Лекция 5.

Стохастическое управление с конечным горизонтом.

Цель этой лекции – описать алгоритм стохастического управления с конечным горизонтом в дискретном времени, называемый также алгоритмом динамического программирования. Начнём с постановки задачи управления ***детерминированными*** системами, потом рассмотрим ***стохастические*** системы. Затем опишем алгоритм и рассмотрим ряд примеров.

Динамические системы с дискретным временем.

Рассмотрим «систему», эволюционирующую во времени. Время предполагается дискретным $n = 0,1,2,\ldots$ Состояние системы в момент n представлено элементом x_n некоторого пространства E. Мы будем говорить, что эта система является ***детерминированной динамической системой,*** если для каждого $n \in N$ найдётся функция $f_n: E \to E$ такая, что

$$x_{n+1} = f_n(x_n).$$

Знания состояния x_n в некоторый момент достаточно для того, чтобы описать его будущую эволюцию (x_{n+1}). Например, так можно моделировать простую механическую систему, где её состоянием является двумерный вектор «положение-скорость».

Представим себе теперь, что в каждый момент ***можно воздействовать на систему.*** Её эволюция тогда может быть описана уравнением

$$x_{n+1} = f_n(x_n, u_n),$$

где u_n, со значениями в некотором множестве C, является ***управляющим параметром,*** который можно выбрать в момент n, чтобы изменить (модифицировать) эволюцию системы. В этом случае говорят, что имеют дело с ***детерминированной управляемой динамической системой.*** Например, если x_n описывает состояние ракеты, то u_n может быть количеством топлива, которое необходимо подавать в двигатель в момент n.

Рассмотрим теперь ***случайные динамические системы.***

Рассмотрим систему, подверженную неконтролируемым ***случайным воздействиям*** $\{\varepsilon_n(\omega), n \in N\}$, определённым на вероятностном пространстве $(\Omega, \mathcal{F}, P)$ и со значениями в некотором множестве W, снабжённом σ- алгеброй $\mathcal{W}$. Начнём с рассмотрения динамической системы без управления. Из-за случайных воздействий состояние системы является случайной величиной, которую мы будем обозначать X_n. Эта с.в. принимает значения в пространстве E, снабжённом σ- алгеброй $\mathcal{E}$.

Уравнение, описывающее эволюцию системы, теперь имеет вид

$$X_{n+1} = \varphi_n(X_n, \varepsilon_n), \forall n \in N, \qquad (1)$$

где отображение $\varphi_n: E \times W \to E$ предполагается измеримым. Предположим также, что с.в. $\{\varepsilon_n, n \in N\}$ ***независимы между собой, имеют один и тот же закон распределения μ на $(W, \mathcal{W})$ и независимы от начального состояния X_0.*** Тогда говорят, что

задана ***случайная динамическая система.*** Положим для $x \in E$ и $A \in \mathcal{E}$

$$P_n(x, A) := P(X_{n+1} \in A | X_n = x) = P(\varphi_n(x, \varepsilon_n(\omega)) \in A) = \mu(w \in W;\ \varphi_n(x, w) \in A) \quad (2)$$

Определение 1. Назовём ***переходной вероятностью или переходным ядром на*** $(\boldsymbol{E}, \boldsymbol{\mathcal{E}})$ семейство

$$P(x, A),\ x \in E, A \in \mathcal{E}$$

такое, что

- При любом фиксированном $x \in E$, $A \mapsto P(x, A)$ – вероятностная мера на $(E, \mathcal{E})$.
- При любом фиксированном $A \in \mathcal{E}$, функция $x \mapsto P(x, A)$ измеримая функция x.

Таким образом, для каждого n семейство P_n, введённое в (2), является переходной вероятностью.

Определение 2. Пусть задано семейство переходных вероятностей $\{P_n(x, A)\}$ на $(E, \mathcal{E})$. Последовательность случайных величин $X_n, n \geq 0$, со значениями в E называется ***неоднородной цепью Маркова с переходным ядром*** $\boldsymbol{P_n(x, A)}$**,** если для всех A $\in \mathcal{E}$ и $n \in$ N

$$E\left(1_A(X_{n+1}) \middle| \sigma(X_0, \dots, X_n)\right) = P\left(X_{n+1} \in A \middle| \sigma(X_0, \dots, X_n)\right) = P_n(X_n, A).$$

Если $P_n(x, A)$ не зависит от n, то говорят, что X_n является ***однородной цепью Маркова*** (или просто цепью Маркова).

Предложение 1. Последовательность $\{X_n, n \in N\}$, определённая в (1), является неоднородной цепью Маркова с переходными вероятностями P_n.

