Сложность вычислений
Экстрактор Тревисана.
Повышение трудности функции.
Лемма о трудном распределении Импальяццо.
XOR-лемма Яо.

Дмитрий Ицыксон

ПОМИ РАН

21 апреля 2011

Экстрактор из псевдослучайного генератора

- Мы строили по трудной функции псевдослучайный генератор.
- Конструкция пользовалась функцией как черным ящиком.
- По схеме, отличающей выход псевдослучайного генератора от равномерного распределения, можно построить схему, вычисляющую f.

Псевдослучайный генератор

Теорема. (Теорема Нисана-Вигдерсона, переформулировка) Для любой правильной функции $S : \mathbb{N} \to \mathbb{N}$, $\exists c > 0, \epsilon > 0$ алгоритмы G и R

- По функции $f : \{0,1\}^\ell \to \{0,1\}$ и строке $z \in \{0,1\}^{c\ell}$, алгоритм $G^f(z)$ работает время $2^{O(\ell)}$ и выдает строку длины $m = (S(\ell))^\epsilon$
- Если $h : \{0,1\}^m \to \{0,1\}$ функция, для которой $\left|\Pr[h(G^f(U_{c\ell})) = 1] - \Pr[h(U_m) = 1]\right| \geq \frac{1}{20}$, тогда существует подсказка a размера $S(\ell)^{1/4}$, что для всех x: $R^h(a, x) = f(x)$ и R работает время $S(\ell)^{1/4}$.

Экстрактор Тревисана

Лемма. $f \in \{0,1\}^n$, $z \in \{0,1\}^{c\ell}$, $\ell = \log n$. G — генератор из предыдущей теоремы, $S = k$. $f : \{0,1\}^\ell \to \{0,1\}$. Тогда $\mathrm{Ext}(f, z) \mapsto G^f(z)$ — это $(k, \frac{1}{5})$-экстрактор.

Доказательство.

- Пусть D — это (n,k)-источник, пусть h $\frac{1}{10}$-различает $\mathrm{Ext}(D, U_{c\ell})$ от U_m.
- С вероятностью $\geq \frac{1}{20}$ по $f \leftarrow D$ функция h $\frac{1}{20}$-различает $G^f(U_{c\ell})$ от U_m. Назовем такие f плохими.
- Для каждой плохой функции есть подсказка a размера $k^{1/4}$, что f вычислим $x \mapsto R^h(a,x)$.
- Значит, плохих f не больше, чем $2^{k^{1/4}}$.
- $D(\text{плохие } f) \leq 2^{k^{1/4}} 2^{-k} << \frac{1}{20}$. Противоречие.

План

$$H^{\rho}_{avg}(f) = \max\{S \mid \forall C, |C| \leq S \implies \Pr_{x \leftarrow U_n}[C(x) = f(x)] < \rho\}$$

$$H_{wrs}(f) = H^{1}_{avg},\ H_{avg}(f) = \max\left\{S \mid H^{\frac{1}{2}+\frac{1}{S}}_{avg}(f) \geq S\right\}$$

- По f' с большой $H^{1-\delta}_{avg}(f')$ строим f'' с большой $H^{\frac{1}{2}+\epsilon}_{avg}(f')$.
  - *XOR*-лемма Яо
- По f с большой $H_{wrs}(f)$ строим f' с большой $H^{1-\delta}_{avg}(f')$.
  - Коды, исправляющие ошибки

Лемма Импальяццо

- Пусть $H^{1-\delta}_{avg}(f) \geq S$
- Каждая схема размера $\leq S$ ошибается хотя бы на $\delta 2^n$ входах
  1. Для каждой схемы эти трудные входы свои
  2. Трудные входы — общие для всех схем

Лемма. (Лемма Импальяццо о трудном распределении) Пусть $H^{1-\delta}_{avg}(f) \geq S$, тогда существует такое распределение H на $\{0,1\}$, что

