключения не является дифференцируемой.

Упражнение 1.3. Доказать недифференцируемость функции Беллмана $V(x_1, x_2) = -T(x_1, x_2)$ на линии переключения AOB .

Таким образом, в данном примере, который ранее был полностью изучен на основе *принципа максимума*, **Предположение 1.2** схемы *динамического программирования* не выполняется во всех точках линии переключения. Для любого ненулевого начального состояния заключительный участок оптимальной траектории проходит по линии переключения.

Упражнение 1.4. Проверить, что в открытой области

$$I_- = \{(x_1, x_2) : x_1 + \frac{1}{2}\text{sign}(x_2)x_2^2 > 0\},$$

расположенной выше линии переключения, функция Беллмана $V(x_1, x_2)$ удовлетворяет уравнению Беллмана

$$\max_{|u| \leq 1} (V'_{x_1}(x_1, x_2) \cdot x_2 + V'_{x_2}(x_1, x_2) \cdot u) = 1.$$

Решение. В области I_- имеем:

$$x_1 + \frac{1}{2}x_2^2 > 0,$$

$$V'_{x_1}(x_1, x_2) = -T'_{x_1}(x_1, x_2) = -\frac{\partial}{\partial x_1} \left(x_2 + 2\sqrt{x_1 + \frac{1}{2}x_2^2} \right) = -\frac{1}{\sqrt{x_1 + \frac{1}{2}x_2^2}},$$

$$V'_{x_2}(x_1, x_2) = -T'_{x_2}(x_1, x_2) = -\frac{\partial}{\partial x_2} \left(x_2 + 2\sqrt{x_1 + \frac{1}{2}x_2^2} \right) = -1 - \frac{x_2}{\sqrt{x_1 + \frac{1}{2}x_2^2}},$$

$$V'_{x_2}(x_1, x_2) < 0.$$

Справедливость последнего неравенства при $x_2 \geq 0$ очевидна, если же $x_2 < 0$, то $\text{sign}(x_2) = -1$, и условие $(x_1, x_2) \in I_-$ принимает вид $x_1 - \frac{1}{2}x_2^2 > 0$, поэтому

$$\begin{aligned} V'_{x_2}(x_1, x_2) &= -\frac{\sqrt{x_1 + \frac{1}{2}x_2^2} + x_2}{\sqrt{x_1 + \frac{1}{2}x_2^2}} = \\ &= \frac{-x_1 + \frac{1}{2}x_2^2}{\sqrt{x_1 + \frac{1}{2}x_2^2}[\sqrt{x_1 + \frac{1}{2}x_2^2} - x_2]} < 0. \end{aligned}$$

И, наконец, привлекая доказанное неравенство $V'_{x_2}(x_1, x_2) < 0$, $(x_1, x_2) \in I_-$, получаем:

$$\begin{aligned} &\max_{|u| \leq 1} (V'_{x_1}(x_1, x_2) \cdot x_2 + V'_{x_2}(x_1, x_2) \cdot u) = \\ &= V'_{x_1}(x_1, x_2) \cdot x_2 + \max_{|u| \leq 1} V'_{x_2}(x_1, x_2) \cdot u = \{V'_{x_2} < 0\} = \\ &= V'_{x_1}(x_1, x_2) \cdot x_2 + V'_{x_2}(x_1, x_2) \cdot (-1) = \end{aligned}$$

$$= -\frac{x_2}{\sqrt{x_1 + \frac{1}{2}x_2^2}} + \left(-1 - \frac{x_2}{\sqrt{x_1 + \frac{1}{2}x_2^2}}\right) \cdot (-1) \equiv 1.$$

Итак, доказано, что в области I_- функция $V(x_1, x_2)$ удовлетворяет уравнению Беллмана.

Упражнение 1.5. Проверить, что в открытой области

$$I_+ = \{(x_1, x_2) : x_1 + \frac{1}{2}\text{sign}(x_2)x_2^2 < 0\},$$

расположенной ниже линии переключения, функция Беллмана $V(x_1, x_2)$ удовлетворяет уравнению Беллмана

$$\max_{|u| \leq 1} (V'_{x_1}(x_1, x_2) \cdot x_2 + V'_{x_2}(x_1, x_2) \cdot u) = 1.$$

Оптимальный по быстродействию синтез в этом примере определяется формулой

$$u(x_1, x_2) = \begin{cases} -1, & (x_1, x_2) \in I_- \cup AO, \\ +1, & (x_1, x_2) \in I_+ \cup BO, \end{cases}$$

где

AO — полупарабола $x_1 = \frac{1}{2}x_2^2, x_2 < 0$,

BO — полупарабола $x_1 = -\frac{1}{2}x_2^2, x_2 > 0$.

1.6.6 Пример 1.8.

1.7 Связь метода динамического программирования с принципом максимума

Покажем, как из подхода динамического программирования может быть выведен принцип максимума для задачи быстродействия (P). Эти рассуждения проводятся при выполнении следующего дополнительного предположения о гладкости функции Беллмана.

Предположение 1.3. При всех $x \neq x_1$ определена и непрерывна матрица вторых производных (гессиан)

$$V''(x) = \left(\frac{\partial^2 V(x)}{\partial x^i \partial x^j} \right)_{i,j=1}^n$$

функции Беллмана $V(x) = -T(x)$, где $T(x)$ — оптимальное время перехода из начального состояния x в конечное состояние x_1 .

Введем в рассмотрение функцию

$$B(x, u) = (V'(x), f(x, u)) = \sum_{i=1}^n \frac{\partial V(x)}{\partial x^i} f^i(x, u), \quad x \neq x_1. \quad (31)$$

Привлекая обозначение (31), запишем уравнение Беллмана в виде

$$\max_{u \in U} B(x, u) = 1. \quad (32)$$

Градиент функции (31) по аргументу x имеет вид

$$B'_x(x, u) = V''(x)f(x, u) + f'^*(x, u)V'(x). \quad (33)$$

Действительно, имеем:

$$\frac{\partial B(x, u)}{\partial x^j} = \sum_{i=1}^n \frac{\partial^2 V(x)}{\partial x^i \partial x^j} f^i(x, u) + \sum_{i=1}^n \frac{\partial f^i(x, u)}{\partial x^j} \frac{\partial V(x)}{\partial x^i}, \quad j = 1, \dots, n. \quad (34)$$

Формулы (34) для частных производных влекут формулу (33), дающую компактное векторное представление градиента функции B по аргументу x .

