# Лекция 4

# Непересекающиеся NP-пары (07.10.2010)

**(Конспект: С. Капулкин)**

## 4.1 Основные определения и свойства

**Определение 4.1.** Будем называть непересекающимися **NP**-парами пары множеств $(A, B)$, где $A, B \in \mathbf{NP}$ и $A \cap B = \emptyset$.

Для **NP**-пары $(A, B)$ ставится задача разделения. По данному слову $x$ требуется определить, принадлежит ли $x$ множеству $A$ или множеству $B$. В первом случае возвращаем единицу, во втором ноль. Если $x$ не лежит ни в $A$, ни в $B$, можно вернуть произвольное значение.

**Замечание 4.1.** Если $A = \overline{B}$, то задача разделения эквивалентна задаче распознавания языка из $\mathbf{NP} \cap \mathbf{co\text{-}NP}$.

Определим сведение для задачи разделения **NP**-пар.

**Определение 4.2.** $(A, B) \to (C, D)$, если существует полиномиально вычислимая функция $f$, такая что $f(A) \subseteq C, f(B) \subseteq D$.

Вопрос существования полных (относительно только что определённого сведения) **NP**-пар является открытым.

**Упражнение 4.1.** Доказать, что из существования полной **NP**-пары следует существование такой полной пары $(A, B)$, что $A, B$ — **NP**-полные.

**Пример 4.1** (**NP**-пара криптосистемы)**.** Рассмотрим криптосистему, передающую сообщения, кодируя их побитно. В ней нули и единицы закодированы словами так, чтобы по данному слову противник не мог бы быстро (за полиномиальное время) определить, какой бит кодирует это слово (в частности, слова эти выбираются вероятностным алгоритмом, всякий раз новое). При этом, зная подсказку (ключ), получатель может быстро определить, код какого бита был ему передан.

Пусть $A$ — множество слов, кодирующих 0, $B$ — множество слов, кодирующих 1. Тогда $(A, B)$ — **NP**-пара. Действительно, $A, B \in \mathbf{NP}$, т.к. существует подсказка

(а именно, случайные биты, использованные кодирующим алгоритмом), с помощью которой можно быстро распознать $A$ и $B$. При этом $A \cap B = \emptyset$, иначе в криптосистеме случались бы ошибки раскодирования.

**Пример 4.2** (Каноническая **NP**-пара системы доказательств)**.**

$$A = \mathtt{SAT}_* = \{(F, 1^t) | F \in \mathtt{SAT}\},$$

множество всех выполнимых формул с дописанным к ним произвольным количеством единиц.

$$B = REF_\Pi = \{(F, 1^t) | F \in \overline{\mathtt{SAT}},\ \exists w\ (|w| < t \wedge \Pi(F, w) = 1)\},$$

множество всех тавтологий с дописанными к ним $t$ единицами так, что длина кратчайшего $\Pi$-доказательства $w$ для $F$ не превосходит $t$.

Очевидно, что множество $A \in \mathbf{NP}$; множество $B \in \mathbf{NP}$, т.к. подсказкой является доказательство $w$, а паддинг в $t$ единичек позволяет нам проверить подсказку за полиномиальное от длины входа время.

## 4.2 Разделимость канонической NP-пары и слабая автоматизируемость

**Определение 4.3.** Система доказательств $\Pi$ автоматизируема, если существует алгоритм, выдающий по входу из языка его доказательство за полиномиальное время от длины кратчайшего доказательства и длины входа.

(В частности, длина выданного доказательства не может оказаться слишком большой.)

**Определение 4.4.** Система доказательств $\Pi$ слабо автоматизируема, если автоматизируема некоторая не менее сильная система доказательств $\Pi' : \Pi' \leq \Pi$.

