Pseudorandom generators

29 April, 2021 13:30

- **One-time Pad:**

→ encrypting a message $m \in \{0,1\}^n$ using one-time pad $k \in \{0,1\}^n$ picked at random

$$m \xrightarrow{Enc} m \oplus k = e \xrightarrow{DEC} m = e \oplus k$$

provably secure if $\underline{k}$ is used only once.

→ inefficient

Want: $g : \{0,1\}^\ell \to \{0,1\}^n$

pick $s \in \{0,1\}^\ell$ at random

$$m \xrightarrow{Enc} m \oplus g(s) \xrightarrow{Dec} m = e \oplus g(s)$$

g ... "pseudorandom generator"

Q: when is it secure?

- $f : \{0,1\}^n \quad \{0,1\}$ ... test

$g : \quad \to \quad$ PRG

Def: g fools f if $\qquad \varepsilon = \frac{1}{n} \quad \varepsilon = \frac{1}{n^{\omega(1)}}$

$$\left| \Pr_{z \in \{0,1\}^n}[f(z) = 1] - \Pr_{s \in \{0,1\}^\ell}[f(g(s)) = 1] \right| \le \varepsilon$$

- Want g which fools all f from some class of fcn's,
e.g. all polynomially computable f:

  - f computable probabilistically in time $m(n) = n^{O(1)}$

$$\ell = O(\lg n)$$

$$g : \{0,1\}^{\ell(n)} \to \{0,1\}^{m(n)}$$

s.t.
$$\forall n \ \forall x \in \{0,1\}^n \ \left| \Pr_{z \in \{0,1\}^{m(n)}} [f(x,z) = 1] - \Pr_{s \in \{0,1\}^{\ell(n)}} [f(x, g(s)) = 1] \right| \le \frac{1}{4}.$$

→ derandomization, combinatorial constructions, ...
→ cryptography

- PRG is good ≈ one cannot predict next bit of its output.

$$g : \{0,1\}^{\ell} \to \{0,1\}^{\ell+1}$$

- PRG's fooling BPP computations exist
$\iff \exists f \in E$ which requires ckts of size $2^{\delta n}$ for some $\delta > 0$.

Ex: Randomized alg. for STCONN (reachability)

Input: undirected graph G, $s, t \in V(G)$

Output: Is there a path from s to t?

Output: Is there a path from s to t ?

Alg.: for $16n^3$ steps repeat:

if $s = t$ then output YES.

$s \leftarrow$ random neighbor of s

output NO.

If $s \rightsquigarrow t$ then $\Pr[\mathrm{Alg}(G,s,t) = \mathrm{YES}] \geq \frac{2}{3}$

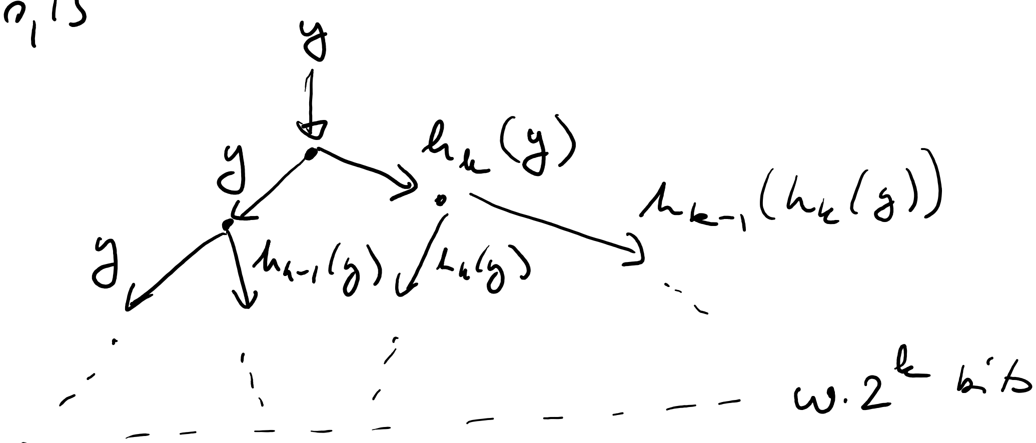
If $s \not\rightsquigarrow t$ then $\Pr[\mathrm{Alg}(G,s,t) = \mathrm{NO}] = 1$.

