МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение высшего образования
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»

Кафедра математического анализа

Динамическое программирование и уравнение Беллмана

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студентки 4 курса 421 группы

направления ***02.03.01 Математика и компьютерные науки***

механико-математического факультета

Кирилиной Ксении Георгиевны

Научный руководитель

доцент, к.ф.-м.н., доцент		В. Г. Тимофеев
должность, уч.степень, уч.звание	подпись, дата	инициалы, фамилия

Заведующий кафедрой

и. о. зав. кафедрой, к.ф.-м.н., доцент		Е. В. Разумовская
должность, уч.степень, уч.звание	подпись, дата	инициалы, фамилия

Саратов 2022

Содержание

Введение

Динамическое программирование - это вычислительный метод решения задач оптимального управления определенной структуры. Динамическое программирование возникло и сформировалось в 1950-1953 гг. благодаря работам Р. Беллмана.

Динамическое программирование определяет оптимальное решение в многомерной задаче оптимального управления путем ее декомпозиции на этапы, каждый из которых представляет подзадачу относительно одной переменной.

1 Динамическое программирование

Опишем подход к вариационным задачам с помощью динамического программирования.

Рассмотрим задачу минимизации функционала I(y) = $\int\limits_a^b F(x,y,y')dx$,

где функция y подчинена граничному условию $y(a) = c$. Минимальное значение будет функцией начального значения a, переменной x и граничного значения c функции y. Здесь c измеряет начальное состояние системы, а a определяет продолжительность процесса.

Введем теперь функцию f(a, c) = $\min_y I(y)$

Итак, мы включили предложенную выше частную задачу, где a и c - постоянные, в семейство задач, возникающих, когда параметры a и c меняются в областях $-\infty < a < b$ и $-\infty < c < \infty$.

Начнем с вывода уравнения для функции $f(a,c)$, а затем покажем, как это уравнение приводит к некоторым дальнейшим результатам.

В силу аддитивности интеграла $\int\limits_a^b = \int\limits_a^{a+\Delta} + \int\limits_{a+\Delta}^b$

из принципа оптимальности немедленно получаем функциональное уравнение f(a, c) = $\min_{y[a,a+\Delta]}[\int\limits_a^{a+\Delta} F(x,y,y')dx + f(a+\Delta, c(y))]$

где минимизация производится по всем функциям y, определенным на промежутке $a \leq x \leq a+\Delta$ и удовлетворяющим условиям $y(a) = c$ и $c(y) = y(a+\Delta)$.

Мы используем это соотношение двояким образом: в аналитических рассуждениях - устремляя Δ к 0, а при численных расчетах - взяв Δ малым, но отличным от нуля.

2 Уравнение Беллмана

Принцип Беллмана дает достаточные условия оптимальности процесса в задаче оптимального управления. Он базируется на следующем ключевом факте:

Если кривая $x^*(t)$ является оптимальной траекторией в задаче управления динамической системой на отрезке времени $[t_0, T]$, с некоторым начальным условием $x(t_0) = x_0$, то для любого момента времени $\tau \in [t_0, T]$ оптимальным решением задачи управления системой на отрезке времени $[\tau, T]$ с начальным условием $x(\tau) = x^*(\tau)$ будет являться участок той же самой траектории $x^*(t)$.

Рассмотрим задачу оптимального управления в виде:

$$J(x(\cdot), u(\cdot)) = \int_{t_0}^{t_1} F(t, x(t), u(t))dt + \Phi_0(t_1, x(t_1)) \to \max \tag{2.1}$$

$$\dot{x}(t) = f(t, x(t), u(t)), \quad x(t_0) = x_0 \tag{2.2}$$

$$u(t) \in U_t \tag{2.3}$$

и пусть J^* - значение функционала на оптимальном ее решении $x^*(t), u^*(t)$.