Пусть

$$(x(t), u(t)), \quad t_0 \leq t \leq t_1, \quad x(t) \neq x_1, \quad t < t_1, \quad (35)$$

— оптимальная пара в рассматриваемой задаче быстродействия (P) . Тогда можно записать соотношения

$$B(x(t), u(t)) \equiv 1, \quad t_0 \leq t < t_1, \quad (36)$$

$$B(x, u(t)) \leq 1, \quad \forall x \neq x_1, \quad (37)$$

которые показывают, что гладкая по x функция $B(x, u(t))$ достигает максимального значения по первому аргументу при $x = x(t)$. Следовательно, имеет место равенство

$$B'_x(x, u(t))|_{x=x(t)} = 0 \quad (38)$$

(необходимое условие максимума — обращение в нуль градиента). Из (38) и (33) получаем:

$$V''(x(t))f(x(t), u(t)) = -f'_x{}^*(x(t), u(t))V'(x(t)). \quad (39)$$

Введем в рассмотрение n -мерную векторную функцию

$$\psi(t) \equiv V'(x(t)), \quad t_0 \leq t < t_1. \quad (40)$$

Дифференцирование соотношения (40) по времени t позволяет, привлекая дифференциальное уравнение управляемого движения и соотношения (39), (40), получить

$$\begin{aligned} \dot{\psi}(t) &= \frac{d}{dt} V'(x(t)) = V''(x(t))\dot{x}(t) = \{(P)\} = \\ &= V''(x(t))f(x(t), u(t)) = \{(39)\} = \\ &= -f'_x{}^*(x(t), u(t))V'(x(t)) = \{(40)\} = \\ &= -f'_x{}^*(x(t), u(t))\psi(t). \end{aligned} \quad (41)$$

Крайние части соотношений (41) показывают, что векторная переменная $\psi(t)$ является решением сопряженного уравнения

$$\dot{\psi} = -f'_x{}^*(x(t), u(t))\psi. \quad (42)$$

Уравнение Беллмана дает условие максимума

$$(\psi(t), f(x(t), u(t))) = \max_{v \in U} (\psi(t), f(x(t), v)),$$

которое может быть записано в терминах функции Гамильтона-Понтрягина

$$H(x, \psi, u) = (\psi, f(x, u))$$

в хорошо известной форме

$$H(x(t), \psi(t), u(t)) = \max_{u \in U} H(x(t), \psi(t), u). \quad (43)$$

Известная теорема, [1], о необходимых условиях оптимальности в форме принципа максимума Понтрягина для задачи быстрогодействия (P) формулируется следующим образом.

Теорема 1.3. *Если пара (35) образует оптимальное решение задачи быстрогодействия (P), то существует такая непрерывная ненулевая n -мерная векторная функция $\psi(t)$, что функции $x(t), \psi(t), u(t)$ удовлетворяют системе дифференциальных уравнений*

$$\dot{x} = H'_\psi \equiv f(x, u), \quad (44)$$

$$\dot{\psi} = -H'_x \equiv -f'_x{}^*(x, u)\psi \quad (45)$$

и условию максимума (43).

Уравнение (44) является дифференциальным уравнением управляемого движения, а уравнение (45) — сопряженным уравнением. Эта теорема получена выше, исходя из положений динамического программирования, при выполнении **Предположений 1.1, 1.2, 1.3** о функции Беллмана $V(x)$. При этом дано конструктивное описание оптимальной сопряженной переменной $\psi(t)$ в терминах градиента функции Беллмана на оптимальной траектории, см. формулу (40). Заметим, что утверждение теоремы 1.3, см. [1], а также рассмотренный выше пример 1.7, имеет место и без предположений о гладкости функции Беллмана.

Замечание 1.3. Перепишем формулу (40) в терминах функции оптимального времени:

$$\psi(t) = -T'(x(t)).$$

Таким образом, оптимальная сопряженная переменная $\psi(t)$ совпадает с антиградиентом оптимального времени на оптимальной траектории $x(t)$. В частности, при $t = t_0$ получаем очень важное соотношение для начального значения сопряженной переменной:

$$\psi(t_0) = -T'(x_0).$$

Лекция 2

2 Метод динамического программирования в задаче оптимального управления с интегральным функционалом

2.1 Постановка задачи

Рассмотрим задачу оптимального управления с интегральным функционалом

$$\begin{cases} \dot{x} = f(x, u), & x \in \mathbb{R}^n, u \in U, \quad t_0 \leq t \leq t_1, \\ x(t_0) = x_0, x(t_1) = x_1, \\ L = \int_{t_0}^{t_1} f^0(x(t), u(t)) dt \rightarrow \min. \end{cases} \quad (P)$$

Здесь $f(x, u)$ — n -мерная векторная функция с координатами $f^1(x, u), \dots, f^n(x, u)$. Начальный момент времени t_0 и конечный момент времени t_1 считаются заданными, $t_0 < t_1$. Начальная точка x_0 траектории и ее конечная точка x_1 фиксированы. Класс допустимых управлений состоит из кусочно-непрерывных функций времени со значениями в области управления U . Предполагается, что функции $f^i(x, u), i = 0, \dots, n$, и их частные производные $\frac{\partial f^i(x, u)}{\partial x^j}, i = 0, \dots, n, j = 1, \dots, n$, определены и непрерывны на прямом произведении $\mathbb{R}^n \times U$.

