



Билеты по методам оптимизации

Собрано из методички Новиковой и ее лекций.

Сделано пчелом с АСВК в 2022 году

Учтите, что некоторые вещи были расписаны,
а некоторые полностью взяты с методички и
сокращены в блоке ниже параграфа, так как
не на все хватило времени $\backslash_(_)_/$.

Внимание: данный PDF включает в себя все \LaTeX исходники.

Чтобы извлечь прикрепленные файлы, воспользуйтесь:

- Linux: утилитой `pdfdetach`
- Windows/Macos: AdobeAcrobat.

Пожалуйста – исправляйте, дополняйте, улучшайте :)

Оглавление

1	Билет 1	4
1	Определения: индивидуальной и массовой задачи, кодировки задачи, алгоритма решения массовой задачи, временной сложности алгоритма.	4
2	Формула градиентного метода в задаче безусловной минимизации.	4
2	Билет 2	6
1	Задачи распознавания свойств. Классы P и NP.	6
2	Формула метода Ньютона в задаче безусловной минимизации	8
3	Билет 3	9
1	Теорема об экспоненциальной временной оценке для задач из класса NP	9
2	Идея метода штрафов	9
4	Билет 4	11
1	Определение полиномиальной сводимости. Класс NPC. Теорема Кука (б/д)	11
2	Геометрическое описание симплекс-метода	12
5	Билет 5	13
1	Критерий NP-полноты. Док-во NP-полноты задачи БЛН	13
2	Методы глобальной минимизации	13
6	Билет 6	16
1	Док-во NP-полноты задачи 3-выполнимость. NP-трудные задачи	16
2	Формула градиентного метода в задаче безусловной минимизации	16
7	Билет 7	18
1	Класс co-NP. Пример задачи, допускающей хорошую характеристику. Доказательство утверждения о взаимоотношении классов NPC и co-NP	18
2	Формула метода Ньютона в задаче безусловной минимизации	18
8	Билет 8	20
1	Взаимоотношение классов P, NP и NPC, NP и co-NP. Класс PSPACE	20
2	Полиномиальный алгоритм округления ε_1 -приближенного решения системы линейных неравенств	20
9	Билет 9	21
1	Псевдополиномиальные алгоритмы. Пример для задачи о рюкзаке	21
2	Идея метода эллипсоидов	22
10	Билет 10	24
1	Сильная NP-полнота. Теорема о связи сильной NP-полноты задачи с существованием псевдополиномиального алгоритма ее решения	24
2	Идея метода штрафов	24
11	Билет 11	26
1	Определение комбинаторной задачи оптимизации и приближенного алгоритма ее решения. Утверждение о разнице между приближенным и точным оптимумом для задачи о рюкзаке (ЗР)	26
2	Идея метода Ньютона	27
12	Билет 12	28
1	Определение ε -приближенного алгоритма и полностью полиномиальной приближенной схемы (ПППС). Связь между существованием ПППС и псевдополиномиальностью	28
2	Теорема оптимальности для разложимых функций	28

13 Билет 13	30
1 Теорема об отсутствии ПППС для задач оптимизации, соответствующих сильно NP-полным задачам распознавания свойств	30
2 Геометрическая идея симплекс-метода	30
14 Билет 14	31
1 Определение озЛП. Принцип граничных решений. Алгебраическая и битовая сложность ЛП. Результаты о сложности для задач, близких к ЛП	31
2 Идея метода ветвей и границ. Пример для задачи БЛП	32
15 Билет 15	35
1 Теорема о границах решений задач ЛП с целыми коэффициентами	35
2 Метод ветвей и границ для ЦЛП. Различные стратегии метода	35
16 Билет 16	37
1 Теорема о мере несовместности систем линейных неравенств с целыми коэффициентами	37
2 Метод ветвей и границ для глобальной минимизации Липшицевых функций	37
17 Билет 17	39
1 Следствия систем линейных неравенств. Аффинная лемма Фаркаша (б/д)	39
2 Понятие о временной сложности алгоритмов	39
18 Билет 18	40
1 Лемма Фаркаша о неразрешимости	40
2 Понятие о недетерминированно полиномиальных задачах	40
19 Билет 19	42
1 Теорема двойственности ЛП	42
2 Метод динамического программирования для БЛП с неотрицательными коэффициентами	42
20 Билет 20	44
1 Сведение озЛП к однородной системе уравнений с ограничением $x > 0$, $x \neq 0$	44
2 Применение метода динамического программирования для понижения размерности разложимой оптимизационной задачи	44
21 Билет 21	47
1 Классификация задач математического программирования. Преимущества выпуклого случая	47
2 Полиномиальный алгоритм округления ε_1 -приближенного решения системы линейных неравенств	48
22 Билет 22	49
1 Необходимые условия локального минимума при ограничениях-неравенствах для дифференцируемых функций	49
2 Идея метода Кармаркара	50
23 Билет 23	52
1 Понятие о регулярности ограничений-неравенств в задаче математического программирования	52
2 Описание метода эллипсоидов	53

24 Билет 24	55
1 Теорема о целочисленности решения задачи ЛП с целыми коэффициентами для вполне унимодулярных матриц ограничений	55
2 Оценка сложности метода эллипсоидов поиска ε_2 -приближенного решения озЛП. . . .	56
25 Типичные задачи	58

Билет 1

1 Определения: индивидуальной и массовой задачи, кодировки задачи, алгоритма решения массовой задачи, временной сложности алгоритма.

Определение. Массовая задача (проблема) Π определяется:

1. общим списком всех параметров задачи (свободных параметров, значения которых не заданы)
2. формулировкой свойств, которым должен удовлетворять ответ (решение задачи).

Определение. Индивидуальная задача $I \in \Pi$ - частный случай массовой задачи Π при конкретно заданных параметрах.

Определение. Множество D - множество всех возможных значений параметров, заданных первым пунктом определения Π .

Определение. Алгоритм A решает массовую задачу Π , если для любой индивидуальной задачи $I \in \Pi$ алгоритм A применим к I (т.е. останавливается за конечное число шагов) и дает решение этой задачи I .

Определение. Схема кодирования (кодировка) задачи задается на некотором алфавите $\Sigma = \{\sigma_i\}$, Σ^* - все слова конечной длины в алфавите Σ .

Определение. Длиной слова $\sigma = \sigma_{i_1}\sigma_{i_2}\dots\sigma_{i_n}$, $\sigma_{i_j} \in \Sigma \ \forall i_j$ называется число $n \in \mathbb{N}$, $n = |\sigma|$ - количество букв в слове.

Определение. Кодировкой задачи Π называется такое отображение $e : \Pi \rightarrow \Sigma^*$, которое любой индивидуальной задаче $I \in \Pi$ ставит в соответствие ее код $e(I) = \sigma \in \Sigma^*$ и:

1. Возможно однозначное декодирование – $\forall I_1 \neq I_2, e(I_1) \neq e(I_2)$
2. e, e^{-1} полиномиально вычислимы – $\exists A$ - алгоритм, реализующий e, e^{-1} и полином $p(\cdot)$, для которого $\forall I \in \Pi$ время определения $e(I)$ и $e^{-1}(e(I))$ не превосходит $p(|e(I)|)$.
3. Кодировка неизбыточна – для любой другой кодировки e' , удовлетворяющей 1 и 2 найдется полином $p'(\cdot)$ такой, что $\forall I \in \Pi, |e(I)| < p'(|e'(I)|)$.
Например $|e(I)| = \log_2(I)$, $|e'(I)| = \log_{10}(I)$, понятно, что $\log_{10} < \log_2$, но $\exists p'(x) = 10 \cdot x$ и тогда $|e(I)| < p'(|e'(I)|)$, $\forall I \in \Pi$.

Определение. Обозначим $t_A(\sigma)$ время работы над словом $\sigma \in \Sigma^*$ (число шагов) алгоритма A до остановки. Временной сложностью алгоритма A решения массовой задачи Π назовем функцию $T_A(\cdot)$, определяемую как

$$T_A(n) = \max_{\sigma \in \Sigma^* : |\sigma| < n} t_A(\sigma) \quad \forall n \in \mathbb{Z}_+$$

2 Формула градиентного метода в задаче безусловной минимизации.

Общими методами локальной оптимизации (для произвольного, не обязательно выпуклого, случая) являются *методы локального спуска*.

Определение. Вектор $h \in \mathbb{R}^n$ называется *направлением убывания* функции f в точке x , если $f(x + \alpha h) < f(x)$ для всех достаточно малых $\alpha > 0$.

Утверждение 3. Пусть f дифференцируема в точке x . Тогда, если $\langle \text{grad} f(x), h \rangle < 0$, то h — направление убывания функции f в точке x , а если h — направление убывания функции f в точке x , то $\langle \text{grad} f(x), h \rangle \leq 0$.

Доказательство. Из условия дифференцируемости f имеем для достаточно малых $\alpha > 0$: $f(x + \alpha h) - f(x) = \langle \text{grad} f(x), \alpha h \rangle + o(\alpha) = \alpha \{ \langle \text{grad} f(x), h \rangle + o(\alpha)/\alpha \}$. Очевидно, последняя добавка не изменит знака выражения в фигурных скобках, если скалярное произведение строго отрицательно или строго положительно. Отсюда автоматически вытекает требуемое утверждение. ■

Замечание. Направление локального убывания дифференцируемой функции должно составлять острый угол с ее градиентом, который является наилучшим направлением убывания.

Определение. Точка x называется *стационарной точкой*, если $\text{grad} f(x) = 0$.

Отметим, что в условной оптимизации равенство нулю градиента уже не является необходимым условием минимума. Но в более простом случае $X = \mathbb{R}^n$ можно, двигаясь небольшими шагами в направлении антиградиента функции f в текущей точке, прийти в стационарную точку, как правило, локального минимума. Так мы получаем идею *градиентного метода безусловной минимизации*

Определение. Формула градиентного метода безусловной минимизации задается итеративной процедурой

$$x^{t+1} = x^t - \alpha_t \text{grad} f(x^t), \quad t = 1, 2, \dots, \quad \forall x^1 \in \mathbb{R}^n.$$

Определение. Параметр α_t называется *шаговым множителем* и может выбираться, исходя из различных соображений, разными способами.

Способы выбора шагового множителя:

1. *Пассивные способы* — $\{\alpha_t\}$ выбирается заранее.

- Постоянный шаг — $\alpha_t = \alpha_0$ для достаточно малых α_0 .
- Убывающий шаг (если α_0 не известно или при наличии помех)
 $\alpha_t \downarrow 0, \quad \sum \alpha_t = \infty$ (то есть ряд расходится), $\sum \alpha_t^2 < \infty$, например $\alpha_t = 1/t$.

2. *Адаптивные способы* — $\{\alpha_t\}$ зависит от реализующейся $\{x^t\}$.

- Метод скорейшего спуска — $\alpha_t \in \text{Arg min}_{\alpha > 0} f(x^t - \alpha \text{grad} f(x^t))$.
- Метод дробления шага (деления пополам) — если $f(x^{t+1}) > f(x^t)$, то возврат к t -й итерации с $\alpha_t := \alpha_t/2$. (Возможно и увеличение шага при стабильном убывании f , т.е. приближенный скорейший спуск.)
- Правило Армихо — путем дробления шага добиваемся для α_t выполнения условия

$$f(x^t - \alpha_t \text{grad} f(x^t)) - f(x^t) \leq -\varepsilon \alpha_t \|\text{grad} f(x^t)\|^2$$

В общем случае дифференцируемой, ограниченной снизу f можно получить сходимость градиентного метода к множеству стационарных точек, а при дополнительных предположениях доказывается (за исключением варианта с убывающим шагом) *линейная скорость сходимости*, которая в выпуклых задачах означает $\|x^{t+1} - x^*\| \leq q \|x^t - x^*\|$ для некоторого $0 < q < 1$.

Указанная линейная оценка объясняется тем, что в процессе минимизации градиентным методом используется линейная аппроксимация целевой функции на каждом шаге. Более высокую скорость сходимости получают для методов, основанных на квадратичной аппроксимации, в предположении дважды дифференцируемости f . Типичным примером здесь является *метод Ньютона*.

Билет 2

1 Задачи распознавания свойств. Классы P и NP.

Определение. Массовая задача Π где второй пункт формулировки определяется как
2. Допускается только два ответа - ДА и НЕТ,
называется *задачей распознавания свойств*.

Определение. Решением индивидуальной задачи $I \in \Pi$ распознавания свойств является правильное распознавание, принадлежит ли она к задачам с ответом ДА.

Практически не ограничивает класс решаемых задач. Любая задача минимизации может быть представлена в виде задачи распознавания свойств. Для этого необходимо ввести еще один параметр B , который будет определять значение *оптимума*.

Пример: существует ли маршрут длины, не превышающий B ?

Определение. Алгоритм A решения массовой задачи Π в формулировке распознавания свойств – ДМТ (детерминированная машина Тьюринга) с входным алфавитом Σ и конечными состояниями q_Y (ДА) и q_N (НЕТ).

Определение. Язык массовой задачи Π в формулировке распознавания свойств с кодировкой e – множество всех слов каждой правильной индивидуальной задачи $I \in Y(\Pi)$:

$$L(\Pi, e) = e(Y(\Pi)) = \{\sigma \in \Sigma^* \mid \Sigma - \text{алфавит } e, \sigma = e(I), I \in Y(\Pi)\}$$

(Y – множество тех индивидуальных задач, ответ на которых – ДА).

Определение. Языком алгоритма A называется множество принимаемых алфавитом слов (с которыми на входе A останавливается в состоянии q_Y): $L(A) = \{\sigma \in \Sigma^* \mid \Sigma - \text{алфавит } A, \text{ и } A(\sigma) = q_Y\}$

Определение. Алгоритм A *решает* массовую задачу Π в формулировке распознавания свойств с кодировкой e , если $L(A) = L(\Pi, e)$ (язык алгоритма совпадает с языком массовой задачи) и $\forall \sigma \in \Sigma^*$, A останавливается.

Определение. Класс *полиномиально разрешимых задач* P :

$P \doteq \{L(\Pi, e) \mid \exists A \text{ решающий } \Pi \text{ с кодировкой } e, \exists p(\cdot) - \text{полином} : T_A(n) < p(n) \quad \forall n \in \mathbb{Z}_+\}$.

Определение. Задача Π называется *полиномиально разрешимой*, если $\exists e$ – кодировка такая, что $L(\Pi, e) \in P$. Будем писать $\Pi \in P$.

Замечание. Определение классов идет для языков, а не задач. Задачи принадлежат некоторым языкам.

Замечание. Класс P относится только к задачам распознавания свойств (имеется ввиду, что для других задач мы не можем писать $\Pi \in P$, хоть существуют полиномиальные задачи не в формулировке распознавания свойств).

Замечание. Существуют задачи, для которых доказана их неполиномиальность: 1) алгоритмически неразрешимые; 2) с записью решения, длина которого превосходит полином; 3) с временной сложностью больше полинома.

Пример: полиномиальной считается задача распознавания четности целого числа.

Определение. Недетерминированной машиной Тьюринга (НДМТ) \hat{A} определяется как набор обычных – детерминированных – машин Тьюринга (ДМТ) $A(S)$ с алфавитом Σ , где S пробегает все слова из Σ^* :

$$\hat{A} = \{A(S)\}_{S \in \Sigma^*}$$

Замечание. Слова S в определении НДМТ можно проинтерпретировать как подсказки к решению (догадки), тогда ДМТ $A(S)$ проверяет для входного слова σ подсказку S и в случае правильности останавливается в состоянии ДА. НДМТ \hat{A} проверяет для входного слова σ все возможные подсказки, и если хоть одна правильная догадка существует, то НДМТ останавливается с ответом ДА.

Замечание. проверить решимость задачи с подсказкой на НДМТ можно за полином времени (длина подсказки тоже должна не превосходить полинома от длины входа).

Определение. НДМТ \hat{A} останавливается, когда останавливается первая из ДМТ $A(S)$, принимающая входное слово (останавливается в состоянии ДА).

Определение. Язык НДМТ - множество слов, принимаемых хотя бы одной ДМТ $A(S) \in \hat{A}$:

$$\hat{L}(\hat{A}) = \{\sigma \in \Sigma^* \mid \exists S \in \Sigma^* : \sigma \in L(A(S))\}$$

Замечание. НДМТ не может остановиться с ответом НЕТ (она будет работать вечно).

Определение. НДМТ \hat{A} решает массовую задачу Π с кодировкой e , если их языки совпадают: $L(\Pi, e) = \hat{L}(\hat{A})$.

Определение. Время работы НДМТ над словом σ как минимальное из времен работы над входом σ ДМТ $A(S)$, принимающих σ с учетом времени прочтения слова S (т.е. его длины).

$$\hat{t}_{\hat{A}}(\sigma) = \min_{\{S \mid \sigma \in L_{A(S)}\}} \{|S| + t_{A(S)}(\sigma)\}.$$

Определение. Временной сложностью НДМТ \hat{A} решения массовой задачи Π назовем функцию $\hat{T}_{\hat{A}}(\cdot) : \forall n \in \mathbb{Z}_+$

$$\hat{T}_{\hat{A}}(n) = \max_{\sigma \in \hat{L}(\hat{A}) : |\sigma| < n} \hat{t}_{\hat{A}}(\sigma) = \max_{\{\sigma \in \hat{L}(\hat{A}) : |\sigma| < n\}} \min_{\{S \mid \sigma \in L_{A(S)}\}} \{|S| + t_{A(S)}(\sigma)\}$$

Определение. Класс недетерминированных полиномиальных задач **NP**: такие языки, для которых существует НДМТ решающая задачу Π с кодировкой e , что временная сложность этой НДМТ будет полиномом от любой длины входы нашей задачи:

$$\mathbf{NP} = \{L(\Pi, e) \mid \exists \hat{A} - \text{НДМТ, реш. задачи } \Pi \text{ с кодировкой } e, \exists p(\cdot) - \text{полином} : \hat{T}_{\hat{A}} < p(n) \forall n \in \mathbb{Z}_+\}$$

Определение. Недетерминированной задачей называется такая задача Π , если для нее существует такая кодировка e , что $L(\Pi, e) \in \mathbf{NP}$. Будем писать $\Pi \in \mathbf{NP}$.

Утверждение 1. $\mathbf{P} \subseteq \mathbf{NP}$.

Замечание. доказать, что включение строгое/не строгое невозможно.

Замечание. Задача, для которой можно доказать, что она не полиномиальна, можно доказать, что она и не из **NP**. Если же задача из класса **NP**, то доказать, что она не полиномиальна нельзя.

Пример: Недетерминированной полиномиальной задачей является КМ, т.к. проверить догадку (маршрут) можно за полином времени.

Определение. Дополнительной задачей $\bar{\Pi}$ к массовой задаче Π распознавания свойств называется такая задача, которая задает обратный вопрос, нежели Π . Формально: $D(\bar{\Pi}) = D(\Pi)$, $Y(\bar{\Pi}) \setminus Y(\Pi)$.

Определение. $\text{co-P} = \{\bar{\Pi} \mid \Pi \in \text{P}\}$.

Определение. $\text{co-NP} = \{\bar{\Pi} \mid \Pi \in \text{NP}\}$

Утверждение 2. $\text{co-P} = \text{P}$.

Замечание. аналогичного утверждения про co-NP и NP дать нельзя.

2 Формула метода Ньютона в задаче безусловной минимизации

Для задач условной минимизации, например $\min_{x \in [1,2]} x^2$ предложенные методы нуждаются в модификации. В частности, для приведенного примера, когда множество X имеет достаточно простую структуру, указанные выше формулы совмещаются с процедурой проектирования на X на каждом шаге метода. Так приходим к методу *проекции градиента*

$$x^{t+1} = \text{Pr}_X\{x^t - \alpha_t \text{grad} f(x^t)\}, \quad t = 1, 2, \dots, \quad \forall x^1 \in \mathbb{R}^n$$

Пусть $f \in \mathbf{C}^2(\mathbb{R}^n)$, разложим функцию f в ряд Тейлора в окрестности текущей точки x^t :

$$f(x) - f(x^t) = \langle \text{grad} f(x^t), x - x^t \rangle + \frac{1}{2} \langle f''(x^t)(x - x^t), x - x^t \rangle + o(\|x - x^t\|^2).$$

Выберем x^{t+1} из условия минимизации квадратичной аппроксимации $f(x)$ в точке x^t , т.е. квадратичной части приращения $f(x) - f(x^t)$, получим метод Ньютона:

$$x^{t+1} = x^t - (f''(x^t))^{-1} \cdot \text{grad} f(x^t), \quad t = 1, 2, \dots$$

где начальное приближение x^1 должно находиться достаточно близко к точке оптимума x^* . В таком случае (и при дополнительных предположениях, более сильных, чем для приведенной ранее оценки скорости сходимости градиентного метода) для метода Ньютона будет справедлива *квадратичная скорость сходимости*

$$\|x^{t+1} - x^*\| \leq Q \|x^t - x^*\|^2, \quad \text{т.е.} \quad \|x^{t+1} - x^*\| \leq \frac{1}{Q} (Q \|x^1 - x^*\|)^{2^t}$$

что предполагает $\|x^1 - x^*\| < 1/Q$ (оценку для Q см., например, в [5, с. 192]). Еще раз подчеркнем, что градиентный метод в отличие от ньютоновского сходится при любом начальном приближении. Из определения метода Ньютона также следует требование невырожденности матрицы вторых производных (гессиана) функции f .

Нетрудно видеть, что полученная формула метода Ньютона решения задач безусловной минимизации совпадает с формулой метода Ньютона решения системы уравнений $\text{grad} f(x) = 0$, соответствующей необходимым условиям экстремума.