1. $\forall x \in \{0,1\}^n, H(x) \leq \frac{1}{\delta} 2^{-n}$ (плотность H равна δ)
2. $\forall$ схемы C размера $\leq \frac{\epsilon^2 S}{100n}$: $\Pr_{x \leftarrow H}[C(x) = f(x)] \leq \frac{1}{2} + \epsilon$

MIN-MAX теорема

q_1	a_{11}	a_{12}	.	.	a_{1m}
q_2	a_{21}	a_{22}	.	.	a_{2m}
.	.	.	.	.	.
.	.	.	.	.	.
q_n	a_{n1}	a_{n2}	.	.	a_{nm}
	p_1	p_2	.	.	p_m

- $\sum_{i=1}^{m} p_i = 1$, $\sum_{i=1}^{n} q_i = 1$
- a_{ij} — выигрыш строчного игрока
- $qAp = \sum_{i,j} p_i q_j a_{ij}$ — матожидание выигрыша

Теорема. Существуют такие стратегии p^*, q^*, что для всех стратегий p, q выполняются неравенства:

$$qAp^* \leq q^*Ap^* \leq q^*Ap$$

Доказательство леммы Импальяццо

Лемма. (Лемма Импальяццо о трудном распределении) Пусть $H^{1-\delta}_{avg}(f) \geq S$, тогда существует такое распределение H на $\{0,1\}$, что

1. $\forall x \in \{0,1\}^n, H(x) \leq \frac{1}{\delta} 2^{-n}$ (плотность H равна δ)
2. $\forall$ схемы C размера $\leq \frac{\epsilon^2 S}{100n}$: $\Pr_{x \leftarrow H}[C(x) = f(x)] \leq \frac{1}{2} + \epsilon$

Доказательство.

- Дима выбирает схему C размера $\leq S' = \frac{\epsilon^2 S}{100n}$, Эдик выбирает распределение H плотности δ.
- Эдик платит Диме $\Pr_{x \leftarrow H}[C(x) = f(x)]$ рублей.
- Распределение на распределениях плотности δ — это распределение плотности δ

Продолжение доказательства

- Дима выбирает схему C размера $\leq S' = \frac{\epsilon^2 S}{100n}$, Эдик выбирает распределение H плотности δ.
- Эдик платит Диме $\Pr_{x \leftarrow H}[C(x) = f(x)]$ рублей.
- Распределение на распределениях плотности δ — это распределение плотности δ
- Пусть $\mathcal{C}, H^*$ — это решение игры.
- $\forall H$ плотности δ $\Pr_{C \leftarrow \mathcal{C}, x \leftarrow H}[C(x) = f(x)] \geq \Pr_{C \leftarrow \mathcal{C}, x \leftarrow H^*}[C(x) = f(x)] \geq \frac{1}{2} + \epsilon$
- $x \in \{0,1\}^n$ плохой, если $\Pr_{C \leftarrow \mathcal{C}}[C(x) = f(x)] \leq \frac{1}{2} + \frac{\epsilon}{2}$
- Плохих x меньше $\delta 2^n$
- $C_1, C_2, \ldots, C_t \leftarrow \mathcal{C}$, $t = \frac{50n}{\epsilon^2}$, $C = majority(C_1, C_2, \ldots, C_n)$.

Продолжение доказательства

- $\forall H$ плотности δ $\Pr_{C \leftarrow \mathcal{C}, x \leftarrow H}[C(x) = f(x)] \geq \frac{1}{2} + \epsilon$
- $x \in \{0,1\}^n$ плохой, если $\Pr_{C \leftarrow \mathcal{C}}[C(x) = f(x)] \leq \frac{1}{2} + \frac{\epsilon}{2}$
- Плохих x меньше $\delta 2^n$
- $C_1, C_2, \ldots, C_t \leftarrow \mathcal{C}$, $t = \frac{50n}{\epsilon^2}$, $C = majority(C_1, C_2, \ldots, C_n)$.
- (Оценки Чернова в аддитивной форме) $X_i \in \{0,1\}$ — независимые одинаково распределенные случайные величины. $\mathrm{E}[X_i] = \mu$. Тогда $\Pr\left[\left|\frac{\sum_{i=1}^n X_i}{n} - \mu\right| > \varepsilon\right] \leq 2e^{-2\varepsilon^2 n}$
- Для хорошего x $\Pr[C(x) = f(x)] \geq 1 - 2e^{-25n}$
- Хороших x не более 2^n, поэтому найдется схема C, которая вычисляет f на всех хороших x.
- Значит, $\Pr_{x \leftarrow U_n}[C(x) = f(x)] \geq 1 - \delta$
- $|C| \leq tS' + cn = \frac{S}{2} + cn < S$, противоречие!!!