Теперь для произвольного момента времени $\tau \in [t_0, T]$ и произвольной точки фазового пространства y положим в задаче (2.1) - (2.3) $t_0 = \tau, x(\tau) = y$. Функцию $J^*(\tau, y)$, равную значению функционала на оптимальном решении такой задачи, будем называть *функцией Беллмана* или *функцией выигрыша.*

Отметим, что $J^* = J^*(t_0, x_0)$.

Исследуем теперь изменение функции $J^*(t, x)$ с течением времени вдоль оптимальной траектории системы, то есть, при $x = x^*(t)$.

Рассмотрим малое приращение времени dt. За это время система перейдет в новое состояние

$$x^*(t+dt) \approx x^*(t) + dx^*(t),$$

где, из (2.2),

$$dx^*(t) = f(t, x^*(t), u^*(t))dt$$

Изменение значения функционала (2.1) на отрезке $[t, t+dt]$, может происходить только за счет интегральной его части и приближенно составляет

$$\int_t^{t+dt} F(t, x^*(t), u^*(t))dt \approx F(t, x^*(t), u^*(t))dt,$$

а оставшаяся часть, согласно принципу оптимальности Беллмана, будет равна $J^*(t+dt, x^*(t+dt))$. Таким образом, получено следующее рекуррентное соотношение:

$$J^*(t, x^*(t)) \approx F(t, x^*(t), u^*(t))dt + J^*(t+dt, x^*(t+dt)). \tag{2.4}$$

Теперь, пользуясь оптимальностью $u^*(t)$, можем переписать (2.4) следующим образом:

$$J^*(t, x(t)) \approx \max_{u(t)\in U_t} \{F(t, x(t), u(t))dt + J^*(t+dt, x(t+dt))\} \tag{2.5}$$

Далее, в предположении дифференцируемости $J^*(t, x)$ по своим аргументам, переходя к пределу при $dt \to 0$ и учитывая (2.2), получим следующее соотношение:

$$-\frac{\partial J^*(t, x)}{\partial t} = \max_{u(t)\in U_t} \{F(t, x(t), u(t)) + \frac{\partial J^*(t, x)}{\partial x} f(t, x(t), u(t))\} \tag{2.6}$$

Соотношение (2.6) представляет собой дифференциальное уравнение в частных производных первого порядка для определения функции $J^*(t, x)$. Оно называется *уравнением Беллмана в дифференциальной форме*.

Краевым условием для данного уравнения является оптимальное значение функционала при $t = t_1$, равное терминальному члену:

$$J^*(t_1, x(t_1)) = \Phi_0(t_1, x(t_1)). \tag{2.7}$$

Как правило, аналитическое решение уравнения (2.6) найти довольно сложно или вовсе невозможно. Поэтому прибегают к дискретизации задачи (2.1) - (2.3) с последующим ее численным решением. Дискретная задача формулируется следующим образом:

$$J(x(\cdot), u(\cdot)) = \sum_{i=0}^{N-1} F(t_i, x_i, u_i)\Delta t_i + \Phi_0(x_N) \to \max \tag{2.8}$$

$$x_{i+1} = f(x_i, u_i), x_0 \quad - \text{задано} \tag{2.9}$$

$$u_i \in U_i, \tag{2.10}$$

Отметим, что в дискретной задаче состояние системы будет описываться вектором $x = (x_0, x_1, \ldots, x_N) \in \mathbf{R}^{N+1}$, а управление - вектором $u = (u_0, u_1, \ldots, u_{N-1}) \in \mathbf{R}^N$.

Для (2.8) - (2.10) уравнение Беллмана будет иметь следующий вид:

$$J_i^*(x_i) = \max_{u_i \in U_i}\{F(t_i, x_i, u_i)\Delta t_i + J_{i+1}^*(f(x_i, u_i))\}, \tag{2.11}$$

с краевым условием

$$J_N^*(x_N) = \Phi_0(x_N)$$

Решение задачи (2.11) при заданных краевых условиях производится последовательным решением уравнения (2.11) для шагов $i = N - 1, N - 2, \ldots, 0$ (обратный ход метода Беллмана). При этом на каждом шаге получается оптимальное управление u_i^* как функция от текущего состояния системы x_i.

