Глава 13

Динамическое программирование

Динамическое программирование является одним из двух современных направлений в теории задач управления [1]. Метод динамического программирования может применяться непосредственно при решении общей задачи управления [2]:

$$\max_{\{\mathbf{u}(t)\}} \mathbf{J} = \int_{t_0}^{t_1} I(\mathbf{x}, \mathbf{u}, t)\, dt + F(\mathbf{x}_1, t_1)$$

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{f}(\mathbf{x}, \mathbf{u}, t) \\ \mathbf{x}(t_0) &= \mathbf{x}_0 \\ \mathbf{x}(t_1) &= \mathbf{x}_1 \\ \{\mathbf{u}(t)\} &\in U. \end{aligned} \tag{13.0.1}$$

Сущность подхода динамического программирования состоит в следующем: данная конкретная задача управления «погружается» в более широкий класс задач, которые характеризуются рядом параметров; затем с помощью центрального принципа —«принципа оптимальности»— определяется основное рекуррентное соотношение, связывающее задачи из этого класса. Если выполнены некоторые дополнительные предположения относительно гладкости участвующих в рассмотрении функций, то из главного рекуррентного соотношения вытекает основное дифференциальное уравнение в частных производных — уравнение

---

[1] Основная литература по динамическому программированию [1, 2, 3, 4, 5, 6, 7].

[2] Более полно общая задача управления рассматривалась в гл. 11.

Беллмана,— решая которое можно найти решение вышеупомянутого широкого класса задач.

Вслед за этим, как частный случай, определяется и решение данной конкретной задачи.

13.1. ПРИНЦИП ОПТИМАЛЬНОСТИ И УРАВНЕНИЕ БЕЛЛМАНА

Формулировка принципа *оптимальности*:

«Оптимальное поведение обладает тем свойством, что, каковы бы ни были первоначальное состояние и решение (т. е. управление) в начальный момент, последующие решения должны составлять оптимальное поведение относительно состояния, получающегося в результате первого решения [1].

На рис. 13.1 дана иллюстрация этого принципа на примере задачи с одной фазовой координатой. Кривая $x^*(t)$ $(t_0 \leqslant t \leqslant t_1)$ — это фазовая траектория, соответствующая оптимальному управлению, при этом предполагается, что начальное состояние и конечное — фиксированы. Вся траектория разделена на две части: ① и ② относительно момента времени τ. Согласно принципу оптимальности, траектория ②, определенная при $\tau \leqslant t \leqslant t_1$, должна представлять собой оптимальную траекторию по отношению к начальному состоянию $x(\tau)$. Следовательно, вторая часть оптимальной траектории сама по себе должна быть оптимальной траекторией, вне зависимости от того, что происходило с системой до того, как она пришла к состоянию, являющемуся начальным для второй части траектории.

Предположим, что общая задача управления (13.0.1) имеет решение. Максимальное значение целевого функционала задачи с начальным состоянием x и начальным временем t

$$J^*(\mathbf{x},\ t) \tag{13.1.2}$$

[1] См. работу Беллмана [1]. Доказательство необходимости принципа оптимальности можно легко получить, рассуждая от противного. Арис [8] выразил этот принцип в следующих словах: «Если вы не используете наилучшим образом то, чем вы располагаете, то вы никогда не распорядитесь наилучшим образом и тем, что вы могли бы иметь в дальнейшем».

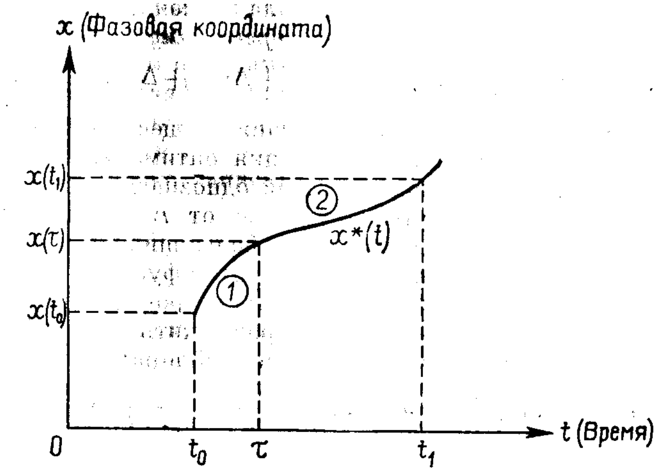

Рис. 13.1. Согласно принципу оптимальности, отдельный участок оптимальной траектории ② также представляет собой оптимальную траекторию.

назовем *функцией оптимального поведения* [1]. Тем самым задача оказывается «погруженной» в более широкий класс задач, характеризуемых значениями $n+1$ начальных параметров. Оптимальное значение целевой функции данной конкретной задачи (13.0.1) имеет, таким образом, вид

$$J^* = J^*(\mathbf{x}_0,\ t_0). \tag{13.1.3}$$

Если $J^*(\mathbf{x},\ t)$ является функцией оптимального поведения для задачи с начальным состоянием x в момент t, то, согласно принципу оптимальности, $J^*(\mathbf{x}+\Delta\mathbf{x},\ t+\Delta t)$ является фукнцией оптимального поведения для второй части оптимальной тректории с начальным моментом времени $t+\Delta t$ и начальным состоянием $\mathbf{x}+\Delta\mathbf{x}$. Поскольку прирост функции оптимального поведения на протяжении всего промежутка времени между t и $t+\Delta t$ может происходить только за счет изменения подынтегральной функции, то он равен $I(\mathbf{x},\ \mathbf{u},\ t)\,\Delta t$. Значения функции оптимального поведения на всем интервале времени, начавшемся в момент t, представляют собой оптимальную сумму

[1] Отметим, что в то время как J представляет собой функционал, зависящий от управления $\{\mathbf{u}(t)\}$, J^* является функцией, зависящей от $n+1$ параметров: $\mathbf{x}$ и t.