Доказательство. Заметим, что X_n является функцией $X_0, \varepsilon_0, \dots, \varepsilon_{n-1}$. Это легко показать индукцией по n: для $n = 0$ это очевидно, и если $X_n = f_n(X_0, \varepsilon_0, \dots, \varepsilon_{n-1})$, то

$$X_{n+1} = \varphi_n(X_n, \varepsilon_n) = \varphi_n(f_n(X_0, \varepsilon_0, \dots, \varepsilon_{n-1}), \varepsilon_n).$$

Отсюда следует, что ε_n не зависит от σ- алгебры, порождённой $X_0, \dots, X_n$, то есть от $\sigma(X_0, \dots, X_n)$. Сама с.в. X_n, очевидно, измерима относительно $\sigma(X_0, \dots, X_n)$. Используя известное свойство условных математических ожиданий (мы его доказывали для с.в., принимающих конечное или счётное множество значений и для с.в., имеющих плотность, но оно верно и в общем случае), получим

$$E\big(1_A(X_{n+1})\big|\sigma(X_0, \dots, X_n)\big)(\omega) =$$

$$E\left(1_A\big(\varphi_n(X_n, \varepsilon_n)\big)\middle|\sigma(X_0, \dots, X_n)\right)(\omega)$$

$$E\left(1_A\big(\varphi_n(x, \varepsilon_n)\big)\right)|_{x=X_n(\omega)} = P(\varphi_n(x, \varepsilon_n) \in A)|_{x=X_n(\omega)} =$$

$$P_n(X_n(\omega), A).$$

Последнее равенство вытекает из (2). Предложение тем самым доказано.

На самом деле можно доказать и обратное: ***любая*** неоднородная цепь Маркова задаётся некоторой стохастической динамической системой, но мы это не будем использовать и доказывать это не будем.

Введём важное обозначение: если P переходная вероятность на E и $f: E \to \boldsymbol{R}^+$ измерима, определим **оператор** $\mathbb{P}$ следующим образом

$$\mathbb{P}f(x) = \int_E f(y)P(x, dy).$$

Напомним несколько фактов из анализа, необходимых для дальнейшего.

Теорема Леви о монотонной сходимости.

Следующая теорема позволяет совершать предельный переход под знаком интеграла.

Теорема. *Пусть последовательность измеримых функций является почти наверное неубывающей, то есть*

$$f_1(\omega) \le f_2(\omega) \le \cdots \le f_n(\omega) \le \cdots$$

почти наверное. Положим

$$f(\omega) = \lim_{n\to\infty} f_n(\omega).$$

Тогда

$$\lim_{n\to\infty} \int_{\Omega} f_n \, d\mu = \int_{\Omega} \lim_{n\to\infty} f_n(\omega) d\mu = \int_{\Omega} f d\mu.$$

(здесь интеграл в правой части может быть и бесконечным).

Предложение 2. Если $X_n, n \in N$, неоднородная цепь Маркова для заданного семейства переходных вероятностей $P_n(x, A)$, то для любой измеримой неотрицательной $f: E \to R^+$

$$E\big(f(X_{n+1})\big|\sigma(X_0, \dots, X_n)\big) = \int_E f(y) P_n(X_n, dy) = \mathbb{P}_n f(X_n).$$

Доказательство. Для $f = 1_A$, $A \in \mathcal{E}$, это просто определение $P_n(X_n, A)$ (см. Определение 2). По линейности это соотношение верно для любой ступенчатой функции $f = \sum_{k=1}^{n} a_k 1_{A_k}$. Затем, используя теорему Леви о монотонной сходимости, докажем Предложение 2 для любой измеримой ***положительной*** функции (так как она является пределом возрастающей последовательности

ступенчатых функций). Переход к измеримым функциям f любого знака следует из представления $f = f^+ - f^-$.

Случайные динамические системы с управлением.

Рассмотрим теперь ***динамическую систему с управлением, подверженную случайным воздействиям*** $\{\varepsilon_n, n \in N\}$. Динамика такой системы описывается соотношением

$$X_{n+1} = \varphi_n(X_n, U_n, \varepsilon_n), \forall n \in N, \qquad (3)$$

где отображение

$$\varphi_n : E \times C \times W \to E$$

предполагается измеримым, U_n - управление (возможно, являющееся случайной величиной) со значениями в $(C, \mathcal{C})$. Интерпретируя индекс n как время, мы требуем, чтобы управление в момент n, т.е. U_n, ***было измеримой функцией «предыстории»*** $(X_0, \dots, X_n)$, то есть выбор управления U_n делается в момент n ***и не опирается на знание будущего.***

Определение 3. Говорят, что $\{U_n, n \in N\}$ является ***последовательностью управлений***, если для любого $n \geq 0$, U_n является $\sigma(X_0, \dots, X_n)$ – измеримой, или, что эквивалентно (в силу Леммы Дуба-Дынкина), если существует последовательность измеримых функций (преобразований) $\nu_n : E^{n+1} \to C$ таких, что

$$U_n = \nu_n(X_0, \dots, X_n).$$

Последовательность преобразований $(\nu_n), n = 0,1,2, \dots$ называется ***стратегией*** или ***политикой.***

Резюмируем: В момент n нам доступно состояние X_n. В зависимости от этого состояния и от предыдущих ему состояний можно выбирать управление U_n, которое будет нам помогать

реализовать нашу цель. Напротив, случайное воздействие ε_n не является X_n - измеримым: оно неизвестно в момент n, и это воздействие, которое будет, скорее всего, нам мешать.