Ниже обсуждается метод динамического программирования применительно к задаче оптимального управления (P). Основной конструкцией метода динамического программирования в задаче (P) является уравнение Беллмана

$$V'_t(x, t) + \max_{u \in U} \{-f^0(x, u) + (V'_x(x, t), f(x, u))\} = 0.$$

Это уравнение можно записать в следующей компактной форме

$$\max_{u \in U} \mathcal{B}(x, t, u) = 0,$$

где

$$\mathcal{B}(x, t, u) = V'_t(x, t) - f^0(x, u) + (V'_x(x, t), f(x, u)).$$

К уравнению Беллмана присоединяется граничное условие

$$V(x_1, t_1) = 0.$$

Обратим внимание на то, что в этой задаче управления функция Беллмана $V(x, t)$ зависит от двух переменных: n -мерной переменной x и одномерной переменной (времени) t . Напомним, что в задаче быстрогодействия функция Беллмана зависела только от $x \in \mathbb{R}^n \setminus \{x_1\}$. При рассмотрении задачи (P) с другими краевыми условиями в конце процесса граничное условие для функции Беллмана определенным образом модифицируется.

2.2 Погружение задачи (P) в семейство аналогичных задач (P_α) . Введение функции Беллмана

Наряду с задачей (P) рассмотрим семейство задач управления

$$\begin{cases} \dot{x} = f(x, u), & x \in \mathbb{R}^n, u \in U, \quad t \in [\tau, t_1], \\ x(\tau) = y, x(t_1) = x_1, \\ L_\alpha = \int_{\tau}^{t_1} f^0(x(t), u(t)) dt \rightarrow \min. \end{cases} \quad (P_\alpha)$$

Семейство задач (P_α) зависит от параметра $\alpha = (y, \tau)$. Задача (P_α) при $\alpha = (x_0, t_0)$ превращается в задачу (P) .

Пусть

$$T(y, \tau) = \min \int_{\tau}^{t_1} f^0(x(t), u(t)) dt = \min L_\alpha$$

— минимальное значение функционала в задаче (P_α) . Тогда

$$V(y, \tau) = -T(y, \tau)$$

— функция Беллмана.

2.3 Проверка принципа оптимальности ("хвост" оптимального процесса оптимален)

Пусть

$$(x(t), u(t)), \quad t_0 \leq t \leq t_1, \quad (1)$$

— оптимальный процесс в задаче (P) . Здесь $u(t)$ — допустимое управление, $x(t)$ — отвечающая этому управлению траектория, $x(t_0) = x_0$, $x(t_1) = x_1$. Возьмем любую точку $\tau \in [t_0, t_1]$ и рассмотрим "хвост" процесса (1)

$$(x(t), u(t)), \quad \tau \leq t \leq t_1, \quad (2)$$

с начальным условием $x|_{t=\tau} = x(\tau) \equiv y$.

Проверим оптимальность пары (2) в задаче (P_α) , где $\alpha = (y = x(\tau), \tau)$. Проверка выполняется методом от противного. Допустим, что это не так. Тогда существует допустимое управление

$$\hat{u}(t), \tau \leq t \leq t_1,$$

с лучшим значением функционала; отвечающую этому управлению траекторию обозначим

$$\hat{x}(t), \tau \leq t \leq t_1; \quad \hat{x}(t)|_{t=\tau} = x(\tau) = y, \hat{x}(t)|_{t=t_1} = x_1,$$

при этом

$$\int_{\tau}^{t_1} f^0(\hat{x}, \hat{u}) dt < \int_{\tau}^{t_1} f^0(x, u) dt.$$

Построим новое управление

$$u_*(t) = \begin{cases} u(t), & t \in (t_0, \tau), \\ \hat{u}(t), & t \in [\tau, t_1], \end{cases}$$

и отвечающую ему траекторию

$$x_*(t) = \begin{cases} x(t), & t \in (t_0, \tau), \\ \hat{x}(t), & t \in [\tau, t_1], \end{cases} \quad x_*(t_0) = x_0, x_*(t_1) = x_1.$$

Для функционала L на паре $(x_*(t), u_*(t))$ имеем:

$$\begin{aligned} \int_{t_0}^{t_1} f^0(x_*, u_*) dt &= \left(\int_{t_0}^{\tau} + \int_{\tau}^{t_1} \right) f^0(x_*, u_*) dt = \\ &= \int_{t_0}^{\tau} f^0(x, u) dt + \int_{\tau}^{t_1} f^0(\hat{x}, \hat{u}) dt < \\ &< \int_{t_0}^{t_1} f^0(x, u) dt. \end{aligned}$$

Последний интеграл представляет оптимальное значение функционала, и поэтому полученное неравенство противоречит оптимальности пары (1). *Принцип оптимальности* Беллмана для задачи (P) обоснован.

Обратим внимание на то, что управление $u_*(t)$, $t_0 \leq t \leq t_1$, является допустимым, так как класс допустимых управлений состоит из кусочно-непрерывных функций со значениями в области управления U .

2.4 Формулировка предположений о функции Беллмана. Вывод уравнения Беллмана. Необходимые условия оптимальности

2.4.1 Предположения 2.1, 2.2 о функции Беллмана

Сформулируем предположения о функции Беллмана $V(y, \tau)$.

Предполагаем, что функция Беллмана $V(y, \tau)$ определена и непрерывна на некотором множестве $W \subset \mathbb{R}^{n+1}$ значений (y, τ) , $y \in \mathbb{R}^n$, $\tau \in [t_0, t_1]$, причем точка

$$(x_1, t_1) \in \partial W$$

— граничная точка множества W . Для функции V выполняется граничное условие

$$V(x_1, t_1) = 0.$$

Аргументы функции Беллмана будем обозначать далее буквами x и t .

Предположение 2.1. Функция $V(x, t)$ определена и непрерывна на множестве $W \subset \mathbb{R}^n \times [t_0, t_1]$ с непустой внутренностью

$$W^0 = \text{int } W \neq \emptyset,$$

$$(x_1, t_1) \in \partial W, \quad V(x_1, t_1) = 0.$$

Предположение 2.2. Функция Беллмана имеет в непустом открытом множестве W^0 непрерывные частные производные первого порядка

$$V'_x(x, t), V'_t(x, t), \quad (x, t) \in W^0.$$

2.4.2 Необходимые условия оптимальности. Вывод уравнения Беллмана

Пусть

$$(x(t), u(t)), \quad t_0 \leq t \leq t_1,$$

— оптимальный процесс в задаче (P) и пусть выполняется включение

$$(x(t), t) \in W^0 \quad \forall t \in (t_0, t_1).$$

Обсудим *необходимые условия оптимальности*. Возьмем точку $t \in (t_0, t_1)$, и малое положительное число Δt , так что $t + \Delta t \in (t_0, t_1)$. Считаем, что t — точка непрерывности управления $u(t)$. На основании принципа оптимальности вдоль оптимального процесса имеем:

$$\begin{aligned} T(x(t), t) &= \int_t^{t+\Delta t} f^0(x(s), u(s)) ds + \int_{t+\Delta t}^{t_1} f^0(x(s), u(s)) ds = \\ &= \int_t^{t+\Delta t} f^0(x(s), u(s)) ds + T(x(t + \Delta t), t + \Delta t). \end{aligned}$$