На втором этапе по полученным функциям $u_i^*(x_i)$ производится *синтез оптимального управления* для задачи с конкретным начальным условием x_0.

Таким образом, метод динамического программирования, в отличие от рассмотренных выше необходимых условий, дававших оптимальное управление как функцию времени $u^*(t)$(*программное управление*), позволяет определять оптимальное управление как функцию состояния системы $u^*(t, x)$(*синтезированное управление*), что дает возможность отыскивать решение сразу для целого класса задач с различными начальными условиями.

Далее будем считать, что в функционал задачи время не входит явно. Положим шаг Δt_i равным 1. Введем понятие *горизонта планирования* как количества шагов, оставшихся до завершения управления. Обозначим

$$V_k(x) = J^*_{N-k}(x),$$

т. е. максимальный выигрыш, который можно получить за k шагов, если начать из состояния x. В этом случае рекуррентное соотношение для $V_k(x)$ принимает вид:

$$V_k(x) = \max_{u \in U}\{F(x, u) + V_{k-1}(f(x, u))\}, \tag{2.12}$$

с краевым условием: $V_0(x) = \Phi_0(x)$.

3 Задача распределения ресурса

Имеется некоторый ресурс в объеме $a > 0$, который необходимо распределить между N агентами, так, чтобы максимизировать их суммарную полезность, если функция полезности i-го агента

$$F_i(u_i) = lnu_i,$$

где u_i - объем ресурса, получаемый i-м агентом. (Считаем, что агенты как-то перенумерованы.) Решение. В формальной постановке задача имеет вид:

$$J(u) = \sum_{i=1}^{N} lnu_i \to \max; \tag{3.1}$$

$$\sum_{i=1}^{N} u_i \leq a; \quad a > 0$$

Приведем ее к задаче оптимального управления. Для этого необходимо выделить переменную, являющуюся аналогом времени (номера шага) в задаче оптимального управления, горизонта планирования, а также параметры состояния и управления в каждый момент времени.

Пусть номером шага в задаче является номер агента i, для которого принимается решение о распределении ресурса. Тогда величина u_i будет являться управлением на i-м шаге. Введем параметр состояния системы x_i как объем ресурса, имеющийся к i-му шагу $(i = \overline{1, N})$. Тогда, из условия задачи получаем

$$x_{i+1} = x_i - u_i; \quad x_1 = a. \tag{3.2}$$

Так как может быть распределено ресурса не более, чем имеется в наличии, то имеет место ограничение на управление

$$0 \leq u_i \leq x_i. \tag{3.3}$$

Таким образом, (3.1) - (3.3) представляет собой задачу оптимального управления в дискретном времени. Решим ее с использованием принципа Беллмана. Обозначим через $V_k(x)$ значение функции выигрыша, когда горизонт планирования равен k, т. е. ресурс x распределяется между k агентами (не важно, что последними, так как все агенты имеют одинаковые функции полезности).

Рассмотрим последний шаг в нашей задаче, который имеет место после того, как ресурс полностью распределен между всеми агентами. Согласно краевому условию функция Беллмана V_0 на этом шаге равна

$$V_0(x) = \Phi_0(x) \equiv 0$$

Рассмотрим теперь ситуацию, когда ресурс должен быть распределен одному агенту. В этом случае горизонт планирования $k = 1$ и рекуррентное соотношение (2.12) принимает вид

$$V_1(x) = \max_{0 \leq u \leq x} lnu + V_0(x-u) = \max_{0 \leq u \leq x} lnu = lnx,$$

откуда $u_N^*(x) = x$.

