Leftover Hash Lemma

Joshua Cooper

March 2, 2009

Lemma 1. *Let $m = k - 2\log(1/\epsilon)$. Then, for every (n,k)-source X,*

$$\Delta(H(X) \circ H, U_m \circ H) < \epsilon,$$

where H is a randomly chosen function from a pairwise independent $\{0,1\}^n$ to $\{0,1\}^m$ hash function family $\mathcal{H}$ parameterized by $\{0,1\}^{n+m}$.

Proof. Let $\mathbf{p} \in \mathbb{R}^{2^{2m+n}}$ denote the probability vector corresponding to a random choice of $x \in \{0,1\}^n$ and $h \in \mathcal{H}$ (represented as a length $n+m$ bit string). We wish to show that

$$\frac{1}{2}|\mathbf{p} - \mathbb{1}|_1 < \epsilon,$$

where $\mathbb{1}$ corresponds to the uniform distribution. It is possible to write $\mathbf{p} = \mathbb{1} + \mathbf{w}$ for some $\mathbf{w} \perp \mathbb{1}$, so we may apply the Pythagorean theorem and Cauchy-Schwarz to conclude

$$\frac{1}{2}|\mathbf{p} - \mathbb{1}|_1 \leq \frac{1}{2}2^{m+n/2}\|\mathbf{p} - \mathbb{1}\|_2 = \frac{1}{2}2^{m+n/2}\sqrt{\|\mathbf{p}\|_2^2 - \|\mathbb{1}\|_2^2}. \tag{1}$$

Now, $\|\mathbb{1}\|_2^2 = 2^{-2m-n}$, so it remains to compute $\|\mathbf{p}\|_2^2$. It is not hard to see that

$$\|\mathbf{p}\|_2^2 = \mathbf{P}_{\substack{x,x' \in X \\ h,h' \in \mathcal{H}}} \left(h(x) \circ h = h'(x') \circ h'\right).$$

Since the suffixes of the two strings strings must agree, this probability can be written

$$\|\mathbf{p}\|_2^2 = 2^{-n-m} \cdot \mathbf{P}_{\substack{x,x' \in X \\ h \in \mathcal{H}}} \left(h(x) \circ h = h(x') \circ h\right).$$

There are two ways that the event "$h(x) \circ h = h(x') \circ h$" can occur: $x = x'$ or $x \neq x'$ but $h(x) = h(x')$. The former has probability at most 2^{-k}, since X is an (n, k)-source. (Here we are using the easily derived fact that this probability is maximized by a flat source.) The latter probability is exactly equal to 2^{-m}, since $\mathcal{H}$ is a pairwise independent hash function family. Therefore,

$$\|\mathbf{p}\|_2^2 \leq 2^{-n-m}(2^{-k} + 2^{-m}) = 2^{-n-m-k} + 2^{-n-2m}.$$

Plugging this into (1) yields

$$\begin{aligned}
\Delta(H(X) \circ H, U_m \circ H) &\leq \frac{1}{2} 2^{m+n/2} \sqrt{2^{-n-m-k}} \\
&= 2^{m+n/2-1-n/2-m/2-k/2} \\
&= 2^{m/2-k/2-1} \\
&= 2^{-\log 1/\epsilon - 1} = \epsilon/2 < \epsilon.
\end{aligned}$$

□