Рассмотрим теперь вычисления, аналогичные уже проделанным ранее для случайной динамической системы в отсутствие управления.

Предложение 3. Положим $P_n^{(u)}(x, A) := P(\varphi_n(x, u, \varepsilon_n) \in A)$ для $x \in E, u \in C,\ A \in \mathcal{E}\ u\ n \geq 0$. Тогда

$$E\big(1_A(X_{n+1})\big|\sigma(X_0, \dots, X_n)\big)(\omega) = P_n^{(U_n)}(X_n(\omega), A).$$

Доказательство. Полностью аналогично доказательству Предложения 1. В самом деле, применяя рекуррентно соотношение (3), получим, что случайный вектор (X_n, U_n) является $\sigma(X_0, \varepsilon_0, \dots, \varepsilon_{n-1})$ – измеримым (мы здесь используем то, что $U_n = \nu_n(X_0, \dots, X_n)$). Вектор (X_n, U_n) измерим относительно $\sigma(X_0, \dots, X_n)$ (управление U_n измеримо относительно этой σ - алгебры по определению), тогда как ε_n не зависит от этой σ - алгебры. Снова пользуясь тем же свойством условных м.о., что и в Предложении 1, получим

$$E\big(1_A(X_{n+1})\big|\sigma(X_0, \dots, X_n)\big)(\omega) =$$

$$E\left(1_A\big(\varphi_n(X_n, U_n, \varepsilon_n)\big)\middle|\sigma(X_0, \dots, X_n)\right)(\omega) =$$

$$E(1_A\big(\varphi_n(x, u, \varepsilon_n)\big)|_{x=X_n(\omega), u=U_n(\omega)} =$$

$$P(\varphi_n(x, u, \varepsilon_n) \in A)|_{x=X_n(\omega), u=U_n(\omega)} =$$

$$P_n^{U_n}(X_n(\omega), A).$$

Дадим теперь общее определение.

Определение 4. Назовём ***марковской моделью с управлением*** семейство

$$P_n^{(u)}(x, A), n \in N, x \in E, u \in C, A \in \mathcal{E}$$

переходных вероятностей на E таких, что для фиксированного $A \in \mathcal{E}$ функция $(x, u) \mapsto P_n^{(u)}(x, A)$ является измеримой.

Если $P_n^{(u)}(x, A)$ не зависит от n, мы будем писать $P^{(u)}(x, A)$.

Если задана марковская модель с управлением (в смысле Определения 4), то с каждой стратегией $\nu = (\nu_0, \nu_1, \dots)$ можно связать процесс X_n следующим образом: X_0 выберем произвольным образом. Далее, для $n = 1,2, \dots$ будем рассуждать по индукции. При фиксированной траектории

$$(X_0, \dots, X_n)$$

случайная величина X_{n+1} будет иметь распределение $P_n^{(U_n)}(X_n, \cdot)$, где $U_n = \nu_n(X_0, \dots, X_n)$, что означает следующее

$$E\big(1_A(X_{n+1})\big|\sigma(X_0, \dots, X_n)\big)(\omega) = P_n^{(U_n)}(X_n(\omega), A)$$

для всех $A \in \mathcal{E}$. Определение корректно, поскольку U_n зависит только от $X_0, \dots, X_n$. Обозначим через P_x^ν (чтобы явно подчеркнуть зависимость от стратегии (ν)) закон распределения построенного процесса при условии, что $X_0 = x$ с вероятностью 1, т.е. мы с вероятностью 1 начинаем из точки x. Для математического ожидания будем использовать обозначение E_x^ν.

Аналогично Предложению 2 доказывается следующая Лемма.

Лемма 1. Для любой ограниченной измеримой функции $f: E \to \boldsymbol{R}^+$

$$E_x^\nu\big(f(X_{n+1})\big|\sigma(X_0, \dots, X_n)\big) = \int f(y) P_n^{U_n}(X_n, dy) = \mathbb{P}_n^{(U_n)} f(X_n).$$

Лемма легко обобщается на измеримые ограниченные функции f, принимающие значения любого знака, если воспользоваться представлением $f = f^+ - f^-$.