Билет 3

1 Теорема об экспоненциальной временной оценке для задач из класса NP

Теорема 1.1. Для любой недетерминированно полиномиальной задачи существует ДМТ, решающая ее с экспоненциальной временной сложностью, т.е. $\forall \Pi \in \mathbf{NP} \exists p(\cdot) \text{ — полином — и ДМТ } A:$

$$A \text{ решает } \Pi \text{ и } T_A(n) < 2^{p(n)} \quad \forall n \in \mathbb{Z}_+.$$

Доказательство. Так как $\Pi \in \mathbf{NP}$, то для любого слова σ из языка задачи Π (то есть для всех слов, для которых ответ на задачу ДА) существует правильная догадка S полиномиальной длины: $|S| < p_1(|\sigma|)$, $p_1(\cdot)$ — полином, и существует ДМТ $A(S) : t_{A(S)}(\sigma) < p_2(|\sigma|)$, $p_2(\cdot)$ — полином. Построим ДМТ A , которая работает над любым входным словом $\sigma \in \Sigma^*$ (с любой индивидуальной задачей $I \in \Pi$) следующим образом: рассматриваются все слова S из Σ^* длины меньше $p_1(|\sigma|)$ и делается не более $p_2(|\sigma|)$ шагов с каждой ДМТ $A(S)$. Если очередная ДМТ останавливается в состоянии q_Y (т.е. соответствующая догадка оказалась правильной), считаем слово σ принятым и работу ДМТ A законченной; если ни одна из ДМТ $A(S)$ не остановилась за отведенное время или остановилась в состоянии q_N , то заканчиваем работу ДМТ A и приписываем ей конечное состояние q_N . В последнем случае ДМТ A делает наибольшее число шагов, и это число меньше $p_2(|\sigma|) \cdot |\Sigma|^{p_1(|\sigma|)}$ (второй сомножитель равен числу проверяемых догадок, $|\Sigma|$ — число символов в алфавите Σ). Отсюда следует утверждение теоремы. ■

Замечание. оценка в конце приводится следующим образом

$$p_2(|\sigma|) \cdot |\Sigma|^{p_1(|\sigma|)} = 2^{\log_2(p_2(|\sigma|) \cdot |\Sigma|^{p_1(|\sigma|)})} = 2^{\log_2 p_2(|\sigma|)} \cdot 2^{p_1(|\sigma|) \log_2 |\Sigma|} \sim 2^{p(|\sigma|)} \sim 2^{p(n)}$$

2 Идея метода штрафов

Для задач условной минимизации, например $\min_{x \in [1,2]} x^2$ предложенные методы (град. спуска и Ньютона) нуждаются в модификации. В частности, для приведенного примера, когда множество X имеет достаточно простую структуру, указанные выше формулы совмещаются с процедурой проектирования на X на каждом шаге метода. Так приходим к методу *проекции градиента*

$$x^{t+1} = \text{Pr}_X \{x^t - \alpha_t \text{grad} f(x^t)\}, \quad t = 1, 2, \dots, \quad \forall x^1 \in \mathbb{R}^n$$

То есть делаем шаг градиентным методом, смотрим — выходит ли он за наши ограничения, и если он вышел за ограничения, то берем проекцию этой (“вылетевшей”) точки x_c на наше множество как $\min_{x \in X} \|x - x_c^{t+1}\|$. Работает, если проекцию найти просто (что не всегда так).

Для более сложных множеств X , допустим, задаваемых ограничениями неравенствами

$$X \doteq \{x \in \mathbb{R}^n \mid g_i(x) \leq 0 \quad \forall i \in M\}, \quad (3)$$

универсальным способом освобождения от ограничений является их штрафование. А именно для достаточно большой константы $C > 0$ вместо задачи условной минимизации (1),(3) рассматривают задачу безусловной минимизации оштрафованной целевой функции

$$\min_{x \in \mathbb{R}^n} \left\{ f(x) + C \sum_{i \in M} [g_i^+(x)]^p \right\}, \quad \text{где} \quad \sum_{i \in M} [g_i^+(x)]^p -$$

это *функция штрафа (штрафная функция)* для ограничений неравенств, $g^+(\cdot) \doteq \max[0, g(\cdot)]$ — *срезка* g , параметр штрафа $p \geq 1$. (Другие виды функций штрафа см. в [4,5].) В условиях непрерывности

функций f , g_i , непустоты X и ограниченности множества Лебега функции f можно доказать, что с ростом константы штрафа

$$\lim_{C \uparrow \infty} \min_{x \in \mathbb{R}^n} \{f(x) + C \sum_{i \in M} [g_i^+(x)]^p\} = f^* \quad (4)$$

Если $p = 1$ (функция-срезка и, следовательно, штрафная функция является острой), то $\exists C^* : \min\{f(x) + C^* \sum_{i \in M} g_i^+(x)\} = f^*$ (существует *точный штраф*). Однако при $p > 1$ — *гладкий штраф* подобное равенство означало бы несущественность ограничений $x \in X$ (точка безусловного минимума и так находится в X).

Нельзя свести безусловную к минимизации для гладкого штрафа, но (4) позволяет итеративно комбинировать метод штрафов и градиентный методов: $\forall x^1 \in \mathbb{R}^n$

$$x^{t+1} = x^t - \alpha_t \{ \text{grad} f(x^t) + C_t p \sum_{i \in M} [g_i^+(x^t)]^{p-1} \text{grad} g_i(x^t) \}, \quad t = 1, 2, \dots,$$

Эта процедура сходится при определенных соотношениях между $\{\alpha_t\}$ и $\{C_t\}$, в частности для убывающего шага при $\sum \alpha_t^2 C_t^2 < \infty$ (например, $\alpha_t = 1/t$, $C_t < \sqrt{t}$).

Это был *внешний штраф*. Для гладкого случая он всегда будет немного вылезать за ограничения. Есть еще *внутренний штраф* или *барьер*, который приоритизирует нахождения решения в ограничениях, нежели его точность. В частности, это метод *Кармаркара*, когда задача (9) ($Py = q$, $y \geq \bar{0}$), эквивалента задаче условной минимизации

$$\min_{x \geq \bar{0}, \sum x_j = N} p(x)$$

сводится к безусловной минимизации специальной барьерной функции $k(x)$, не позволяющий методу Ньютона выйти за ограничения $x > 0$, если в этих ограничениях выбрано начальное приближение.

Билет 4

1 Определение полиномиальной сводимости. Класс NPC. Теорема Кука (б/д)

Определение. Будем говорить, что массовая задача распознавания свойств Π' с кодировкой e' полиномиально сводится к задаче Π с кодировкой e , если любая индивидуальная задача $I' \in \Pi'$ может быть сведена за полиномиальное от ее длины время к некоторой $I \in \Pi$ с сохранением ответа. Формально:

$$\begin{aligned} \exists f : e'(D(\Pi')) &\rightarrow e(D(\Pi)), \quad - \text{функция сводимости, т.ч.} \\ f(e'(Y(\Pi'))) &= e(Y(\Pi)), \quad \text{то есть} \\ \forall \sigma' \in e'(Y(\Pi')) : f(\sigma') &\in e(Y(\Pi)) \quad \text{и} \quad \forall \sigma'' \in e'(D(\Pi') \setminus Y(\Pi')) : f(\sigma'') \in e(D(\Pi) \setminus Y(\Pi)), \\ \text{и } \exists A_f &- \text{ДМТ, реализующая } f \text{ за полином времени, т.е.} \\ \exists p_f(\cdot) : \forall \sigma \in e'(D(\Pi')), &t_{A_f}(\sigma) < p_f(|\sigma|) \end{aligned}$$

В случае, когда соответствующие кодировки не избыточны, будем использовать термин полиномиальной сводимости по отношению к самим задачам без указания кодировок. Обозначение: $\Pi' \propto \Pi$.

Утверждение 3. Если $\Pi_1 \propto \Pi_2$ и $\Pi_2 \propto \Pi_3$, то $\Pi_1 \propto \Pi_3$.

Утверждение 4. Если $\Pi' \propto \Pi$ и $\Pi \in \mathbf{P}$, то и $\Pi' \in \mathbf{P}$.

Доказательство. Обозначим A - ДМТ, решающую Π с полиномиальной временной сложностью, и построим ДМТ A' , решающую Π' с полиномиальной временной сложностью, как суперпозицию ДМТ A и A_f : $A' = A \circ A_f$, т.е. сначала к любому входному слову $\sigma' \in e'(D(\Pi'))$ применяется A_f , а потом к получившемуся слову $\sigma = f(\sigma')$ (длиной не более $p_f(|\sigma'|)$) применяется A . Временная сложность A' : $T_{A'}(\cdot) \leq T_{A_f}(\cdot) + T_A(p_f(\cdot))$ — полином.

Утверждение 5. Если $\Pi' \propto \Pi$ и $\Pi \in \mathbf{NP}$, то и $\Pi' \in \mathbf{NP}$.

Доказательство. Аналогично заменой ДМТ на НДМТ.

Определение. Массовая задача Π называется **NP-полной** (**NP-complete** — **NPC**) или *универсальной*, если $\Pi \in \mathbf{NP}$ и $\forall \Pi' \in \mathbf{NP} \quad \Pi' \propto \Pi$. (то есть любая недетерминированно полиномиальная задача сводится к Π).

Замечание. грубо говоря, $\mathbf{NPC} = \max_{\propto} \mathbf{NP}$.

Утверждение. Класс **NPC** не пуст.

Утверждение 6. Если $\mathbf{P} \cap \mathbf{NPC} \neq \emptyset$, то $\mathbf{P} = \mathbf{NP}$. А если $\mathbf{NPC} \cap (\mathbf{NP} \setminus \mathbf{P}) \neq \emptyset$, то $\mathbf{NPC} \subseteq \mathbf{NP} \setminus \mathbf{P}$.

То есть, если бы удалось найти полиномиальный алгоритм решения хоть одной **NPC** задачи, то были бы построены полиномиальные алгоритмы решения всех **NPC** и **NP** задач.

А если для какой-либо **NPC** задачи доказать отсутствие полиномиального алгоритма ее решения, то это не только дает строгое включение $\mathbf{P} \subset \mathbf{NP}$, но и влечет за собой доказательство невозможности построения полиномиального алгоритма любой задачи из **NPC**.

Определение. Задача выполнимости (ВЫП): выяснить выполнимость конъюнктивной нормальной формы (КНФ) - конъюнкции конечного числа дизъюнктивных функций, то есть дизъюнкций булевых переменных z_i или их отрицания \bar{z}_i .

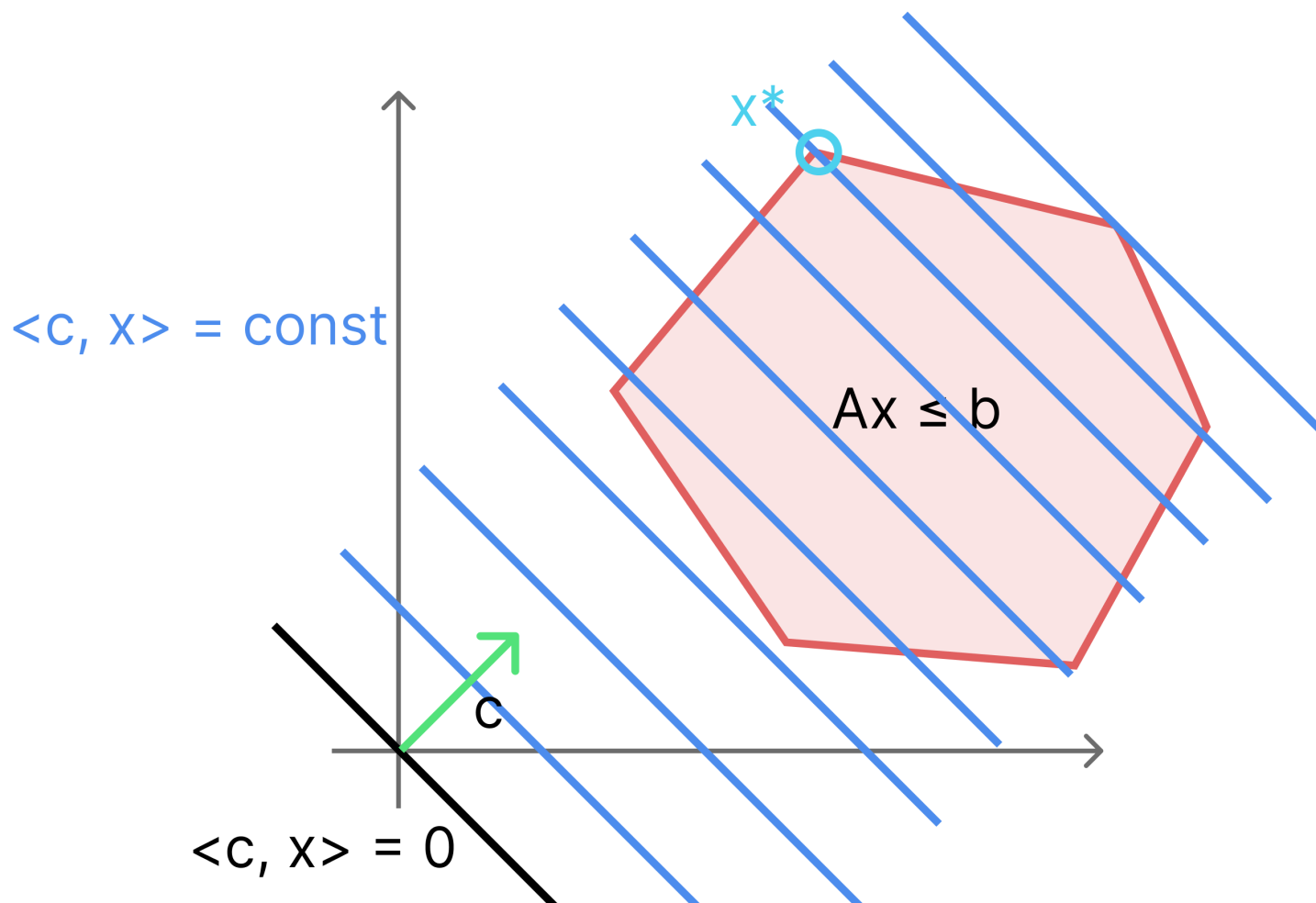
Кратко: для КНФ K на входе, определить существует ли z^0 - набор входных значений, т.ч. $K(z^0) = 1$.

Теорема 1.2 (т. Кука) (б/д). $\mathbf{ВЫП} \in \mathbf{NPC}$.

2 Геометрическое описание симплекс-метода

Перебор всех вершин по возрастанию целевой функции $\langle c, x \rangle$.

$\max \langle c, x \rangle = d = \langle c, x^* \rangle$ -- решение



Билет 5

1 Критерий NP-полноты. Док-во NP-полноты задачи БЛН

Теорема 1.3 (критерий NP-полноты). $\Pi \in \text{NPC} \iff \Pi \in \text{NP}$, и $\exists \Pi' \in \text{NPC} : \Pi' \propto \Pi$ (Π - массовая задача распознавания свойств).

Доказательство. $\forall \Pi'' \in \text{NP} : \Pi'' \propto \Pi' \text{, тогда из } \Pi' \propto \Pi \Rightarrow \Pi'' \propto \Pi \Rightarrow \Pi \in \text{NPC.} \blacksquare$

Определение. ЦЛН - задача о существовании целочисленного решения системы линейных неравенств с целыми коэффициентами.

Утверждение. $\text{ВЫП} \propto \text{ЦЛН}$.

Доказательство. $]Q_p = z_{i_1} \vee z_{i_2} \vee \dots \vee z_{i_n} \vee \overline{z_{j_1}} \vee \overline{z_{j_2}} \vee \dots \vee \overline{z_{j_m}}$ - элементарная дизъюнкция. $K = Q_1 \wedge Q_2 \wedge \dots \wedge Q_s$ - КНФ.

Тогда $\forall p = \overline{1, s}$ в Q_p заменим знаки \vee на $+$, далее $\overline{z_j}$ на $(1 - z_j)$, а в K заменим знаки \wedge на \cdot . Составив систему $\{K \geq 1, z_i \geq 0, z_i \leq 1, \forall i\}$, получим ЦЛН.

Утверждение 7. $\text{ЦЛН} \in \text{NPC}$.

Доказательство. 1) $\text{ЦЛН} \in \text{NP}$, так как для проверки ответа достаточно просмотреть выполнение каждого неравенства, что не превосходит полинома от длины записи всех коэффициентов.

2) $\text{ВЫП} \propto \text{ЦЛН}$ — уже доказано выше.

Утверждение 8. $\text{БЛН} \in \text{NPC}$ (следствие из двух последних утверждений).

2 Методы глобальной минимизации

1. Как уже отмечалось ранее, задачи глобальной оптимизации (т.е. в невыпуклом случае задачи оптимизации вообще) являются переборными. Переборные алгоритмы не эффективны (в расчете на худшую задачу), поэтому успех в решении каждой конкретной задачи существенным образом зависит от способа организации перебора. Если мы готовы оставить возможность или невозможность решения нашей задачи на волю случая, то естественно использовать случайный перебор. Этот способ перебора обычно является самым простым и, как правило, экономит память. Для задачи поиска глобального минимума ему соответствует следующий *метод Монте-Карло*.

Сокращение. Эффективного алгоритма на худшую задачу глобальной оптимизации нет, они все переборные. Успех решения конкретной задачи зависит от способа перебора. Естественно брать случайный перебор. Для задачи поиска глобального минимума — *метод Монте-Карло*.

Пусть решается задача (1) из §8, где (для упрощения изложения) множество ограничений X — единичный n -мерный куб. Выбираем в соответствии с равномерным распределением на X случайные точки x^t , в которых вычисляем значение целевой функции, запоминаем текущее наименьшее значение — *рекорд* — и реализующую его точку. Тогда $\forall \varepsilon > 0$ вероятность

$$\mathbf{P}(|\min_t f(x^t) - f^*| > \varepsilon) \rightarrow 0 \text{ при } t \rightarrow \infty$$

Сходимость такого метода будет довольно медленной. При этом не известно, на каком расстоянии от точки минимума находится полученная реализация.

Сокращение. Сходимость предложенного метода будет довольно медленной и зависит от начального приближения. Стоит требовать, чтобы рассматриваемая функция была Липшицевой и искать не точное, ε -приближенное решение.

Определение. Функция $f(x)$ называется Липшицевой на X с константой L , если $\forall x, x' \in X : |f(x) - f(x')| \leq L\|x - x'\|$. Обозначается $f \in \text{Lip}(X, L)$.

Сузим класс рассматриваемых задач (1), предположив вдобавок к предыдущему, что функция цели липшицева на X с константой L : $f \in \text{Lip}(X, L)$, т.е. $|f(x) - f(x')| \leq L\|x - x'\| \forall x, x' \in X$. И не рассчитывая найти точное решение, зададимся подходящим $\varepsilon > 0$ с целью поиска ε -приближенного решения $x^\varepsilon : f(x^\varepsilon) \leq f(x^*) + \varepsilon$. (На близость x^ε и x^* никаких условий не накладывается.)

Теперь мы можем применять методы детерминированного перебора. Пассивный (не использующий при выборе очередной точки информацию, полученную для предыдущих) способ поиска приводит к полному перебору: разобьем X на подкубы X^j так, чтобы $\forall x, x' \in X^j : \|x - x'\| \leq \delta \doteq \varepsilon/L$, в каждом X^j берем произвольную точку x^j и полагаем

$$f(x^\varepsilon) \doteq \min_j f(x^j).$$

Очевидно, x^ε и есть искомое ε -приближенное решение. (Действительно, $\forall j, \forall x \in X^j$ $f(x^\varepsilon) \leq f(x^j) \leq f(x) + \varepsilon$ по условию Липшица, и, в частности, для $x = x^*$ имеем $f(x^\varepsilon) \leq f(x^*) + \varepsilon$ — соответствие с определением.) Однако сторона каждого j -го подкуба равна $\varepsilon/(L\sqrt{n})$, а всего подкубов и, следовательно, вычислений значений целевой функции будет $(L\sqrt{n}/\varepsilon)^n$ в любом случае, что не мыслимо даже для десятка переменных. Поэтому разрабатываются методы последовательного перебора, позволяющие учитывать уже вычисленные значения и адаптироваться к нехудшему случаю.

Сокращение. Пассивный способ поиска приводит к полному перебору. Разбив на соответствие подкубы мы действительно находим решение, удовлетворяющие ε -приближению, но приходится делать $(L\sqrt{n}/\varepsilon)^n$ вычислений, что даже при 10 переменных очень плохо.

Предположим, что уже вычислены значения функции в точках x^1, \dots, x^{j-1} и рекордным оказалось значение $f(x^r) = R$. Тогда, если $f(x^j) < f(x^r)$, то обновляем рекорд $r := j$, $R := f(x^j)$, а если $f(x^j) > f(x^r)$, то можно не вычислять значений функции на множестве $T_j(R) \doteq \{x \in X : \|x - x^j\| \leq (f(x^j) - R)/L\}$, так как это не даст нового рекорда (ибо $\forall x \in T_r$ $f(x^j) - f(x) \leq L\|x - x^j\| \leq f(x^j) - f(x^r)$, т.е. $f(x) \geq f(x^r) = R$, и значит, среди них нет глобально-оптимального решения). Обновление рекорда в принципе позволяет “отбросить” аналогичные множества $T_i(R)$ для $i = 1, \dots, j-1$.

Естественно, в T_i , T_j могут попасть и точки x^k с уже вычисленным значением $f(x^k)$ (которые таким образом вычислялись зря). Поэтому хотелось бы так организовать перебор, чтобы по возможности уменьшить число подобных “зряшных” вычислений. К сожалению, оптимальной стратегии организации перебора для многомерных задач нет. Использование случайных точек x^i приводит к проблеме хранения и обновления сложного множества $\cup T_i(R)$ заведомо не оптимальных точек. Метод послыного перебора дает возможность сокращения лишь по одной переменной. Для задач большой размерности предлагается (различными авторами) следующий метод перебора по схеме *ветвей и границ*.

Сокращение. Если предположить, что часть $(j-1)$ значений уже вычислено, то последовательным обновлением рекордного значения можно отбросить часть новых точек на рассмотрение. Бывает так, что какие-то из уже вычисленных точек отбрасываются. Стратегии, минимизирующей вычисления таких точек для многомерных задач, — нет. Более того, нужно хранить и обновлять сложное множество заведомо не оптимальных точек.

2. Метод ветвей и границ (МВГ) для глобальной минимизации.

Пусть x^1 — центр куба X . Вычисляем $f(x^1)$ и присваиваем это значение рекорду $R := f(x^1)$. Разбиваем куб на 2^n одинаковых подкубов X^{1i} со стороной $1/2$ и вычисляем значения целевой функции в их центрах: $f(x^{1i})$, $i = 1, \dots, 2^n$, обновляя по ходу вычислений значение рекорда $R := \min_i f(x^{1i})$. Проверяем выполнение условия $X^{1i} \subseteq T_{1i}(R)$ для $i = 1, \dots, 2^n$ и отбрасываем соответствующие подкубы. Каждый из оставшихся разбиваем на 2^n одинаковых подкубов X^{2ij} со стороной $1/4$ и поступаем, как прежде. На любом шаге у нас формируется множество \mathbf{K} “кубиков” со сторонами 2^{-l} , $l \geq 2$, целое. Правило выбора очередного кубика для разбиения называется *правилом ветвления* — возможные варианты приводятся ниже. Кубики со стороной не больше $\varepsilon/(L\sqrt{n})$ исключаются из множества \mathbf{K} — дробление кубика заканчивается. Также исключаются кубики, попавшие в множество $T_k(R)$ (с индексом k — номером кубика) для текущего значения рекорда, — *правило отсечения ветвей*. Рекорд обновляется при получении меньшего значения целевой функции (*правило получения границ, т.е. оценок*). Значения целевой функции вычисляются в центре каждого нового подкубика, включаемого в \mathbf{K} после разбиения выбранного для этого кубика. Алгоритм останавливается, когда \mathbf{K} пусто.