**Утверждение 4.1.** *Если каноническая* **NP***-пара системы* $\Pi$ *разделима за полиномиальное время, то* $\Pi$ *слабо автоматизируема.*

*Доказательство.* Пусть у нас есть алгоритм $B$, который разделяет **NP**-пару $(\mathtt{SAT}_*, REF_\Pi)$ за полиномиальное время. Построим искомую автоматизируемую систему доказательств.

Доказательством строки $x$ будем считать строку $1^t$, для которой $(x, 1^t) \in REF_\Pi$ (ясно, что эта система не менее сильна, чем $\Pi$, ведь наименьшее такое $t$ — это длина кратчайшего $\Pi$-доказательства по определению $REF_\Pi$). Чтобы породить такое доказательство, автоматизирующий алгоритм $A$ на входе $x$ будет дописывать к слову по 1 единичке и применять разделяющий алгоритм, пока тот не сообщит, что $(x, 1^k) \in REF_\Pi$ для некоторого $k$. (Заметим, что алгоритм $A$ может остановиться даже раньше, чем $k$ сравняется с длиной кратчайшего доказательства для $x$, поскольку $B$ имеет право отвечать всё, что угодно, для входа не из $\mathtt{SAT}_* \cup REF_\Pi$.) Время работы алгоритма $A$ ограничено $O(t \cdot T_B^2(x, 1^t) = poly(|x| + t)$, где $t$ — длина кратчайшего доказательства для $x$ в $\Pi$.

Проверяющий алгоритм просто запускает $B$ и убеждается, что $B$ не утверждает принадлежности входа $\mathtt{SAT}_*$. $\square$

## 4.3 Моделирование систем vs сводимость NP-пар

**Теорема 4.1.** *Пусть $S$ и $W$ — системы доказательств, и $S \le W$. Тогда* $(\mathtt{SAT}_*, REF_W) \to (\mathtt{SAT}_*, REF_S)$.

*Доказательство.* Поскольку $S \le W$, имеется полином $p$, ограничивающий размер доказательств в $S$ относительно размера доказательств в $W$ и длины входа. Искомое сведение: $R(F, 1^t) = (F, 1^{p(t,|F|)})$. □

Утверждение, обратное теореме 4.1, неверно: из сводимости **NP**-пар моделируемость систем доказательств не следует. Приведём контрпример, подтверждающий это.

**Пример 4.3.** Рассмотрим системы доказательств $CP$ и $CP^2 = CP + \{a_{x,y} = x \wedge y | x, y$ — старые переменные$\}$. То есть система доказательств $CP^2$ образована из системы $CP$ добавлением новых переменных $a_{x,y}$ для всех попарных коньюнкций, что выражается следующими новыми аксиомами, которые являются непосредственным переводом соответствующих дизъюнкций:

$$a_{x,y} < x$$
$$a_{x,y} < y$$
$$x + y \le a_{x,y} + 1$$

Докажем, что у $CP$ и $CP^2$ одинаковые **NP**-пары, но при этом $CP^2 < CP$.

**Утверждение 4.2.** *$CP$ и $CP^2$ образуют эквивалентные канонические* **NP***-пары.*

*Доказательство.* Так как доказательство в $CP$ является доказательством в $CP^2$, $(\mathtt{SAT}_*, REF_{CP}) \to (\mathtt{SAT}_*, REF_{CP^2})$. Обратное сведение добавляет к формуле новые аксиомы (в виде дизъюнкций, а не неравенств), соответствующие новым переменным. □

Докажем, что $CP^2 < CP$. Для этого достаточно предоставить бесконечную последовательность формул, для которых есть доказательства полиномиальной длины в $CP^2$, а для доказательств в $CP$ существует нижняя экспоненциальная оценка. В качестве такой формулы рассмотрим следующее утверждение: граф, содержащий клику размера $n$, нельзя правильно раскрасить в $n-1$ цвет.