Как показывает доказываемое ниже Предложение 4, эти свойства полностью определяют закон распределения случайного процесса X_n. Для доказательства нам понадобится теорема о монотонном классе функций.

Теорема о монотонном классе функций.

Теорема. *Пусть $\mathcal{K}$ - совокупность ограниченных измеримых вещественнозначных функций на Ω, замкнутая относительно произведений (т.е. если $f, g \in \mathcal{K}$, то $f \cdot g \in \mathcal{K}$) и пусть $\mathcal{B}$ -σ-алгебра, порождённая функциями из $\mathcal{K}$. Пусть $\mathcal{H} \supset \mathcal{K}$ - векторное пространство (над R) ограниченных измеримых вещественнозначных функций на Ω,* ***содержащее константы и замкнутое относительно монотонных пределов,*** *то есть, если $(f_n) \subset \mathcal{H}$, $sup_n sup_\omega |f_n(\omega)| < +\infty$ и если*

$$0 \le f_1 \le \cdots \le f_n \le \cdots$$

то $f := lim_n f_n \in \mathcal{H}$. Тогда $\mathcal{H}$ содержит ***все*** *ограниченные $\mathcal{B}$ - измеримые вещественнозначные функции на Ω.*

Замечание. Можно показать, что в условиях Теоремы векторное пространство $\mathcal{H}$ замкнуто относительно равномерной сходимости. Используя теорему о монотонном классе, докажем следующее предложение.

Предложение 4. *Если $\Phi: E^n \to \boldsymbol{R}^+$ ограничена и измерима, то*

$$E_x^v\big(\Phi(X_1, X_2, \dots, X_n)\big) =$$

$$\int_E \int_E \dots \int_E \Phi(x_1, x_2, \dots, x_n) P_0^{\nu_0(x)}(x, dx_1) P_1^{\nu_1(x,x_1)}(x_1, dx_2) \dots$$

$$\dots P_{n-1}^{\nu_{n-1}(x,\dots,x_{n-1})}(x_{n-1}, dx_n). \qquad (4)$$

Замечание. Из Предложения 4 легко найти совместный закон распределения вектора $(X_1, X_2, \dots, X_n)$ беря в качестве $\Phi(x_1, x_2, \dots, x_n)$ индикатор множества $A \subset R^n$:

$$P_x\big((X_1, X_2, \dots, X_n) \in A\big) =$$

$$\int_A P_0^{\nu_0(x)}(x, dx_1)\, P_1^{\nu_1(x,x_1)}(x_1, dx_2) \dots P_{n-1}^{\nu_{n-1}(x,\dots,x_{n-1})}(x_{n-1}, dx_n).$$

Доказательство Предложения 4. Доказательство проводится индукцией по числу переменных n. При $n = 1$ утверждение сводится к Лемме 1. Рассмотрим в качестве $\mathcal{H}$ совокупность тех Φ: $E^n \to R^+$, для которых равенство (4) имеет место, это векторное пространство функций, содержащее константы, и эта совокупность замкнута относительно монотонных пределов в силу теоремы Леви о монотонной сходимости. В качестве совокупности $\mathcal{K}$ можно взять произведения $\Phi(x_1, x_2, \dots, x_n) = f_1(x_1) \cdot \dots \cdot f_n(x_n)$. Очевидно, что такие $\Phi(x_1, x_2, \dots, x_n)$ замкнуты относительно произведений:

$$f_1(x_1) \cdot \dots \cdot f_n(x_n) \cdot g_1(x_1) \cdot \dots \cdot g_n(x_n) =$$
$$\big(f_1(x_1) \cdot g_1(x_1)\big) \cdot \dots \cdot \big(f_n(x_n) \cdot g_n(x_n)\big).$$

Из теоремы о монотонном классе функций тогда следует, что теорему достаточно доказать для таких произведений, то есть для случая, когда Ф является произведением

$$\Phi(X_1, X_2, \dots, X_n) = f_1(X_1) \cdot \dots \cdot f_n(X_n).$$

Это будет означать, что $\mathcal{K} \subset \mathcal{H}$. Мы знаем (Лемма 1), что

$$E_x^{\nu}(f_n(X_n)\big|\sigma(X_0, \dots, X_{n-1})\big) = \mathbb{P}_{n-1}^{(U_{n-1})} f_n(X_{n-1}).$$