Algorithm runs in randomized logarithmic space $= RL$.

Nisan's PRG for RL:

$h_1, h_2, h_3, \dots h_k : \{0,1\}^w \to \{0,1\}^w$ pair-wise ind. hash fcn's.

$y \in \{0,1\}^w$



$G_0(y) = y$

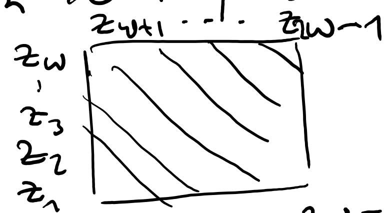
$G_k(y, h_1, h_2, \dots h_k) = G_{k-1}(y, h_1, h_2, \dots h_{k-1}) \circ G_{k-1}(h_k(y), h_1 \dots h_{k-1})$

$h : \{0,1\}^w \to \{0,1\}^w$ ... pair-wise ind.

e.g. 1) $h_{A,b}(x) = Ax + b$ $\qquad A \in \{0,1\}^{w \times w}$, $b \in \{0,1\}^b$

random

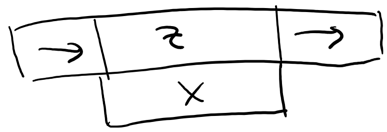
2) Convolution $\approx$ Toeplitz matrix



$= A_z$

pick random $z \in \{0,1\}^{2w-1}$ $\qquad b \in \{0,1\}^w$

$$h_{z,b}(x) = A_z x + b.$$



$(x, z_{i \dots i+w-1}) + b_i$

$W = O(\lg n) \qquad k = O(\lg n)$

- $A, B \subseteq \{0,1\}^w \qquad \alpha = \frac{|A|}{2^w} \qquad \beta = \frac{|B|}{2^w}$

$$C_h = \{x \in \{0,1\}^w;\ x \in A \ \& \ h(x) \in B\}$$

$C_h \approx A \times B \qquad \gamma_h = \frac{|C_h|}{2^w}$

want: $\gamma_h \approx \alpha \cdot \beta$

Claim: 1) $\mathbb{E}_h[\gamma_h] = \alpha \cdot \beta$

(***) 2) $\Pr[|\gamma_h - \alpha \cdot \beta| > \varepsilon] \le \frac{\alpha\beta}{\varepsilon^2 \cdot 2^w} \le \frac{1}{\varepsilon^2 \cdot 2^w}$

(***) 2) $\Pr_h\left[\,|\gamma_h - \alpha\cdot\beta| > \varepsilon\,\right] = \frac{1}{\varepsilon^2\cdot 2^w} - \varepsilon^2\cdot 2^w$

<u>Pf:</u> 1) $\forall x \in \{0,1\}^w$ r.v. $Y_x = \begin{cases} 1 & h(x)\in B \\ 0 & \text{else}\end{cases}$

$$|C_h| = \sum_{x\in A} Y_x$$

$$\mathbb{E}_h[\gamma_h] = \mathbb{E}_h\left[\frac{|C_h|}{2^w}\right] = \frac{1}{2^w}\cdot\sum_{x\in A}\mathbb{E}[Y_x]$$

$$= \frac{|A|}{2^w}\cdot\beta = \alpha\cdot\beta$$

2) $$\mathbb{E}\left[(\gamma_h - \alpha\cdot\beta)^2\right] = \mathbb{E}\left[\left(\frac{1}{2^w}\sum_{x\in A} Y_x - \alpha\cdot\beta\right)^2\right]$$

$$= \mathbb{E}\left[\frac{1}{(2^w)^2}\left(\sum_{x\in A} Y_x\right)^2 - \frac{2}{2^w}\alpha\cdot\beta\cdot\sum_{x\in A} Y_x + (\alpha\beta)^2\right]$$