Аналогично, при горизонте планирования $k = 2$ имеем:

$$V_2(x) = \max_{0 \leq u \leq x} lnu + V_1(x-u) \max_{0 \leq u \leq x} lnu + ln(x-u).$$

Максимум выражения в фигурных скобках по $u \in [0, x]$ достигается при $u^*(x) = \frac{x}{2}$.

Покажем далее, что для горизонта $k = 0, \ldots, N$ оптимальное управление на шаге (N + 1 - k) и функция Беллмана горизонта k имеют вид: u$^*_{N+1-k}(x) = \frac{x}{k}, \quad V_k = kln\frac{x}{k}$ Предположим, что это верно на некотором шаге (N + 1 - k). Определим оптимальное управление и функцию Беллмана горизонта k:

$$V_{k+1}(x) = \max_{0 \leq u \leq x} lnu + V_k(x-u) = \max_{0 \leq u \leq x} lnu + kln\frac{x-u}{k}.$$

Обозначим

$$A(u) = lnu + kln\frac{x-u}{k}.$$

Условия первого порядка максимума функции $A(u_{N-k})$ имеют вид:

$$\frac{dA}{du} = \frac{1}{u} - \frac{k}{x-u} = 0,$$

откуда

$$u^*_{N-k}(x) = \frac{x}{k+1}, \quad V_{k+1}(x) = (k+1)ln\frac{x}{k+1}.$$

Таким образом, определен общий вид оптимального управления для произвольного шага в задаче. Теперь проведем синтез оптимального управления для задачи с N агентами и начальным объемом ресурса, равным a:

$$u^*_1(x_1) = \frac{x_1}{N} = \frac{a}{N}; \quad x_2 = x_1 - u^*_1 = a - \frac{a}{N} = \frac{a(N-1)}{N};$$

$$u^*_2(x_2) = \frac{x_2}{N-1} = \frac{a}{N}; \quad x_3 = x_2 - u^*_2 = \frac{a(N-1)}{n} - \frac{a}{N} = \frac{a(N-2)}{N}$$

$\dots$

$$u^*_k(x_k) = \frac{x_k}{N+1-k} = \frac{a}{N}; \quad x_{k+1} = x_k - u^*_k = \frac{a(N+1-k)}{N} - \frac{a}{N} = \frac{a(N-k)}{N}$$

$\dots$

Таким образом, в данной задаче оптимальным является равномерное распределение ресурса между агентами:

$$u^* = (\frac{a}{N}, \frac{a}{N}, \dots, \frac{a}{N}).$$

Заключение

Метод динамического программирования позволяет найти оптимальное решение в многошаговом процессе принятия решения. Применяется для решения задач оптимального управления определенной структуры.

Преимущество подхода, при котором оптимальное решение в многомерной задаче путем ее декомпозиции на этапы, каждый из которых представляет подзадачу относительно одной переменной, состоит в том, что вместо многомерной задачи на каждом этапе решаются одномерные оптимизационные задачи.

Список используемой литературы

1 Беленький, В. З. Оптимальное управление: принцип максимума и динамическое программирование. М. : РЭШ, 2001 - 114 с.

2 Беллман, Р., Дрейфус, С. Прикладные задачи динамического программирования. М. : Наука, 1965. - 458 с.

3 Кормен, Т. Х., Лейзерсон, Ч. И., Ривест, Р.Л., Штайн, К. Алгоритмы: построение и анализ. — 2-е изд. — М.: Вильямс, 2005. — 1296 с.

4 Афанасьев, В.Н., Колмановский, В.Б., Носов, В.Р. Математическая теория конструирования систем управления. — 3-е изд., испр. и доп. — М.: Высш. шк., 2003. — 614 с.

5 Зайцев, Г. Ф. Теория автоматического управления и регулирования.— 2-е изд., перераб. и доп.— К.: Выща шк. Головное изд-во, 1989.— 431 с.

6 Понтрягин, Л.С., Болтянский, В.Г., Гамкрелидзе, Р.В., Мищенко, Е.Ф. Математическая теория оптимальных процессов. М., Наука, 1969.