Поэтому, пользуясь свойствами условного математического ожидания, получим:

$$E_x^{\nu}\big(\Phi(X_1, X_2, \dots, X_n)\big) = E_x^{\nu}[f_1(X_1) \cdot \dots \cdot f_n(X_n)] =$$

$$E_x^{\nu}\{E_x^{\nu}[f_1(X_1) \cdot \dots \cdot f_n(X_n)|\sigma(X_0, \dots, X_{n-1})]\} =$$

$$E_x^{\nu}[f_1(X_1) \cdot \dots \cdot f_{n-1}(X_{n-1}) E_x^{\nu}(f_n(X_n)|\sigma(X_0, \dots, X_{n-1}))] =$$

$$E_x^{\nu}\left[f_1(X_1) \cdot \dots \cdot f_{n-1}(X_{n-1}) \mathbb{P}_{n-1}^{(U_{n-1})} f_n(X_{n-1})\right].$$

Предположим, что Предложение 4 верно для любой функции от числа переменных $\leq n-1$, применим её к ограниченной измеримой функции $(n-1)$ -ой переменной следующего вида

$$g(x_1, \dots, x_{n-1}) = f_1(x_1) \cdot \dots \cdot f_{n-1}(x_{n-1}) \mathbb{P}_{n-1}^{\nu_{n-1}(x, x_1, \dots, x_{n-1})} f_n(x_{n-1}).$$

Получим

$$E_x^{\nu}\big(\Phi(X_1, X_2, \dots, X_n)\big) = E_x^{\nu}\big(g(X_1, \dots X_{n-1})\big) =$$

$$\int \dots \int [f_1(x_1) \cdot \dots \cdot f_{n-1}(x_{n-1}) \cdot$$

$$\cdot\ \mathbb{P}_{n-1}^{\nu_{n-1}(x,\dots,x_{n-1})} f_n(x_{n-1})\Big] P_0^{\nu_0(x)}(x, dx_1) P_1^{\nu_1(x,x_1)}(x_1, dx_2)\ \dots$$

$$\times P_{n-2}^{\nu_{n-2}(x,\dots,x_{n-2})}(x_{n-2}, dx_{n-1}),$$

откуда получаем желаемое равенство, поскольку, по определению оператора $\mathbb{P}$

$$\mathbb{P}_{n-1}^{\nu_{n-1}(x,\dots,x_{n-1})} f_n(x_{n-1}) = \int f_n(x_n)\, P_{n-1}^{\nu_{n-1}(x,\dots,x_{n-1})}(x_{n-1}, dx_n).$$

Таким образом, (4) верно для всех функций, являющихся произведениями $f_1(X_1) \cdot \dots \cdot f_n(X_n)$, а, следовательно, по теореме о монотонном классе, и для всех ограниченных измеримых функций от n переменных. Предложение 4 доказано.

Если пространство E счетное, то для каждого переходного ядра P положим

$$P(x, y) := P(x, \{y\}), x, y \in E.$$

Совместный закон распределения (дискретный) вектора $(X_1, X_2, \dots, X_n)$ имеет следующий простой вид.

Лемма 2. Если E счётно, то

$$P_x^{(\nu)}(X_0 = x, X_1 = x_1, \dots, X_n = x_n) =$$

$$P_0^{(u_0)}(x, x_1) P_1^{(u_1)}(x_1, x_2) \dots P_{n-1}^{(u_{n-1})}(x_{n-1}, x_n),$$

где $u_0, u_1, \dots, u_{n-1}$ значения управлений $U_0, U_1, \dots, U_{n-1}$ когда $X_0 = x, X_1 = x_1, \dots, X_{n-1} = x_{n-1}$.

Марковские стратегии.

По определению, последовательность управлений $\{U_n\}$ такова, что каждая U_n является измеримой функцией $X_0, \dots, X_n$. Мы увидим, что часто можно ограничиться случаем, когда U_n является функцией только X_n, то есть $U_n = \nu_n(X_n)$, где $\nu_n: E \to C$.

Определение 5. Назовём ***марковской стратегией*** такую последовательность управлений $\{U_n\}$, что каждая U_n является измеримой функцией только от X_n: $U_n = \nu_n(X_n)$, где $\nu_n: E \to C$. Если к тому же ν_n не зависит от n, то стратегию называют ***стационарной марковской***.

Следующее предложение обобщает Предложение 1 на случай систем с управлением.

Предложение 5. Если последовательность $\{U_n\}$ управлений соответствует марковской стратегии ν_n, то последовательность X_n, определённая в (3), является неоднородной цепью Маркова с переходными вероятностями

$$P_n(x, A) = P_n^{(\nu_n(x))}(x, A)$$

Если ни ν_n, ни $P_n^{(u)}(x, A)$ не зависят от n, то X_n является однородной цепью Маркова.

В дальнейшем для краткости будем использовать обозначение: $P_n^{\nu}(x, A) = P_n^{(\nu(x))}(x, A)$.

Динамическое программирование.