Отсюда, привлекая функцию V , получаем равенство

$$\frac{V(x(t + \Delta t), t + \Delta t) - V(x(t), t)}{\Delta t} = \frac{1}{\Delta t} \int_t^{t+\Delta t} f^0(x(s), u(s)) ds.$$

Переходя в нем к пределу при $\Delta t \rightarrow +0$, получаем при сделанных предположениях

$$\frac{d}{dt} V(x(t), t) = f^0(x(t), u(t)),$$

или

$$(V'_x(x(t), t), \dot{x}(t)) + V'_t(x(t), t) = f^0(x(t), u(t)),$$

или

$$(V'_x(x(t), t), f(x(t), u(t))) + V'_t(x(t), t) = f^0(x(t), u(t)).$$

Таким образом, вдоль оптимального процесса выполняется равенство

$$(V'_x(x(t), t), f(x(t), u(t))) + V'_t(x(t), t) - f^0(x(t), u(t)) = 0 \quad \forall t \in (t_0, t_1). \quad (3)$$

Заметим, что условие (3) доказано для любой точки t непрерывности управления $u(t)$. Однако оно имеет место и в любой точке $\theta \in (t_0, t_1)$ конечного скачка управления. Действительно, принимая во внимание соглашение $u(\theta - 0) = u(\theta)$ о непрерывности слева допустимых управлений в точке разрыва и выполняя в (3) предельный переход $t \rightarrow \theta - 0$, получаем утверждение (3) в полном объеме. Итак, нами получены *необходимые условия оптимальности* в форме (3).

С помощью принципа оптимальности аналогичным образом (соответствующие выкладки приводятся ниже, см. также раздел "Метод динамического программирования в задаче быстрогодействия") получаем неравенство

$$V'_t(x, t) + (V'_x(x, t), f(x, u)) - f^0(x, u) \leq 0 \quad \forall (x, t) \in W^0, \forall u \in U. \quad (4)$$

Из (3) и (4) следует, что

$$\max_{u \in U} [V'_t(x, t) + (V'_x(x, t), f(x, u)) - f^0(x, u)] = 0, \quad (x, t) \in W^0, \quad (5)$$

или

$$V'_t(x, t) + \max_{u \in U} [-f^0(x, u) + (V'_x(x, t), f(x, u))] = 0, \quad (x, t) \in W^0. \quad (6)$$

Последнее уравнение называется *уравнением Беллмана*. К этому уравнению присоединяется граничное условие¹

$$V(x_1, t_1) = 0. \quad (7)$$

Таким образом, доказана

Теорема 2.1. В Предположениях 2.1, 2.2 функция Беллмана $V(x, t)$ удовлетворяет условиям (6), (7). Для оптимальной пары $(x(t), u(t))$, $t_0 \leq t \leq t_1$, имеет место равенство (3).

В заключение приведем обоснование неравенства (4). При любых $v \in U$, малом $\Delta t > 0$, $(x, t) \in W^0$, t — точка непрерывности управления $u(t)$, принцип оптимальности позволяет записать неравенство

$$T(x, t) \leq \int_t^{t+\Delta t} f^0(y(s), v) ds + T(y(t + \Delta t), t + \Delta t),$$

где $y(s)$, $t \leq s \leq t + \Delta t$, $y(s)|_{s=t} = x$, — траектория, отвечающая постоянному управлению v . При малых $\Delta t > 0$ будет $(y(t + \Delta t), t + \Delta t) \in W^0$. Записанное выше неравенство в терминах функции V принимает вид

$$\frac{V(y(t + \Delta t), t + \Delta t) - V(x, t)}{\Delta t} \leq \frac{1}{\Delta t} \int_t^{t+\Delta t} f^0(y(s), v) ds \quad \forall v \in U$$

и, в результате предельного перехода $\Delta t \rightarrow +0$, приводит к утверждению (4).

2.5 Уравнение Беллмана и достаточные условия оптимальности

Теорема 2.2. Пусть

- 1) $V(x, t)$ — гладкое решение уравнения Беллмана (6), удовлетворяющее условию (7),
- 2) вдоль допустимого процесса

$$(x(t), u(t)), t_0 \leq t \leq t_1, \quad x(t_0) = x_0, x(t_1) = x_1, \quad (x(t), t) \in W^0, t_0 < t < t_1, \quad (8)$$

выполняется условие (3).

Тогда процесс (8) оптимален. Оптимальное значение функционала в задаче (P) равно $-V(x_0, t_0)$.

Доказательство теоремы 2.2. Нужно доказать, что значение функционала L для процесса (8)

$$L_{(8)} = \int_{t_0}^{t_1} f^0(x(t), u(t)) dt$$

при условии (3) является минимальным. Возьмем любой другой допустимый процесс

$$(\hat{x}(t), \hat{u}(t)), t_0 \leq t \leq t_1, \quad \hat{x}(t_0) = x_0, \hat{x}(t_1) = x_1, \quad (\hat{x}(t), t) \in W^0, t_0 < t < t_1, \quad (9)$$

¹ Условие (7) понимается в следующем смысле: для числа $\Delta > 0$ и любого допустимого процесса $(y(t), v(t))$, $t_1 - \Delta \leq t \leq t_1$, $y(t_1) = x_1$; $(y(t), t) \in W^0$, $t_1 - \Delta < t < t_1$;

$$\lim_{t \rightarrow t_1 - 0} V(y(t), t) = 0 \equiv V(x_1, t_1).$$

и запишем неравенство (4), полагая в нем $x = \hat{x}(t)$, $u = \hat{u}(t)$:

$$V'_t(\hat{x}(t), t) + (V'_x(\hat{x}(t), t), f(\hat{x}(t), \hat{u}(t))) \leq f^0(\hat{x}(t), \hat{u}(t)).$$