Сокращение. МВГ разбивает исходный куб на граф-дерева подкубов, итеративно исключая бесполезные (которые попали в $T_s(R)$).

Указанная терминология и название метода определяются тем, что визуально данная схема перебора представляется в виде графа-дерева, корневая вершина которого соответствует кубу X , вершины первого яруса — подкубам X^{1i} , вершины второго яруса — кубикам X^{2ij} , подсоединенным к своим *порождающим* вершинам X^{1i} 1-го яруса, и т.д. Если кубик исключается из \mathbf{K} , его вершина *закрывается* — из нее не будут идти ветви на следующий ярус. Если кубик еще не включен в \mathbf{K} , его вершина еще *не раскрыта*. Порядок закрытия вершины определяется правилом отсечения (своим для каждой массовой задачи — см. также в §11), порядок раскрытия — правилом ветвления (своим для каждой индивидуальной задачи). Различают два вида правил ветвления по типу построения дерева решений (выбора вершин для раскрытия): “в ширину”, когда сначала раскрываются все вершины одного яруса до перехода к следующему, и “в глубину” — всякий раз раскрывается лишь одна (обычно с лучшим значением рекорда) вершина на ярусе до конца ветви. На практике реализуют некоторую смесь, например, первое правило, пока хватает машинной памяти (в \mathbf{K} не слишком много элементов), затем переключаемся на второе. Предпочтительность той или иной стратегии ветвления оценивается каждым вычислителем по-своему, исходя из главной задачи метода ветвей и границ — быстрее получить лучший рекорд, чтобы отсечь больше ветвей.

Сокращение. Порядок исключения (закрытия) вершин, ненужных для вычисления кубов, — правило отсечения — зависит от конкретной массовой задачи, а правило раскрытия (вычисления) новых кубов — правило ветвления — от конкретной индивидуальной задачи. Правил ветвления два: в ширину (раскрыть все кубы яруса), в глубину (раскрыть только один с наилучшим рекордом). На практике используют смесь.

В рассматриваемой задаче есть хороший способ улучшения рекорда — локальная оптимизация (см. в §8). Ее имеет смысл проводить из текущей точки, в которой произошло обновление рекорда, например, делая несколько шагов градиентного метода. При этом расположение кубиков менять не надо, просто увеличивается шанс сокращения перебора (отбрасывания больших кубиков).

Отметим, что в худшем случае $f = \text{const}$ ($UT_i = \emptyset$) — не удастся отбросить ни одной точки x — и приходим к полному перебору; т.е. указанная в п.1 экспоненциальная оценка точна на классе всех липшицевых функций.

Сокращение. Имеет смысл улучшать рекорд с помощью локальной оптимизации: при обновлении рекорда сделать пару шагов градиентным методом.

Билет 6

1 Док-во NP-полноты задачи 3-выполнимость. NP-трудные задачи

Определение. Задача 3-ВЫП – подзадача задачи ВЫП, где в каждой дизъюнктивной функции КНФ ровно по 3 переменных.

Утверждение. Так как $D(3\text{-ВЫП}) \subset D(\text{ВЫП})$ и $\text{ВЫП} \in \mathbf{NP} \Rightarrow 3\text{-ВЫП} \propto \text{ВЫП} \Rightarrow 3\text{-ВЫП} \in \mathbf{NP}$.

Утверждение 9. $\text{ВЫП} \propto 3\text{-ВЫП}$.

Доказательство. Достаточно доказать, что любую дизъюнктивную функцию от k переменных можно свести к КНФ с дизъюнкциями от 3 переменных, тогда утверждение очевидно.

Пусть $f^j(z_{j_i} \mid i = \overline{1, k})$ – дизъюнктивная функция от k переменных. Представим z_{j_i} и $\overline{z_{j_i}}$ как y_i для упрощения записи. В таком случае $f^j = y_1 \vee y_2 \vee \dots \vee y_k$. Введем дополнительные переменные u^j и представим f^j в виде КНФ как:

$$K^j = (y_1 \vee y_2 \vee u_1^j) \wedge (y_3 \vee \overline{u}_1^j \vee u_2^j) \wedge \dots \wedge (y_{k-2} \vee \overline{u}_{k-4}^j \vee u_{k-3}^j) \wedge (y_{k-1} \vee y_k \vee \overline{u}_{k-3}^j)$$

Данная замена не эквивалентна. Если $f^j = 0$, то построенная КНФ $K^j = 0 \quad \forall u$, однако если $f^j = 1$, то $\exists u : K^j = 1$. Но этого достаточно для сохранения ответа на вопрос о существовании выполняющего набора.

Замечание. если f^j зависит от $k = 1, 2$ переменных, то способствующие u^j переменные являются фиктивными, а f^j имеет вид:

$$f^j|_{k=1} = y_1 = (y_1 \vee u_1 \vee u_2) \wedge (y_1 \vee u_1 \vee \overline{u}_2) \wedge (y_1 \vee \overline{u}_2 \vee u_1) \wedge (y_1 \vee \overline{u}_2 \vee \overline{u}_1)$$

$$f^j|_{k=2} = y_1 \vee y_2 = (y_1 \vee y_2 \vee u_1) \wedge (y_1 \vee y_2 \vee \overline{u}_1)$$

Утверждение. $3\text{-ВЫП} \in \mathbf{NPC}$ (следует из двух предыдущих утверждений).

Определение. Массовая задача Π' сводится по Тьюрингу к задаче Π , значит, что из существования полиномиального алгоритма решения Π следует существование полиномиального алгоритма и для Π' . Обозначение: $\Pi' \propto_T \Pi$.

Определение. В класс **NP-трудных** (**NP-hard**) задач входят следующие массовые задачи Π :

1. Π – распознавания свойств, для которой доказано, что $\Pi' \propto \Pi$, $\Pi' \in \mathbf{NPC}$, но не показано, что $\Pi \in \mathbf{NP}$.
2. Π – оптимизации, для которой способствующая задача Π распознавания свойств **NPC**.
3. Π (любая массовая), к которой сводится по Тьюрингу какие-либо **NPC** задачи Π' — $\Pi' \in \mathbf{NPC} : \Pi' \propto_T \Pi$.

Определение. Задачи Π (любые массовые), для которых $\exists \Pi' \in \mathbf{NPC} : \Pi' \propto_T \Pi$ и $\exists \Pi'' \in \mathbf{NP} : \Pi \propto_T \Pi''$ называются из класса **NP-эквивалентных**.

2 Формула градиентного метода в задаче безусловной минимизации

Общими методами локальной оптимизации (для произвольного, не обязательно выпуклого, случая) являются *методы локального спуска*.

Определение. Вектор $h \in \mathbb{R}^n$ называется *направлением убывания* функции f в точке x , если $f(x + \alpha h) < f(x)$ для всех достаточно малых $\alpha > 0$.

Утверждение 3. Пусть f дифференцируема в точке x . Тогда, если $\langle \text{grad} f(x), h \rangle < 0$, то h — направление убывания функции f в точке x , а если h — направление убывания функции f в точке x , то $\langle \text{grad} f(x), h \rangle \leq 0$.

Доказательство. Из условия дифференцируемости f имеем для достаточно малых $\alpha > 0$: $f(x + \alpha h) - f(x) = \langle \text{grad} f(x), \alpha h \rangle + o(\alpha) = \alpha \{ \langle \text{grad} f(x), h \rangle + o(\alpha)/\alpha \}$. Очевидно, последняя добавка не изменит знака выражения в фигурных скобках, если скалярное произведение строго отрицательно или строго положительно. Отсюда автоматически вытекает требуемое утверждение. ■

Замечание. Направление локального убывания дифференцируемой функции должно составлять острый угол с ее градиентом, который является наилучшим направлением убывания.

Определение. Точка x называется *стационарной точкой*, если $\text{grad} f(x) = 0$.

Отметим, что в условной оптимизации равенство нулю градиента уже не является необходимым условием минимума. Но в более простом случае $X = \mathbb{R}^n$ можно, двигаясь небольшими шагами в направлении антиградиента функции f в текущей точке, прийти в стационарную точку, как правило, локального минимума. Так мы получаем идею *градиентного метода безусловной минимизации*.

Определение. Формула градиентного метода безусловной минимизации задается итеративной процедурой

$$x^{t+1} = x^t - \alpha_t \text{grad} f(x^t), \quad t = 1, 2, \dots, \quad \forall x^1 \in \mathbb{R}^n.$$

Определение. Параметр α_t называется *шаговым множителем* и может выбираться, исходя из различных соображений, разными способами.

Способы выбора шагового множителя:

1. *Пассивные способы* — $\{\alpha_t\}$ выбирается заранее.

- Постоянный шаг — $\alpha_t = \alpha_0$ для достаточно малых α_0 .
- Убывающий шаг (если α_0 не известно или при наличии помех)
 $\alpha_t \downarrow 0, \quad \sum \alpha_t = \infty$ (то есть ряд расходится), $\sum \alpha_t^2 < \infty$, например $\alpha_t = 1/t$.

2. *Адаптивные способы* — $\{\alpha_t\}$ зависит от реализующейся $\{x^t\}$.

- Метод скорейшего спуска — $\alpha_t \in \text{Arg min}_{\alpha > 0} f(x^t - \alpha \text{grad} f(x^t))$.
- Метод дробления шага (деления пополам) — если $f(x^{t+1}) > f(x^t)$, то возврат к t -й итерации с $\alpha_t := \alpha_t/2$. (Возможно и увеличение шага при стабильном убывании f , т.е. приближенный скорейший спуск.)
- Правило Армихо — путем дробления шага добиваемся для α_t выполнения условия

$$f(x^t - \alpha_t \text{grad} f(x^t)) - f(x^t) \leq -\varepsilon \alpha_t \|\text{grad} f(x^t)\|^2$$

В общем случае дифференцируемой, ограниченной снизу f можно получить сходимость градиентного метода к множеству стационарных точек, а при дополнительных предположениях доказывается (за исключением варианта с убывающим шагом) *линейная скорость сходимости*, которая в выпуклых задачах означает $\|x^{t+1} - x^*\| \leq q \|x^t - x^*\|$ для некоторого $0 < q < 1$.

Указанная линейная оценка объясняется тем, что в процессе минимизации градиентным методом используется линейная аппроксимация целевой функции на каждом шаге. Более высокую скорость сходимости получают для методов, основанных на квадратичной аппроксимации, в предположении дважды дифференцируемости f . Типичным примером здесь является *метод Ньютона*.

Билет 7

1 Класс co-NP. Пример задачи, допускающей хорошую характеристику. Доказательство утверждения о взаимоотношении классов NPC и co-NP

Определение. Дополнительной задачей $\bar{\Pi}$ к массовой задаче Π распознавания свойств называется такая задача, которая задает обратный вопрос, нежели Π . Формально: $D(\bar{\Pi}) = D(\Pi)$, $Y(\bar{\Pi}) \setminus Y(\Pi)$.

Определение. $\text{co-P} = \{\bar{\Pi} \mid \Pi \in \text{P}\}$.

Определение. $\text{co-NP} = \{\bar{\Pi} \mid \Pi \in \text{NP}\}$

Утверждение. $\text{co-P} = \text{P}$ (состояние ДМТ из q_Y за один шаг заменим на q_N).

Замечание. аналогичного утверждения про co-NP и NP дать нельзя.

Определение. Задачей, имеющей *хорошую характеристику* называется такая массовая задача распознавания свойств, для которого выполнено: $\Pi \in \text{NP} \cap \text{co-NP}$.

Пример: задача простые числа (ПЧ) имеет доказанную хорошую характеристику. А совсем недавно был найден способствующий алгоритм.

Утверждение. из $\text{P} = \text{co-P}$ следует $\text{P} \subseteq \text{NP} \cap \text{co-NP}$.

Утверждение 10. Если для некоторой $\Pi \in \text{NPC}$, $\bar{\Pi} \in \text{NP}$, то $\text{NP} = \text{co-NP}$.

Доказательство. Так как $\Pi \in \text{NPC}$, то $\forall \Pi' \in \text{NP}$ справедливо $\Pi' \propto \Pi \Rightarrow \bar{\Pi}' \propto \bar{\Pi}$ (полиномиальное сведение осуществляется по той же функции, что и для $\Pi' \propto \Pi$). Так как $\bar{\Pi} \in \text{NP} \Rightarrow \bar{\Pi}' \in \text{NP}$, а с учетом произвольности $\Pi' \in \text{NP}$ имеет, что $\text{co-NP} \subseteq \text{NP}$. Аналогичное рассуждения проводим для $\bar{\Pi} = \Pi$, получаем обратное включение.

Замечание. в настоящее время гипотеза $\text{NP} = \text{co-NP}$ кажется нереальной.

Замечание. а оно ли имеется ввиду?

2 Формула метода Ньютона в задаче безусловной минимизации

Для задач условной минимизации, например $\min_{x \in [1,2]} x^2$ предложенные методы нуждаются в модификации. В частности, для приведенного примера, когда множество X имеет достаточно простую структуру, указанные выше формулы совмещаются с процедурой проектирования на X на каждом шаге метода. Так приходим к методу *проекции градиента*

$$x^{t+1} = \text{Pr}_X\{x^t - \alpha_t \text{grad} f(x^t)\}, \quad t = 1, 2, \dots, \quad \forall x^1 \in \mathbb{R}^n$$

Пусть $f \in \text{C}^2(\mathbb{R}^n)$, разложим функцию f в ряд Тейлора в окрестности текущей точки x^t :

$$f(x) - f(x^t) = \langle \text{grad} f(x^t), x - x^t \rangle + \frac{1}{2} \langle f''(x^t)(x - x^t), x - x^t \rangle + o(\|x - x^t\|^2).$$

Выберем x^{t+1} из условия минимизации квадратичной аппроксимации $f(x)$ в точке x^t , т.е. квадратичной части приращения $f(x) - f(x^t)$, получим метод Ньютона:

$$x^{t+1} = x^t - (f''(x^t))^{-1} \cdot \text{grad} f(x^t), \quad t = 1, 2, \dots$$

где начальное приближение x^1 должно находиться достаточно близко к точке оптимума x^* . В таком случае (и при дополнительных предположениях, более сильных, чем для приведенной ранее оценки скорости сходимости градиентного метода) для метода Ньютона будет справедлива *квадратичная скорость сходимости*

$$\|x^{t+1} - x^*\| \leq Q \|x^t - x^*\|^2, \quad \text{т.е.} \quad \|x^{t+1} - x^*\| \leq \frac{1}{Q} (Q \|x^1 - x^*\|)^{2^t}$$

что предполагает $\|x^1 - x^*\| < 1/Q$ (оценку для Q см., например, в [5, с. 192]). Еще раз подчеркнем, что градиентный метод в отличие от ньютоновского сходится при любом начальном приближении. Из определения метода Ньютона также следует требование невырожденности матрицы вторых производных (гессиана) функции f .

Нетрудно видеть, что полученная формула метода Ньютона решения задач безусловной минимизации совпадает с формулой метода Ньютона решения системы уравнений $\text{grad} f(x) = 0$, соответствующей необходимым условиям экстремума.

Билет 8

1 Взаимоотношение классов P, NP и NPC, NP и co-NP. Класс PSPACE

Определение. Класс **PSPACE** – общий класс массовых задач (не обязательно распознавания свойств), для которых существует алгоритмы, требующие не более чем полиномиальной памяти.

Замечание. Вопрос о равенстве **P** и **PSPACE** является открытым.

Определение. Класс **PSPACE-полных** задач – задачи, к которым полиномиально сводятся любые задачи из **PSPACE**.

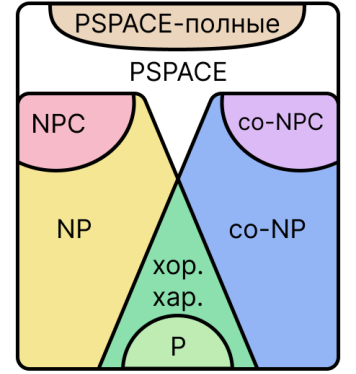


Диаграмма взаимосвязи классов сложности.

2 Полиномиальный алгоритм округления ε_1 -приближенного решения системы линейных неравенств

1. Имея ε -приближенное решение $Ax \leq b$ (1) с $\varepsilon \leq \varepsilon_1 = 1/[(n+2)\Delta(A)]$, можно (на основании теоремы 2, §5, см. билет 16) быть уверенным в существовании точного решения системы линейных неравенств. Оказываются, процедура получения x^0 из x^{ε_1} является полиномиальной. Соответствующий алгоритм округления ε_1 -приближенного решения системы (1) до точного был указан Л. Г. Хачияном и состоит в следующем.

Определение. Алгоритм округления – полиномиальный алгоритм, с помощью которого из ε_1 -приближенного решения можно получить точное.

Присвоим $x^1 := x^{\varepsilon_1}$ и подставим x^1 в (1). Разобьем множество $M \doteq \{1, \dots, m\}$ индексов неравенств в системе на два подмножества

$$M(x^1) \doteq \{i : |\langle a_i, x^1 \rangle - b_i| \leq \varepsilon_1\}, \\ M \setminus M(x^1) \doteq \{i : \langle a_i, x^1 \rangle - b_i \leq -\varepsilon_1\}$$

Найдем решение x'^1 системы равенств $A_{M(x^1)}x = b_{M(x^1)}$ (существует по теореме 2). Пусть x'^1 не является точным решением (1), т.е. в x'^1 не выполнилось i -е неравенство для какого-либо $i \notin M(x^1)$. Тогда введем множество индексов невыполненных неравенств $M^+ \doteq \{i | \langle a_i, x'^1 \rangle > b_i\} \subseteq M \setminus M(x^1)$ и рассмотрим на отрезке $[x^1, x'^1]$ ближайшую к x'^1 точку, в которой еще выполнены все неравенства для $i \in M^+$ (в x^1 они выполнены с ε_1 -запасом). А именно определим

$$\tau \doteq \min_{i \in M^+} \frac{b_i - \langle a_i, x^1 \rangle}{\langle a_i, x'^1 \rangle - \langle a_i, x^1 \rangle}, \quad i_1 \doteq \arg \min_{i \in M^+} \frac{b_i - \langle a_i, x^1 \rangle}{\langle a_i, x'^1 \rangle - \langle a_i, x^1 \rangle}$$

и присвоим $x^2 := (1 - \tau)x^1 + \tau x'^1$. Имеем $M(x^2) \supseteq M(x^1) \cup \{i_1\}$, ибо неравенства с индексами из $M(x^1)$ ε_1 -приближенно выполнялись как равенства на всем отрезке $[x^1, x'^1]$, а неравенство с индексом $i_1 \in M^+$, не выполненное в точке x'^1 , выполняется в x^2 как равенство по построению. Таким образом, $M(x^2) \supset M(x^1)$, но $|M(x)| \leq m$, поэтому, повторяя указанную процедуру с заменой x^1 на x^2 и т.д., придем не более чем через $\max(n, m)$ шагов к тому, что решение x' соответствующей системы равенств окажется x^0 – решением (1).

С учетом полиномиальности задачи решения систем уравнений предложенный алгоритм округления полиномиален.

Билет 9

1 Псевдополиномиальные алгоритмы. Пример для задачи о рюкзаке

Определение. Задача о рюкзаке (ЗР) – из набора вещей собрать рюкзак ограниченного объема с наибольшей полезностью:

$$\max_{z : z_j \in \{0,1\} \quad \forall j \in \overline{1,n}} \left\{ \sum_{j=1}^n c_j z_j \mid \sum_{j=1}^n w_j z_j \leq K \right\}$$

Здесь $c_j \in \mathbb{Z}_+$ – полезность (ценность), $w_j \in \mathbb{Z}_+$ – объем (вес) вещи, а переменная z_j определяет – класть или не класть ее в рюкзак, $K \in \mathbb{Z}_+$ – максимальный объем (вес) рюкзака.

В формулировке распознавания свойств – можно ли набрать рюкзак ценностью не меньше B .

Замечание. ЗР – частный случай БЛН.

Утверждение. ЗР \in NPC.

Метод динамического программирования для решения ЗР

Произвольная индивидуальная задача $I \in$ ЗР погружается в семейство задач поиска

$$F_i(E) \doteq \max_{z : z_j \in \{0,1\} \quad \forall j = \overline{i,n}} \left\{ \sum_{j=i}^n c_j z_j \mid \sum_{j=i}^n w_j z_j \leq K - E \right\},$$

$F_1(0)$ – значение ЗР. И решаются все задачи данного семейства по рекуррентным формулам, где i убывает с n до 1. А именно, положим $F_i(E) \doteq 0 \quad \forall E \geq K, \quad \forall i$. Имеем $\forall E = \overline{0, K-1}$:

$$F_n(E) = \begin{cases} 0 & E > K - w_n \\ c_n & E \leq K - w_n \end{cases}$$

и $\forall i = n-1, \dots, 2$:

$$F_i(E) = \max\{F_{i+1}(E), c_i + F_{i+1}(E + w_i)\} \doteq \max_{z_i \in \{0,1\}} \{c_i z_i + F_{i+1}(E + w_i z_i)\}$$

$$F_1(0) = \max\{F_2(0), c_1 + F_2(w_1)\}$$

Число итераций предложенного алгоритма равно nK и того же порядка будет его временная сложность. Отметим, что алгоритм не является полиномиальным, ибо для записи числа K требуется порядка $\log_2 K$ символов; является переборным – перебирает все варианты заполненности рюкзака.

Определение. Обозначим через $\text{num}(I)$ максимальное по модулю целое число (или 0), фигурирующее при задании числовых параметров для индивидуальной задачи I , и через $|I| \doteq |e(I)|$ – длину записи I .

Определение. Алгоритм A решения массовой задачи Π (не обязательно распознавания свойств) называется *псевдополиномиальным*, если для некоторого полинома $p(\cdot, \cdot)$ выполнено $t_A(e(I)) < p(|I|, \text{num}(I)) \quad \forall I \in \Pi$.

Пример: алгоритм динамического программирования для задачи о рюкзаке (ЗР) является псевдополиномиальным.