Это утверждение мы запишем в виде формулы следующим образом. Условие, что граф содержит $n$-клику, запишем в виде гомоморфизма $q : K_n \to G$: отображения вершин, которое переводит вершины $K_n$ в вершины $G$ так, что если между двумя вершинами $K_n$ было ребро, то оно будет и $G$. При этом мы запрещаем графу $G$ иметь петли. Второй гомоморфизм $r : G \to K_{n-1}$ будет описывать раскраску $G$ в $n-1$ цвет. Этот гомоморфизм так же переводит вершины графа $G$ в вершины полного графа $K_{n-1}$ с сохранением ребер. Вершины $K_{n-1}$ соответствуют цветам. Если мы отправляем ребро из $G$ в $K_{n-1}$, то мы правильно красим его, поскольку иначе в $K_{n-1}$ появится петля.

В новых терминах, утверждение о графе будет звучать следующим образом: не существует одновременно двух гомоморфизмов $q : K_n \to G$, $r : G \to K_{n-1}$. Осталось записать записать это утверждение в виде формулы.

Пусть $G = (V, E)$, $|V| = m$. Будем использовать переменные

- $q_{ki}$ — $k$-ая вершина $K_n$ переведена в $i$-ую вершину $G$,
- $p_{ij}$ — $\{i, j\} \in E$,
- $r_{i,l}$ — $i$-ая вершина $G$ покрашена в $l$-ый цвет.

Запишем условия.

1. Каждая вершина клики отправлена на свое место в граф: $\sum_{i=1}^{m} q_{ki} \geq 1$.
2. ... на свое персональное место: это можно записать, как $q_{ki} + q_{k'i} \leq 1$ $(k \neq k')$, но воспользовавшись приемом из доказательства принципа Дирихле, можно из этого сразу получить $\sum_{k=1}^{n} q_{ki} \leq 1$.
3. ... и только одно: аналогичным образом $\sum_{i=1}^{m} q_{ki} \leq 1$.
4. Между двумя вершинами клики есть ребро, и оно окажется в графе $G$: $q_{ki} + q_{k'j} \leq p_{ij} + 1, (k \neq k', i < j)$.
5. Каждая вершина покрашена: $\sum_{l=1}^{n-1} r_{il} \geq 1$.
6. Корректность раскраски: $p_{ij} + r_{il} + r_{jl} \leq 2$.

**Теорема 4.2.** *Формула, предложенная выше, имеет полиномиальное доказательство в $CP^2$.*

*Доказательство.* Построим полиномиальное доказательство формулы в $CP^2$. Идея доказательства в том, что композиция $q$ и $r$ противоречит принципу Дирихле. Введем обозначение: $x_{kl}$ — $k$-ая вершина $K_n$ покрашена в $l$-ый цвет; $x_{kl}$ можно выразить в виде следующей формулы:

$$x_{kl} = \sum_{i=1}^{m} q_{ki} r_{il}. \tag{4.1}$$

В доказательстве, которое мы предложим, вместо $q_{ki}r_{il}$ нужно подставить новые переменные $a_{q_{ki},r_{il}}$, имеющиеся в $CP^2$, но для удобства, чтобы не работать со столь громоздкими обозначениями, в тексте их использовать не будем.

Стоит заметить, что $x_{kl}$ не является новой переменной. Это лишь обозначение формулы 4.1, которую мы и будем подставлять в доказательство на место $x_{kl}$. Однако $x_{kl}$ принимает лишь значение 0 и 1, поэтому в доказательстве *теоремы* мы можем обращаться с ней как с переменной. Действительно, то, что $x_{kl} \geq 0$, следует из аксиом и построения формулы. То что $x_{kl} \leq 1$, можно получить из аксиом и формул $q_{ki}r_{il} \leq q_{ki}$ и условия 1.