Рассмотрим управляемую ***марковскую*** модель (см. Определение 4). Число N, если оно фиксировано, называют иногда ***горизонтом.*** ***Требуется найти последовательность управлений таким***

образом, чтобы минимизировать функцию потерь (издержек). Эта функция имеет вид

$$E\left(\sum_{k=0}^{N-1} c_k(X_k, U_k) + \gamma(X_N)\right),$$

где $c_k: E \times C \to R \cup \{+\infty\}, \gamma: E \to R \cup \{+\infty\}$. Заметим, что на последнем, N-ом шаге, издержки зависят только от состояния X_N и не зависят от управления, поскольку на последнем шаге управление не применяется. Мы увидим, что эта задача имеет решение в виде алгоритма, который легко реализовать на компьютере. Это было сделано Р. Беллманом в 1953 г. Этот алгоритм имеет различные названия: ***алгоритм динамического программирования, алгоритм обратной рекурсии, принцип оптимальности Беллмана.***

Положим

$$J(x) = \min_\nu E_x^\nu \left(\sum_{k=0}^{N-1} c_k(X_k, U_k) + \gamma(X_N)\right),$$

где минимум берётся по всем стратегиям $\nu = (\nu_0, \nu_1, \dots, \nu_{N-1})$.

Теорема 1. (алгоритм обратной рекурсии). Положим $J_N(x) = \gamma(x)$, затем найдём последовательно ***обратным ходом*** для $n = N - 1, N - 2, \dots, 1, 0,$

$$J_n(x) = \min_{u \in C} \left\{c_n(x, u) + \mathbb{P}_n^{(u)} J_{n+1}(x)\right\}. \qquad (5)$$

*(*напомним, что $\mathbb{P}_n^{(u)} f(x) = \int f(y)\, P_n^{(u)}(x, dy)$). Предположим, что все интегралы существуют и что при каждом n минимум по u достигается в некоторой точке $\nu_n(x)$. Тогда ***на последнем этапе*** мы получим искомый минимум, т.е. $J_0(x) = J(x)$, *а* $U_n =$

$\nu_n(X_n), n = 0, \dots, N-1,$ дают соответствующую *оптимальную* ***марковскую*** *стратегию.*

Доказательство. Рассмотрим произвольное управление ν. Для всех $n < N$ имеем с учётом Леммы 1

$$E_x^{\nu}\left((J_n(X_n) - J_{n+1}(X_{n+1}))\middle|\sigma(X_0, \dots, X_n)\right) = J_n(X_n) - \mathbb{P}_n^{(U_n)} J_{n+1}(X_n).$$

Используя свойства условного математического ожидания, и то, что $J_0(x) = E_x^{\nu} J_0(X_0)$ (напомним, что процесс с вероятностью 1 начинается в точке x, т. е. $X_0 = x$), получим

$$J_0(x) = E_x^{\nu}\left[\sum_{n=0}^{N-1}\left(J_n(X_n) - J_{n+1}(X_{n+1})\right)\right] + E_x^{\nu}[J_N(X_N)] =$$

$$E_x^{\nu}\left[E_x^{\nu}\left[\sum_{n=0}^{N-1}\left(J_n(X_n) - J_{n+1}(X_{n+1})\right)|\sigma(X_0, \dots, X_n)\right]\right] + E_x^{\nu}[J_N(X_N)]$$

$$= E_x^{\nu}\left[\sum_{n=0}^{N-1}(J_n(X_n) - \mathbb{P}_n^{(U_n)} J_{n+1}(X_n))\right] + E_x^{\nu}[\gamma(X_N)]$$

$$\leq E_x^{\nu}\left[\sum_{n=0}^{N-1} c_n(X_n, U_n) + \gamma(X_N)\right],$$

поскольку, согласно (5), $J_n(x) - \mathbb{P}_n^{(u)} J_{n+1}(x) \leq c_n(x, u)$. Равенство достигается, тогда и только тогда, когда $U_n = \nu_n(X_n)$, для всех $n = 0, \dots, N-1$. Теорема доказана.

Интерпретируя этот алгоритм, говорят, что в момент n индивидуум минимизирует сумму своих издержек в этот момент и то, что он может знать в этот момент о своих будущих издержках. Заметим, что в этом алгоритме ***оптимальное управление является***

марковской стратегией (U_n зависит только от X_n). Заменяя везде в доказательстве min на max, немедленно получим

Теорема 2. (алгоритм динамического программирования для максимума).

Для $N-1, N-2, \dots, 1, 0$, последовательно находим

$$J_n(x) = max_{u \in C}\left\{c_n(x,u) + \mathbb{P}_n^{(u)} J_{n+1}(x)\right\}.$$

(напомним, что $\mathbb{P}_n^{(u)} f(x) = \int f(y)\, P_n^{(u)}(x, dy)$). Предположим, что все интегралы существуют и что максимум по u достигается в некоторой точке $\nu_n(x)$. Тогда

$$J_0(x) = \max_{\nu} E_x^{\nu}\left(\sum_{k=0}^{N-1} c_k(X_k, U_k) + \gamma(X_N)\right),$$

а $U_n = \nu_n(X_n), n = 0, \dots, N-1$, соответствующая оптимальная марковская стратегия.