2 Идея метода эллипсоидов

Аналогичный алгоритм (как и для ε_1 окр.) имеется и для округления ε_2 -приближенного решения озЛП $x_{\varepsilon_2}^*$ до точного x^* (см. [3, с. 21]). Поэтому для построения полиномиального алгоритма решения озЛП осталось указать полиномиальный алгоритм поиска ε_2 -приближенного решения озЛП в шаре $\|x\| \leq n^{1/2}\Delta$ или удостоверения, что такого решения нет (по теоремам 1,2* из §5). Требуемый алгоритм, основанный на *методе эллипсоидов*, который предложили в 1976–77 гг. Д. Б. Юдин и А. С. Немировский и (независимо) Н. З. Шор, приводится в следующих пунктах.

Здесь и далее $\Delta \doteq \Delta(D)$, где матрица D задается таблицей (3) (симплекс таблицей).

2. Пусть E — произвольный эллипсоид в \mathbb{R}^n с центром ξ и ненулевого объема $\text{vol}E$. Рассмотрим $(n-1)$ -мерную плоскость, заданную вектором g нормали и проходящую через центр ξ эллипсоида E . Обозначим через $E^-(g)$ один из двух полуэллипсоидов, на которые разбивает E данная плоскость, $E^-(g) = E \cap \{x \mid \langle g, x - \xi \rangle \leq 0\}$.

Утверждение 1. Полуэллипсоид $E^-(g)$ эллипсоида E можно целиком заключить в новый эллипсоид E' , имеющий объем, строго меньший E ,

$$\frac{\text{vol}E'}{\text{vol}E} < e^{-1/(2n+2)}, \quad (*)$$

и E' можно вычислить по $E^-(g)$ за $O(n^2)$ арифметических операций.

Доказательство. Пусть E — единичный шар с центром в точке $\bar{0}$: $E = \{x \in \mathbb{R}^n : \|x\| \leq 1\}$, а $E^-(g) = E \cap \{x_n \geq 0\}$. Поместим центр E' в точку $\xi' = (0, \dots, 0, \frac{1}{n+1})$, тогда

$$E' = \{x \mid (x_1^2 + \dots + x_{n-1}^2)/\beta^2 + (x_n - \frac{1}{n+1})^2/\alpha^2 \leq 1\},$$

где $\alpha \doteq 1 - 1/(n+1) < e^{-1/(n+1)}$, $\beta^2 \doteq 1 + 1/(n^2 - 1) < e^{1/(n^2-1)}$.

Отношение объемов равно произведению полуосей $\alpha\beta^{n-1} < e^{-1/(2n+2)}$, откуда получаем (*), ибо любой эллипсоид можно превратить в шар аффинным преобразованием координат, сохраняющим объем. Действительно, будем представлять произвольный эллипсоид E с помощью его центра ξ и матрицы Q ($n \times n$), задающей указанное преобразование: $E = \{x \mid x = \xi + Qy, \|y\| \leq 1\}$. Обозначим $\eta \doteq Q^T g / \|Q^T g\|$, где верхний индекс T — знак транспонирования. Тогда ξ' и Q' , представляющие эллипсоид E' минимального объема, описанный вокруг полуэллипсоида $E^-(g)$, пересчитываются по формулам

$$\xi' = \xi - \frac{1}{n+1}Q\eta, \quad Q' = \frac{n}{\sqrt{(n^2-1)}}\{Q + (\sqrt{\frac{n-1}{n+1}} - 1)Q\eta\eta^T\}$$

за $O(n^2)$ арифметических операций.

Метод эллипсоидов получения ε -приближенного решения озЛП.

Положим $\varepsilon := \varepsilon_2 \doteq 1/(2n^2\Delta^3)$. Введем множество ε -приближенных решений озЛП в шаре радиуса $R \doteq n^{1/2}\Delta$ с центром в $\bar{0}$:

$$X_\varepsilon^* \doteq \{x \mid \langle a_i, x \rangle \leq b_i + \varepsilon \quad \forall i = \overline{1, m}, \quad \langle c, x \rangle \geq d^* - \varepsilon, \quad \|x\| \leq R\}$$

Выберем указанный выше шар в качестве начальной итерации для эллипсоида $E \supset X_\varepsilon^*$. Рассмотрим произвольную итерацию.

Проверяем, является ли центр ξ эллипсоида E ε -приближенным решением. Если да, то алгоритм заканчивает свою работу, в противном случае строим эллипсоид E' для очередной итерации как минимальный по объему эллипсоид, содержащий полуэллипсоид $E^-(g)$, где вектор g определяется следующим образом. Так как $\xi \notin X_\varepsilon^*$, то либо

$$1^0) \exists i : \langle a_i, \xi \rangle > b_i + \varepsilon, \text{ и тогда } g := a_i, \text{ либо}$$

$$2^0) \langle c, \xi \rangle < d^* - \varepsilon \text{ и } g := -c.$$

Убедимся, что при этом $X_\varepsilon^* \subset E'$. Действительно, для варианта 1^0

$\forall x \in X_\varepsilon^* \langle a_i, x \rangle \leq b_i + \varepsilon < \langle a_i, \xi \rangle$, т.е. $X_\varepsilon^* \subset E \cap \{x \mid \langle a_i, x - \xi \rangle \leq 0\} = E^-(a_i) \subset E'$; и аналогично получим для варианта 2^0

$$X_\varepsilon^* \subset E \cap \{x \mid \langle c, x - \xi \rangle \geq 0\} = E^-(-c) \subset E'$$

Теперь с $E := E'$ возвращаемся к началу итерации (на новый шаг).

Оценим число итераций метода эллипсоидов. Покажем, что X_ε^* содержит шар радиуса $r/2$, где $r \doteq \varepsilon/(hn^{1/2}) < R$, $h \geq |a_{ij}|, |c_j|$ (h высота задачи). Пусть x^* — точное решение в X_ε^* . Из $\|x^* - x\| \leq r$ следует $|\langle a_i, x \rangle - \langle a_i, x^* \rangle| \leq \|a_i\| \|x^* - x\| \leq hn^{1/2}r = \varepsilon \quad \forall i \in M$ и $|\langle c, x \rangle - \langle c, x^* \rangle| \leq \|c\| \|x^* - x\| \leq hn^{1/2}r$, т.е. указанный выбор r гарантирует, что все такие x будут ε -приближенными решениями. Поскольку $\|x^*\| \leq R$, то множество тех из рассматриваемых x , для которых $\|x\| \leq R$ (т.е. пересечение шаров радиуса r и R , включающее центр первого), содержит шар радиуса $r/2$. Этот шар и принадлежит X_ε^* . Таким образом, объем X_ε^* больше объема n -мерного шара радиуса $r/2$. Значит, объем эллипсоида, построенного последним, например E^k для k -й итерации, не должен оказаться меньше объема этого шара. Отсюда и из утверждения 1 получаем для k соотношение

$$\left(\frac{r}{2R}\right)^n \leq \frac{\text{vol} X_\varepsilon^*}{\text{vol} E^1} \leq \frac{\text{vol} E^k}{\text{vol} E^1} < e^{-k/(2n+2)}$$

из которого k (по определению r, R, ε, h и Δ) не превосходит

$$2n^2 \ln(Rnh/\varepsilon) < 2n^2 \ln(2n^{3.5}\Delta^5) < 10n^2 \ln(n\Delta)$$

Билет 10

1 Сильная NP-полнота. Теорема о связи сильной NP-полноты задачи с существованием псевдополиномиального алгоритма ее решения

Определение. Обозначим через $\text{num}(I)$ максимальное по модулю целое число (или 0), фигурирующее при задании числовых параметров для индивидуальной задачи I , и через $|I| \doteq |e(I)|$ – длину записи I .

Определение. *Полиномиальным сужением* массовой задачи Π называется множество индивидуальных задач, числовые параметры которых не превосходят полинома от длины входа:

$$\Pi_{p(\cdot)} = \{I \in \Pi \mid \text{num}(I) < p(|I|)\}$$

Определение. Массовая задача Π распознавания свойств называется *сильно NP-полной*, если ее полиномиальное сужение **NP-полно**, то есть $\exists p(\cdot)$ – полином, т.ч. $\Pi_{p(\cdot)} \in \text{NPC}$.

Замечание. сильно NPC задачи наиболее трудные для счета среди всех NP, в оптимизационной постановке нет эффективных алгоритмов даже приближенного решения. Решаются разбиением на подзадачи, а они уже перебором методом ветвей и границ.

Пример: задачи ВЫП, 3-ВЫП, БЛН, ЦЛН и КМ являются сильно NPC.

Теорема 1.4. Если $P \neq NP$, то ни для какой сильно NPC задачи не существует псевдополиномиального алгоритма решения.

Доказательство. Проведем от противного. Пусть ДМТ A решает сильно NPC задачу Π и $\forall I \in \Pi$: $t_A(e(I)) < p'(|I|, \text{num}(I))$ для полинома $p'(\cdot, \cdot)$. Тогда $\forall I \in \Pi_{p(\cdot)}$ $t_A(e(I)) < p'(|I|, p(|I|)) = p''(|I|)$, т.е. $\Pi_{p(\cdot)} \in P$ – противоречие с $\Pi_{p(\cdot)} \in \text{NPC}$ или утверждением 6 о пересечении P и NPC (см. билет 4).

2 Идея метода штрафов

Для задач условной минимизации, например $\min_{x \in [1,2]} x^2$ предложенные методы (град. спуска и Ньютона) нуждаются в модификации. В частности, для приведенного примера, когда множество X имеет достаточно простую структуру, указанные выше формулы совмещаются с процедурой проектирования на X на каждом шаге метода. Так приходим к методу *проекции градиента*

$$x^{t+1} = \text{Pr}_X\{x^t - \alpha_t \text{grad} f(x^t)\}, \quad t = 1, 2, \dots, \quad \forall x^1 \in \mathbb{R}^n$$

То есть делаем шаг градиентным методом, смотрим – выходит ли он за наши ограничения, и если он вышел за ограничения, то берем проекцию этой (“вылетевшей”) точки x_c на наше множество как $\min_{x \in X} \|x - x_c^{t+1}\|$. Работает, если проекцию найти просто (что не всегда так).

Для более сложных множеств X , допустим, задаваемых ограничениями неравенствами

$$X \doteq \{x \in \mathbb{R}^n \mid g_i(x) \leq 0 \quad \forall i \in M\}, \quad (3)$$

универсальным способом освобождения от ограничений является их штрафование. А именно для достаточно большой константы $C > 0$ вместо задачи условной минимизации (1),(3) рассматривают задачу безусловной минимизации оштрафованной целевой функции

$$\min_{x \in \mathbb{R}^n} \left\{ f(x) + C \sum_{i \in M} [g_i^+(x)]^p \right\}, \quad \text{где} \quad \sum_{i \in M} [g_i^+(x)]^p -$$

это *функция штрафа* (*штрафная функция*) для ограничений неравенств, $g^+(\cdot) \doteq \max[0, g(\cdot)]$ — *срезка* g , параметр штрафа $p \geq 1$. (Другие виды функций штрафа см. в [4,5].) В условиях непрерывности функций f, g_i , непустоты X и ограниченности множества Лебега функции f можно доказать, что с ростом константы штрафа

$$\lim_{C \uparrow \infty} \min_{x \in \mathbb{R}^n} \{f(x) + C \sum_{i \in M} [g_i^+(x)]^p\} = f^* \quad (4)$$

Если $p = 1$ (функция-срезка и, следовательно, штрафная функция является острой), то $\exists C^* : \min\{f(x) + C^* \sum_{i \in M} g_i^+(x)\} = f^*$ (существует *точный штраф*). Однако при $p > 1$ — *гладкий штраф* подобное равенство означало бы несущественность ограничений $x \in X$ (точка безусловного минимума и так находится в X).

Нельзя свести безусловную к минимизации для гладкого штрафа, но (4) позволяет итеративно комбинировать метод штрафов и градиентный методов: $\forall x^1 \in \mathbb{R}^n$

$$x^{t+1} = x^t - \alpha_t \{\text{grad} f(x^t) + C_t p \sum_{i \in M} [g_i^+(x^t)]^{p-1} \text{grad} g_i(x^t)\}, \quad t = 1, 2, \dots,$$

Эта процедура сходится при определенных соотношениях между $\{\alpha_t\}$ и $\{C_t\}$, в частности для убывающего шага при $\sum \alpha_t^2 C_t^2 < \infty$ (например, $\alpha_t = 1/t, C_t < \sqrt{t}$).

Это был *внешний штраф*. Для гладкого случая он всегда будет немного вылезать за ограничения. Есть еще *внутренний штраф* или *барьер*, который приоритизирует нахождения решения в ограничениях, нежели его точность. В частности, это метод *Кармаркара*, когда задача (9) ($Py = q, y \geq \bar{0}$), эквивалента задаче условной минимизации

$$\min_{x \geq \bar{0}, \sum x_j = N} p(x)$$

сводится к безусловной минимизации специальной барьерной функции $k(x)$, не позволяющий методу Ньютона выйти за ограничения $x > 0$, если в этих ограничениях выбрано начальное приближение.

Билет 11

1 Определение комбинаторной задачи оптимизации и приближенного алгоритма ее решения. Утверждение о разнице между приближенным и точным оптимумом для задачи о рюкзаке (ЗР)

Определение. Задачи дискретной (комбинаторной) оптимизации – индивидуальные задачи I массовой задачи оптимизации Π , для которых решение есть произвольная реализация оптимума:

$$\text{Opt}_{\Pi}(I) \doteq \max_{z \in S_{\Pi}(I)} f_{\Pi}(I, z)$$

То есть такая точка $z^*(I) \in S_{\Pi}(I)$, для которой $f_{\Pi}(I, z^*(I)) = \text{Opt}_{\Pi}(I)$. Здесь $S_{\Pi}(I) \subseteq \mathbb{Z}^{n(I)}$ – область допустимых значений дискретной переменной z , $f_{\Pi}(I, \cdot) : S_{\Pi}(I) \rightarrow \mathbb{Z}$ – целевая функция индивидуальной задачи I оптимизации.

Замечание. Например в ЗР $S_{\Pi}(I)$ – ограничения на рюкзак.

Замечание. В постановке может быть как \max , так и \min .

Утверждение. Если $\mathbf{P} \neq \mathbf{NP}$, то для задачи оптимизации не существует полиномиального алгоритма, если ее задача распознавания свойств из класса \mathbf{NPC} .

Определение. σ_S – компоненты входного слова $\sigma = e(I)$, определяющие параметры индивидуальной задачи $I \in \Pi$, которые задают допустимую область (ограничения задачи).

Определение. σ_f – компоненты входного слова $\sigma = e(I)$, определяющие параметры индивидуальной задачи $I \in \Pi$, которые задают функцию цели.

Пример: для ЗР $f_{\text{ЗР}}(\sigma, z) = \langle c, z \rangle$

($\langle \cdot, \cdot \rangle$ – скалярное произведение), $S_{\text{ЗР}}(\sigma) = \{z = (z_1, \dots, z_n) \mid z_j \in \{0, 1\} \ \forall j, \langle w, z \rangle \leq K\}$, а

$\sigma_S = (n, w, K)$, $\sigma_f = c$.

Определение. Алгоритм A решает задачу оптимизации, если $\forall I \in \Pi$ алгоритм A применим (т.е. останавливается) и дает $z^*(I)$ – решение задачи I (либо что $S_{\Pi}(I) = \emptyset$ и решения не существует).

Определение. Алгоритм A называется *приближенным алгоритмом решения массовой задачи оптимизации* Π , если $\forall I \in \Pi$ он находит некоторую точку из допустимой области $z_A(I) \in S_{\Pi}(I)$ (если $S_{\Pi}(I) \neq \emptyset$), принимаемую за приближенное решение. Значение $f_{\Pi}(I, z_A(I))$ называется *приближенным значением оптимума* и обозначается – $A(I)$.

Замечание. близость к оптимальному решению (оптимум) не подразумевается.

Утверждение 11. Если $\mathbf{P} \neq \mathbf{NP}$, то ни для какой константы $C > 0 \nexists A$ – полиномиальный алгоритм решения ЗР с оценкой $|\text{Opt}_{\text{ЗР}}(I) - A(I)| \leq C \ \forall I \in \text{ЗР}$.

Доказательство. Проведем от противного. Пусть найдены такие C и A . Построим алгоритм A' следующим образом: $\forall I \in \text{ЗР}$ домножим все коэффициенты c_j на $C + 1$ – получим индивидуальную задачу $I' \in \text{ЗР}$, к которой применим алгоритм A и разделим полученный ответ на $C + 1$, т.е. $A'(I) = A(I')/(C + 1)$.

Очевидно, $\text{Opt}_{\text{ЗР}}(I') = (C + 1)\text{Opt}_{\text{ЗР}}(I)$ и из полиномиальности алгоритма A вытекает полиномиальность A' . При этом его точность равна

$$|\text{Opt}_{\text{ЗР}}(I) - A'(I)| = \frac{|\text{Opt}_{\text{ЗР}}(I') - A(I')|}{C + 1} \leq \frac{C}{C + 1} < 1$$

То есть равна нулю (так как все значения целевой функции целые).

Получили полиномиальный алгоритм точного решения ЗР. Проверка $\text{Opt}_{\text{ЗР}}(I) \geq B$ полиномиальна, значит, построили и полиномиальный алгоритм решения ЗР в постановке распознавания свойств, что с учетом универсальности последней противоречит утверждению о пересечении **P** и **NPC** (см. билет 4). ■

2 Идея метода Ньютона

Пусть $f \in \mathbf{C}^2(\mathbb{R}^n)$, разложим функцию f в ряд Тейлора в окрестности текущей точки x^t :

$$f(x) - f(x^t) = \langle \text{grad} f(x^t), x - x^t \rangle + \frac{1}{2} \langle f''(x^t)(x - x^t), x - x^t \rangle + o(\|x - x^t\|^2).$$

Выберем x^{t+1} из условия минимизации квадратичной аппроксимации $f(x)$ в точке x^t , т.е. квадратичной части приращения $f(x) - f(x^t)$, получим метод Ньютона:

$$x^{t+1} = x^t - (f''(x^t))^{-1} \text{grad} f(x^t), \quad t = 1, 2, \dots,$$

где начальное приближение x^1 должно находиться достаточно близко к точке оптимума x^* . В таком случае (и при дополнительных предположениях, более сильных, чем для приведенной ранее оценки скорости сходимости градиентного метода) для метода Ньютона будет справедлива *квадратичная скорость сходимости*

$$\|x^{t+1} - x^*\| \leq Q \|x^t - x^*\|^2, \quad \text{т.е.} \quad \|x^{t+1} - x^*\| \leq \frac{1}{Q} (Q \|x^1 - x^*\|)^{2^t}$$

что предполагает $\|x^1 - x^*\| < 1/Q$ (оценку для Q см., например, в [5, с. 192]). Еще раз подчеркнем, что градиентный метод в отличие от ньютоновского сходится при любом начальном приближении. Из определения метода Ньютона также следует требование невырожденности матрицы вторых производных (гессиана) функции f .

Нетрудно видеть, что полученная формула метода Ньютона решения задач безусловной минимизации совпадает с формулой метода Ньютона решения системы уравнений $\text{grad} f(x) = 0$, соответствующей необходимым условиям экстремума.

Билет 12

1 Определение ε -приближенного алгоритма и полностью полиномиальной приближенной схемы (ПППС). Связь между существованием ПППС и псевдополиномиальностью

Определение. Приближенный алгоритм A решения массовой задачи Π оптимизации называется ε -приближенным алгоритмом решения Π для некоторого $\varepsilon > 0$, если:

$$\forall I \in \Pi \quad \frac{|\text{Opt}_{\Pi}(I) - A(I)|}{|\text{Opt}_{\Pi}(I)|} < \varepsilon$$

То есть относительная погрешность не превосходит ε .

Теорема 1.5 (связь пс-пол. и ПППС) (б/д). Пусть для массовой задачи Π оптимизации:

1. $\exists A$ - псевдополиномиальный алгоритм.
2. $\forall I \in \Pi$ справедливо $|\text{Opt}_{\Pi}(I)| < p_1(|I|, \text{num}(I))$ и $\text{num}(I) < p_2(|I|, \text{Opt}_{\Pi}(I))$ для некоторых полиномов p_1, p_2 .
3. $\forall \sigma = e(I), I \in \Pi$: параметры σ_S и σ_f (см. билет 11) не пересекаются; и $\forall z \in S_{\Pi}(\sigma)$ функция цели $f_{\Pi}(\sigma, z)$ линейно зависит от параметров σ_f .

Тогда $\exists p(\cdot, \cdot)$ - полином, т.ч. $\forall \varepsilon > 0 \quad \exists A_{\varepsilon}$ - ε -приближенный алгоритм решения Π с временной сложностью $T_{A_{\varepsilon}}(|I|) < p(|I|, 1/\varepsilon)$.

Определение. Полностью полиномиальными приближенными схемами (ПППС) называются алгоритмы $\{A_{\varepsilon}\}$ определенные в теореме 1.5 (выше).

Замечание. ПППС - лучшее для NP-hard.

2 Теорема оптимальности для разложимых функций

Определение. Функция f называется *разложимой* на f_1 и f_2 , если она разделяема на f_1, f_2 и функция f_1 монотонно не убывает (\nearrow) по последнему аргументу.

Теорема 12.1 (оптимальности для разложимых функций).

$$\min_{(x,y)} f(x, y) = \min_x f_1(x, \min_y f_2(y)),$$

и точно так же для \max .

Доказательство проведем для случая минимума.

(Равенство будет вытекать из пары противоположных неравенств.)

$$\text{По определению минимума} \quad \min_{x,y} f_1(x, f_2(y)) \leq f_1(x^0, f_2(y^0)) \quad \forall x^0, y^0$$

$$\text{а значит и для} \quad y^0 := \arg \min_y f_2(y), \quad x^0 := \arg \min_x f_1(x, f_2(y^0)),$$

что доказывает неравенство " \leq ". Аналогично, в силу неубывания f_1 по последнему аргументу, $f_1(x', \min_y f_2(y)) \leq f_1(x', f_2(y')) \quad \forall x', y'$.

$$\text{Положим} \quad y' := \arg \min_y f_1(x', f_2(y)), \quad x' := \arg \min_x \{\min_y f_1(x, f_2(y))\}.$$

Поскольку повторный \min равен двойному, в правой части получили $\min_{x,y} f(x,y)$, чем доказали и неравенство “ \geq ”. ■

Для задачи условной оптимизации теорема оптимальности для разложимых функций переписывается следующим образом:

$$\min_{(x,y) \in \Omega} f(x,y) = \min_{x : Y(x) \neq \emptyset} f_1(x, \min_{y \in Y(x)} f_2(y)) \quad (5)$$

где $Y(x) = \{y \mid (x,y) \in \Omega\}$.