вкладов двух частей этого интервала времени. Таким образом, приходим к *основному рекуррентному соотношению*

$$J^*(\mathbf{x}, t) = \max_{\{\mathbf{u}(t)\}} [I(\mathbf{x}, \mathbf{u}, t)\,\Delta t + J^*(\mathbf{x}+\Delta\mathbf{x}, t+\Delta t)]. \qquad (13.1.4)$$

В динамическом программировании существенную роль играет предположение, что функция оптимального поведения $J^*(\mathbf{x}, t)$ представляет собой однозначную и непрерывно дифференцируемую функцию от $n+1$ переменных. Иначе говоря, решения задач более широкого класса являются однозначными и непрерывными функциями относительно изменений начальных параметров [1]. Благодаря этому предположению можно разложить $J^*(\mathbf{x}+\Delta\mathbf{x}, t+\Delta t)$ в точке $(\mathbf{x}, t)$ по формуле Тейлора:

$$J^*(\mathbf{x}+\Delta\mathbf{x}, t+\Delta t) = J^*(\mathbf{x}, t) + \frac{\partial J^*}{\partial \mathbf{x}}\Delta\mathbf{x} + \frac{\partial J^*}{\partial t}\Delta t + \ldots \qquad (13.1.5)$$

где $\partial J^*/\partial\mathbf{x}$ — это вектор-строка

$$\frac{\partial J^*}{\partial \mathbf{x}} = \left(\frac{\partial J^*}{\partial x_1}, \frac{\partial J^*}{\partial x_2}, \ldots, \frac{\partial J^*}{\partial x_n}\right). \qquad (13.1.6)$$

Подставляя (13.1.5) в (13.1.4), получаем

$$0 = \max_{\{\mathbf{u}(t)\}}\left[I(\mathbf{x}, \mathbf{u}, t) + \frac{\partial J^*}{\partial \mathbf{x}}\frac{\Delta\mathbf{x}}{\Delta t} + \frac{\partial J^*}{\partial t} + \ldots\right]. \qquad (13.1.7)$$

Переход к пределу при $\Delta t \to 0$ приводит к соотношению

$$-\frac{\partial J^*}{\partial t} = \max_{\{\mathbf{u}(t)\}}\left[I(\mathbf{x}, \mathbf{u}, t) + \frac{\partial J^*}{\partial \mathbf{x}}\mathbf{f}(\mathbf{x}, \mathbf{u}, t)\right], \qquad (13.1.8)$$

так как

$$\lim_{\Delta t\to 0}\frac{\Delta\mathbf{x}}{\Delta t} = \dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}, t). \qquad (13.1.9)$$

[1] Во многих задачах эти предположения о гладкости не выполняются, кроме того, заранее вообще неизвестно, будут ли они выполнены в данной конкретной задаче. См. [9]. В качестве примера решения, которое не является гладкой функцией параметров, рассмотрим задачу об отыскании геодезических линий (линий, соответствующих кратчайшим расстояниям между двумя точками) на сфере. Такой линией является дуга большого круга. В частности, кратчайший путь, соединяющий две точки земного экватора, проходит именно по экватору. Предположим теперь, что начальная точка перемещается по экватору, удаляясь от конечной точки. В конце концов будет достигнуто такое положение начальной точки, при котором кратчайший путь, соединяющий эту точку с конечной, будет соответствовать движению в направлении, противоположном использованному ранее. Производная кратчайшего расстояния относительно положения начальной точки (измеряемого, например, долготой точки) претерпевает разрыв в указанной точке.

Уравнение (13.1.8) является основным дифференциальным уравнением в частных производных, используемым в динамическом программировании. Оно называется *уравнением Беллмана* [1]. Так как второй член в квадратных скобках представляет собой скалярное произведение вектора-строки $\partial G^*/\partial\mathbf{x}$ и вектора-столбца $\mathbf{f}(\mathbf{x}, \mathbf{u}, t)$, то уравнение Беллмана можно записать как

$$-\frac{\partial J^*}{\partial t} = \max_{\{\mathbf{u}(t)\}}\left[I(\mathbf{x}, \mathbf{u}, t) + \sum_{j=1}^{n}\frac{\partial J^*}{\partial x_j}f_j(\mathbf{x}, \mathbf{u}, t)\right]. \qquad (13.1.10)$$

С уравнением Беллмана связано в качестве граничного условия ограничение, налагаемое на конечное состояние:

$$J^*(\mathbf{x}(t_1), t_1) = F(\mathbf{x}_1, t_1). \qquad (13.1.11)$$

Это условие показывает, что значение функции оптимального поведения для задачи, начальным моментом и начальным состоянием которой являются соответственно конечный момент и конечное состояние, равно значению функции $F(\cdots)$, рассчитанному в данный момент времени при данном состоянии.

Если бы уравнение Беллмана было решено, то мы получили бы функцию оптимального поведения и, следовательно, оптимальное значение целевой функции для исходной задачи можно было бы определить, как частное значение этой функции при указанных в формулировке задачи конкретных начальных условиях. Однако в общем случае это уравнение в частных производных первого порядка, как правило, нелинейное, не имеет аналитического решения. В принципе можно решать уравнения Беллмана, представленные в виде разностных схем, на цифровых элект-

[1] Пусть $\{\mathbf{u}^*(t)\}$ является решением задачи максимизации правой части уравнения Беллмана и функция $H(\mathbf{x}, \partial J^*/\partial\mathbf{x}, t)$ определена следующим образом:

$$H\left(\mathbf{x}, \frac{\partial J^*}{\partial \mathbf{x}}, t\right) = I(\mathbf{x}, \mathbf{u}^*, t) + \frac{\partial J^*}{\partial \mathbf{x}}f(\mathbf{x}, \mathbf{u}^*, t).$$

Тогда приходим к дифференциальному уравнению в частных производных $H\left(\mathbf{x}, \frac{\partial J}{\partial \mathbf{x}}, t\right) + \frac{\partial J^*}{\partial t} = 0$, называемому уравнением Гамильтона — Якоби. См. дополнительно к основной литературе, указанной в примечании на стр. 395, Гельфанд и Фомин [10] и Хестенс [11].