Детерминированный случай.

Алгоритм динамического программирования даёт интересный результат уже для детерминированной модели, то есть для модели $x_{n+1} = f_n(x_n, u_n)$ (нет случайных воздействий ε_n). Его используют для дискретных моделей, когда нет гладкости, в случае гладких моделей можно использовать и другие методы оптимизации (вариационное исчисление, градиентный метод).

Рассмотрим детерминированную схему с управлением

$$x_{n+1} = f_n(x_n, u_n), n = 0,1, \dots, N-1, x_n \in R^d, u_n \in U \subseteq R^r.$$

Такие системы возникают в экономике, когда планирование идёт по годам, месяцам и.т.д. Функционал качества определяется так:

$$J = \sum_{k=0}^{N-1} c_k(x_k, u_k) + \gamma(x_N).$$

В экономических терминах c_k - затраты в состоянии x_k при реализации управления x_k. Заданы начальные данные $x_0 \in R^d$. Задача ставится следующим образом:

$$J \to \min$$

Прямой путь решения задачи следующий:

$$u_0 \to x_1 = x_1(x_0, u_0) \to^{u_1} x_2 = x_2(x_0, u_0, u_1) \to \cdots$$

Если подставить, то получим выражение:

$$J = J(x_0; u_0, u_1, \dots, u_{N-1}).$$

То есть ***функционал качества выражен как функция от начального условия и от всех управлений на каждом шаге с нулевого до $N-1$ - го***. Получили задачу оптимизации для функции J от $N \times r$ переменных (напомним, что $u_n \in U \subseteq R^r$). Но, как правило, в прикладных задачах N велико. Следовательно, получили задачу большой размерности. Поэтому прямой путь неприемлем. Метод дискретного динамического программирования сводит задачу к ***последовательности задач оптимизации размерности r***. Будем использовать принцип оптимальности Беллмана. Смысл принципа оптимальности: ***остаток оптимальной траектории оптимален.*** Принцип Беллмана справедлив не всегда, но для аддитивных функционалов (вроде нашей функции издержек) он справедлив.

Начнём процедуру ***попятного (обратного) движения***. Пусть мы находимся в предпоследней точке x_{N-1}, тогда

$$J_{N-1} := c_{N-1}(x_{N-1}, u_{N-1}) + \gamma(x_N) =$$

$$c_{N-1}(x_{N-1}, u_{N-1}) + \gamma(f_{N-1}(x_{N-1}, u_{N-1})).$$

Поскольку остаток оптимальной траектории оптимален, находим требуемое управление на предпоследнем шаге

$$u_{N-1}: \min_{u_{N-1} \in U} J_{N-1}.$$

Но точка оптимальной траектории x_{N-1} нам пока не известна, поэтому решение будет зависеть от x_{N-1}.

$$u_{N-1} = u_{N-1}(x_{N-1}).$$

Обозначим $\min_{u_{N-1} \in U} J_{N-1} := S_{N-1}(x_{N-1})$.

Следующий шаг. Пусть мы в точке $N-2$. Рассмотрим соответствующий остаток оптимальной траектории:

$$J_{N-2} := c_{N-2}(x_{N-2}, u_{N-2}) + c_{N-1}(x_{N-1}, u_{N-1}) + \gamma(x_N).$$

Находим минимум

$$\min_{u_{N-1} \in U, u_{N-2} \in U} J_{N-2} =$$

$$\min_{u_{N-2} \in U} \left\{ c_{N-2}(x_{N-2}, u_{N-2}) + \min_{u_{N-1} \in U} [c_{N-1}(x_{N-1}, u_{N-1}) + \gamma(x_N)] \right\}.$$

Но мы уже решили задачу минимизации выражения в квадратных скобках. Следовательно,

$$\min_{u_{N-1} \in U, u_{N-2} \in U} J_{N-2} = \min_{u_{N-2} \in U} \{ c_{N-2}(x_{N-2}, u_{N-2}) + S_{N-1}(x_{N-1}) \} =$$

$$\min_{u_{N-2} \in U} \{ c_{N-2}(x_{N-2}, u_{N-2}) + S_{N-1}(f_{N-2}(x_{N-2}, u_{N-2})) \}.$$