Замечание. Указанная теорема используется для понижения размерности оптимизационных задач и в методе ДП.

Билет 13

1 Теорема об отсутствии ПППС для задач оптимизации, соответствующих сильно NP-полным задачам распознавания свойств

Теорема 1.6. Если для Π оптимизации соответствующая ей Π распознавания свойств является сильно **NPС** и $\exists p'(\cdot)$ - полином, т.ч. $|\text{Opt}_{\Pi}(I)| < p'(\text{num}(I)) \quad \forall I \in \Pi$, то при условии, что **P** \neq **NP**, ПППС для Π не существует.

Доказательство. Проведем от противного. Пусть ПППС существует. Построим алгоритм A' следующим образом: $\forall I \in \Pi$, A' вызывает A_{ε} с $\varepsilon = 1/(p'(\text{num}(I)) + 1)$.

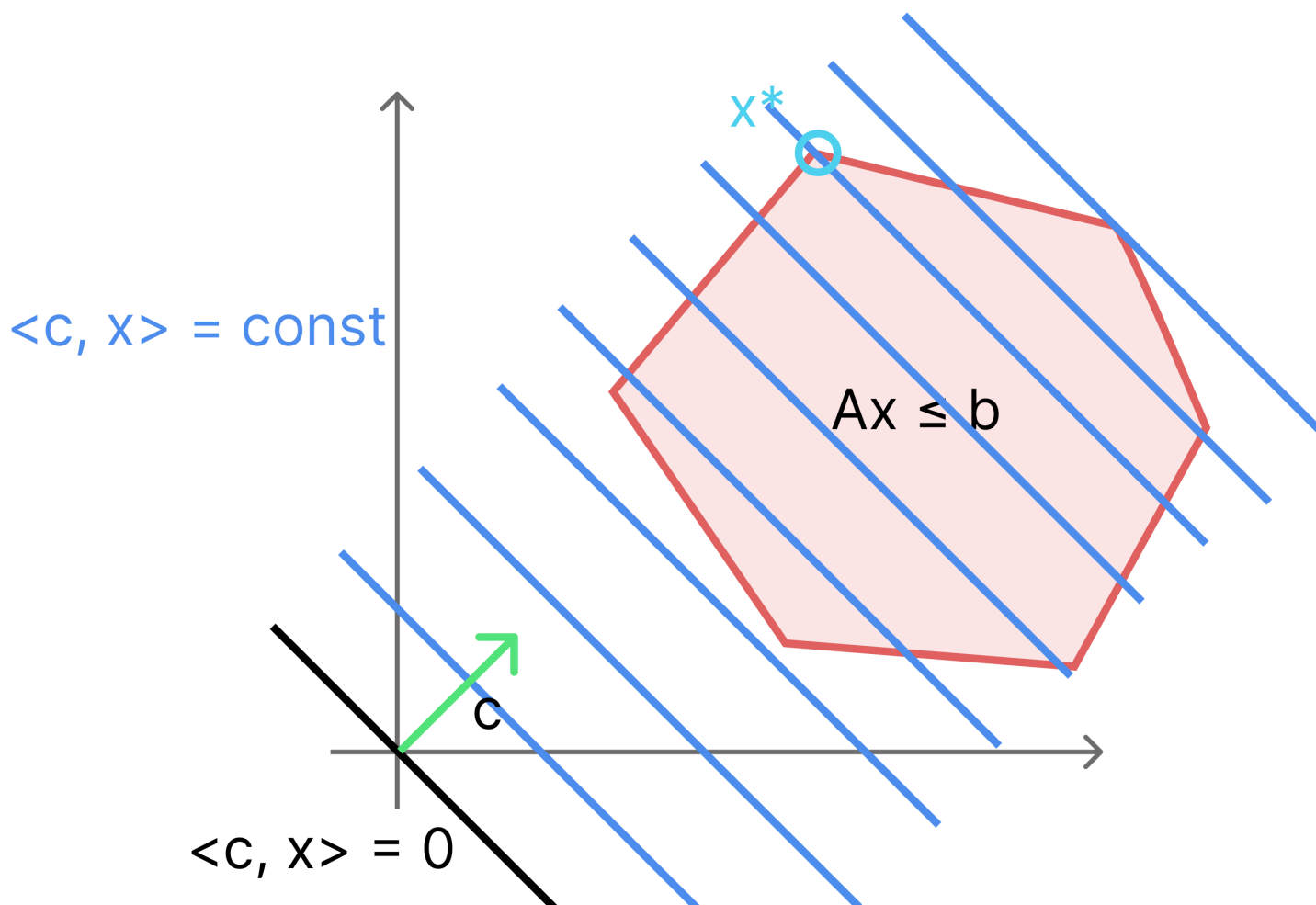
Тогда по определению ε -приближенного алгоритма A_{ε} :

$$|\text{Opt}_{\Pi}(I) - A'(I)| < \frac{|\text{Opt}_{\Pi}(I)|}{p'(\text{num}(I)) + 1} < \frac{p'(\text{num}(I))}{p'(\text{num}(I)) + 1}$$

по условию теоремы. Но в левой части полученного неравенства было целое число, которое оказывается равным нулю как неотрицательное, меньшее 1. Таким образом, алгоритм A' точен, причем $T_{A'}(|I|) = T_{A_{\varepsilon}}(|I|) < p(|I|, p'(\text{num}(I)) + 1)$ по определению ПППС. Следовательно, алгоритм A' псевдополиномиален, что противоречит теореме 1.4 (см билет 10). ■

2 Геометрическая идея симплекс-метода

$$\max \langle c, x \rangle = d = \langle c, x^* \rangle \quad \text{-- решение}$$



Билет 14

1 Определение озЛП. Принцип граничных решений. Алгебраическая и битовая сложность ЛП. Результаты о сложности для задач, близких к ЛП

Определение. *Линейное программирование (ЛП)* изучает теорию, приложение и методы решения конечных систем линейных неравенств с конечным число вещественных неизвестных $A \in \mathbb{Z}^{m \times n}$, $b \in \mathbb{Z}^m$, $x \in \mathbb{R}^n$, A – не содержит нулевых строк:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \leq b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \leq b_2 \\ \dots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \leq b_m \end{cases} \quad (1)$$

Определение. *Основная задача ЛП (озЛП)* состоит в нахождении такого решения x системы линейных неравенств $Ax \leq b$, т.ч. оно максимизирует заданную линейную функцию $\langle c, x \rangle = c_1x_1 + c_2x_2 + \dots + c_nx_n$.

$$d^* = \max_{x \in \mathbb{R}^n : Ax \leq b} \langle c, x \rangle \quad (2)$$

Замечание. грубо говоря - эта задача заключается в нахождении верхушки параллелепипеда (полиэдра).

Определение. озЛП с n неизвестными и m ограничениями называется задачей размерности (n, m) и задается таблицей своих коэффициентов:

$$\begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \dots & \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \\ c_1 & c_2 & \dots & c_n & 0 \end{vmatrix} \quad (3)$$

Замечание. умение решать озЛП предполагает умение решать $Ax \leq b$ *duh*

Замечание. таблица коэффициентов озЛП называется *симплекс таблицей*

Определение. *Каноническая задача ЛП:* $\min_{Ax=b, x \geq \bar{0}} \langle c, x \rangle$ или $\max_{Ax=b, x \geq \bar{0}} \langle c, x \rangle$.

Замечание. ее можно представить в виде озЛП.

Определение. *Принцип граничных решений:* если озЛП имеет решение, то найдется такая подматрица A_I матрицы A , что любое решение системы уравнений $A_I x = b_I$ ($\{a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n = b_i \mid i \in I\}$) реализует максимум в озЛП.

Замечание. если A_I невырожденная, то она угловая точка.

Замечание. метод Гаусса требует не более полинома 3 степени от m, n , но число возможных подматриц растет экспоненциально.

Утверждение. Если угловая точка задачи ЛП существует, то озЛП разрешима, причем в ней.

Утверждение. задачу ЛП можно свести к дискретной принципом граничных решений.

Определение. *Алгебраическая сложность* - функция, оценивающая число арифметических операций.

Определение. *Битовая сложность* - функция, оценивающая число арифметических операций с битами (или с конечными порциями - размером регистра машины) цифровой записи параметров индивидуальной задачи ЛП в зависимости от длины входного слова, то есть от n, m и длин l кодов чисел.

Утверждение. Битовая сложность алгоритма соответствует его временной сложности.

Утверждение. Задача ЛП $\in \mathbf{P}$.

Утверждение. Задача ЛН $\in \mathbf{P}$ (следует из ЛП $\in \mathbf{P}$).

Утверждение. Аналогичные задачи с дополнительным ограничением целочисленности или булевости решения **NP-полны**: БЛН и ЦЛН $\in \mathbf{NPC}$, т.е. полиномиальные алгоритмы для них вряд ли будут построены.

2 Идея метода ветвей и границ. Пример для задачи БЛП

2. *Метод ветвей и границ для ЦЛП.* Рассматривается задача

$$\max_{z \in \mathbf{Z}^n: Az \leq b} \langle c, z \rangle, \quad (1)$$

решением которой является целочисленный вектор z^* .

В корневой вершине метода вместо задачи (1) решается озЛП

$$\max_{x \in \mathbf{R}^n: Ax \leq b} \langle c, x \rangle, \quad (2)$$

решением которой является вектор x^0 . Если x^0 оказался целочисленным, то $z^* := x^0$ — решение задачи (1) закончено. Иначе $\exists x_j^0 \notin \mathbf{Z}$ и осуществляем ветвление по j -й компоненте следующим образом.

Из вершины выходят две ветви, и на новом ярусе к ограничениям озЛП, решаемой в порождающей вершине, добавляется ограничение $x_j \leq \lfloor x_j^0 \rfloor$ для 1-й ветви или $x_j \geq \lceil x_j^0 \rceil$ для 2-й ветви. Значение максимума в исходной задаче ЦЛП (1), очевидно, равно максимальному из значений подзадач ЦЛП на каждой ветви. Но, как и ранее, вместо подзадачи ЦЛП рассматривается подзадача без ограничения целочисленности.

Такая озЛП и решается в очередной порожжденной вершине в случае ее раскрытия, обозначим решение через x^k .

- Если x^k — целочисленное, то вершина закрывается, а значение $\langle c, x^k \rangle$ функции цели сравнивается с рекордом для его обновления или, по первому разу, присваивается рекорду, и точка x^k — допустимая точка в задаче (1) — запоминается. После получения рекорда может быть закрыта любая раскрытая вершина, для которой оптимальное значение целевой функции окажется меньше рекорда. Действительно, поскольку максимум по большему множеству не меньше максимума по меньшему, то значение озЛП дает оценку сверху (*границу*) значения соответствующей целочисленной подзадачи, и когда верхняя оценка не превышает рекорда, бессмысленно пытаться увеличить рекорд на данной ветви.

Другим случаем закрытия вершины (отсечения ветви) является неразрешимость поставленной озЛП и, следовательно, той же подзадачи ЦЛП.

- Если x^k — нецелочисленное, то $\exists x_i^k \notin \mathbf{Z}$, и осуществляем ветвление по i -й компоненте описанным выше способом. Процедура заканчивается после закрытия всех вершин, тогда значение (1) равно

текущему рекорду, либо рекорд остался неопределенным и задача (1) не имеет решения.

Выбор стратегии ветвления в ЦЛП играет не меньшую роль, чем в глобальной оптимизации. Отсутствие рекорда приводит к лишнему перебору, но процедура ветвления “в глубину” может вместо рекорда дать несовместную систему ограничений. Кроме того, для нескольких нецелых компонент x^k не понятно, по какой из них лучше осуществлять ветвление: по новой, которая не рассматривалась на предыдущих ярусах, или сначала перебрать все допустимые целые значения одной из компонент (по аналогии с БЛП — см. ниже). Последняя стратегия имеет смысл при наличии двусторонних ограничений на переменные.

Существует частный класс задач ЦЛП, в которых ограничение целочисленности оказывается несущественным.

Определение. Матрица называется *вполне унимодулярной*, если определитель любой ее невырожденной квадратной подматрицы равен по модулю 1.

Утверждение 11.1. Если матрица ограничений разрешимой задачи ЛП с целыми коэффициентами вполне унимодулярна, то у нее существует целочисленное решение.

Доказательство. Очевидно из принципа граничных решений (§5) и правила Крамера (см. доказательство теоремы 1 §5).

Утверждение 11.2. Матрица A вполне унимодулярна тогда и только тогда, когда для любого целочисленного вектора $b \in \mathbb{Z}^m$ все вершины многогранника $Ax \leq b, x \geq \bar{0}$ являются целочисленными.

Доказательство. В одну сторону аналогично предыдущему, в другую сторону см. ссылку в [2, с. 333].

Таким образом, вполне унимодулярными матрицами ограничений в принципе ограничивается класс задач ЦЛП, эквивалентных ЛП и, следовательно, допускающих эффективное решение. Отметим, что указанный класс, хотя и чрезвычайно узок с формальной точки зрения (элементами матрицы A могут быть только 0, 1 и -1, причем по большей части 0), соответствует достаточно широкому классу практических задач оптимизации на графах и сетях (одно- и двухпродуктовые сети, двудольные графы и т.п.).

Приведем без доказательства еще одно полезное утверждение, позволяющее в некоторых случаях получать приближенное решение ЦЛП путем решения ЛП.

Утверждение 11.3. Если все элементы симплекс-таблицы a_{ij}, b_i, c_j натуральные числа, то для любого решения x^0 задачи ЛП

$$\max_{Ax \leq b, x \geq \bar{0}} \langle c, x \rangle$$

вектор $[x^0]$, составленный из компонент $[x_j^0]$, будет допустимым в данной задаче. При этом для решения x^* соответствующей задачи ЦЛП

$$\max_{Ax \leq b, x \in \mathbb{Z}_+^n} \langle c, x \rangle$$

очевидна оценка $|\langle c, [x^0] \rangle - \langle c, x^* \rangle| \leq \langle c, \bar{1} \rangle$.

Условие положительности исходных данных выполняется для некоторых экономических задач. Такой же результат можно получить для ряда многопродуктовых потоковых задач на сетях и других линейных задач максимизации с положительным c , в которых допустимое множество вместе с любой точкой x содержит и все x' с компонентами $x'_j \in [0, x_j]$. Однако поиск x^* по $[x^0]$ может потребовать перебора 2^n вариантов округления компонент x^0 .

К сожалению, в общем случае и перебора всех возможных вариантов округления компонент решения непрерывной задачи ЛП оказывается недостаточно для получения решения ЦЛП (например, при $n = 2$, если для положительного c рассмотреть систему ограничений $-9x_1 + 10x_2 \leq 0$, $-8x_1 + 10x_2 \leq -1$). Таким образом, поиск решения ЦЛП может потребовать очень большого перебора целочисленных точек, и возникает та же, что и в §10, задача организации перебора с целью попытаться его сократить в случае не самой плохой задачи. Одним из достаточно употребительных методов перебора здесь является метод ветвей и границ, который для ЦЛП будет рассмотрен в п.2. Другие методы см. в [2,6].

3. Метод ветвей и границ для БЛП. Частным случаем задачи (1) ЦЛП является задача БЛП

$$\max_{z \in \mathbf{B}^n: Az \leq b} \langle c, z \rangle, \quad (3)$$

решение которой — вектор z^0 из булева куба.

Из результатов §2 (утверждения 8) вытекает **NP-трудность** БЛП и, следовательно, правомерность использования переборных схем для решения (3). В §12 будет показана схема динамического программирования для БЛП с неотрицательными коэффициентами, а для произвольных задач (3) применима схема предыдущего пункта, которая несколько упрощается за счет дополнительного ограничения $0 \leq z_i \leq 1$, превращающего ЦЛП в БЛП. А именно, после замены \mathbb{Z}^n на \mathbf{B}^n , при ветвлении в новые подзадачи добавляется вместо ограничений неравенств условие равенства 0 (для одной ветви) или 1 (для другой) той переменной, по которой осуществляется ветвление. Таким образом указанная переменная становится булевой во всех нижних ярусах, т.е. по ней не придется вновь проводить ветвление, а значит, на n -м ярусе решение (3) будет закончено. Число раскрываемых вершин (или решений подзадач ЛП) при этом не превысит 2^{n+1} , что, конечно, тоже немало, но заметно меньше, чем для ЦЛП (сравнимо со случаем, предусмотренным утверждением 11.3).

1 Теорема о границах решений задач ЛП с целыми коэффициентами

Теорема 5.1 (о границах решений). Если озЛП размерности (n, m) с целыми коэффициентами разрешима, то у нее существует рациональное решение $x^* \in \mathbb{Q}$ в шаре $\|x\| \leq n^{1/2} \Delta([A|b])$ и значением озЛП $d^* \doteq \langle c, x^* \rangle$ является рациональное число $t/s \in \mathbb{Q}$ со знаменателем, ограниченным величиной $\Delta(A)$.

Доказательство. На основании принципа граничных решений $\exists A_I \subseteq A$: по правилу Крамера $|x_j^*| = |\det A_I^j / \det A_I| \leq \Delta([A|b])$, ибо $|\det A_I| \geq 1$, а определитель матрицы A_I^j , полученной из A_I заменой j -го столбца на $\pm b_I$, не превышает по модулю $\Delta([A|b])$. Отсюда для евклидовой нормы x^* получаем требуемую оценку. С учетом целочисленности вектора $c \in \mathbb{Z}$ знаменатель d^* может быть выбран равным знаменателю $x_j^* \forall j$, и 2-е утверждение теоремы следует из определения $\Delta(A) \geq |\det A_I|$. ■

2 Метод ветвей и границ для ЦЛП. Различные стратегии метода

2. Метод ветвей и границ для ЦЛП. Рассматривается задача

$$\max_{z \in \mathbf{Z}^n: Az \leq b} \langle c, z \rangle, \quad (1)$$

решением которой является целочисленный вектор z^* .

В корневой вершине метода вместо задачи (1) решается озЛП

$$\max_{x \in \mathbf{R}^n: Ax \leq b} \langle c, x \rangle, \quad (2)$$

решением которой является вектор x^0 . Если x^0 оказался целочисленным, то $z^* := x^0$ — решение задачи (1) закончено. Иначе $\exists x_j^0 \notin \mathbf{Z}$ и осуществляем ветвление по j -й компоненте следующим образом.

Из вершины выходят две ветви, и на новом ярусе к ограничениям озЛП, решаемой в порождающей вершине, добавляется ограничение $x_j \leq \lfloor x_j^0 \rfloor$ для 1-й ветви или $x_j \geq \lceil x_j^0 \rceil$ для 2-й ветви. Значение максимума в исходной задаче ЦЛП (1), очевидно, равно максимальному из значений подзадач ЦЛП на каждой ветви. Но, как и ранее, вместо подзадачи ЦЛП рассматривается подзадача без ограничения целочисленности.

Такая озЛП и решается в очередной порожжденной вершине в случае ее раскрытия, обозначим решение через x^k .

- Если x^k — целочисленное, то вершина закрывается, а значение $\langle c, x^k \rangle$ функции цели сравнивается с рекордом для его обновления или, по первому разу, присваивается рекорду, и точка x^k — допустимая точка в задаче (1) — запоминается. После получения рекорда может быть закрыта любая раскрытая вершина, для которой оптимальное значение целевой функции окажется меньше рекорда. Действительно, поскольку максимум по большему множеству не меньше максимума по меньшему, то значение озЛП дает оценку сверху (*границу*) значения соответствующей целочисленной подзадачи, и когда верхняя оценка не превышает рекорда, бессмысленно пытаться увеличить рекорд на данной ветви.

Другим случаем закрытия вершины (отсечения ветви) является неразрешимость поставленной озЛП и, следовательно, той же подзадачи ЦЛП.

- Если x^k — нецелочисленное, то $\exists x_i^k \notin \mathbf{Z}$, и осуществляем ветвление по i -й компоненте описанным выше способом. Процедура заканчивается после закрытия всех вершин, тогда значение (1) равно

текущему рекорду, либо рекорд остался неопределенным и задача (1) не имеет решения.

Выбор стратегии ветвления в ЦЛП играет не меньшую роль, чем в глобальной оптимизации. Отсутствие рекорда приводит к лишнему перебору, но процедура ветвления “в глубину” может вместо рекорда дать несовместную систему ограничений. Кроме того, для нескольких нецелых компонент x^k не понятно, по какой из них лучше осуществлять ветвление: по новой, которая не рассматривалась на предыдущих ярусах, или сначала перебрать все допустимые целые значения одной из компонент (по аналогии с БЛП — см. ниже). Последняя стратегия имеет смысл при наличии двусторонних ограничений на переменные.

Билет 16

1 Теорема о мере несовместности систем линейных неравенств с целыми коэффициентами

Определение. Точка x^ε называется ε -приближенным решением системы линейных неравенств $Ax \leq b$, если $\langle a_i, x^\varepsilon \rangle \leq b_i + \varepsilon \quad \forall i = \overline{1, m}$, где a_i — i -я строка матрицы A , или в матричной записи, обозначая e — вектор-столбец из единиц

$$Ax^\varepsilon \leq b + \varepsilon e \quad (1_\varepsilon)$$

Теорема 5.2 (о мере несовместности). Если система ЛН $Ax \leq b$ имеет ε_1 -приближенное решение для $\varepsilon_1 \doteq 1/[(n+2)\Delta(A)]$, то эта система разрешима, т.е. имеет точное решение x^0 .

Доказательство. Обозначим через ε^* минимальное ε , при котором система (1_ε) имеет решение (по условию $\varepsilon^* \leq \varepsilon_1$):

$$\varepsilon^* \doteq \min_{(x, \varepsilon) : Ax \leq b + \varepsilon e} \varepsilon.$$

Допустим, что утверждение теоремы не верно, тогда $\varepsilon^* > 0$. Задача определения ε^* является (с учетом равенства $\min(\cdot) = -\max(-\cdot)$) озЛП с целевым вектором $c = (0, \dots, 0, -1)$, $n+1$ переменными (x, ε) и ограничениями $Ax - \varepsilon e \leq b$.