XOR-лемма

Теорема. (Яо, 1982). $f : \{0,1\}^n \to \{0,1\}$,
$f^{\oplus k} : \{0,1\}^{nk} \to \{0,1\}$:
$f^{\oplus k}(x_1, x_2, \ldots, x_k) = f(x_1) \oplus f(x_2) \oplus \cdots \oplus f(x_k)$. Тогда
$\forall \delta > 0, \forall \epsilon > 2(1-\delta)^k$ выполняется $H_{avg}^{\frac{1}{2}+\epsilon}(f^{\oplus k}) \geq \frac{\epsilon^2}{400n} H_{avg}^{1-\delta}(f)$.

Доказательство. Пусть найдется схема C размера $S' = \frac{\epsilon^2}{400n} S$, что

$$\Pr_{(x_1,x_2,\ldots,x_k) \leftarrow U_n^k} \left[C(x_1, x_2, \ldots, x_k) = \bigoplus_{i=1}^{k} f(x_k) \right] \geq \frac{1}{2} + \epsilon$$

- Пусть H — трудное распределение из леммы Импальяццо для $\frac{\epsilon}{2}$. $H(x) \leq \frac{1}{\delta} 2^{-n}$.
- $U_n = (1-\delta)G + \delta H$
- $U_n^2 = (1-\delta)^2 G^2 + (1-\delta)\delta GH + \delta(1-\delta)HG + \delta^2 H^2$
- $U_n^k = (1-\delta)^k G^k + (1-\delta)^{k-1}\delta G^{k-1} H + \cdots + \delta^k H^k$

Продолжение доказательства

- $U_n = (1-\delta)G + \delta H$
- $U_n^k = (1-\delta)^k G^k + (1-\delta)^{k-1}\delta G^{k-1}H + \cdots + \delta^k H^k$
- $\mathcal{D}$ — распределение на $\{0,1\}^{nk}$.
  $P_{\mathcal{D}} = \Pr_{x \leftarrow \mathcal{D}}\left[C(x_1, x_2, \ldots, x_k) = \bigoplus_{i=1}^k f(x_k)\right]$
- $\frac{1}{2} + \epsilon \le P_{U_n^k} = (1-\delta)^k P_{G^k} + (1-\delta)^{k-1}\delta P_{G^{k-1}H} + \cdots + \delta^k P_{H^k}$
- $\frac{1}{2} + \frac{\epsilon}{2} \le (1-\delta)^{k-1}\delta P_{G^{k-1}H} + \cdots + \delta^k P_{H^k}$, $\boxed{\epsilon > 2(1-\delta)^k}$
- $P_{G^{(i)}H^{(k-i)}} > \frac{1}{2} + \frac{\epsilon}{2}$
- $\Pr_{x \leftarrow G^{(i)}H^{(k-i)}}\left[C(x_1, x_2, \ldots, x_k) \bigoplus_{i=1}^k f(x_k)\right] > \frac{1}{2} + \frac{\epsilon}{2}$
- $\Pr_{x_\ell \leftarrow H}\left[C(x_1, x_2, \ldots, x_k) \oplus \bigoplus_{i \neq \ell} f(x_i) = f(x_\ell)\right] > \frac{1}{2} + \frac{\epsilon}{2}$
- Противоречие с леммой Импальяццо для схемы $C(x_1, x_2, \ldots, x_k) \oplus \bigoplus_{i \neq \ell} f(x_i)$.