Но точка x_{N-2} оптимальной траектории пока неизвестна, поэтому мы можем только найти оптимальное управление на шаге $N-2$ как функцию x_{N-2}:

$$u_{N-2} = u_{N-2}(x_{N-2}).$$

Пусть

$$\min_{u_{N-1}\in U, u_{N-2}\in U} J_{N-2} := S_{N-2}(x_{N-2}).$$

Замечаем закономерность:

$$S_{N-k} = \min_{u_{N-k}\in U}\{\{c_{N-k}(x_{N-k}, u_{N-k}) + S_{N-k+1}(f_{N-k}(x_{N-k}, u_{N-k}))\}\},$$

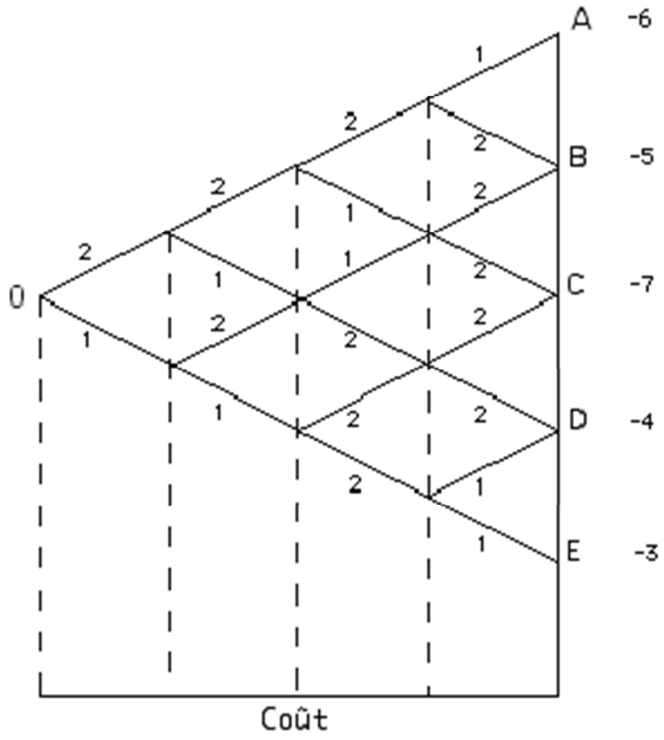
$u_{N-k} = u_{N-k}(x_{N-k})$ – оптимальное управление на шаге $N-k$.

Полученное уравнение называется рекуррентным уравнением Беллмана, S_i - функции Беллмана. Двигаясь таким образом ***из конца в начало,*** на последнем этапе получим $u_0 = u_0(x^0)$. Но x^0 известна по постановке задачи. Начинаем этап ***прямого движения***: по x^0 вычисляем $u_0 = u_0(x^0)$. Далее, вычисляем $x_1 = f_1(x^0, u_0)$ и $u_1 = u_1(x_1)$. Получили, что оптимальное управление на первом шаге u_1 теперь известно. И так далее, узнаем всю оптимальную траекторию $x_2, \dots, x_N$. Ещё раз подчеркнём, что используя алгоритм Беллмана, ***мы решаем N задач минимизации, каждая из которых r - мерна, то есть вместо минимизации функции $N \times r$ переменных мы решаем серию задач минимизации функций r переменных.***

Классический иллюстративный пример применения детерминированного алгоритма – нахождение самого «выгодного» пути до сокровища.

Предположим, что каждый из городов A,B,C,D,E обладает сокровищем (размеров 6, 5, 7, 4 и 3). Исходя из города O и зная стоимость проезда по отрезкам пути, указанным на рисунке, надо определить оптимальный путь (т.е. такой, что остаток после затрат на весь путь и получения в конечном пункте сокровища будет

максимальным). В этом случае N=5, каждое состояние системы связываем с вершиной графа, а управление – со стрелками, выходящими из неё. При выбранных вершине и стрелке, т.е. паре (x_n, u_n) функция $c_n(x_n, u_n)$ - это стоимость «проезда» по этой стрелке. Значения $\gamma(x_N)$ - это сокровища городов. Есть 3 оптимальных пути. Максимальный остаток после прохождения любого из трех оптимальных путей равен 1.



Фундаментальное замечание относительно сложности задачи. Имеется, очевидно, другое решение задачи о сокровищах: вычислить стоимости всех путей и выбрать путь (или пути), имеющие минимальную стоимость. Предположим, что эта поездка за сокровищами длится не 5 единиц времен, как в примере, а N=5000 единиц времени. Тогда число возможных путей будет порядка 2^N , то есть примерно 10^{1500} (так как $2^{10} \sim 10^3$) и задача

становится невыполнимой даже с привлечением мощных компьютеров. Напротив, алгоритм Беллмана приводит к числу операций порядка $N + (N-1) + \cdots + 1 \sim \frac{N^2}{2}$, и алгоритм становится реализуемым на компьютере.