Считая t точкой непрерывности управления $\hat{u}(t)$, последнее неравенство можно записать в форме

$$\frac{d}{dt} V(\hat{x}(t), t) \leq f^0(\hat{x}(t), \hat{u}(t)). \quad (10)$$

Проведем, наконец, заключительную выкладку:

$$\begin{aligned} L_{(8)} &= \int_{t_0}^{t_1} f^0(x(t), u(t)) dt = \{(3)\} = \\ &= \int_{t_0}^{t_1} \frac{d}{dt} [V(x(t), t)] dt = \{\text{формула Ньютона-Лейбница}\} = \\ &= V(x(t), t)|_{t=t_0}^{t=t_1} = V(x_1, t_1) - V(x_0, t_0) = \{(9)\} = V(\hat{x}(t), t)|_{t=t_0}^{t=t_1} = \\ &= \{\text{формула Ньютона-Лейбница}\} = \int_{t_0}^{t_1} \frac{d}{dt} [V(\hat{x}(t), t)] dt \leq \{(10)\} \leq \\ &\leq \int_{t_0}^{t_1} f^0(\hat{x}(t), \hat{u}(t)) dt = L_{(9)} \equiv L(\hat{x}, \hat{u}). \end{aligned}$$

Полученное неравенство доказывает оптимальность процесса (8). В силу (7) из предыдущей выкладки получаем, что оптимальное значение функционала $L_{(8)} = -V(x_0, t_0)$. Теорема 2.2 доказана.

Таким образом, метод динамического программирования применительно к задаче управления (P) рассмотрен как с точки зрения *необходимых условий оптимальности*, так и с точки зрения *достаточных условий оптимальности*.

2.6 Схема решения задачи управления (P) с интегральным функционалом методом динамического программирования

2.6.1 а) Ищется гладкое решение $V(x, t)$ уравнения Беллмана

$$V'_t(x, t) + \max_{u \in U} [-f^0(x, u) + (V'_x(x, t), f(x, u))] = 0, \quad (x, t) \in W^0,$$

с граничным условием $V(x_1, t_1) = 0$.

Априори существование такой функции $V(x, t)$ неизвестно. Область W^0 также заранее неизвестна. Даже при известном факте существования такого решения вопрос о конструктивном его нахождении остается открытым. Предположим, что это решение $V(x, t)$ с непрерывными производными $V'_t(x, t)$, $V'_x(x, t)$ найдено. Тогда оптимальное значение функционала равно $-V(x_0, t_0)$.

2.6.2 *b) Ищется максимизирующая точка $u(x, t)$ из условия максимума:*

$$-f^0(x, u) + (V'_x(x, t), f(x, u)) = \max_{u \in U}, \quad (x, t) \in W^0.$$

Предположим, что максимизирующая точка (максимизатор)

$$u(x, t) = \operatorname{argmax}_{u \in U} [-f^0(x, u) + (V'_x(x, t), f(x, u))], \quad (x, t) \in W^0,$$

является единственной при каждом (x, t) . Функция $u(x, t)$, является оптимальным управлением в форме синтеза (обратной связи), зависящим от времени. Обычно при решении задач пункты а) и б) этой схемы не разделяются и выполняются параллельно.

2.6.3 *c) Составляется замкнутая система*

$$\dot{x} = f(x, u(x, t)) \equiv F(x, t), \quad (x, t) \in W^0.$$

Решение замкнутой системы $x(t, x_0)$, при начальном условии $x(t_0, x_0) = x_0$ удовлетворяет конечному условию $x(t_1 - 0, x_0) = x_1$ и является оптимальной траекторией задачи (P). Оптимальное управление в программной форме для задачи (P) имеет вид

$$u(t) = u(x(t, x_0), t),$$

где $u(x, t)$ — оптимальное управление в форме синтеза.

2.7 Связь метода динамического программирования с принципом максимума

2.7.1 Формулировка принципа максимума для задачи управления (P)

Покажем, что необходимые условия оптимальности в форме принципа максимума для задачи (P) можно вывести из соображений динамического программирования. Это будет сделано при выполнении **Предположений 2.1, 2.2**, которые приведены выше, и еще одного дополнительного **Предположения 2.3** о гладкости функции Беллмана, которое будет сформулировано ниже.

Сейчас напомним формулировку принципа максимума для задачи (P). Вводятся: вектор

$$\tilde{\psi} = \begin{pmatrix} \psi_0 \\ \psi \end{pmatrix} \in \mathbb{R}^{n+1}, \quad \psi_0 \in \mathbb{R}^1, \psi \in \mathbb{R}^n,$$

функция Гамильтона-Понтрягина

$$K = (\tilde{\psi}, \tilde{f}) = \psi_0 f^0(x, u) + (\psi, f(x, u)),$$

записывается сопряженное уравнение

$$\dot{\tilde{\psi}} = -K'_x,$$

которое равносильно системе

$$\dot{\psi}_0 = 0, \dot{\psi} = -f'^*_x(x, u)\psi - \psi_0(f^0)'_x(x, u).$$

Если $\psi_0 = -1$, то n -мерная компонента сопряженного уравнения принимает вид

$$\dot{\psi} = -f'_x(x, u)\psi + (f^0)'_x(x, u). \quad (11)$$

Вдоль оптимального процесса с некоторой нетривиальной сопряженной переменной выполняется условие максимума

$$K = \max_{u \in U},$$

или, в более подробной записи,

$$K(\tilde{x}(t), \tilde{\psi}(t), u(t)) = \max_{v \in U} K(\tilde{x}(t), \tilde{\psi}(t), v).$$

Условие максимума при $\psi_0 = -1$ принимает вид

$$-f^0(x(t), u(t)) + (\psi(t), f(x(t), u(t))) = \max_{v \in U} [-f^0(x(t), v) + (\psi(t), f(x(t), v))]. \quad (12)$$

2.7.2 Вывод принципа максимума из динамического программирования при дополнительном предположении о гладкости функции Беллмана