Следовательно, по теореме 2.1, ε^* может быть представлена в виде дроби со знаменателем, не превышающим $\Delta([A] - e) \leq (n+1)\Delta(A)$, т.е. $\varepsilon^* \geq 1/[(n+1)\Delta(A)] > \varepsilon_1$ — пришли к противоречию с определением ε^* . ■

Замечание. Аналогичное утверждение справедливо и для озЛП.

Замечание. В $\Delta([A] - e)$ идет разложение по последнему столбцу, откуда и получается неравенство.

Определение. Точка x_ε^* называется ε -приближенным решением озЛП, если она является ε -приближенным решением системы и реализует максимум в ε -точности: $\langle a_i, x_\varepsilon^* \rangle \leq b_i + \varepsilon \quad \forall i = \overline{1, m}$ и $\langle c, x_\varepsilon^* \rangle \geq d^* - \varepsilon$.

Теорема 2* (о мере несовместности) (б/д). Если озЛП имеет ε_2 -приближенное решение для $\varepsilon_2 \doteq 1/(2n^2\Delta^3(A))$, то эта задача имеет точное решение x^* .

2 Метод ветвей и границ для глобальной минимизации Липшицевых функций

Определение. Функция $f(x)$ называется Липшицевой на X с константой L , если $\forall x, x' \in X : |f(x) - f(x')| \leq L\|x - x'\|$. Обозначается $f \in \text{Lip}(X, L)$.

2. Метод ветвей и границ (МВГ) для глобальной минимизации.

Пусть x^1 — центр куба X . Вычисляем $f(x^1)$ и присваиваем это значение рекорду $R := f(x^1)$. Разбиваем куб на 2^n одинаковых подкубов X^{1i} со стороной $1/2$ и вычисляем значения целевой функции в их центрах: $f(x^{1i})$, $i = 1, \dots, 2^n$, обновляя по ходу вычислений значение рекорда $R := \min_i f(x^{1i})$. Проверяем выполнение условия $X^{1i} \subseteq T_{1i}(R)$ для $i = 1, \dots, 2^n$ и отбрасываем соответствующие подкубы. Каждый из оставшихся разбиваем на 2^n одинаковых подкубов X^{2ij} со стороной $1/4$ и поступаем, как прежде. На любом шаге у нас формируется множество \mathbf{K} “кубиков” со сторонами 2^{-l} , $l \geq 2$, целое. Правило выбора очередного кубика для разбиения называется *правилом ветвления* — возможные варианты приводятся ниже. Кубики со стороной не больше $\varepsilon/(L\sqrt{n})$ исключаются из множества \mathbf{K} — дробление кубика заканчивается. Также исключаются кубики, попавшие в множество $T_k(R)$ (с индексом k — номером кубика) для текущего значения рекорда,

— *правило отсечения ветвей*. Рекорд обновляется при получении меньшего значения целевой функции (*правило получения границ, т.е. оценок*). Значения целевой функции вычисляются в центре каждого нового подкубика, включаемого в \mathbf{K} после разбиения выбранного для этого кубика. Алгоритм останавливается, когда \mathbf{K} пусто.

Сокращение. МВГ разбивает исходный куб на граф-дерева подкубов, итеративно исключая бесполезные (которые попали в $T_s(R)$).

Указанная терминология и название метода определяются тем, что визуально данная схема перебора представляется в виде графа-дерева, корневая вершина которого соответствует кубу X , вершины первого яруса — подкубам X^{1i} , вершины второго яруса — кубикам X^{2ij} , подсоединенным к своим *порождающим* вершинам X^{1i} 1-го яруса, и т.д. Если кубик исключается из \mathbf{K} , его вершина *закрывается* — из нее не будут идти ветви на следующий ярус. Если кубик еще не включен в \mathbf{K} , его вершина еще *не раскрыта*. Порядок закрытия вершины определяется правилом отсечения (своим для каждой массовой задачи — см. также в §11), порядок раскрытия — правилом ветвления (своим для каждой индивидуальной задачи). Различают два вида правил ветвления по типу построения дерева решений (выбора вершин для раскрытия): “в ширину”, когда сначала раскрываются все вершины одного яруса до перехода к следующему, и “в глубину” — всякий раз раскрывается лишь одна (обычно с лучшим значением рекорда) вершина на ярусе до конца ветви. На практике реализуют некоторую смесь, например, первое правило, пока хватает машинной памяти (в \mathbf{K} не слишком много элементов), затем переключаемся на второе. Предпочтительность той или иной стратегии ветвления оценивается каждым вычислителем по-своему, исходя из главной задачи метода ветвей и границ — быстрее получить лучший рекорд, чтобы отсечь больше ветвей.

Сокращение. Порядок исключения (закрытия) вершин, ненужных для вычисления кубов, — правило отсечения — зависит от конкретной массовой задачи, а правило раскрытия (вычисления) новых кубов — правило ветвления — от конкретной индивидуальной задачи. Правил ветвления два: в ширину (раскрыть все кубы яруса), в глубину (раскрыть только один с наилучшим рекордом). На практике используют смесь.

В рассматриваемой задаче есть хороший способ улучшения рекорда — локальная оптимизация (см. в §8). Ее имеет смысл проводить из текущей точки, в которой произошло обновление рекорда, например, делая несколько шагов градиентного метода. При этом расположение кубиков менять не надо, просто увеличивается шанс сокращения перебора (отбрасывания больших кубиков).

Отметим, что в худшем случае $f = \text{const}$ ($\cup T_i = \emptyset$) — не удастся отбросить ни одной точки x — и приходим к полному перебору; т.е. указанная в п.1 экспоненциальная оценка точна на классе всех липшицевых функций.

Сокращение. Имеет смысл улучшать рекорд с помощью локальной оптимизации: при обновлении рекорда сделать пару шагов градиентным методом.

Билет 17

1 Следствия систем линейных неравенств. Аффинная лемма Фаркаша (б/д)

Здесь и далее $M = \overline{1, m} = \{1, 2, \dots, m\}$.

Определение. Линейное неравенство

$$\langle c, x \rangle \leq d \quad (4)$$

является *следствием* разрешимой системы линейных неравенств $Ax \leq b$, если для любого x , удовлетворяющего $Ax \leq b$, выполнено (4).

Способ получения неравенств-следствий довольно прост: выберем произвольные $\lambda_i \geq 0 \quad \forall i \in M$, домножим на λ_i каждое i -е неравенство системы $Ax \leq b$ и сложим; получим для вектора

$$c = \sum_{i \in M} \lambda_i a_i \text{ и любого числа } d \geq \sum_{i \in M} \lambda_i b_i$$

что (4) будет следствием (1). Оказывается, других следствий у ЛН не бывает.

Лемма Фаркаша (аффинная) (б/д). Линейное неравенство (4) является следствием разрешимой в вещественных переменных системы ЛН $Ax \leq b$ тогда и **только тогда**, когда существует вектор $\lambda \in \mathbb{R}^m$:

$$c = \sum_{i \in M} \lambda_i a_i, \quad d \geq \sum_{i \in M} \lambda_i b_i, \quad \lambda_i \geq 0 \quad \forall i \in M \quad (5)$$

Замечание. Применяется в МП и везде, где есть задачи на двойственность.

Фан факт: лемма *аффинная* потому что операции сложения и домножение на числа – аффинные.

2 Понятие о временной сложности алгоритмов

Определение. Обозначим $t_A(\sigma)$ время работы над словом $\sigma \in \Sigma^*$ (число шагов) алгоритма A до остановки. *Временной сложностью* алгоритма A решения массовой задачи Π назовем функцию $T_A(\cdot)$, определяемую как

$$T_A(n) = \max_{\sigma \in \Sigma^* : |\sigma| < n} t_A(\sigma) \quad \forall n \in \mathbb{Z}_+$$

Определение. Время работы НДМТ над словом σ как минимальное из времен работы над входом σ ДМТ $A(S)$, принимающих σ с учетом времени прочтения слова S (т.е. его длины).

$$\hat{t}_A(\sigma) = \min_{\{S \mid \sigma \in L_{A(S)}\}} \{|S| + t_{A(S)}(\sigma)\}.$$

Определение. Временной сложностью НДМТ \hat{A} решения массовой задачи Π назовем функцию $\hat{T}_{\hat{A}}(\cdot) : \forall n \in \mathbb{Z}_+$

$$\hat{T}_{\hat{A}}(n) = \max_{\sigma \in \hat{L}(\hat{A}) : |\sigma| < n} \hat{t}_{\hat{A}}(\sigma) = \max_{\{\sigma \in \hat{L}(\hat{A}) : |\sigma| < n\}} \min_{\{S \mid \sigma \in L_{A(S)}\}} \{|S| + t_{A(S)}(\sigma)\}$$

Билет 18

1 Лемма Фаркаша о неразрешимости

Лемма Фаркаша (о неразрешимости). Система ЛН $Ax \leq b$ неразрешима тогда и только тогда, когда разрешима система

$$\sum_{i \in M} \lambda_i a_i = \bar{0}, \quad \sum_{i \in M} \lambda_i b_i \leq -1, \quad \lambda_i \geq 0 \quad \forall i \in M \quad (6)$$

Доказательство. Пусть $Ax \leq b$ неразрешима, тогда из разрешимости системы $\langle a_i, x \rangle + x_{n+1} \leq b_i \quad \forall i \in M$ должно следовать, что $x_{n+1} \leq -\varepsilon < 0$, т.е. следствием этой системы является неравенство $\langle (0, \dots, 0, 1/\varepsilon), (x, x_{n+1}) \rangle \leq -1$ и из аффинной леммы Фаркаша получаем (6) (а также в дополнение $\sum \lambda_i = 1/\varepsilon$).

Если же (6) разрешима, то указанное выше неравенство $\langle \bar{0}, x \rangle \leq -1$ оказывается следствием $Ax \leq b$ и должно выполняться для всех x , удовлетворяющих $Ax \leq b$, значит, таких не существует. ■

2 Понятие о недетерминированно полиномиальных задачах

Определение. *Недетерминированной машиной Тьюринга (НДМТ) \hat{A} определяется как набор обычных – детерминированных – машин Тьюринга (ДМТ) $A(S)$ с алфавитом Σ , где S пробегает все слова из Σ^* :*

$$\hat{A} = \{A(S)\}_{S \in \Sigma^*}$$

Замечание. Слова S в определении НДМТ можно проинтерпретировать как подсказки к решению (догадки), тогда ДМТ $A(S)$ проверяет для входного слова σ подсказку S и в случае правильности останавливается в состоянии ДА. НДМТ \hat{A} проверяет для входного слова σ все возможные подсказки, и если хоть одна правильная догадка существует, то НДМТ останавливается с ответом ДА.

Замечание. проверить решимость задачи с подсказкой на НДМТ можно за полином времени (длина подсказки тоже должна не превосходить полинома от длины входа).

Определение. НДМТ \hat{A} останавливается, когда останавливается первая из ДМТ $A(S)$, принимающая входное слово (останавливается в состоянии ДА).

Определение. Язык НДМТ – множество слов, принимаемых хотя бы одной ДМТ $A(S) \in \hat{A}$:

$$\hat{L}(\hat{A}) = \{\sigma \in \Sigma^* \mid \exists S \in \Sigma^* : \sigma \in L(A(S))\}$$

Замечание. НДМТ не может остановиться с ответом НЕТ (она будет работать вечно).

Определение. НДМТ \hat{A} решает массовую задачу Π с кодировкой e , если их языки совпадают: $L(\Pi, e) = \hat{L}(\hat{A})$.

Определение. Время работы НДМТ над словом σ как минимальное из времен работы над входом σ ДМТ $A(S)$, принимающих σ с учетом времени прочтения слова S (т.е. его длины).

$$\hat{t}_{\hat{A}}(\sigma) = \min_{\{S \mid \sigma \in L_{A(S)}\}} \{|S| + t_{A(S)}(\sigma)\}.$$

Определение. Временной сложностью НДМТ \hat{A} решения массовой задачи Π назовем функцию $\hat{T}_{\hat{A}}(\cdot) : \forall n \in \mathbb{Z}_+$

$$\hat{T}_{\hat{A}}(n) = \max_{\sigma \in \hat{L}(\hat{A}) : |\sigma| < n} \hat{t}_{\hat{A}}(\sigma) = \max_{\{\sigma \in \hat{L}(\hat{A}) : |\sigma| < n\}} \min_{\{S \mid \sigma \in L_{A(S)}\}} \{|S| + t_{A(S)}(\sigma)\}$$

Определение. Класс недетерминированных полиномиальных задач **NP**: такие языки, для которых существует НДМТ решающая задачу Π с кодировкой e , что временная сложность этой НДМТ будет полиномом от любой длины входы нашей задачи:

$$\mathbf{NP} = \{L(\Pi, e) \mid \exists \hat{A} - \text{НДМТ, реш. задачи } \Pi \text{ с кодировкой } e, \exists p(\cdot) - \text{полином} : \hat{T}_{\hat{A}} < p(n) \forall n \in \mathbb{Z}_+\}$$

Определение. Недетерминированной задачей называется такая задача Π , если для нее существует такая кодировка e , что $L(\Pi, e) \in \mathbf{NP}$. Будем писать $\Pi \in \mathbf{NP}$.

Утверждение 1. $\mathbf{P} \subseteq \mathbf{NP}$.

Замечание. доказать, что включение строгое/не строгое невозможно.

Замечание. Задача, для которой можно доказать, что она не полиномиально, можно доказать, что она и не из **NP**. Если же задача из класса **NP**, то доказать, что она не полиномиальна нельзя.

Пример: Недетерминированной полиномиальной задачей является КМ, т.к. проверить догадку (маршрут) можно за полином времени.

Определение. Дополнительной задачей $\bar{\Pi}$ к массовой задаче Π распознавания свойств называется такая задача, которая задает обратный вопрос, нежели Π . Формально: $D(\bar{\Pi}) = D(\Pi)$, $Y(\bar{\Pi}) \setminus Y(\Pi)$.

Определение. $\mathbf{co-P} = \{\bar{\Pi} \mid \Pi \in \mathbf{P}\}$.

Определение. $\mathbf{co-NP} = \{\bar{\Pi} \mid \Pi \in \mathbf{NP}\}$

Утверждение 2. $\mathbf{co-P} = \mathbf{P}$.

Замечание. аналогичного утверждения про **co-NP** и **NP** дать нельзя.

1 Теорема двойственности ЛП

Определение. Двойственной к задаче ЛП на максимум (max) с ограничениями неравенствами в форме озЛП (2) называется следующая задача ЛП на минимум (min) с ограничениями в канонической форме:

$$\min \left\{ \sum_{i \in M} \lambda_i b_i \mid \sum_{i \in M} \lambda_i a_i = c, \lambda_i \geq 0 \quad \forall i \in M \right\}, \quad \text{или в краткой записи} \quad \min_{\lambda A=c, \lambda \geq \bar{0}} \langle \lambda, b \rangle \quad (7)$$

Для того, чтобы построить двойственную к произвольной задаче ЛП, надо представить ее в форме озЛП, применить формулу (7), а затем вернуться к обозначениям исходной задачи.

Теорема 4.4 (двойственности ЛП). Задача ЛП разрешима тогда и только тогда, когда разрешима двойственная к ней. В случае разрешимости оптимальные значения целевых функций в обеих задачах совпадают, т.е. $d^* = d^{**}$, где d^* – значение (2), d^{**} – значение (7).

Доказательство. Проведем для случая озЛП, поскольку любая задача ЛП адекватно представляется в такой форме.

Пусть задача (2) разрешима, тогда (4) является следствием (1) $\forall d \geq d^*$ и не является $\forall d < d^*$, что по аффинной лемме Фаркаша эквивалентно разрешимости (5) при $d \geq d^*$ и неразрешимости (5) при $d < d^*$, т.е. $d^* = \min\{d \mid (5)\}$, а это и есть значение (7).

И наоборот, из разрешимости (7) следует неразрешимость (6), ибо в противном случае min в (7) обращался бы в $-\infty$ (так как прибавление решения (6) к решению (7) дает допустимую точку и уменьшает значение целевой функции (7)). Отсюда получаем разрешимость (1) по лемме Фаркаша о неразрешимости. Кроме того, разрешимость (7) означает разрешимость (5) для любого $d \geq d^{**}$, так что (4) оказывается следствием (1) для $d \geq d^{**}$, и поэтому d^{**} ограничивает сверху значение (2), т.е. максимум в (2) достигается. Таким образом получили разрешимость задачи (2) и можем вернуться к началу доказательства для установления равенства $d^* = d^{**}$. ■

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \leq b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \leq b_2 \\ \dots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \leq b_m \end{cases} \quad (1)$$

$$d^* = \max_{x \in \mathbb{R}^n : Ax \leq b} \langle c, x \rangle \quad (2)$$

$$\langle c, x \rangle \leq d \quad (4)$$

$$c = \sum_{i \in M} \lambda_i a_i, \quad d \geq \sum_{i \in M} \lambda_i b_i, \quad \lambda_i \geq 0 \quad \forall i \in M \quad (5)$$

$$\sum_{i \in M} \lambda_i a_i = \bar{0}, \quad \sum_{i \in M} \lambda_i b_i \leq -1, \quad \lambda_i \geq 0 \quad \forall i \in M \quad (6)$$

$$\min_{\lambda A=c, \lambda \geq \bar{0}} \langle \lambda, b \rangle. \quad (7)$$

2 Метод динамического программирования для БЛП с неотрицательными коэффициентами

2. Примерами разложимых функций могут служить min, max, сумма, произведение (с неотрицательными коэффициентами) и т.п. Исходно метод ДП использовался для оптимизации динамических систем, что нашло отражение в применяемой терминологии. Так, \mathcal{E} соответствует физическому

пространству состояний (возможных координат траектории движения), x_i — управлению в момент времени t_i , воздействие управления на траекторию определяется функцией перехода в следующее состояние, на конечное состояние наложены ограничения принадлежности к \mathcal{E}_T , начальное состояние фиксировано; $f_i(x_i, E)$ — стоимость управления системой, находящейся в состоянии E , f — стоимость всей траектории E_1, \dots, E_{n-1} .

$$F_i(E_{i-1}) = \min_{x_i \in X_i} f_i(x_i, F_{i+1}(h_i(x_i, E_{i-1}))) \quad \forall E_{i-1} \in \mathcal{E} \quad (11)$$

Соотношение (11) означает минимизацию стоимости “хвоста” траектории в каждый момент времени, что согласуется с принципом оптимальности, сформулированным Р. Бэллманом: оптимальная политика управления такова, что для любого начального состояния и любых решений (по выбору управления), принятых на начальных шагах, оставшиеся решения образуют оптимальную политику, начинающуюся с состояния, возникшего в результате этих решений. (Отметим, что в случае строгой монотонности f таким образом можно получить любое решение, в случае нестрогой монотонности — хотя бы одно).

Проиллюстрируем применение метода ДП на примере решения задач БЛП с неотрицательными коэффициентами (элементами симплекс-таблицы). Итак, вернемся к задаче (3)

$$F^* = \max_{z \in \mathbf{B}^n: Az \leq b} \langle c, z \rangle$$

в предположении $a_{ij}, b_i, c_j \in \mathbb{Z}_+$. Обозначим через \bar{a}_j j -й столбец матрицы A . Рассмотрим семейство задач поиска

$$F_k(E) \doteq \max_{z: z_j \in \{0,1\} \quad \forall j=k,\dots,n} \sum_{j=k}^n c_j z_j$$

$$\sum_{j=k}^n \bar{a}_j z_j \leq b - E,$$

где $E \in \mathcal{E} \doteq \{E \in \mathbb{Z}_+^m \mid E_i \leq b_i \quad \forall i = \overline{1, m}\}, \quad k = \overline{1, n}$.

Очевидно, $F^* = F_1(0)$. Возвратное уравнение в данном случае:

$$F_k(E) = \max\{F_{k+1}(E), c_k + F_{k+1}(E + \bar{a}_k)\},$$

$$F_n(E) = \begin{cases} c_n, & E \leq b - \bar{a}_n, \\ 0 & \text{иначе.} \end{cases}$$

Находим $\forall E \in \mathcal{E}$ $F_n(E)$ и соответствующие $x_n(E)$, затем для $k = n-1, \dots, 2$ определяем $F_k(E)$ и реализующие их $x_k(E)$ из возвратного уравнения, вычисляем $F_1(0)$, $x_1(0)$ и далее $x_2(E^1), \dots, x_n(E^{n-1})$ в зависимости от того, какие состояния E^1, \dots, E^{n-1} были в конечном счете использованы при вычислении $F_1(0)$, если посмотреть по всем шагам алгоритма.

Число шагов предложенного алгоритма равно n и на n -м шаге рассматривается $\min\{(b_1 + 1) \cdot \dots \cdot (b_m + 1), 2^{n-1}\}$ состояний, на $(n-1)$ -м — минимум из левой части (равной $|\mathcal{E}|$) и 2^{n-2} и т.п. Так что при больших b метод ДП решает примерно столько же задач, сколько МВГ в худшем случае, однако решаемые задачи здесь проще (проверка ограничений вместо ЛП). Подчеркнем, что процедура ДП не дает способов сокращения перебора, тогда как удачный выбор стратегии ветвления в МВГ (например, на основе имеющейся у вычислителя дополнительной информации или эвристических соображений) позволяет (хотя и не гарантированно) решать задачи большей размерности. Отметим также отсутствие ограничения неотрицательности коэффициентов для работы МВГ. В принципе, возможно комбинирование обеих схем (см. [6]).

Билет 20

1 Сведение озЛП к однородной системе уравнений с ограничением $x > 0, x \neq 0$

Утверждение 7.5. Задача ЛП эквивалентна поиску неотрицательного ненулевого решения однородной системы линейных уравнений.

Доказательство. На основании утверждения 4 озЛП сводится к некоторой системе ЛН (с целыми коэффициентами) относительно вектора вещественных неизвестных y :

$$\hat{P}y = \hat{q}, \quad y \geq \bar{0}, \quad (9)$$

Пусть \hat{P} – матрица $(K \times (N-1))$. Введем параметр \hat{R} , мажорирующий координаты какого-то решения (9) (по теореме о границах решений), если система (9) разрешима. Добавим к (9) неравенство $\langle y, e \rangle = y_1 + \dots + y_{N-1} \leq N\hat{R}$, которое превратим в равенство с помощью дополнительной переменной y_N : $\langle \hat{y}, e \rangle = y_1 + \dots + y_{N-1} + y_N = N\hat{R}$, а (9) переписется как $[\hat{P}|\bar{0}]\hat{y} = \hat{q}, \quad \hat{y} \geq \bar{0}$. Теперь сделаем замену переменных $x := \hat{y}/\hat{R}$ и обозначим $P \doteq N\hat{R}[\hat{P}|\bar{0}] - [\hat{q}|\hat{q}|\dots|\hat{q}]$. Придем к однородной системе $Px = \bar{0}$ с дополнительными ограничениями $x = (x_1, \dots, x_N) \geq \bar{0}, \langle x, e \rangle = N$, что соответствует системе $Px = \bar{0}, \quad x \geq \bar{0}, \quad \langle x, e \rangle > 0$ с решениями-лучами $tx^0 \quad \forall t > 0$, любое из которых пересчитывается в решение исходной системы. ■

Вопрос. Что за решения-лучи?