$$= \frac{1}{2^{2w}}\sum_{x,y\in A}\mathbb{E}[Y_x\cdot Y_y] - 2\cdot\alpha\cdot\beta\cdot\underbrace{\mathbb{E}[\gamma_h]}_{\alpha\cdot\beta} + (\alpha\beta)^2$$

$$= \frac{1}{2^w}\sum_{x,y\in A}\mathbb{E}[Y_x\cdot Y_y] - (\alpha\beta)^2$$

$$= \frac{1}{2^{2w}}\left(\sum_{\substack{x,y\in A\\ x\neq y}}\mathbb{E}[Y_x\cdot Y_y] + \sum_{x\in A}\underbrace{\mathbb{E}[Y_x]}_{\beta}\right) - (\alpha\beta)^2$$

$$= \underbrace{\frac{1}{2^{2w}}\left(|A|\cdot(|A|-1)\right)\cdot\beta^2}_{\leq(\alpha\beta)^2} + \frac{\alpha\beta}{2^w} - (\alpha\beta)^2$$

$$\leq \frac{\alpha\beta}{2^w}$$

$$\Pr\left[\,|\gamma_h - \alpha\beta| \geq \varepsilon\,\right] = \Pr_h\left[(\gamma_h - \alpha\cdot\beta)^2 \geq \varepsilon^2\right]$$

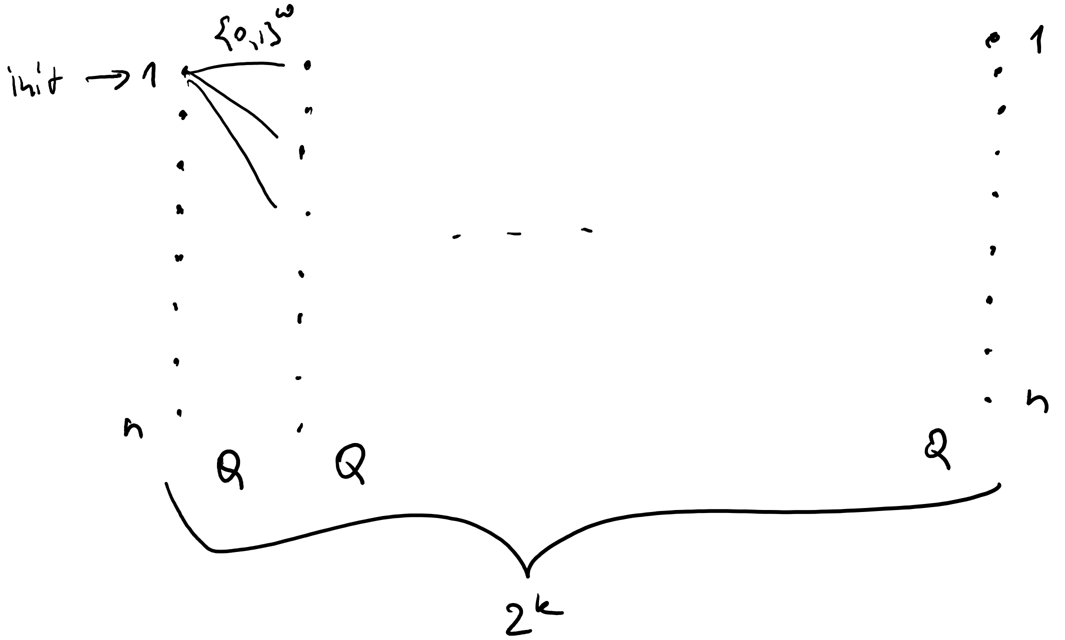
$$\Pr_h\left[\,|\gamma_h - \alpha\beta| \geq \varepsilon\,\right] = \Pr_h\left[(\gamma_h - \alpha\cdot\beta) \geq \varepsilon\right]$$

$$\leq \frac{\alpha\beta}{2^w}\cdot\frac{1}{\varepsilon^2}$$

markov $\quad \Pr\left[X \geq k E(X)\right] \leq \frac{1}{k}$

$$k\cdot E[X] = \varepsilon^2 \rightarrow \frac{1}{k} = \frac{E(X)}{\varepsilon^2}$$