Присоединим к Предположениям 2.1, 2.2 о функции Беллмана следующее

Предположение 2.3. Для функции Беллмана $V(x, t)$ на множестве W^0 определены и непрерывны частные производные второго порядка

$$V''_{xx}(x, t), V''_{xt}(x, t).$$

Пусть

$$(x(t), u(t)), t_0 \leq t \leq t_1, \quad x(t_0) = x_0, x(t_1) = x_1, \quad (x(t), t) \in W^0, t \in (t_0, t_1), \quad (13)$$

— оптимальный процесс в задаче (P). Получим для этого процесса необходимые условия оптимальности в форме принципа максимума с параметром $\psi_0 = -1$, исходя из необходимых условий оптимальности в форме динамического программирования. Введем в рассмотрение n -мерную векторную функцию $\psi(t)$, полагая

$$\psi(t) = V'_x(x(t), t). \quad (14)$$

Будет показано, что векторная переменная (14) удовлетворяет сопряженному уравнению (11) и, вместе с оптимальным процессом (13), удовлетворяет условию максимума (12). Считая t точкой непрерывности управления $u(t)$, найдем производную векторной переменной (14) по времени t :

$$\begin{aligned} \dot{\psi}(t) &= V''_{xx}(x(t), t)\dot{x}(t) + V''_{xt}(x(t), t) = \\ &= V''_{xx}(x(t), t)f(x(t), u(t)) + V''_{xt}(x(t), t). \end{aligned} \quad (15)$$

В силу Теоремы 2.1 вдоль оптимального процесса (13) выполняются соотношения (3) и (4):

$$V'_t(x(t), t) + (V'_x(x(t), t), f(x(t), u(t))) - f^0(x(t), u(t)) = 0, \quad t \in (t_0, t_1), \quad (16)$$

$$V'_t(x(t), t) + (V'_x(x(t), t), f(x(t), u)) - f^0(x(t), u) \leq 0, \quad t \in (t_0, t_1), u \in U. \quad (17)$$

Если ввести обозначение

$$\mathcal{B}(x, t, u) = V'_t(x, t) + (V'_x(x, t), f(x, u)) - f^0(x, u), \quad (18)$$

то уравнение Беллмана (6) запишется в форме

$$\max_{u \in U} \mathcal{B}(x, t, u) = 0. \quad (19)$$

Поэтому, в силу (16) - (19), имеем:

$$\mathcal{B}(x(t), t, u(t)) = 0, \quad (20)$$

$$\mathcal{B}(x, t, u(t)) \leq 0 \quad \forall (x, t) \in W^0. \quad (21)$$

Соотношения (20), (21) показывают, что функция $\mathcal{B}(x, t, u(t))$ достигает максимального значения по первому аргументу при $x = x(t)$. Поэтому градиент этой функции по первому аргументу обращается в нуль:

$$\mathcal{B}'_x(x, t, u(t))|_{x=x(t)} = 0, \quad t \in (t_0, t_1). \quad (22)$$

Вычислим градиент функции $\mathcal{B}$ по аргументу x . Имеем:

$$\mathcal{B}'_x(x, t, u) = V''_{tx}(x, t) - (f^0)'_x(x, u) + V''_{xx}(x, t)f(x, u) + f'^*(x, u)V'_x(x, t). \quad (23)$$

Условие максимума (22) в силу выражения (23) для градиента $\mathcal{B}'_x$ принимает вид

$$\begin{aligned} 0 &= V''_{tx}(x(t), t) + V''_{xx}(x(t), t)f(x(t), u(t)) + \\ &+ f'^*(x(t), u(t))V'_x(x(t), t) - (f^0)'_x(x(t), u(t)). \end{aligned} \quad (24)$$

Первые два слагаемых в правой части (24), в силу (15) и сделанных предположений о гладкости функции Беллмана, равны производной $\dot{\psi}(t)$. Поэтому, привлекая (14) и (15), соотношение (24) можно переписать в форме

$$0 = \dot{\psi}(t) + f'^*(x(t), u(t))\psi(t) - (f^0)'_x(x(t), u(t)).$$

Последнее соотношение говорит о том, что функция (14) удовлетворяет сопряженному уравнению (11):

$$\dot{\psi}(t) = -f'^*(x(t), u(t))\psi(t) + (f^0)'_x(x(t), u(t)).$$

Из соотношений (20), (21) извлекается условие максимума (12).

Таким образом, необходимые условия оптимальности в форме принципа максимума получены из необходимых условий оптимальности в форме динамического программирования. Еще раз подчеркнем, что принцип максимума обосновывается при менее жестких предположениях, чем метод динамического программирования, так что необходимые условия оптимальности в форме принципа максимума [1] в полном объеме из динамического программирования не следуют.

2.8 Пример решения задачи управления методом динамического программирования

Пример 2.1. Рассмотрим задачу управления с интегральным функционалом

$$\begin{cases} \dot{x} = Ax + bu, & x \in \mathbb{R}^n, u \in \mathbb{R}^1, \quad 0 \leq t \leq T, \\ x(0) = x_0, x(T) = 0, \\ L = \frac{1}{2} \int_0^T u(t)^2 dt \rightarrow \min_{u(\cdot)}. \end{cases} \quad (25)$$

Здесь $A \in \mathbb{R}^{n \times n}$ — известная матрица, $b, x_0 \in \mathbb{R}^n$ — известные векторы, $T > 0$ — заданная длительность процесса управления. Пара $\{(A, b)\}$ подчиняется ограничению

$$\det(b|Ab| \dots |A^{n-1}b) \neq 0 \quad (\text{условие управляемости}). \quad (26)$$

Решение задачи (25), (26) методом динамического программирования. Уравнение Беллмана имеет вид:

$$V'_t(x, t) + \max_{u \in \mathbb{R}^1} \left\{ -\frac{1}{2}u^2 + (V'_x(x, t), Ax + bu) \right\} = 0, \quad (x, t) \in W^0. \quad (27)$$

Максимизатор в уравнении (27) легко вычисляется:

$$u = b^* V'_x(x, t).$$

Поэтому уравнение (27) принимает вид

$$V'_t(x, t) + (V'_x(x, t), Ax) + \frac{1}{2}(b^* V'_x(x, t))^2 = 0. \quad (28)$$

Проверим, что решением последнего уравнения служит функция

$$V(x, t) = \frac{1}{2}x^* W^{-1}(t, T)x, \quad t < T, \quad (29)$$

— квадратичная форма относительно переменной x с матрицей $W^{-1}(t, T)$, где обращаемая матрица

$$W(t, T) = \int_T^t e^{(\sigma-T)A} b b^* e^{(\sigma-T)A^*} d\sigma \quad (30)$$

— симметричная $(n \times n)$ -матрица, отрицательно определенная при $t < T$ в силу условия управляемости (26).