ронно-вычислительных машинах с большим быстродействием. Однако даже современные электронно-вычислительные машины, работающие с высокой скоростью счета, не обладают такой машинной памятью, которая позволяла бы найти достаточно хорошее приближение к решению, даже если система имеет сравнительно скромную *размерность*[1]. Беллман называет это препятствие «проклятием размерности» («curse of dimensionality»).

13.2. ДИНАМИЧЕСКОЕ ПРОГРАММИРОВАНИЕ И ВАРИАЦИОННОЕ ИСЧИСЛЕНИЕ

Задача динамического программирования является более общей, нежели задача классического вариационного исчисления. Если рассматривать классическую задачу вариационного исчисления как частный случай задачи динамического программирования, то из необходимого условия динамического программирования — выполнения уравнения Беллмана — должны вытекать необходимые условия вариационного исчисления, включая уравнения Эйлера, условия Лежандра, условие Вейерштрасса и условия Вейерштрасса — Эрдмана для точки излома [1, 2, 12, 13].

Классическая задача вариационного исчисления является частным случаем задачи динамического программирования (13.0.1), в которой

$$\dot{\mathbf{x}} = \mathbf{u}, \quad \Omega = E^n, \tag{13.2.1}$$

т. е. в этой задаче управляющими параметрами являются скорости изменения фазовых координат во времени, причем

[1] Задачи динамического программирования требуют, чтобы память машины содержала Q^n ячеек памяти, где Q — это размер сетки, т. е. число дискретных значений, принимаемых каждой из фазовых координат. Если, например, каждая фазовая координата разбита на 100 дискретных значений, а $n = 4$, то память должна состоять из 100 млн. ячеек. Поскольку оперативная память большинства современных машин не содержит 100 млн. ячеек, то вычислительные процедуры динамического программирования зависят от информации, хранимой во внешней памяти машины на дисках или лентах. Существуют, правда, несколько способов уменьшения трудностей, вызываемых размерностью задачи. См. Беллман и Дрейфус [3].

на значения управляющих параметров не накладывается никаких ограничений. Уравнение Беллмана принимает вид

$$-\frac{\partial J^*}{\partial t} = \max_{\{\dot{\mathbf{x}}\}} \left[I(\mathbf{x}, \dot{\mathbf{x}}, t) + \frac{\partial J^*}{\partial \mathbf{x}} \dot{\mathbf{x}} \right]. \tag{13.2.2}$$

Необходимым условием существования максимума выражения, стоящего в квадратных скобках, является выполнение соотношения

$$\frac{\partial}{\partial \dot{\mathbf{x}}} \left[I(\mathbf{x}, \dot{\mathbf{x}}, t) + \frac{\partial J^*}{\partial \mathbf{x}} \dot{\mathbf{x}} \right] = 0 \tag{13.2.3}$$

или соотношения

$$\frac{\partial I}{\partial \dot{\mathbf{x}}} = -\frac{\partial J^*}{\partial \mathbf{x}}, \tag{13.2.4}$$

поскольку $\partial J^*/\partial \mathbf{x}$ не зависит от $\dot{x}$.

Найдем производную по t от $\partial I/\partial \dot{\mathbf{x}}$.
Учитывая тот факт, что $\partial J^*/\partial \mathbf{x}$ зависит от $\mathbf{x}$ и от t, и вводя обозначения

$$\frac{\partial^2 J^*}{\partial t\, \partial \mathbf{x}} = \left(\frac{\partial^2 J^*}{\partial t\, \partial x_1}, \ldots, \frac{\partial^2 J^*}{\partial t\, \partial x_n} \right)$$

$$\frac{\partial^2 J^*}{\partial \mathbf{x}^2} = \begin{pmatrix} \frac{\partial^2 J^*}{\partial x_1^2} & \cdots & \frac{\partial^2 J^*}{\partial x_1\, \partial x_n} \\ \vdots & & \\ \frac{\partial^2 J^*}{\partial x_n\, \partial x_1} & \cdots & \frac{\partial^2 J^*}{\partial x_n^2} \end{pmatrix}, \tag{13.2.5}$$

получаем, что

$$\frac{d}{dt}\left(\frac{\partial I}{\partial \dot{\mathbf{x}}} \right) = -\frac{d}{dt}\left(\frac{\partial J^*}{\partial \mathbf{x}} \right) = -\frac{\partial^2 J^*}{\partial t\, \partial \mathbf{x}} - (\dot{\mathbf{x}})' \frac{\partial^2 J^*}{\partial \mathbf{x}^2}. \tag{13.2.6}$$

Однако из уравнения Беллмана следует, что

$$-\frac{\partial}{\partial \mathbf{x}}\left(\frac{\partial J^*}{\partial t} \right) = \frac{\partial I}{\partial \mathbf{x}} + (\dot{\mathbf{x}})' \frac{\partial^2 J^*}{\partial \mathbf{x}^2}. \tag{13.2.7}$$

Сопоставим (13.2.6) и (13.2.7). Учитывая, что в данном случае величина смешанных частных производных второго порядка не зависит от порядка дифференцирования, приходим к уравнению Эйлера для вариационного исчисления:

$$\frac{\partial I}{\partial \mathbf{x}} - \frac{d}{dt}\left(\frac{\partial I}{\partial \dot{\mathbf{x}}} \right) = 0. \tag{13.2.8}$$

Условие Лежандра следует непосредственно из необходимого условия второго порядка для существования максимума выражения в квадратных скобках в (13.2.2), состоящего в том, что матрица

$$\frac{\partial^2}{\partial \dot{\mathbf{x}}^2}\left[I(\mathbf{x}, \dot{\mathbf{x}}, t)+\frac{\partial J^*}{\partial \mathbf{x}}\dot{\mathbf{x}}\right] \tag{13.2.9}$$

является отрицательно полуопределенной или отрицательно определенной.