- We say h is (ε, A, B)-independent if $|\gamma_h - \alpha\cdot\beta| < \varepsilon$.
- consider "branching pgm"/graph:



Q represent the transitions of a lg-space machine between different configurations (config's $\{1, \ldots, n\}$)

Want: $\Pr_{z\in\{0,1\}^{w\cdot 2^k}}\left[Q^{2^k}(z) = 1\right] \approx \Pr_{\substack{y\in\{0,1\}^w \\ h_1, h_2, \ldots h_k}}\left[Q^{2^k}(G_k(y, h_1, h_2, \ldots h_k)) = 1\right]$

For distribution D on $\{0,1\}^{w2^k}$

$Q(D)$ is the $n\times n$ matrix of transition pbbtics, i.e.

$$Q(D)_{i,j} = \Pr[D \text{ takes us from } i \text{ to } j]$$

For matrix M def. $\|M\| = \max_i \sum_j |M_{i,j}|$

... the ℓ_1-norm of largest row

Claim: $\|M+N\| \le \|M\| + \|N\|$

$\|M\cdot N\| \le \|M\|\cdot\|N\|$

Pf: (EXC). $\blacksquare$

· $(h_1,\dots h_k)$ are (ε, Q)-good if

$$\|Q(G_k(U_w, h_1,\dots h_k)) - Q(U_{w\cdot 2^k})\| \le \varepsilon$$

uniform on $\{0,1\}^w$ uniform on $\{0,1\}^{w2^k}$

Thm: $\forall Q: [n]\times\{0,1\}^w \to [n]$

$$\Pr_{h_1,h_2,\dots h_k}\left[(h_1,h_2,\dots h_k) \text{ is not } ((2^k-1)\varepsilon, Q)\text{-good}\right] \le \frac{n^7\cdot k}{\varepsilon^2\cdot 2^w}$$

Pf: induction on k:

$k=0$ ... trivo

$\left[k=1 \quad \dots \text{ by claim } (\cancel{***}) \quad \le \frac{n^2}{\varepsilon^2\cdot 2^w}\right]$

$k > 0$

$h_1, h_2, \dots h_k$ picked at random

def. $B_{i,j}^{h_1 \dots h_{k-1}} = \{ x \in \{0,1\}^w, \ G_{k-1}(x, h_1, h_2, \dots h_{k-1}) \text{ takes } i \xrightarrow{Q} j \}$

Consider event:

1) $(h_1, h_2, \dots h_{k-1})$ is $((2^{k-1}-1)\varepsilon, Q)$ - good

2) $\forall\, i, l, j$: h_k is $\left(\frac{\varepsilon}{n^2}, B_{i,l}^{h_1 \dots h_{k-1}}, B_{l,j}^{h_1 \dots h_{k-1}}\right)$ - independent

Claim: 1) & 2) $\Rightarrow$ $(h_1 \dots h_k)$ is $((2^k-1)\varepsilon, Q)$ - good

$$\| Q(G_k(U_w, h_1, \dots h_k)) - Q(U_{w \cdot 2^k}) \|$$

$$\le \| Q(G_k(U_w, h_1 \dots h_k)) - (Q(G_{k-1}(U_w, h_1, \dots h_{k-1}))^2 \| \quad \overset{\le \varepsilon}{(*)}$$

triangle ineq.

$$+ \| (Q(G_{k-1}(U_w, h_1, \dots h_{k-1})))^2 - Q(U_{w \cdot 2^k}) \| \quad \overset{(**)}{\le (2^k-2)\varepsilon}$$

to bound $(*)$ consider some fixed i & arbitrary j:

$$Q(G_k(U_w, h_1, \dots h_k))_{i,j} = \sum_l \Pr_x \left[x \in B_{il}^{h_1 \dots h_{k-1}} \ \& \ h_k(x) \in B_{lj}^{h_1 \dots h_{k-1}} \right]$$

$$\left(Q(G_{k-1}(U_w, h_1, \dots h_{k-1}))^2\right)_{i,j} = \sum_l \Pr_x \left[x \in B_{il}^{h_1 \dots h_{k-1}} \right] \cdot \Pr_y \left[y \in B_{lj}^{h_1 \dots h_{k-1}} \right]$$

by 2) $$\left| \Pr_x \left[x \in B_{il}^{h_1 \dots h_{k-1}} \ \& \ h_k(x) \in B_{lj}^{h_1 \dots h_{k-1}} \right] - \Pr_x \left[x \in B_{il}^{h_1 \dots h_{k-1}} \right] \cdot \Pr_y \left[y \in B_{lj}^{h_1 \dots h_{k-1}} \right] \right| \le \frac{\varepsilon}{n^2}$$