Опишем условие успешной раскраски $K_n$ в $n-1$ цвет:

$$x_{kl} + x_{k'l} \leq 1 \quad — \quad \text{каждая вершина покрашена в свой цвет,} \tag{4.2}$$

$$\sum_{l=1}^{n-1} x_{kl} \geq 1 \quad — \quad \text{каждая вершина в какой-нибудь цвет покрашена.} \tag{4.3}$$

Нам нужно построить вывод этих формул, далее мы сможем воспользоваться полиномиальным доказательством принципа Дирихле относительно “переменных” $x_{kl}$ в $CP$.

Распишем условие 4.3: $\sum_{l=1}^{n-1} x_{kl} = \sum_{i=1}^{m}\sum_{l=1}^{n-1} q_{ki}r_{il} = \sum_{i=1}^{m} q_{ki}(\sum_{l=1}^{n-1} r_{il})$. Здесь можно заметить, что по условиям 1 и 5 $\sum_{l=1}^{n-1} r_{il} \geq 1$ и $\sum_{i=1}^{m} q_{ki} \geq 1$ выводится искомое неравенство. Но данный вывод не в $CP^2$ (используется умножение сумм), построим вывод в $CP^2$.

| | |
|---|---|
| $q_{ki}r_{il} \geq q_{ki} + r_{il} - 1$ | аксиомы для переменных $a_{q_{ki}r_{il}}$ - немного перепишем |
| $q_{ki}r_{il} + \overline{q_{ki}} \geq r_{il}$ | суммируем по $l$, условие 5 |
| $\sum_{l=1}^{n-1} q_{ki}r_{il} + (n-1)\overline{q_{ki}} \geq 1$ | прибавим $n-2$ раз $\sum_{l=1}^{n-1} q_{ki}r_{il} \geq 0$ |
| $(n-1)\sum_{l=1}^{n-1} q_{ki}r_{il} + (n-1)\overline{q_{ki}} \geq 1$ | поделим с округлением на $n-1$ |
| $\sum_{l=1}^{n-1} q_{ki}r_{il} + \overline{q_{ki}} \geq 1$ | суммируем по $i$ |
| $\sum_{i=1}^{m}\sum_{l=1}^{n-1} q_{ki}r_{il} + \sum_{i=1}^{m}\overline{q_{ki}} \geq m$ | что можно переписать, перейдя от $\overline{q_{ki}}$ к $q_{ki}$ |
| $\sum_{i=1}^{m}\sum_{l=1}^{n-1} q_{ki}r_{il} \geq 1$ | здесь мы еще воспользовались условием 1 |

Итак, вывод для неравенства 4.3 представлен.

Распишем неравенство 4.2. $x_{kl} + x_{k'l} = \sum_{i=1}^{m} q_{ki}r_{il} + \sum_{j=1}^{m} q_{k'j}r_{jl} \leq 1$. Чтобы доказать, что указанная сумма не превосходит единицу, достаточно показать, что попарная сумма всех слагаемых не превосходит единицу. Далее, так же, как в доказательстве принципа Дирихле, получим неравенство для всей суммы. Рассмотрим отдельно различные виды попарных сумм.

Сумма для разных $k$ и разных $i$: $q_{ki}r_{il} + q_{k'j}r_{jl}$:

| | |
|---|---|
| $\dfrac{q_{ki} + q_{k'j} \leq p_{ij} + 1 \quad p_{ij} + r_{il} + r_{jl} \leq 2}{q_{ki} + q_{k'j} + r_{il} + r_{jl} \leq 3}$ | условия 4 и 6 |
| $\dfrac{q_{ki}r_{il} \leq q_{ki} \quad q_{ki}r_{il} \leq r_{il}}{2q_{ki}r_{il} \leq q_{ki} + r_{il}}$ | аксиомы для новых переменных, складываем |
| $2q_{ki}r_{il} + 2q_{k'j}r_{jl} \leq q_{ki} + r_{il} + q_{k'j} + r_{jl}$ | складываем для разных $k$ и $j$ |
| $2q_{ki}r_{il} + 2q_{k'j}r_{jl} \leq 3$ | с учетом предыдущего неравенства |
| $q_{ki}r_{il} + q_{k'j}r_{jl} \leq 1$ | делим на два |