Поскольку $\partial J^*/\partial \mathbf{x}$ не зависит от $\dot{\mathbf{x}}$, то это условие заключается в том, что матрица

$$\frac{\partial^2 I}{\partial \dot{\mathbf{x}}^2} \tag{13.2.10}$$

является отрицательно полуопределенной или отрицательно определенной, а это есть не что иное, как условие Лежандра.

Если $\{\dot{\mathbf{x}}(t)\}$ представляет собой решение уравнения Беллмана, то для любого вектора-столбца $\dot{\mathbf{z}}$

$$I(\mathbf{x}, \dot{\mathbf{x}}, t)+\frac{\partial J^*}{\partial \mathbf{x}}\dot{\mathbf{x}} \geqslant I(\mathbf{x}, \dot{\mathbf{z}}, t)+\frac{\partial J^*}{\partial \mathbf{x}}\dot{\mathbf{z}}. \tag{13.2.11}$$

Преобразуем это выражение, воспользовавшись равенством (13.2.4):

$$I(\mathbf{x}, \dot{\mathbf{z}}, t)-I(\mathbf{x}, \dot{\mathbf{x}}, t)-\frac{\partial I}{\partial \dot{\mathbf{x}}}(\mathbf{x}, \dot{\mathbf{x}}, t)(\dot{\mathbf{z}}-\dot{\mathbf{x}}) \leqslant 0. \tag{13.2.12}$$

Полученное неравенство представляет собой условие Вейерштрасса.

Наконец, условия Вейерштрасса — Эрдмана для точки излома можно получить из следующих уравнений:

$$\begin{gathered}\frac{\partial I}{\partial \dot{\mathbf{x}}}=-\frac{\partial J^*}{\partial \mathbf{x}}\\ I-\frac{\partial I}{\partial \dot{\mathbf{x}}}\dot{\mathbf{x}}=I+\frac{\partial J^*}{\partial \mathbf{x}}\dot{\mathbf{x}}=-\frac{\partial J^*}{\partial t}.\end{gathered} \tag{13.2.13}$$

Так как функции $\partial J^*/\partial \mathbf{x}$ и $\partial J^*/\partial t$ являются непрерывными, то

$$\frac{\partial I}{\partial \dot{\mathbf{x}}} \text{ и } \left(I-\frac{\partial I}{\partial \dot{\mathbf{x}}}\dot{\mathbf{x}}\right) \tag{13.2.14}$$

непрерывны в точке излома, т. е. выполнены условия Вейерштрасса — Эрдмана.

Таким образом, подход динамического программирования позволяет получить необходимые условия для классической задачи вариационного исчисления. Динамическое программирование можно применять также и для исследования задач вариационного исчисления при наличии ограничений, рассмотренных в разделе 12.4. Так, например, уравнение Беллмана для изопериметрической задачи, т. е. задачи с ограничением вида

$$\int_{t_0}^{t_1} G(\mathbf{x}, \dot{\mathbf{x}}, t)=c, \tag{13.2.15}$$

принимает следующую форму:

$$-\frac{\partial J^*}{\partial t}=\max_{\{\dot{\mathbf{x}}\}}\left[I(\mathbf{x}, \dot{\mathbf{x}}, t)+\frac{\partial J^*}{\partial \mathbf{x}}\dot{\mathbf{x}}+\frac{\partial J^*}{\partial c}G(\mathbf{x}, \dot{\mathbf{x}}, t)\right]. \tag{13.2.16}$$

Из этого выражения можно вывести те же условия, что и при постановке задачи в обычном для вариационного исчисления виде, поскольку множителем Лагранжа здесь является величина изменения оптимального значения функционала, вызванного изменением константы ограничения c, т. е.

$$y=\frac{\partial J^*}{\partial c}. \tag{13.2.17}$$

Вообще говоря, частные производные функции оптимального поведения можно интерпретировать как множители Лагранжа, измеряющие чувствительность решения.

13.3. РЕШЕНИЕ МНОГОШАГОВЫХ ЗАДАЧ ОПТИМИЗАЦИИ МЕТОДОМ ДИНАМИЧЕСКОГО ПРОГРАММИРОВАНИЯ

Во многих динамических задачах время рассматривается не как непрерывная, а как дискретная величина. Задачи такого типа, называемые *многошаговыми задачами оптимизации*, можно решать методом динамического программирования [1, 14, 8, 15, 16].