$\Rightarrow$ [illegible]

$$\Rightarrow \left| Q\left(G_k(U_w, h_1 \dots h_k)\right)_{i,j} - \left[Q\left(G_{k-1}(U_w, h_1 \dots h_{k-1})\right)^2\right]_{i,j}\right| \le \frac{\varepsilon}{n}$$

$$\Rightarrow \quad (*) \le \varepsilon$$

to bound $(**)$

$$Q(U_{w\cdot 2^k}) = \left(Q(U_{w\cdot 2^{k-1}})\right)^2$$

Hence, $(**) = \| \underbrace{Q\left(G_{k-1}(U_w, h_1, \dots, h_{k-1})\right)}_{M}^2 - \left(\underbrace{Q(U_{w\, 2^{k-1}})}_{N}\right)^2 \|$

$$= \|M^2 - N^2\| = \|M^2 - MN + MN - N^2\|$$

$$\le \|M\| \cdot \|M - N\| + \|M - N\| \cdot \|N\|$$

by 1) $\quad \|M - N\| \le (2^{k-1} - 1)\varepsilon$

$\|M\|, \|N\| \le 1$ as M & N are matrices of probabilities

$$\Rightarrow \quad (**) \le 2 \cdot (2^{k-1} - 1)\varepsilon = (2^k - 2)\varepsilon .$$

Hence, it remains to bound the probability of 1) & 2) not happening. By I.H. prob of $\neg$1) is $\le \frac{n^7 \cdot (k-1)}{\varepsilon^2 \cdot 2^w}$.

For fixed $h_1, \dots h_{k-1}$,

By "Mixing" Claim:

$$\Pr[\neg 2] \le n^3 \cdot \frac{n^4}{\varepsilon^2 \cdot 2^w} = \frac{n^7}{\varepsilon^2 \cdot 2^w}$$

The claim follows. $\boxtimes$

Corollary: $RL \subseteq DTISP(n^{O(1)}, \log^2 n)$

Pf: Pick the hash fcn's $h_1, h_2, \dots h_k$ one by one while always checking that the next one satisfies 2) from the above proof.

Given $h_1, \dots h_\ell$ we can calculate $\forall i, j$

$$Q(G_\ell(u_w, h_1, \dots h_\ell))_{i,j}$$

in $O(\log n)$ space so checking the condition 2) can be done in log-space. ∎

Thm: (Saks-Zhou '96): $RL \subseteq DSPACE(\log^{3/2} n)$

idea: ... pick $\sqrt{\log n}$ hash fcn & reuse them $\sqrt{\log n}$ times.

Application:

(D. Sivakumar '02)

Discrepancy: Given $A_1, \dots A_m \subseteq [n]$

Find $D \subseteq [n]$ s.t.

$$\forall i: \ |(A_i \cap D) - (A_i \setminus D)| \leq \sqrt{|A_i| \cdot \log m}.$$

→ random bits will give a good set D w.h.p.

→ use Nisan's PRG to generate D.

2. It takes $O(m \lg n)$ space to check the condition for all $i = 1, \ldots m$ if D is given "on-the-fly".

Solution: Design m $O(\lg n)$-space tests for each i individually. Nisan's PRG will work for each of them w.p. $\geq 1 - \frac{1}{m^2}$ so it will work for all of them at once w.p. $\geq 1 - \frac{1}{m}$.

→ find the appropriate hash funcs deterministically like in $RL \subseteq DTISP(n^{O(1)}, \lg^2 n)$. $\boxtimes$

→ Johnson-Lindenstrauss lemma, ... using similar technique.

streaming, ...