Множества W^0, W имеют вид:

$$W^0 = \mathbb{R}^n \times (0, T),$$

$$(x = 0, T) \in \partial W,$$

$$W = \mathbb{R}^n \times [0, T) \cup \{(x = 0, T)\}.$$

Имеем:

$$V'_t(x, t) = \frac{1}{2}x^* \left[\frac{d}{dt} W^{-1}(t, T) \right] x, \quad (31)$$

$$V'_x(x, t) = W^{-1}(t, T)x. \quad (32)$$

Упражнение 2.1. Проверить правило дифференцирования обратной матрицы

$$\frac{d}{dt} S^{-1}(t) = -S^{-1}(t) \left[\frac{d}{dt} S(t) \right] S^{-1}(t). \quad (33)$$

Дифференцирование матричной функции (30) по верхнему пределу интегрирования t дает:

$$W'_t(t, T) = e^{(t-T)A} b b^* e^{(t-T)A^*}. \quad (34)$$

Тогда, на основании (33), (34), имеем:

$$\begin{aligned} \frac{d}{dt}W^{-1}(t, T) &= -W^{-1}(t, T)W'_t(t, T)W^{-1}(t, T) = \\ &= -W^{-1}(t, T)e^{(t-T)A}bb^*e^{(t-T)A^*}W^{-1}(t, T). \end{aligned} \quad (35)$$

Для первого члена уравнения (28) из (35), (31) получаем следующее выражение:

$$V'_t(x, t) = -\frac{1}{2}x^*W^{-1}(t, T)e^{(t-T)A}bb^*e^{(t-T)A^*}W^{-1}(t, T)x. \quad (36)$$

Второй член уравнения (28) в силу (32) может быть записан в виде:

$$(V'_x(x, t), Ax) = \frac{1}{2}x^* [W^{-1}(t, T)A + A^*W^{-1}(t, T)] x. \quad (37)$$

Третий член уравнения (28) представим в форме:

$$\frac{1}{2}(b^*V'_x(x, t))^2 = \frac{1}{2}x^*W^{-1}(t, T)bb^*W^{-1}(t, T)x. \quad (38)$$

Подстановка выражений (36) - (38) в уравнение (32) дает:

$$\begin{aligned} &\frac{1}{2}x^* \left[-W^{-1}(t, T)e^{(t-T)A}b b^*e^{(t-T)A^*}W^{-1}(t, T) + \right. \\ &\left. + W^{-1}(t, T)A + A^*W^{-1}(t, T) + W^{-1}(t, T)b b^*W^{-1}(t, T) \right] x = 0 \end{aligned}$$

или

$$\begin{aligned} &\frac{1}{2}x^*W^{-1}(t, T) \left[-e^{(t-T)A}bb^*e^{(t-T)A^*} + \right. \\ &\left. + AW(t, T) + W(t, T)A^* + bb^* \right] W^{-1}(t, T)x = 0. \end{aligned}$$

Отсюда ясно, что для доказательства того, что функция (29) является решением уравнения Беллмана (32), достаточно доказать матричное равенство

$$-e^{(t-T)A}b b^*e^{(t-T)A^*} + AW(t, T) + W(t, T)A^* + bb^* = 0. \quad (39)$$

Докажем (39). Привлекая определение (30) матрицы $W(t, T)$, последовательно находим:

$$\begin{aligned} &AW(t, T) + W(t, T)A^* = \\ &= \int_T^t \left[Ae^{(\sigma-T)A}b b^*e^{(\sigma-T)A^*} + e^{(\sigma-T)A}b b^*e^{(\sigma-T)A^*}A^* \right] d\sigma = \\ &= \int_T^t \frac{d}{d\sigma} \left[e^{(\sigma-T)A}b b^*e^{(\sigma-T)A^*} \right] d\sigma = \left. e^{(\sigma-T)A}b b^*e^{(\sigma-T)A^*} \right|_{\sigma=T}^{\sigma=t} = \\ &= e^{(t-T)A}b b^*e^{(t-T)A^*} - bb^*. \end{aligned}$$

Сравнение крайних членов этой цепочки равенств доказывает (39). Итак, доказано, что функция (29) является решением уравнения Беллмана (32).

Отметим, что для функции (29) при любом $t < T$ выполняется равенство

$$V(x, t)|_{x=0} = 0.$$

Оптимальное управление в задаче (25) определяется равенством

$$u(x, t) = b^* V'_x(x, t) \equiv b^* W^{-1}(t, T)x, \quad (x, t) \in W^0,$$

а оптимальное значение функционала

$$\begin{aligned} L_{op} &= -V(x_0, 0) = -\frac{1}{2}x_0^* W^{-1}(0, T)x_0 = \\ &= -\frac{1}{2}x_0^* \left(\int_T^0 e^{(\sigma-T)A} b b^* e^{(\sigma-T)A} d\sigma \right)^{-1} x_0 = \\ &= \{\sigma - T = -s, d\sigma = -ds, \quad \sigma = T \Rightarrow s = 0, \sigma = 0 \Rightarrow s = T\} = \\ &= \frac{1}{2}x_0^* W_0^{-1}(T)x_0, \end{aligned}$$

где

$$W_0(T) = \int_0^T e^{-sA} b b^* e^{-sA^*} ds$$

— матрица управляемости.

Алгебраическое условие управляемости (26) равносильно положительной определенности матрицы управляемости $W_0(T)$, $T > 0$.