2 Применение метода динамического программирования для понижения размерности разложимой оптимизационной задачи

Для начала рассмотрим задачу оптимизации, записанную в виде

$$f^* = \min_{g(x,y) \in \mathcal{E}_T \subset \mathbb{R}^m} f(x,y), \quad x \in X \subseteq \mathbb{R}^n, \quad y \in Y(x) \subseteq \mathbb{R}^k. \quad (6)$$

Здесь \mathcal{E}_T называется множеством *терминальных состояний* системы по ассоциации с динамическими системами управления, для оптимизации которых было изобретено ДП. Например, для $g = (g_1, \dots, g_m)$, если ограничения задачи заданы в форме $g_i(x,y) \leq 0 \quad \forall i = \overline{1, m}$, то $\mathcal{E}_T = \mathbb{R}_-^m$.

Пусть f разложима (4), а g разделяема:

$$g(x,y) = h_2(y, h_1(x)), \quad h_1: \mathbb{R}^n \rightarrow \mathbb{R}^m, \quad h_2: \mathbb{R}^{k+m} \rightarrow \mathbb{R}^m,$$

тогда введем $\forall x, y \quad E = h_1(x), \quad E' = h_2(y, E)$ и вычисляем g как $g(x,y) = E'$. Функции h_1, h_2 называются *функциями перехода*, векторы E, E' — *состояниями системы*. Множество всех возможных состояний системы обозначается \mathcal{E} и формально задается так:

- 1) $\mathcal{E} \supset h_1(X) \doteq \{h_1(x) | x \in X\}$,
- 2) $\forall E \in \supset h_1(X) \quad \{h_2(y, E) | y \in Y(X)\} \subset \mathcal{E}$

(множество в качестве аргумента означает объединение по всем аргументам из этого множества).

Рассмотрим для (6) семейство задач поиска

$$F_2(E) = \min_{y: h_2(y,E) \in \mathcal{E}_T} f_2(y),$$

которые нужно решать $\forall E \in \mathcal{E}$. По теореме оптимальности

$$f^* = \min_{x \in X} f_1(x, F_2(h_1(x))).$$

Алгоритм ДП:

$\forall E \in \mathcal{E}$ вычисляем $F_n(E)$ из (12), последовательно для $i = n - 1, \dots, 2$ определяем $F_i(E)$ из (11), (10), затем F^* из (9).

Число шагов алгоритма (решений задач одномерной минимизации) будет порядка $n|\mathcal{E}|$. Таким образом метод ДП имеет смысл применять для задач с не очень большим числом состояний ($|\mathcal{E}|$ мало).

1 Классификация задач математического программирования. Преимущества выпуклого случая

Определение. Задача математического программирования (МП) ставится как:

$$\min_{x \in X} f(x) \quad (1)$$

Здесь требуется найти $\arg \min_{x \in X} f(x) \in \text{Arg} \min_{x \in X} f(x)$, т.е.

$$x^* \in X^* \doteq \{x^* \in X \mid f(x^*) \leq f(x) \quad \forall x \in X\}, \quad \text{и} \quad f^* = f(x^*) \quad (2)$$

Любой такой x^* называется *решением* (1); f^* — *значение* (1), или *оптимальное значение* целевой функции f в задаче (1), X — *множество ограничений* или *допустимое множество*.

Замечание. Задача ЛП, как и задача минимизации функции Кармаркара, является частным случаем задачи МП.

Определение. Классификация задачи оптимизации зависит от природы множества X :

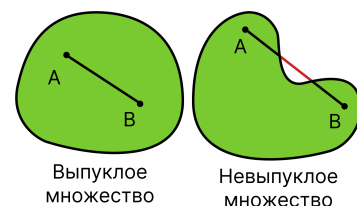
- дискретные (комбинаторные) — X конечно или счетно,
- целочисленные — $x_j \in \mathbb{Z}$,
- булевы — $x_j \in \{0, 1\}$,
- вещественные (непрерывные) — $X \subseteq \mathbb{R}^n$,
- бесконечномерные или в функциональном пространстве, например, когда X — подмножество гильбертова пространства L_2 , и т.п.

Замечание. Мы будем рассматривать в основном задачи с вещественными переменными.

Определение. МП с переменными из \mathbb{R}^n называются *задачами математического программирования* (ЗМП).

Определение. Если $X \subset \mathbb{R}^n$, то говорим о задаче *условной* оптимизации (при условии $x \in X$), иначе ($X = \mathbb{R}^n$) получаем задачу *безусловной* оптимизации.

Определение. Функция f называется *выпуклой* на X , если ее над-график $\text{epigraph}_X f \doteq \{(x, y) \mid y \geq f(x), x \in X\}$ — выпуклое множество. Функция, выпуклая на всей области определения, называется выпуклой. Множество называется выпуклым, если вместе с любыми двумя своими точками оно содержит отрезок, их соединяющий.



Утверждение 2. Любая точка локального минимума выпуклой функции является точкой ее глобального минимума.

Доказательство. Пусть $f(x^0) > f(x^*)$. Тогда $f(x^0) > f(x)$ для всех точек x полуинтервала $(x^0, x^*]$ (по опр. выпуклой функции), а значит, и в некоторой окрестности x^0 — противоречие с опр. локального минимума.

Для решения задач выпуклого программирования применим метод эллипсоидов, причем в гладком случае отсечение полуэллипсоида проводится на основе градиента невыполненного ограничения в полной аналогии с алгоритмом из §6. Поэтому задача поиска ε -приближенного решения задачи выпуклого программирования оказывается полиномиально разрешимой. Для *острых* задач выпуклого программирования — когда функция цели убывает в окрестности минимума не медленнее некоторой линейной функции — можно получить и точное решение.

Утверждение 1. ЦЛН \propto ЗМП.

Доказательство. Поскольку задача ЛН является частным случаем задачи ЛП, то для сведения ЦЛН к ЗМП достаточно представить условие целочисленности переменных в виде ограничений (неравенств) на вещественные переменные, что нетрудно сделать, например, так: $\{x_j \in \mathbb{Z}\}$ эквивалентно $\{x_j \in \mathbb{R} \mid \sin^2(\pi x_j) \leq 0\}$.

2 Полиномиальный алгоритм округления ε_1 -приближенного решения системы линейных неравенств

1. Имея ε -приближенное решение $Ax \leq b$ (1) с $\varepsilon \leq \varepsilon_1 = 1/[(n+2)\Delta(A)]$, можно (на основании теоремы 2, §5, см. билет 16) быть уверенным в существовании точного решения системы линейных неравенств. Оказываясь, процедура получения x^0 из x^{ε_1} является полиномиальной. Соответствующий алгоритм округления ε_1 -приближенного решения системы (1) до точного был указан Л. Г. Хачияном и состоит в следующем.

Определение. Алгоритм округления – полиномиальный алгоритм, с помощью которого из ε_1 -приближенного решения можно получить точное.

Присвоим $x^1 := x^{\varepsilon_1}$ и подставим x^1 в (1). Разобьем множество $M \doteq \{1, \dots, m\}$ индексов неравенств в системе на два подмножества

$$\begin{aligned} M(x^1) &\doteq \{i : |\langle a_i, x^1 \rangle - b_i| \leq \varepsilon_1\}, \\ M \setminus M(x^1) &\doteq \{i : \langle a_i, x^1 \rangle - b_i \leq -\varepsilon_1\} \end{aligned}$$

Найдем решение x'^1 системы равенств $A_{M(x^1)}x = b_{M(x^1)}$ (существует по теореме 2). Пусть x'^1 не является точным решением (1), т.е. в x'^1 не выполнилось i -е неравенство для какого-либо $i \notin M(x^1)$. Тогда введем множество индексов невыполненных неравенств $M^+ \doteq \{i \mid \langle a_i, x'^1 \rangle > b_i\} \subseteq M \setminus M(x^1)$ и рассмотрим на отрезке $[x^1, x'^1]$ ближайшую к x'^1 точку, в которой еще выполнены все неравенства для $i \in M^+$ (в x^1 они выполнены с ε_1 -запасом). А именно определим

$$\tau \doteq \min_{i \in M^+} \frac{b_i - \langle a_i, x^1 \rangle}{\langle a_i, x'^1 \rangle - \langle a_i, x^1 \rangle}, \quad i_1 \doteq \arg \min_{i \in M^+} \frac{b_i - \langle a_i, x^1 \rangle}{\langle a_i, x'^1 \rangle - \langle a_i, x^1 \rangle}$$

и присвоим $x^2 := (1 - \tau)x^1 + \tau x'^1$. Имеем $M(x^2) \supseteq M(x^1) \cup \{i_1\}$, ибо неравенства с индексами из $M(x^1)$ ε_1 -приближенно выполнялись как равенства на всем отрезке $[x^1, x'^1]$, а неравенство с индексом $i_1 \in M^+$, не выполненное в точке x'^1 , выполняется в x^2 как равенство по построению. Таким образом, $M(x^2) \supset M(x^1)$, но $|M(x)| \leq m$, поэтому, повторяя указанную процедуру с заменой x^1 на x^2 и т.д., придем не более чем через $\max(n, m)$ шагов к тому, что решение x' соответствующей системы равенств окажется x^0 – решением (1).

С учетом полиномиальности задачи решения систем уравнений предложенный алгоритм округления полиномиален.

1 Необходимые условия локального минимума при ограничениях-неравенствах для дифференцируемых функций

Определение. Задача МП: $\min_{x \in X} f(x)$ (1)

Определение. $X \doteq \{x \in \mathbb{R}^n \mid g_i(x) \leq 0 \ \forall i \in M\}$ (3)

Определение. Функция f называется *дифференцируемой по Адамару* в точке $x \in \mathbb{R}^n$, если существует вектор $\nabla f(x) \in \mathbb{R}^n$, такой что $\forall y \in \mathbb{R}^n$ выполнено:

$$\lim_{(\tau, y') \rightarrow (+0, y)} \frac{f(x + \tau y') - f(x)}{\tau} = \langle \nabla f(x), y \rangle.$$

Для бесконечномерных задач, когда f — функционал: $E \rightarrow \mathbb{R}^1$, где E некоторое функциональное пространство, требуется: $\nabla f(x) \in E'$ для пространства E' , сопряженного к E , и $x, y \in E$. В гладком случае $\nabla f(x) = \text{grad} f(x)$ и можно положить y' тождественно равным y .

Замечание. Это определение не нужно, достаточно понимать, что f просто n -дифференцируема. А под Δf будем понимать просто $\text{grad} f$.

Определение. Понятие *возможного* или *допустимого* направления в точке $x \in X$ для множества ограничений X как такого вектора y , для которого $\exists \tau^0 > 0 : x + \tau y \in X \ \forall \tau \in [0, \tau^0]$.

Замыкание множества всех допустимых направлений в точке x для X дает определение ниже.

Определение. *Контингентным конусом* к множеству X в точке x называется множество векторов

$$K(X, x) \doteq \{y \mid \exists \{(\tau_t, y^t)\}_{t=1}^\infty : (\tau_t, y^t) \rightarrow (+0, y), \ x + \tau_t y^t \in X \ \forall t\}.$$

Замечание. Очевидно, для $\hat{x} \notin X$ $K(X, \hat{x}) = \emptyset$, а для $x' \in \text{int} X$, $K(X, x') = \mathbb{R}^n$.

Определение. Для $x \in \partial X$ в случае гладкой границы конус $K(X, x)$ называется также *конусом касательных* и соответствует касательным направлениям для ограничений-равенств.

Теорема 9.1 (общий вид необходимых условий локального минимума в задаче (1)).

Пусть функция f дифференцируема по Адамару, $X \subset \mathbb{R}^n$, $X \neq \emptyset$, x^0 — точка локального минимума f в задаче (1), тогда $\forall y \in K(X, x^0) : \langle \nabla f(x^0), y \rangle \geq 0$.

Доказательство. Выберем $y \in K(X, x^0)$. Для соответствующих ему по опр. контингентного конуса $\{\tau_t, y^t\}$ выполнено $x^0 + \tau_t y^t \in X$, и, начиная с достаточно большого t , $x^0 + \tau_t y^t \in X \cap O_\varepsilon(x^0)$ (ибо $\tau_t \rightarrow 0$), следовательно, по определению 1 $f(x^0 + \tau_t y^t) \geq f(x^0)$. В пределе получим

$$\lim_{(\tau, y') \rightarrow (+0, y)} \frac{f(x^0 + \tau y') - f(x^0)}{\tau} = \lim_{(\tau_t, y^t) \rightarrow (+0, y)} \frac{f(x^0 + \tau_t y^t) - f(x^0)}{\tau_t} \geq 0,$$

и требуемое соотношение вытекает из опр. дифференцируемой по Адамаре функции.

Замечание. $O_\varepsilon(x^0)$ — эpsilon окрестность точки x^0 .

Содержательно данные условия означают, что среди допустимых направлений в точке локального минимума не должно быть направлений убывания целевой функции (см. утверждение 3 §8). Однако

в таком общем виде этими условиями неудобно пользоваться.

Конкретизируем полученные условия для ограничений неравенств, когда X задается формулой (3). Введем $\forall x \in X$ множество индексов $J(x) = \{i \in M \mid g_i(x) = 0\}$ — *активных ограничений* в точке x , т.е. таких неравенств из (3), которые в этой точке выполнены как равенства. И определим множество (конус)

$$G(x) \doteq \{y \in \mathbb{R}^n \mid \langle \nabla g_j(x), y \rangle \leq 0 \quad \forall j \in J(x)\}.$$

Определение. Множество X для ограничений неравенств (3) называется *регулярным в точке* $x \in X$, если $G(x) \subseteq K(X, x)$.

Теорема 9.2 (необходимые условия локального минимума с ограничениями неравенствами). Пусть функции $f, g_i \forall i \in M$ дифференцируемы по Адамару, $X \neq \emptyset$, x^0 — точка локального минимума f в задаче (1),(3) и множество X регулярно в точке x^0 . Тогда

$$\exists \lambda_j \geq 0 : \nabla \left\{ f(x^0) + \sum_{j \in J(x^0)} \lambda_j g_j(x^0) \right\} = 0 \quad (5)$$

Доказательство. По теореме 9.1 и из определения регулярности X в x^0 следует, что $\langle \nabla f(x^0), y \rangle \geq 0$ для всех y , удовлетворяющих условию $\langle \nabla g_j(x^0), y \rangle \leq 0 \quad \forall j \in J(x^0)$. Значит, по определению 3 §7, линейное неравенство $\langle \nabla f(x^0), y \rangle \geq 0$ является следствием системы линейных неравенств $\{\langle \nabla g_j(x^0), y \rangle \leq 0 \quad \forall j \in J(x^0)\}$. Приведя это неравенство к стандартному виду $\langle -\nabla f(x^0), y \rangle \leq 0$ и применив аффинную лемму Фаркаша (§7), получим, что

$$\exists \lambda_j \geq 0 : -\nabla f(x^0) = \sum_{j \in J(x^0)} \lambda_j \nabla g_j(x^0) \quad \blacksquare$$

2 Идея метода Кармаркара

Утверждение 7.5. Задача ЛП эквивалентна поиску неотрицательного ненулевого решения однородной системы линейных уравнений.

Доказательство. На основании утверждения 4 озЛП сводится к некоторой системе ЛН (с целыми коэффициентами) относительно вектора вещественных неизвестных y :

$$\hat{P}y = \hat{q}, \quad y \geq \bar{0}, \quad (9)$$

пусть \hat{P} — матрица $(K \times (N-1))$. Введем параметр \hat{R} , мажорирующий координаты какого-то решения (9) (по теореме о границах решений), если система (9) разрешима. Добавим к (9) неравенство $\langle y, e \rangle = y_1 + \dots + y_{N-1} \leq N\hat{R}$,

которое превратим в равенство с помощью дополнительной переменной y_N : $\langle \hat{y}, e \rangle = y_1 + \dots + y_{N-1} + y_N = N\hat{R}$, а (9) перепишется как $[\hat{P}|\bar{0}]\hat{y} = \hat{q}, \quad \hat{y} \geq \bar{0}$. Теперь сделаем замену переменных $x := \hat{y}/\hat{R}$ и обозначим $P \doteq N\hat{R}[\hat{P}|\bar{0}] - [\hat{q}|\hat{q}] \dots [\hat{q}|\hat{q}]$. Придем к однородной системе $Px = \bar{0}$ с дополнительными ограничениями $x = (x_1, \dots, x_N) \geq \bar{0}$, $\langle x, e \rangle = N$, что соответствует системе $Px = \bar{0}, \quad x \geq \bar{0}, \quad \langle x, e \rangle > 0$ с решениями-лучами $tx^0 \quad \forall t > 0$, любое из которых пересчитывается в решение исходной системы.

2. Метод Кармаркара (N. Karmarkar, 1984 г.). Воспользуемся утверждением 5 и обозначениями, введенными при его доказательстве. Пусть $p(x) \doteq (\langle p_1, x \rangle)^2 + \dots + (\langle p_K, x \rangle)^2$, где p_i — строки P . Тогда $p(x) = 0$ эквивалентно $Px = \bar{0}$. Введем функцию Кармаркара

$$k(x) \doteq \frac{[p(x)]^{N/2}}{x_1 x_2 \cdot \dots \cdot x_N}.$$

Применяя теорему 2 и алгоритм округления к задаче решения (9), можно показать, что для точного ее решения достаточно найти такой $\hat{x} > \bar{0}$, для которого $k(\hat{x}) \leq 1/[3(\Delta(\hat{P}))^N]$ [3, с. 25–26].

Сокращение. По сути, используем $Px = \bar{0}$, $x \geq 0$, “порти́м” исходную функцию квадратичной $p(x)$, составляем функцию Кармаркара $k(x)$ и минимизируем уже ее в условии $x > \bar{0}$. Строгое неравенство не даст нам точного решения, но даст нам ε_1 -приближенное, которое потом полиномиально можно свести к точному. Функция $k(x)$ минимизируется итерационными методами (градиентным или Ньютоном).

Полиномиальный алгоритм поиска нужного приближенного \hat{x} приводится в [3, с. 26–28], и мы не будем его описывать. Отметим только, что аналогичный алгоритм может быть построен на основании применения метода Ньютона (см. в разд.3) к задаче минимизации функции Кармаркара или ей подобных. В результате получаем целый класс полиномиальных алгоритмов ЛП, которые на практике оказываются сравнимыми с симплекс-методом, не имея теоретических недостатков последнего. Предложенные алгоритмы строятся на принципиально новой идее: не дискретной, а непрерывной трактовки задачи ЛП, когда вместо перебора конечного числа угловых точек осуществляют поиск решения в исходном пространстве вещественных переменных, и траектории алгоритмов не проходят через угловые точки. Напомним, что метод эллипсоидов также не ориентируется на угловые точки многогранника ограничений. Характерно, что именно такой уход от дискретного программирования позволил построить полиномиальные алгоритмы ЛП. Поэтому далее будет дан некоторый обзор основных подходов к решению непрерывных задач оптимизации.

Замечание. Если бы речь шла о непосредственном поиске точного решения задачи ЛП указанными методами, то нельзя было бы гарантировать конечношаговость (не то, что полиномиальность) соответствующих алгоритмов. Для их применения существенной является возможность остановки в приближенном решении благодаря наличию полиномиального алгоритма округления. Но поскольку для его работы требуется начальное приближение из определенной окрестности решения, зависящей от длины l или высоты h , или длины входа L конкретной задачи ЛП, то и число итераций алгоритмов, базирующихся на рассматриваемом принципе, зависит от числа цифр в записи элементов матрицы ограничений. Так что не удастся использовать данную идею для построения сильнополиномиальных алгоритмов ЛП, кроме как в частных случаях ограниченности элементов матрицы (например, в задачах на графах и сетях, где $a_{ij} = 0, \pm 1$).

1 Понятие о регулярности ограничений-неравенств в задаче математического программирования

В безусловной оптимизации (когда $X = \mathbb{R}^n$) существенную роль играли направления спуска (убывания целевой функции). В условной оптимизации, кроме убывания целевой функции, требуется отслеживать еще и невыход за ограничения. Поэтому вводится понятие *возможного* или *допустимого* направления в точке $x \in X$ для множества ограничений X как такого вектора y , для которого $\exists \tau^0 > 0 : x + \tau y \in X \quad \forall \tau \in [0, \tau^0]$. Замыкание множества всех допустимых направлений в точке x для X дает следующее

Определение. *Контингентным конусом* к множеству X в точке x называется множество векторов

$$K(X, x) \doteq \{y \mid \exists \{(\tau_t, y^t)\}_{t=1}^\infty : (\tau_t, y^t) \rightarrow (+0, y), \quad x + \tau_t y^t \in X \quad \forall t\}.$$

Замечание. Очевидно, для $\hat{x} \notin X$ $K(X, \hat{x}) = \emptyset$, а для $x' \in \text{int} X$, $K(X, x') = \mathbb{R}^n$.

Определение. Для $x \in \partial X$ в случае гладкой границы конус $K(X, x)$ называется также *конусом касательных* и соответствует касательным направлениям для ограничений-равенств.

Теорема 9.1 (общий вид необходимых условий локального минимума в задаче (1)).

Пусть функция f дифференцируема по Адамару, $X \subset \mathbb{R}^n$, $X \neq \emptyset$, x^0 — точка локального минимума f в задаче (1), тогда $\forall y \in K(X, x^0) : \langle \nabla f(x^0), y \rangle \geq 0$.

Доказательство. Выберем $y \in K(X, x^0)$. Для соответствующих ему по опр. контингентного конуса $\{\tau_t, y^t\}$ выполнено $x^0 + \tau_t y^t \in X$, и, начиная с достаточно большого t , $x^0 + \tau_t y^t \in X \cap O_\varepsilon(x^0)$ (ибо $\tau_t \rightarrow 0$), следовательно, по определению 1 $f(x^0 + \tau_t y^t) \geq f(x^0)$. В пределе получим

$$\lim_{(\tau, y') \rightarrow (+0, y)} \frac{f(x^0 + \tau y') - f(x^0)}{\tau} = \lim_{(\tau_t, y^t) \rightarrow (+0, y)} \frac{f(x^0 + \tau_t y^t) - f(x^0)}{\tau_t} \geq 0,$$

и требуемое соотношение вытекает из опр. дифференцируемой по Адамаре функции.