В многошаговых задачах оптимизации время принимает дискретные значения

$$t_0,\ t_0+1,\ t_0+2,\ \ldots,\ t_1. \tag{13.3.1}$$

Состояние системы в момент времени t задается вектором $\mathbf{x}_t$, а *управление* в момент t задается вектором $\mathbf{u}_t$. Состояние в момент $t+1$ задается соотношением

$$\mathbf{x}_{t+1}=\mathbf{f}_t(\mathbf{x}_t, \mathbf{u}_t),\ t=t_0,\ t_0+1,\ \ldots,\ t_1-1, \tag{13.3.2}$$

где $\mathbf{f}_t(\cdots)$ — вектор, составленный из непрерывно дифференцируемых функций текущего состояния и текущих значений управляющих параметров. Предполагается, что фиксировано начальное состояние

$$\mathbf{x}_0. \tag{13.3.3}$$

Требуется найти такую последовательность управляющих векторов

$$\{\mathbf{u}_{t_0}, \mathbf{u}_{t_0+1}, \ldots, \mathbf{u}_{t_1-1}\}, \tag{13.3.4}$$

принадлежащих фиксированной области управления

$$\mathbf{u}_t \in \Omega, \ t = t_0, \ t_0+1, \ \ldots, \ t_1-1, \tag{13.3.5}$$

которая максимизирует целевую функцию

$$J = \sum_{t=t_0}^{t_1-1} I_t(\mathbf{x}_t, \mathbf{u}_t) + F(\mathbf{x}_{t_1}, t_1). \tag{13.3.6}$$

Вполне очевидно, что поставленная таким образом задача аналогична задаче управления (с непрерывным временем).

Подход динамического программирования и в данном случае состоит в том, что решаемая задача «погружается» в более широкий класс задач, описываемых рядом параметров, и вслед за этим с помощью принципа оптимальности определяется основное рекуррентное соотношение. Возьмем в качестве параметров многошаговой задачи оптимизации начальный момент времени и начальное состояние. Тогда функция оптимального поведения равна оптимальному значению целевой функции в задаче с начальным состоянием $\mathbf{x}$ и начальным моментом времени t:

$$J^*(\mathbf{x}, t). \tag{13.3.7}$$

Оптимальное значение целевой функции рассматриваемой задачи равно

$$J^*(\mathbf{x}_0, t_0). \tag{13.3.8}$$

Согласно принципу оптимальности,

$$J^*(\mathbf{x}, t) = \max_{\mathbf{u}_t} [I_t(\mathbf{x}_t, \mathbf{u}_t) + J^*(\mathbf{x}_{t+1}, t+1)]. \tag{13.3.9}$$

Это означает, что оптимальное значение целевой функции в задаче с начальным состоянием $\mathbf{x}$ и начальным временем t равно оптимальному значению суммы двух слагаемых: функции $I_t(\mathbf{x}_t, \mathbf{u}_t)$ в момент t и оптимального значения

функции $J^*(\mathbf{x}_{t+1}, t+1)$ в момент t. Используя уравнение (13.3.2), можно представить рекуррентное соотношение в виде

$$J^*(\mathbf{x}, t) = \max_{u_t} [I_t(\mathbf{x}_t, \mathbf{u}_t) + J^*(\mathbf{f}_t(\mathbf{x}_t, \mathbf{u}_t), t+1)]. \tag{13.3.10}$$

Граничное условие

$$J^*(\mathbf{x}_1, t_1) = F(\mathbf{x}_{t_1}, t_1), \tag{13.3.11}$$

показывает, что оптимальное значение целевой функции в задаче с начальным состоянием $\mathbf{x}_1$ в момент t_1 просто совпадает со значением функции конечных параметров, рассчитанным при $\mathbf{x} = \mathbf{x}_1$, $t = t_1$. Вполне очевидно, что рассмотренная задача аналогична задаче с непрерывным временем.

Другой подход при решении многошаговой задачи оптимизации состоит в том, что в качестве характеристических параметров выбираются не начальное состояние и начальный момент, а начальное состояние и промежуток времени, *остающийся до конечного момента*. В этом случае функцией оптимального поведения является

$$J_\tau^*(\mathbf{x}_{t_1-\tau}), \tag{13.3.12}$$

которая представляет собой оптимальное значение целевой функции для процесса с начальным состоянием $\mathbf{x}_{t_1-\tau}$, развертывающегося в промежутке протяженностью τ. Оптимальное значение целевой функции рассматриваемой нами задачи, соответствующей $\tau = t_1$, равно $J_{t_1}^*(\mathbf{x}_0)$. В этом случае последовательность решений определяется методом динамического программирования в порядке, обратном тому, который рассматривался до сих пор, т. е. начиная с конечного момента времени t_1. Первым членом этой последовательности является $J_0^*(\mathbf{x}_{t_1})$, т. е. оптимальное значение целевой функции с временным промежутком нулевой протяженности, начинающимся (и заканчивающимся) в $\mathbf{x}_{t_1}$. Оптимальное значение целевой функции этой задачи равно значению функции конечных параметров:

$$J_0^*(\mathbf{x}_{t_1}) = F(\mathbf{x}_{t_1}, t_1). \tag{13.3.13}$$

Рассмотрим теперь $J_1^*(\mathbf{x}_{t_1-1})$ — оптимальное значение целевой функции задачи с промежутком, равным одной единице времени, начинающимся в $\mathbf{x}_{t-1}$. Эта задача называется *первым шагом*.