Сумма для разных $k$ и одинаковых $i$: $q_{ki}r_{il} + q_{k'i}r_{il}$:

| | |
|---|---|
| $\dfrac{q_{ki}r_{il} \leq q_{ki} \quad q_{ki} + q_{k'i} \leq 1}{q_{ki}r_{il} + q_{k'i}r_{il} \leq 1}$ | аксиомы для новых переменных и условие 2 |

Сумма для одинаковых $k$ и разных $i$: $q_{ki}r_{il} + q_{kj}r_{jl}$:

| | |
|---|---|
| $\dfrac{q_{ki}r_{il} \leq q_{ki} \quad \sum_{i=1}^{m} q_{ki} \leq 1}{q_{ki}r_{il} + q_{kj}r_{jl} \leq 1}$ | аксиомы для новых переменных и условие 3 |

Итак, вывод для неравенства 4.3 представлен. □

**Замечание 4.2.** Нижняя оценка на длину доказательства данной формулы в $CP$ будет рассмотрена в следующей лекции.

## 4.4 Интерполяционная NP-пара

**Теорема 4.3** (Интерполяционная теорема). *Пусть для любых $\vec{x}, \vec{y}, \vec{z}$ верно что, $A(\vec{x}, \vec{y}) \supset B(\vec{x}, \vec{z})$, где знаком $\vec{a}$ обозначает набор переменных $a_1, a_2, \ldots$; $A$, $B$ — формулы пропозициональной логики. Тогда существует формула (*интерполянт*) $C(\vec{x})$, такая, что $A(\vec{x}, \vec{y}) \supset C(\vec{x})$ и $C(\vec{x}) \supset B(\vec{x}, \vec{z})$.*

*Доказательство.* Будем вычислять значение $C(\vec{x_0})$.

Если $A(\vec{x_0}, \vec{y})$ для некоторых $\vec{y}$ принимает значение 1, то $B(\vec{x_0}, \vec{z})$ для всех $\vec{z}$ должен быть равен 1. Иначе найдутся такие $\vec{y}$ и $\vec{z}$, что $A(\vec{y}) = 1$, $B(\vec{z}) = 0$ и, т.о. импликация в условии теоремы не выполнится. В таком случае $C(\vec{x_0})$ можно взять равным 1.

Если $B(\vec{x_0}, \vec{z})$ для некоторых $\vec{z}$ принимает значение 0, то, аналогично, $A(\vec{x_0}, \vec{y})$ для всех $\vec{y}$ должно быть равным 0. В таком случае $C(\vec{x_0})$ можно взять равным 0.

Остаётся построить по полученной таблице истинности формулу. □

**Замечание 4.3.** Построенная в доказательстве теоремы формула для $C$ перечисляет значения для всех $\vec{x}$ и имеет экспоненциальный размер.

**Определение 4.5** (Интерполяционная **NP**-пара)**.** $I_b = \{(F_0, F_1, \pi) | Vars(F_0) \cap Vars(F_1) = \emptyset,\ \Pi(F_0 \vee F_1, \pi) = 1,\ F_b \neq TAUT\}$ — множество пар формул с доказательством $\pi$ для их дизъюнкции $F_0 \vee F_1$. При этом одна из формул не является тавтологией, а вторая (т.к. переменные у формул разные) является. Интерполяционной **NP**-парой называется пара $(I_0, I_1)$.

Докажем, что определение корректно. Множества не пересекаются, т.к. иначе в пересечении две формулы от разных переменных не являются тавтологиями, а их дизъюнкция — тавтология. Из доказательства $\pi$ же следует, что хотя бы одна формула тавтологией быть должна. $I_b \in \mathbf{NP}$, т.к. подсказкой является выполняющий набор для отрицания формулы $F_b$. Используя выполняющий набор, мы проверяем формулу $F_b$, и далее остается лишь проверить корректность доказательства $\pi$.