Замечание. $O_\varepsilon(x^0)$ — эpsilon окрестность точки x^0 .

Содержательно данные условия означают, что среди допустимых направлений в точке локального минимума не должно быть направлений убывания целевой функции (см. утверждение 3 §8). Однако в таком общем виде этими условиями неудобно пользоваться.

Конкретизируем полученные условия для ограничений неравенств, когда X задается формулой (3). Введем $\forall x \in X$ множество индексов $J(x) = \{i \in M \mid g_i(x) = 0\}$ — *активных ограничений* в точке x , т.е. таких неравенств из (3), которые в этой точке выполнены как равенства. И определим множество (конус)

$$G(x) \doteq \{y \in \mathbb{R}^n \mid \langle \nabla g_j(x), y \rangle \leq 0 \quad \forall j \in J(x)\}.$$

Определение. Множество X для ограничений неравенств (3) называется *регулярным в точке* $x \in X$, если $G(x) \subseteq K(X, x)$.

Теорема 9.2 (необходимые условия локального минимума с ограничениями неравенствами). Пусть функции $f, g_i \quad \forall i \in M$ дифференцируемы по Адамару, $X \neq \emptyset$, x^0 — точка

локального минимума f в задаче (1),(3) и множество X регулярно в точке x^0 . Тогда

$$\exists \lambda_j \geq 0 : \nabla \left\{ f(x^0) + \sum_{j \in J(x^0)} \lambda_j g_j(x^0) \right\} = 0 \quad (5)$$

Доказательство. По теореме 9.1 и из определения регулярности X в x^0 следует, что $\langle \nabla f(x^0), y \rangle \geq 0$ для всех y , удовлетворяющих условию $\langle \nabla g_j(x^0), y \rangle \leq 0 \quad \forall j \in J(x^0)$. Значит, по определению 3 §7, линейное неравенство $\langle \nabla f(x^0), y \rangle \geq 0$ является следствием системы линейных неравенств $\{\langle \nabla g_j(x^0), y \rangle \leq 0 \quad \forall j \in J(x^0)\}$. Приведем это неравенство к стандартному виду $\langle -\nabla f(x^0), y \rangle \leq 0$ и применив аффинную лемму Фаркаша (§7), получим, что

$$\exists \lambda_j \geq 0 : -\nabla f(x^0) = \sum_{j \in J(x^0)} \lambda_j \nabla g_j(x^0) \quad \blacksquare$$

Таким образом, для регулярных ограничений необходимым условием локального минимума в гладкой задаче (1),(3) является равенство нулю дифференциала функции в фигурных скобках в (5) для хоть каких-нибудь $\lambda_j \geq 0$. Чтобы не записывать в явном виде множество активных ограничений, вводят *функцию Лагранжа*

$$L(\lambda, x) \doteq f(x) + \sum_{j \in M} \lambda_j g_j(x) \doteq f(x) + \langle \lambda, \bar{g}(x^0) \rangle$$

(регулярной) задачи (1),(3), где вектор-функция $\bar{g}(\cdot) \doteq (g_j(\cdot) | j \in M)$. Из теоремы 9.2 следует, что равенство нулю дифференциала функции Лагранжа для $\lambda_j \geq 0$ также является необходимым условием локального минимума в регулярной задаче (1),(3), ибо *множители Лагранжа* λ_j , соответствующие неактивным ограничениям, можно взять равными нулю. Последнее условие записывается как

$$\langle \lambda, \bar{g}(x^0) \rangle = 0 \quad (6)$$

и называется *условием дополняющей нежесткости*. Итак, доказана

Теорема 9.3 (принцип оптимальности Лагранжа). В предположениях теоремы 9.2 для задачи (1),(3) существует неотрицательный вектор множителей Лагранжа $\lambda \geq \bar{0}$, такой, что для x^0 выполнены *условия оптимальности*: $\nabla_x L(x^0, \lambda) = \bar{0}$ и (6).

2 Описание метода эллипсоидов

Утверждение 1. Полуэллипсоид $E^-(g)$ эллипсоида E можно целиком заключить в новый эллипсоид E' , имеющий объем, строго меньший E ,

$$\frac{\text{vol} E'}{\text{vol} E} < e^{-1/(2n+2)}, \quad (*)$$

и E' можно вычислить по $E^-(g)$ за $O(n^2)$ арифметических операций.

Доказательство. Пусть E — единичный шар с центром в точке $\bar{0}$: $E = \{x \in \mathbb{R}^n : \|x\| \leq 1\}$, а $E^-(g) = E \cap \{x_n \geq 0\}$. Поместим центр E' в точку $\xi' = (0, \dots, 0, \frac{1}{n+1})$, тогда

$$E' = \{x | (x_1^2 + \dots + x_{n-1}^2)/\beta^2 + (x_n - \frac{1}{n+1})^2/\alpha^2 \leq 1\},$$

где $\alpha \doteq 1 - 1/(n+1) < e^{-1/(n+1)}$, $\beta^2 \doteq 1 + 1/(n^2 - 1) < e^{1/(n^2 - 1)}$.

Отношение объемов равно произведению полуосей $\alpha\beta^{n-1} < e^{-1/(2n+2)}$, откуда получаем (*), ибо

любой эллипсоид можно превратить в шар аффинным преобразованием координат, сохраняющим объем. Действительно, будем представлять произвольный эллипсоид E с помощью его центра ξ и матрицы Q ($n \times n$), задающей указанное преобразование: $E = \{x \mid x = \xi + Qy, \|y\| \leq 1\}$. Обозначим $\eta \doteq Q^T g / \|Q^T g\|$, где верхний индекс T — знак транспонирования. Тогда ξ' и Q' , представляющие эллипсоид E' минимального объема, описанный вокруг полуэллипсоида $E^-(g)$, пересчитываются по формулам

$$\xi' = \xi - \frac{1}{n+1} Q \eta, \quad Q' = \frac{n}{\sqrt{(n^2-1)}} \{Q + (\sqrt{\frac{n-1}{n+1}} - 1) Q \eta \eta^T\}$$

за $O(n^2)$ арифметических операций.

Метод эллипсоидов получения ε -приближенного решения озЛП.

Положим $\varepsilon := \varepsilon_2 \doteq 1/(2n^2 \Delta^3)$. Введем множество ε -приближенных решений озЛП в шаре радиуса $R \doteq n^{1/2} \Delta$ с центром в $\bar{0}$:

$$X_\varepsilon^* \doteq \{x \mid \langle a_i, x \rangle \leq b_i + \varepsilon \quad \forall i = \overline{1, m}, \quad \langle c, x \rangle \geq d^* - \varepsilon, \quad \|x\| \leq R\}$$

Выберем указанный выше шар в качестве начальной итерации для эллипсоида $E \supset X_\varepsilon^*$. Рассмотрим произвольную итерацию.

Проверяем, является ли центр ξ эллипсоида E ε -приближенным решением. Если да, то алгоритм заканчивает свою работу, в противном случае строим эллипсоид E' для очередной итерации как минимальный по объему эллипсоид, содержащий полуэллипсоид $E^-(g)$, где вектор g определяется следующим образом. Так как $\xi \notin X_\varepsilon^*$, то либо

1⁰) $\exists i : \langle a_i, \xi \rangle > b_i + \varepsilon$, и тогда $g := a_i$, либо

2⁰) $\langle c, \xi \rangle < d^* - \varepsilon$ и $g := -c$.

Убедимся, что при этом $X_\varepsilon^* \subset E'$. Действительно, для варианта 1⁰

$\forall x \in X_\varepsilon^* \quad \langle a_i, x \rangle \leq b_i + \varepsilon < \langle a_i, \xi \rangle$, т.е. $X_\varepsilon^* \subset E \cap \{x \mid \langle a_i, x - \xi \rangle \leq 0\} = E^-(a_i) \subset E'$; и аналогично получим для варианта 2⁰

$$X_\varepsilon^* \subset E \cap \{x \mid \langle c, x - \xi \rangle \geq 0\} = E^-(-c) \subset E'$$

Теперь с $E := E'$ возвращаемся к началу итерации (на новый шаг).

Оценим число итераций метода эллипсоидов. Покажем, что X_ε^* содержит шар радиуса $r/2$, где $r \doteq \varepsilon/(hn^{1/2}) < R$, $h \geq |a_{ij}|, |c_j|$ (h — высота задачи). Пусть x^* — точное решение в X_ε^* . Из $\|x^* - x\| \leq r$ следует $|\langle a_i, x \rangle - \langle a_i, x^* \rangle| \leq \|a_i\| \|x^* - x\| \leq hn^{1/2}r = \varepsilon \quad \forall i \in M$ и $|\langle c, x \rangle - \langle c, x^* \rangle| \leq \|c\| \|x^* - x\| \leq hn^{1/2}r$, т.е. указанный выбор r гарантирует, что все такие x будут ε -приближенными решениями. Поскольку $\|x^*\| \leq R$, то множество тех из рассматриваемых x , для которых $\|x\| \leq R$ (т.е. пересечение шаров радиуса r и R , включающее центр первого), содержит шар радиуса $r/2$. Этот шар и принадлежит X_ε^* . Таким образом, объем X_ε^* больше объема n -мерного шара радиуса $r/2$. Значит, объем эллипсоида, построенного последним, например E^k для k -й итерации, не должен оказаться меньше объема этого шара. Отсюда и из утверждения 1 получаем для k соотношение

$$\left(\frac{r}{2R}\right)^n \leq \frac{\text{vol} X_\varepsilon^*}{\text{vol} E^1} \leq \frac{\text{vol} E^k}{\text{vol} E^1} < e^{-k/(2n+2)}$$

из которого k (по определению r, R, ε, h и Δ) не превосходит

$$2n^2 \ln(Rnh/\varepsilon) < 2n^2 \ln(2n^{3.5} \Delta^5) < 10n^2 \ln(n\Delta)$$

Билет 24

1 Теорема о целочисленности решения задачи ЛП с целыми коэффициентами для вполне унимодулярных матриц ограничений

По-видимому, наиболее важным классом задач глобальной оптимизации являются задачи ЦЛП. Эти задачи формулируются как задачи ЛП с дополнительным ограничением целочисленности переменных. Последнее ограничение, какими бы способами от него ни избавляться, “портит” свойство выпуклости (и полиномиальности) задачи ЛП. Например, выразив условие целочисленности в форме ограничений неравенств, рассмотренной в доказательстве утверждения 1 §8, и сняв их методом штрафов, приходим к задаче глобальной оптимизации, имеющей не меньше локальных экстремумов, чем вариантов для целочисленных переменных в исходной ЦЛП. Поэтому на практике удается решать задачи ЦЛП только небольшой размерности или с ограничениями целочисленности не на все, а лишь на несколько переменных.

Существует частный класс задач ЦЛП, в которых ограничение целочисленности оказывается несущественным.

Определение. Матрица называется *вполне унимодулярной*, если определитель любой ее невырожденной квадратной подматрицы равен по модулю 1.

Утверждение 11.1. Если матрица ограничений разрешимой задачи ЛП с целыми коэффициентами вполне унимодулярна, то у нее существует целочисленное решение.

Доказательство. Очевидно из принципа граничных решений (§5) и правила Крамера (см. доказательство теоремы 1 §5).

Утверждение 11.2. Матрица A вполне унимодулярна тогда и только тогда, когда для любого целочисленного вектора b все вершины многогранника $Ax \leq b, x \geq \bar{0}$ являются целочисленными.

Доказательство. В одну сторону аналогично предыдущему, в другую сторону см. ссылку в [2, с. 333].

Таким образом, вполне унимодулярными матрицами ограничений в принципе ограничивается класс задач ЦЛП, эквивалентных ЛП и, следовательно, допускающих эффективное решение. Отметим, что указанный класс, хотя и чрезвычайно узок с формальной точки зрения (элементами матрицы A могут быть только 0, 1 и -1, причем по большей части 0), соответствует достаточно широкому классу практических задач оптимизации на графах и сетях (одно- и двухпродуктовые сети, двудольные графы и т.п.).

Приведем без доказательства еще одно полезное утверждение, позволяющее в некоторых случаях получать приближенное решение ЦЛП путем решения ЛП.

Утверждение 11.3. Если все элементы симплекс-таблицы a_{ij} , b_i , c_j натуральные числа, то для любого решения x^0 задачи ЛП

$$\max_{Ax \leq b, x \geq \bar{0}} \langle c, x \rangle$$

вектор $[x^0]$, составленный из компонент $[x_j^0]$, будет допустимым в данной задаче. При этом для решения x^* соответствующей задачи ЦЛП

$$\max_{Ax \leq b, x \in \mathbb{Z}_+^n} \langle c, x \rangle$$

очевидна оценка $|\langle c, \lfloor x^0 \rfloor \rangle - \langle c, x^* \rangle| \leq \langle c, \bar{1} \rangle$.

Условие положительности исходных данных выполняется для некоторых экономических задач. Такой же результат можно получить для ряда многопродуктовых потоковых задач на сетях и других линейных задач максимизации с положительным c , в которых допустимое множество вместе с любой точкой x содержит и все x' с компонентами $x'_j \in [0, x_j]$. Однако поиск x^* по $\lfloor x^0 \rfloor$ может потребовать перебора 2^n вариантов округления компонент x^0 .

К сожалению, в общем случае и перебора всех возможных вариантов округления компонент решения непрерывной задачи ЛП оказывается недостаточно для получения решения ЦЛП (например, при $n = 2$, если для положительного c рассмотреть систему ограничений $-9x_1 + 10x_2 \leq 0$, $-8x_1 + 10x_2 \leq -1$). Таким образом, поиск решения ЦЛП может потребовать очень большого перебора целочисленных точек, и возникает та же, что и в §10, задача организации перебора с целью попытаться его сократить в случае не самой плохой задачи. Одним из достаточно употребительных методов перебора здесь является метод ветвей и границ, который для ЦЛП будет рассмотрен в п.2. Другие методы см. в [2,6].

Ссылка была на эту теорему:

Теорема 5.1 (о границах решений). Если озЛП (2) размерности (n, m) с целыми коэффициентами разрешима, то у нее существует рациональное решение x^* в шаре $\|x\| \leq n^{1/2} \Delta([A|b])$ и значением озЛП (2) $d^* \doteq \langle c, x^* \rangle$ является рациональное число t/s со знаменателем, ограниченным величиной $\Delta(A)$.

Доказательство. На основании принципа граничных решений $\exists A_I \subseteq A$: по правилу Крамера $|x_j^*| = |\det A_I^j / \det A_I| \leq \Delta([A|b])$, ибо $|\det A_I| \geq 1$, а определитель матрицы A_I^j , полученной из A_I заменой j -го столбца на $\pm b_I$, не превышает по модулю $\Delta([A|b])$. Отсюда для евклидовой нормы x^* получаем требуемую оценку. С учетом целочисленности вектора c знаменатель d^* может быть выбран равным знаменателю $x_j^* \forall j$, и 2-е утверждение теоремы следует из определения $\Delta(A) \geq |\det A_I|$.

2 Оценка сложности метода эллипсоидов поиска ε_2 -приближенного решения озЛП.

Оценим число итераций метода эллипсоидов. Покажем, что X_ε^* содержит шар радиуса $r/2$, где $r \doteq \varepsilon / (hn^{1/2}) < R$, $h \geq |a_{ij}|, |c_j|$ (h высота задачи). Пусть x^* — точное решение в X_ε^* . Из $\|x^* - x\| \leq r$ следует $|\langle a_i, x \rangle - \langle a_i, x^* \rangle| \leq \|a_i\| \|x^* - x\| \leq hn^{1/2}r = \varepsilon \forall i \in M$ и $|\langle c, x \rangle - \langle c, x^* \rangle| \leq \|c\| \|x^* - x\| \leq hn^{1/2}r$, т.е. указанный выбор r гарантирует, что все такие x будут ε -приближенными решениями. Поскольку $\|x^*\| \leq R$, то множество тех из рассматриваемых x , для которых $\|x\| \leq R$ (т.е. пересечение шаров радиуса r и R , включающее центр первого), содержит шар радиуса $r/2$. Этот шар и принадлежит X_ε^* . Таким образом, объем X_ε^* больше объема n -мерного шара радиуса $r/2$. Значит, объем эллипсоида, построенного последним, например E^k для k -й итерации, не должен оказаться меньше объема этого шара. Отсюда и из утверждения 1 получаем для k соотношение

$$\left(\frac{r}{2R}\right)^n \leq \frac{\text{vol} X_\varepsilon^*}{\text{vol} E^1} \leq \frac{\text{vol} E^k}{\text{vol} E^1} < e^{-k/(2n+2)}$$

из которого k (по определению r, R, ε, h и Δ) не превосходит

$$2n^2 \ln(Rnh/\varepsilon) < 2n^2 \ln(2n^{3.5} \Delta^5) < 10n^2 \ln(n\Delta)$$

УПРАЖНЕНИЕ 6. Оценить по порядку битовую длину L входа озЛП: доказать, что $L > O(\ln(n\Delta))$.

Утверждение. $L > O(\ln(n\Delta))$.

Доказательство. Рассмотрим подробнее вход задачи озЛП:

$$D = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \dots & \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \\ c_1 & c_2 & \dots & c_n & 0 \end{vmatrix} \quad (3)$$

В решении опираемся на некоторое преобразование чисел;

P – подматрица D , оценим сверху $\det P$:

$$|\det P| \leq \prod_i \sum_j |a_{ij}| \leq \prod_{i=1}^n \prod_{j=1}^n (|p_{ij}| + L) \leq L$$

Очевидно, что число n в битовой длине входа ограничено $\lceil \log_2(n+1) \rceil$.

Введем обозначения $\text{blen}(x)$ – битовая длина числа x . Тогда

$$\begin{aligned} L &\geq \sum_{i=1}^m \sum_{j=1}^n (\text{blen}(a_{ij} + L)) + \sum_{i=1}^n (\text{blen}(b_i) + L) + \sum_{j=1}^n (\text{blen}(c_j) + L) + \text{blen}(n) = \\ &= \log_2 \prod_{i=1}^m \prod_{j=1}^n (|a_{ij}| + L) + \log_2 \prod_{i=1}^m (|b_i| + L) + \log_2 \prod_{j=1}^n (|c_j| + L) + \log_2(n) = \\ &= \log_2 \left(\prod_{i=1}^m \prod_{j=1}^n (|a_{ij}| + L) \cdot \prod_{k=1}^m (|b_k| + L) \cdot \prod_{k=1}^n (|c_k| + L) \cdot n \right) = \\ &= \log_2 \left(\prod (|d_{ij}| + L) \cdot n \right) = \log_2(\Delta n) = O(\ln(n\Delta)) \end{aligned}$$

То есть получили $L \geq O(\ln(n\Delta))$. ■

Следовательно, число итераций метода эллипсоидов $k < O(n^2)L$, и с учетом $O(n^2 + nm)$ арифметических операций для каждой итерации получим оценку $O(n^3(n+m)L)$ для числа арифметических операций, достаточного методу эллипсоидов при поиске ε_2 -приближенного решения озЛП. Алгоритм округления ε_2 -приближенного решения до точного этой оценки не портит (см.[3, с. 21]). Можно также показать, что при реализации метода эллипсоидов и алгоритма округления все арифметические операции достаточно проводить с числами двоичной длины, ограниченной $O(L)$. При этом ошибки, возникающие за счет конечности числа разрядов (ошибки округлений), поглощаются путем некоторого дополнительного увеличения (“раздутия”) описанного эллипсоида E' на каждой итерации [3, с. 24], что не влияет на порядок оценки для общего числа итераций. В результате временная сложность такой процедуры решения озЛП оказывается полиномом от длины входа и справедлива

Теорема 6.3. Задача ЛП с целыми коэффициентами разрешима за полиномиальное от длины входа время.

Утверждение 6.2. $\text{ЛН} \in \mathbf{P}$ (как следствие предыдущей теоремы).

Типичные задачи

- КМ - задача коммивояжера. Задается матрицей связности между городами.
 $D(KM) = \{C, \{d(c_i, c_j) \in \mathbb{Z}_+ \mid c_i, c_j \in C, i < j\}, B \in \mathbb{Z}_+\}$.
Типы: минимизации, оптимизации (**NP-hard**), распознавания свойств (**NPC**).

- ПЧ - задача распознавания простоты числа.
Типы: распознавания свойств

- ВВП - задача выполнимости конъюнкции конечного числа дизъюнктивных функций. Задается матрицей степеней $\alpha \in \{-1, 0, 1\}^{n \times m}$, таких, что $z^{-1} = \bar{z}$, $z^0 = 0$, $z^1 = z$ Постановка:

$$K = \bigwedge_{i=1}^n \bigvee_{j=1}^m z_j^{\alpha_i^j} \text{ - КНФ, } \exists z^0 \in \{0, 1\}^m : K(z^0) = 1$$

Типы: распознавания свойств (**NPC**)

- БЛН - задача булевы линейные неравенства. Задается матрицей $A \in \mathbb{Z}^{n \times m}$, вектором $b \in \mathbb{Z}^n$ и значениями $z \in \{0, 1\}^m$: $Az \leq b$.
Типы: оптимизации (**NP-hard**), распознавания свойств (**NPC**)

- ЦЛН - задача о существовании целочисленного решения системы линейных неравенств с целыми коэффициентами. Задается матрицей $A \in \mathbb{Z}^{n \times m}$, вектором $b \in \mathbb{Z}^n$ и значениями $x \in \mathbb{Z}^m$: $Ax \leq b$.
Типы: оптимизации (**NP-hard**), распознавания свойств (**NPC**)

- ЗР - задача о рюкзаке. Задается объемом рюкзака K , количеством вещей n , вектором $c \in \mathbb{Z}_+^n$ ценности вещей, вектором $w \in \mathbb{Z}_+^n$ объема вещей и вектором $z \in \{0, 1\}^n$ - класть вещь или нет.
Постановка: $\sum_{j=1}^n c_j z_j \geq B$, $\sum_{j=1}^n w_j z_j \leq K$ (B - минимальная полезность).
Типы: минимизации, оптимизации (**NP-hard**), распознавания свойств (**NPC**).

- ЛН - задача линейные неравенства. Задается матрицей $A \in \mathbb{Z}^{n \times m}$, вектором $b \in \mathbb{Z}^n$ и значениями $x \in \mathbb{R}^m$: $Ax \leq b$.