Оптимальное значение в этой задаче определяется как максимальное значение суммы той части целевой функции, которая соответствует указанному времени — $I_{t_1-1}(\mathbf{x}_{t_1-1}, \mathbf{u}_{t_1-1})$, и оптимального значения задачи с начальным моментом t_1, относительно управляющего вектора $\mathbf{u}_{t_1-1}$, т. е.

$$J_1^*(\mathbf{x}_{t_1-1}) = \max_{\mathbf{u}_{t_1+1}} [I_{t_1-1}(\mathbf{x}_{t_1-1}, \mathbf{u}_{t_1-1}) + J_0^*(\mathbf{x}_{t_1})] \qquad (13.3.14)$$

или, используя (13.3.2),

$$J_1^*(\mathbf{x}_{t_1-1}) = \max_{\mathbf{u}_{t_1-1}} [I_{t_1-1}(\mathbf{x}_{t_1-1}, \mathbf{u}_{t_1-1}) + J_0^*(\mathbf{f}_{t_1-1}(\mathbf{x}_{t_1-1}, \mathbf{u}_{t_1-1}))]. \qquad (13.3.15)$$

Данный выбор управления на первом шаге согласуется с принципом оптимальности, поскольку управление $\mathbf{u}_{t_1-1}$ является оптимальным по отношению к состоянию $\mathbf{x}_{t_1-1}$, которое достигнет в результате $t_1 - 1$ предшествующих выборов управляющих векторов $\mathbf{u}_{t_0}, \mathbf{u}_{t_0+1}, \ldots, \mathbf{u}_{t_1-2}$. Аналогично этому на втором шаге (в задаче с промежутком, равным двум единицам времени)

$$J_2^*(\mathbf{x}_{t_1-2}) = \max_{\mathbf{u}_{t_1-2}} [I_{t_1-2}(\mathbf{x}_{t_1-2}, \mathbf{u}_{t_1-2}) + J_1^*(\mathbf{f}_{t_1-2}(\mathbf{x}_{t_1-2}, \mathbf{u}_{t_1-2}))]. \qquad (13.3.16)$$

Общее рекуррентное соотношение на шаге с номером τ имеет вид

$$J_\tau^*(\mathbf{x}_{t_1-\tau}) = \max_{\mathbf{u}_{t_1-\tau}} [I_{t_1-\tau}(\mathbf{x}_{t_1-\tau}, \mathbf{u}_{t_1-\tau}) + J_{\tau-1}^*(\mathbf{f}_{t_1-\tau}(\mathbf{x}_{t_1-\tau}, \mathbf{u}_{t_1-\tau}))]. \qquad (13.3.17)$$

Оптимальное значение целевой функции рассматриваемой задачи, равное $J_{t_1}^*(\mathbf{x}_0)$, является оптимальным значением последней задачи в последовательности одношаговых задач оптимизации, описываемых функциональными уравнениями (13.3.17) при $\tau = 1, 2, 3, \ldots, t_1$ с граничным условием (13.3.13). Таким образом, многошаговая задача оптимизации методом динамического программирования приведена к последовательности одношаговых задач оптимизации [1].

[1] Численное решение многошаговых задач оптимизации методом динамического программирования на электронных вычислительных машинах, как и в непрерывном случае, затрудняется недостаточным объемом машинной памяти. При таком решении необходимо найти всю последовательность функции $J_\tau^*(\mathbf{x}_{t_1-\tau})$ и хранить ее в памяти машины. Обычно решения находят с помощью некоторых приближений. См. Беллман и Дрейфус [3].

В качестве примера применения подхода динамического программирования к многошаговым задачам оптимизации рассмотрим задачу, в которой требуется найти набор из неотрицательных чисел $u_{t_0}, u_{t_0+1}, \ldots, u_{t_1}$, максимизирующих сепарабельную целевую функцию при условии, что сумма этих чисел равна фиксированному числу [1] c:

$$\max J = \sum_{t=t_0}^{t_1} I_t(u_t)$$
$$u_t \geqslant 0, \quad t = t_0, t_0+1, \ldots, t_1 \qquad (13.3.18)$$
$$\sum_{t=t_0}^{t_1} u_t = c.$$

Можно интерпретировать постоянную c, как общий уровень имеющихся ресурсов и рассматривать ее в качестве параметра задачи. Обозначим через

$$J_\tau^*(c) \qquad (13.3.19)$$

функцию оптимального поведения для процесса, развертывающегося в промежутке протяженности τ и заканчивающегося в момент t_1, с общим запасом ресурсов, равным c. Для процесса на временном промежутке, протяженность которого равна нулю, заканчивающегося при $t = t_1$

$$J_0^*(c) = \max_{u_{t_1}=c} I_{t_1}(u_{t_1}) = I_{t_1}(c). \qquad (13.3.20)$$

Для одношагового процесса, заканчивающегося в момент t_1, надо распределить ресурсы между u_{t_1} и u_{t_1-1}. Согласно принципу оптимальности,

$$J_1^*(c) = \max_{0 \leqslant u_{t_1-1} \leqslant c} [I_{t_1-1}(u_{t_1-1}) + J_0^*(c - u_{t_1-1})], \qquad (13.3.21)$$

так что (из 13.3.20)

$$J_1^*(c) = \max_{0 \leqslant u_{t_1-1} \leqslant c} [I_{t_1}(u_{t_1-1}) + I_{t_1}(c - u_{t_1-1})]. \qquad (13.3.22)$$

Общее рекуррентное соотношение для этой задачи имеет вид

$$J_\tau^*(c) = \max_{0 \leqslant u_{t_1-\tau} \leqslant c} [I_{t_1-\tau}(u_{t_1-\tau}) + J_{\tau-1}^*(c - u_{t_1-\tau})], \qquad (13.3.23)$$

[1] См. работу Беллмана [1], а также книгу Беллмана и Дрейфуса [3]. С формальной точки зрения данная задача сходна с задачей нелинейного программирования с сепарабельной целевой функцией. Использование динамического программирования для решения некоторых задач нелинейного программирования рассматривается в книге Хедли [17].

оно показывает оптимальное распределение ресурсов между $u_{t_1-\tau}$, определяющим значение $I_{t_1-\tau}(u_{t_1-\tau})$, и $c - u_{t_1-\tau}$, определяющим значение $J^*_{\tau-1}(c - u_{t_1-\tau})$.