Для интерполяционной пары также стоит вопрос, возможно ли разделить ее за полиномиальное время.

**Определение 4.6.** Система доказательств обладает свойством эффективной интерполяции, если по доказательству $\pi$ для выражения $G_0(\vec{x}, \vec{y}) \vee G_1(\vec{x}, \vec{z})$ за полиномиальное время можно построить (полиномиальную) схему $C : G_{C(x)} \notin \text{TAUT}$.

Схема $C$, по сути, является интерполянтом (если её переписать как формулу; поскольку мы можем использовать дополнительные переменные, размер возрастёт лишь полиномиально).

Вопрос о разделимости интерполяционной **NP**-пары системы доказательств $\Pi$ “почти эквивалентен” вопросу существования у $\Pi$ свойства эффективной интерполяции.

*Доказательство.* $\Longleftarrow$. Чтобы разделить интерполяционную **NP**-пару, достаточно по $F_0, F_1, \pi$ понять, какая из формул $F_i$ не является тавтологией. По доказательству $\pi$ мы можем за полиномиальное время построить разделяющую схему $C$; за отсутствием общих переменных у $F_0$ и $F_1$ эта схема будет константой (искомым ответом на вопрос о разделении).

$\Longrightarrow$. У нас имеется алгоритм $A$ для “разделения” формул без общих переменных (при наличии доказательства), а нам нужен для формул $G_0$, $G_1$ с общими переменными (тогда мы сможем преобразовать его в схему). Однако нам дан вход $\vec{x}$, который можно без труда подставить в $G_0 \vee G_1$. Остаётся применить $A$ к полученным двум формулам и доказательству дизъюнкции; однако у нас имеется доказательство

$G_0 \vee G_1$ *до* подстановки, а не после. Тем не менее, *если* система устойчива относительно подстановок (т.е. доказательство остаётся верным после подстановки), это не составляет проблемы.

□

**Замечание 4.4.** Все известные системы устойчивы относительно подстановок; более того, любая система очень просто превращается в устойчивую: достаточно разрешить принимать в качестве доказательств доказательства старой системы, снабжённые списком подстановок. Однако формально эквивалентности *для всех систем* всё же нет.

**Определение 4.7** (Reflection property)**.** Reflection property говорит о том, что у данной системы доказательств есть полиномиальное доказательство ее корректности: существует полиномиальное доказательство для "формулы"

$$\Pi(F, \pi) \neq 1 \vee F[A] \neq 1, \tag{4.4}$$

где формула $F$, доказательство $\pi$, набор $A$ заданы векторами булевых переменных нужной длины. Потребуем также, чтобы это доказательство порождалось за полиномиальное время (на вход будут поданы размеры $F$, $A$, $\pi$ в унарной записи). Кавычки вокруг слова "формула" выше означают, что речь идёт о булевой формуле $R(F, A, \pi, \vec{v})$, эквивалентной данной: $\forall\vec{v}\forall F\forall A\forall\pi\ R(\ldots) \equiv (4.4)$, где $\vec{v}$ — вектор дополнительных переменных, а $F, A, \pi$ записаны булевыми переменными некоторым разумным образом.

Формула из определения читается следующим образом: для любых $F$, $\pi$, $A$ верно, что или доказательство $\pi$ не проходит проверку или $A$ не является выполняющим набором. Для корректных систем доказательств данное утверждение всегда верно, однако, оно не обязательно имеет доказательство полиномиальной длины. Также данное нами определение оставляет свободу выбора точной формулировки в виде формулы (аналогично теореме Кука легко видеть, что размер формулировки можно сделать полиномиальным, но всё же это можно сделать по-разному, в частности, с разными дополнительными переменными).

**Следствие 4.1.** *Для систем, устойчивых относительно подстановок и обладающих reflection property, интерполяционная и каноническая пары эквивалентны друг другу.*