В терминах матрицы управляемости оптимальное управление можно записать в виде

$$u(x, t) = -b^* W_0^{-1}(T - t)x.$$

Упражнение 2.2. Исследовать задачу (25) с помощью принципа максимума Понтрягина. Показать, что краевая задача принципа максимума имеет вид

$$\dot{x} = Ax + bb^*\psi, \dot{\psi} = -A^*\psi, \quad x|_{t=0} = x_0, x|_{t=T} = 0,$$

разрешающий эту краевую задачу параметр $p_0 = \psi|_{t=0}$ определяется равенством

$$p_0 = -W_0^{-1}(T)x_0,$$

оптимальная программа

$$u(t) = b^* e^{-tA^*} p_0, \quad 0 \leq t \leq T,$$

оптимальное значение функционала

$$L_{op} = \frac{1}{2}x_0^* W_0^{-1}(T)x_0.$$

2.9 Уравнение Беллмана при других краевых условиях. Пример задачи управления со свободным правым концом траектории

Рассмотрим задачу управления на заданном отрезке времени при свободном правом конце траектории

$$\begin{cases} \dot{x} = f(x, u), & x \in \mathbb{R}^n, u \in U, & t_0 \leq t \leq t_1, \\ x(t_0) = x_0, & x(t_1) \text{ — свободен}, \\ L = \int_{t_0}^{t_1} f^0(x, u) dt \rightarrow \min_{u(\cdot)}. \end{cases}$$

Уравнение Беллмана и граничное условие для функции Беллмана $V(x, t)$ имеют вид

$$\begin{cases} V_t'(x, t) + \max_{u \in U} \{-f^0(x, u) + (V_x'(x, t), f(x, u))\} = 0, \\ V(x, t_1) = 0 \quad \forall x \in \mathbb{R}^n. \end{cases}$$

Пример 2.2. Линейно квадратичная задача оптимального управления. Решение методом динамического программирования

Рассмотрим задачу оптимального управления

$$\begin{cases} \dot{x} = Ax + Bu, & x \in \mathbb{R}^n, u \in \mathbb{R}^r, & 0 \leq t \leq T, \\ x(0) = x_0, & x(T) \text{ — свободен}, \\ L = \int_0^T (x^* Q x + u^* R u) dt \rightarrow \min_u. \end{cases}$$

Здесь $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times r}$, $Q \in \mathbb{R}^{n \times n}$, $R \in \mathbb{R}^{r \times r}$ — известные, не зависящие от времени, матрицы указанных размерностей. Кроме того, матрицы Q, R симметричны, матрица Q — неотрицательно определенная, матрица R — положительно определенная. Областью управления U служит все пространство $\mathbb{R}^r$.

Записываем уравнение Беллмана и граничное условие:

$$V_t'(x, t) + \max_{u \in \mathbb{R}^r} \{-x^* Q x - u^* R u + (V_x'(x, t), Ax + Bu)\} = 0, \quad (40)$$

$$V(x, T) = 0 \quad \forall x. \quad (41)$$

Ищем решение уравнения Беллмана в виде квадратичной формы

$$V(x, t) = -x^* P(t) x, \quad (42)$$

где $P(t)$ — подлежащая определению симметричная $(n \times n)$ -матрица, которая удовлетворяет условию

$$P(T) = 0. \quad (43)$$

При условии (43) граничное условие (41) для функции Беллмана, очевидно, выполнено. Имеем:

$$V_t'(x, t) = -x^* \dot{P}(t) x, \quad (44)$$

$$V_x'(x, t) = -2P(t)x. \quad (45)$$

Градиент максимизируемой в уравнении (40) функции имеет вид

$$\{\dots\}'_u = -2Ru + B^* V_x'(x, t) = -2Ru - 2B^* P(t)x.$$

Приравнявая его нулю, находим максимизатор

$$u = -R^{-1}B^*P(t)x. \quad (46)$$

С помощью (46) вычисляем максимум в уравнении (40) и, привлекая (44), (45), записываем уравнение (40) в виде

$$\begin{aligned} & -x^*\dot{P}(t)x - x^*Qx - (x^*P(t)BR^{-1})R(R^{-1}B^*P(t)x) - \\ & -2(P(t)x, Ax) + 2(P(t)x, BR^{-1}B^*P(t)x) = 0, \end{aligned}$$

или

$$x^*[-\dot{P}(t) - Q - P(t)A - A^*P(t) + P(t)DP(t)]x = 0, \quad D \equiv BR^{-1}B^* \geq 0.$$

Последнее уравнение будет удовлетворено, если в качестве матрицы $P(t)$ взять решение дифференциального уравнения

$$\dot{P} = -Q - PA - A^*P + PDP \quad (47)$$

при начальном условии (43). Уравнение (47) называется *матричным дифференциальным уравнением Риккати*. Можно показать, что матричная задача Коши (47), (43) при сделанных предположениях имеет решение $P(t)$, определенное на всем отрезке $0 \leq t \leq T$.

Итак, функция Беллмана в виде квадратичной формы (42) построена. Матрица $P(t)$ этой формы определяется как решение задачи Коши (47), (43). Оптимальное управление имеет вид нестационарной обратной связи

$$u(x, t) = -R^{-1}B^*P(t)x,$$

линейно зависящей от фазовой переменной x . Оптимальное значение функционала в рассматриваемой линейно-квадратичной задаче управления определяется равенством

$$L_{op} = -V(x_0, 0) = x_0^*P(0)x_0.$$

Литература

1. *Понтрягин Л.С., Болтянский В.Г., Гамкрелидзе Р.В., Мищенко Е.Ф.* Математическая теория оптимальных процессов. М. 1961.
2. *Беллман Р.* Динамическое программирование. М. ИЛ. 1961.
3. *Понтрягин Л.С.* Обыкновенные дифференциальные уравнения. М. 1961.
4. *Арис Р.* Метод динамического программирования в дискретных системах. М. 1970.
5. *Данскин Д.* Максимин. М. ИЛ. 1970.
6. *Киселев Ю.Н.* Оптимальное управление. М. 1986.
7. *Киселев Ю.Н.* Дифференциальные уравнения. 1996. N.4.
8. *Киселев Ю.Н.* Фундаментальная и прикладная математика. 1997. N.3.