Решение задачи отыскивается с помощью общего рекуррентного соотношения (13.3.23) последовательно, начиная с граничного условия (13.3.20) вплоть до шага с номером t_1. Оптимальное значение целевой функции задачи равно $J^*_{t_1}(c)$.

Рассмотрим теперь частную задачу о минимизации суммы квадратов неотрицательных переменных при условии, что их сумма равна фиксированному числу

$$\max J = -\sum_{t=t_0}^{t_1} u_t^2$$
$$u_t \geqslant 0,\ t = t_0,\ t_0+1,\ \ldots,\ t_1 \qquad (13.3.24)$$
$$\sum_{t=t_0}^{t_1} u_t = c.$$

Согласно методу динамического программирования, оптимальное значение целевой функции для задачи с временным промежутком нулевой протяженности равно

$$J_0^*(c) = \max_{u_{t_1}=c} - u_{t_1}^2 = -c^2. \qquad (13.3.25)$$

Первое функциональное уравнение для процесса, развертывающегося на промежутке единичной протяженности, согласно (13.3.21), имеет вид

$$J_1^*(c) = \max_{0\leqslant u_{t_1-1}\leqslant c} [-u_{t_1-1}^2 + J_0^*(c - u_{t_1-1})]. \qquad (13.3.26)$$

Следовательно, используя (13.3.25), можно получить, что

$$J_1^*(c) = \max_{0\leqslant u_{t_1-1}\leqslant c} [-u_{t_1}^2 - (c - u_{t_1-1})^2]. \qquad (13.3.27)$$

Так как частная производная выражения в квадратных скобках в точке максимума равна нулю, то

$$u_{t_1-1} = \frac{1}{2}c, \qquad (13.3.28)$$

что совместимо с ограничением $0 \leqslant u_{t_1-1} \leqslant c$. Таким образом, половина ресурсов должна быть использована в момент t_1, а половина — в момент $t_1 - 1$. Следующее функциональное уравнение имеет вид

$$J_2^*(c) = \max_{0\leqslant u_{t_1-2}\leqslant c} [-u_{t_1-2}^2 + J_1^*(c - u_{t_1-2})^2]. \qquad (13.3.29)$$

Но так как в оптимальной точке $J_1^*(c) = -\frac{1}{2}c^2$, то

$$J_2^*(c) = \max_{0\leqslant u_{t_1-2}\leqslant c} [-u_{t_1-2}^2 - \frac{1}{2}(c - u_{t_1-2})^2]. \qquad (13.3.30)$$

В точке максимума

$$u_{t_1-2} = \frac{1}{3}c. \qquad (13.3.31)$$

Следовательно, одна треть имеющихся ресурсов применяется в момент $t_1 - 2$, а оставшиеся две трети распределяются поровну между моментами $t_1 - 1$ и t_1. Общее решение задачи имеет вид

$$u_{t_0} = u_{t_0+1} = \ldots = u_{t_1} = \frac{c}{(t_1 - t_0) + 1}, \qquad (13.3.32)$$

т. е. для того, чтобы минимизировать сумму квадратов, следует использовать равное количество ресурсов в каждый отдельный момент времени.

ЗАДАЧИ

13-А. Классической задачей управления является задача о брахистохроне: найти такую кривую, соединяющую две точки P и Q, что материальная точка, двигающаяся под действием силы тяжести по этой кривой без трения и с нулевой начальной скоростью в точке P, достигает точки Q за минимальное время. Пусть точка P' лежит на кривой, являющейся решением, между P и Q. Является ли часть этой кривой между P' и Q оптимальной? Если да, то в каком смысле? Что можно сказать по этому поводу относительно части кривой между P и P'?

13-Б. Найдите уравнение Беллмана для следующей задачи:

$$\max_{\{u(t)\}} J = -\int_{t_0}^{t_1} [(x-c)^2 + u^2]\,dt$$
$$\dot{x} = ax + bu$$
$$x(t_0) = x_0$$
$$x(t_1) = x_1.$$

13-В. Найдите уравнение Беллмана для следующей задачи:

$$\max_{\{u(t)\}} J = \int_{t_0}^{t_1} (x-u)\,dt$$

$$\dot{x} = \sqrt{u}$$

$$x(t_0) = x_0$$

$$x(t_1) = x_1$$

$$0 \leqslant u \leqslant x.$$

13-Г. Используя метод динамического программирования, решить задачу управления по минимуму энергии при условии, что уравнения движения линейны:

$$\max_{\{\mathbf{u}(t)\}} J = \int_{t_0}^{t_1} (\mathbf{x}'\mathbf{D}\mathbf{x} + \mathbf{u}'\mathbf{E}\mathbf{u})\,dt$$

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}$$

$$\mathbf{x}(t_0) = \mathbf{x}_0,$$

$$\mathbf{x}(t_1) = \mathbf{x}_1.$$

Здесь D и E — отрицательно определенные матрицы, а A и B — некоторые фиксированные матрицы.

13-Д. Найдите уравнение Беллмана в задаче о минимальном времени перехода из фиксированного начального состояния $\mathbf{x}(t_0) = \mathbf{x}_0$ в точку начала координат $\mathbf{x}(t_1) = \mathbf{0}$, в которой можно выбирать управления

$$\{\mathbf{u}(t)\} \in U, \quad \text{а} \quad \dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}, t).$$

13-Е. Примените результаты решения предыдущей задачи к решению частной задачи о минимальном времени перехода из $(x_1, x_2)'$ в $(0, 0)'$, в которой

$$\dot{x}_1 = V\cos x_3$$

$$\dot{x}_2 = V\sin x_3$$

$$\dot{x}_3 = u$$

$$V = V_0\sqrt{1+\left(\frac{x_2}{a}\right)^2},$$

где V — величина скорости.

13-Ж. Пусть функция оптимального действия в задаче управления типа Лагранжа зависит от начального состояния x и *продолжительности* процесса τ:

$$J^*(\mathbf{x}, \tau) = \max_{\{\mathbf{u}(t)\}} \int_{t_0}^{t_0+\tau} I(\mathbf{x}, \mathbf{u}, t)\,dt$$

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}, t)$$

$$\mathbf{x}(t_0) = \mathbf{x}_0$$

$$\{\mathbf{u}(t)\} \in U,$$

причем оптимальное значение целевой функции равно $J^*(\mathbf{x}_0, t_1 - t_0)$. Найти дифференциальное уравнение в частных производных — уравнение Беллмана — для этой задачи; сравнить полученное уравнение с (13.1.8).

13-К. Используя метод динамического программирования, покажите, что необходимое условие для максимума в классической задаче вариационного исчисления, в которой функция $I(\cdots)$ не зависит от времени t, заключается в выполнении уравнения

$$I - \frac{\partial I}{\partial \dot{\mathbf{x}}}\dot{\mathbf{x}} = \text{const}.$$

13-Л. Вывести условие трансверсальности для задач вариационного исчисления с помощью динамического программирования.

13-М. Рассмотрите следующее обобщение примера, приведенного в разделе 13.3.

$$\max J = -\sum_{t=t_0}^{t_1} w_t u_t^2$$

$$u_t \geqslant 0, \quad t = t_0, t_0+1, \ldots, t_1$$

$$\sum_{t=t_0}^{t_1} u_t = c,$$

где $w_{t_0}, w_{t_0+1}, \ldots, w_{t_1}$ — фиксированные неотрицательные веса.

1. Решите эту задачу методом динамического программирования.
2. Покажите, что решение по методу динамического программирования совпадает с решением этой задачи как задачи нелинейного программирования.

3. Решите частную задачу, в которой
$t_0 = 0,\ t_1 = 2,\ w_{t_0} = 2,\ w_{t_0+1} = 3,\ w_{t_1} = 6,\ c = 100.$

13-Н. Другое обобщение примера из параграфа 13.3:

$$\max J = -\sum_{t=t_0}^{t_1} u_t^{p_t}$$

$$u_t \geqslant 0,\quad t = t_0,\ t_0 + 1,\ \ldots,\ t_1$$

$$\sum_{t=t_0}^{t_1} u_t = c,$$

где $p_{t_0}, p_{t_0+1}, \ldots, p_{t_1}$ — фиксированные положительные постоянные.

1. Решите задачу с помощью динамического программирования.
2. Покажите, что решение по методу динамического программирования совпадает с решением этой задачи, как задачи нелинейного программирования.
3. Решите частную задачу, в которой $t_0 = 0$, $t_1 = 2$, $p_{t_0} = 1$, $p_{t_0+1} = 2$, $p_{t_1} = 3$, $c = 100$.
4. Решите общую задачу при условии, что на управляющие параметры наложены следующие условия:

$$u_t \geqslant 1,\quad t = t_0,\ t_0 + 1,\ \ldots,\ t_1$$

$$\prod_{t=t_0}^{t_1} u_t = c.$$

13-О. Решите задачу:

$$\max J = \sum_{t=t_0}^{t_1} \frac{p_t s_t}{(s_t + u_t)},$$

где p_t и s_t — такие параметры, что

$$p_t \geqslant 0,\quad s_t \geqslant 0,\quad t = t_0,\ t_0 + 1,\ \ldots,\ t_1$$

$$\sum_{t=t_0}^{t_1} p_t = 1,$$

а для управляющих переменных выполняются условия

$$u_t \geqslant 0,\quad t = t_0,\ t_0 + 1,\ \ldots,\ t_1$$

$$\sum_{t=t_0}^{t_1} u_t = c.$$

13-П. Покажите, что если функция $I(\cdot)$ в следующей задаче

$$\max J = \sum_{t=t_0}^{t_1} I(u_t)$$

$$u_t \geqslant 0,\ t = t_0,\ t_0 + 1,\ \ldots,\ t_1$$

$$\sum_{t=t_0}^{t_1} u_t = c$$

является выпуклой, то $\max J = I(c)$.

13-Р. Решите задачу нелинейного программирования

$$\max F(\mathbf{x}) = x_1,\ x_2,\ \ldots,\ x_n = \prod_{j=1}^{n} x_j$$

$$x_j \geqslant 0,\quad j = 1,\ 2,\ \ldots,\ n$$

$$\sum_{j=1}^{n} x_j = a$$

методом динамического программирования.

13-С. Рассмотрим некоторую матрицу $\mathbf{A} = (a_{ij})$. Требуется среди всех возможных траекторий перехода по элементам матрицы от a_{11} до a_{mn} найти траекторию, минимизирующую сумму элементов, через которые она проходит. Перемещения от одного элемента к другому разрешаются только вправо или вниз. Решить задачу методом динамического программирования.

13-Т. Задачу линейного программирования:

найти $\max\limits_{\mathbf{x}} F(\mathbf{x}) = \mathbf{cx}$ при условии, что $\mathbf{Ax} \leqslant \mathbf{b}$, $\mathbf{x} \geqslant \mathbf{0}$,

можно рассматривать как дискретную многошаговую задачу оптимизации. Ее можно решать с помощью принципа оптимальности, взяв за функцию оптимального поведения функцию $F_k^*(b_1, b_2, \ldots, b_m)$, определяемую как решение указанной задачи при дополнительных условиях, что

$$x_{k+1} = x_{k+2} = \ldots = x_n = 0.$$

Найти рекуррентное соотношение и граничное условие для функции оптимального поведения. Может ли указанный метод быть альтернативой симплекс-метода?