Lesson 8: Probabilistic reductions

Theme: Probabilistic reductions and preliminary to Toda's theorem.

1 Probabilistic reduction from SAT to USAT

Let USAT be the following language.

$$\mathsf{USAT} \stackrel{\mathsf{def}}{=} \{\varphi : \varphi \text{ is a boolean formula with unique satisfying assignment}\}$$

Theorem 8.1 (Valiant and Vazirani, 1986) *There is a probabilistic polynomial time algorithm $\mathcal{M}$ such that on input (Boolean) formula φ, the output of $\mathcal{M}$, denoted by $\mathcal{M}(\varphi)$, satisfies the following.*

- *If $\varphi \in \mathsf{SAT}$, then $\mathbf{Pr}[\ \mathcal{M}(\varphi) \in \mathsf{USAT}\ ] \geqslant 3/(16n)$, where n is the number of variables in φ.*
- *If $\varphi \notin \mathsf{SAT}$, then $\mathbf{Pr}[\ \mathcal{M}(\varphi) \in \mathsf{SAT}\ ] = 0$.*

Proof. The algorithm $\mathcal{M}$ works as follows. On input formula φ, do the following.

- Let $x_1, \ldots, x_n$ be the variables in φ.
- Let $x \stackrel{\mathsf{def}}{=} (x_1, \ldots, x_n)$.
- Randomly choose $k \in \{2, \ldots, n+1\}$.
- Randomly choose a hash function $h \in \mathcal{H}_{n,k}$, where $\mathcal{H}_{n,k}$ is pair-wise independent.
- Output the formula $\varphi(x) \wedge (h(x) = 0)$, where 0 is a column vector of zeroes of size k.

Note that the part $h(x) = 0$ can be stated as a boolean formula. If we use the collection $\mathcal{H}_{n,k}$ as in Theorem 8.9, $h(x) = 0$ is of the form: $Ax + b = 0$, which is equivalent to $Ax = b$. This can be written into the following form:

$$\bigwedge_{i=1}^{k} \Big(\big(A_{i,1}x_1 \oplus \cdots \oplus A_{i,n}x_n \big) \ \leftrightarrow\ b_i \Big)$$

Here $\oplus$ denotes the XOR operation. Note that each $A_{i,1}x_1 \oplus \cdots \oplus A_{i,n}x_n$ can be rewritten into formulas using only $\wedge, \vee, \neg$ in quadratic time as follows. Divide it into two halves, rewrite each half (recursively) and combine them with the standard definition of XOR.

Now, we prove the correctness of our algorithm. Obviously, if the input formula φ is not satisfiable, so is the output formula. Suppose φ is satisfiable. Let S be the set of satisfying assignments of φ. With probability $1/n$, the algorithm chooses a value k such that $2^{k-2} \leqslant |S| \leqslant 2^{k-1}$. By Lemma 8.11, the probability that there is a unique $x \in S$ such that $h(x) = 0$ is $\geqslant 3/16$. Thus, the probability that $\mathcal{M}(\varphi) \in \mathsf{USAT}$ is at least $3/(16n)$. ∎

2 The language $\oplus$SAT and the class $\oplus$P

The language $\oplus$SAT is defined as follows.

$$\oplus\mathsf{SAT} \stackrel{\mathsf{def}}{=} \{\varphi : \varphi \text{ is a Boolean formula with } \textit{odd} \text{ number of satisfying assignments}\}$$

The class $\oplus\mathbf{P}$ is defined as follows. A language $L \in \oplus\mathbf{P}$, if there is a polynomial time NTM $\mathcal{M}$ such that for every input word w, $w \in L$ if and only if the number of accepting runs of $\mathcal{M}$ on w is odd number.

We define a few terminology and notations. Let $\sharp\varphi$ denote the number of satisfying assignments of a (Boolean) formula φ. We will define operations $\sim$, $\sqcap$ and $\sqcup$ on formulas such that the following holds.

$$\sharp(\sim\varphi) = \sharp\varphi + 1 \qquad \sharp(\varphi \sqcap \phi) = \sharp\varphi \cdot \sharp\phi \qquad \sharp(\varphi \sqcup \phi) = (\sharp\varphi + 1) \cdot (\sharp\phi + 1) + 1$$

Obviously the following holds.

$$\begin{aligned} \sim\varphi \ &\in \ \oplus\mathsf{SAT} \quad \text{if and only if} \quad \varphi \ \notin \ \oplus\mathsf{SAT} \\ \varphi \sqcap \phi \ &\in \ \oplus\mathsf{SAT} \quad \text{if and only if} \quad \text{both } \varphi, \phi \ \in \ \oplus\mathsf{SAT} \\ \varphi \sqcup \phi \ &\in \ \oplus\mathsf{SAT} \quad \text{if and only if} \quad \text{at least one of } \varphi, \phi \ \in \ \oplus\mathsf{SAT} \end{aligned}$$

These operations are defined as follows.

- For φ with variables $x_1, \ldots, x_n$, we pick a "new" variable z and define $\sim\varphi$ as follows.

$$\sim\varphi \stackrel{\mathsf{def}}{=} \left(\neg z \wedge \varphi\right) \vee \left(z \wedge \bigwedge_{i=1}^{n} x_i\right)$$

- For two formulas φ and ψ, we rename the variables so that the variables in φ and ϕ are disjoint, and define $\varphi \sqcap \psi$ as follows.

$$\varphi \sqcap \phi \stackrel{\mathsf{def}}{=} \varphi \wedge \phi$$

- For two formulas φ and ψ, we rename the variables so that the variables in φ and ϕ are disjoint, and define $\varphi \sqcup \psi$ as follows.

$$\varphi \sqcup \phi \stackrel{\mathsf{def}}{=} \sim(\sim\varphi \sqcap \sim\phi)$$

3 Probabilistic reductions from SAT and $\overline{\mathsf{SAT}}$ to $\oplus$SAT

Theorem 8.1 can be easily extended to obtain reductions from SAT and $\overline{\mathsf{SAT}}$ to $\oplus$SAT.

Lemma 8.2 (Reduction from SAT to $\oplus$SAT) *There is a polynomial time PTM $\mathcal{M}$ that on input formula φ and a positive integer m (in unary), outputs a formula, denoted by $\mathcal{M}(\varphi, m)$, such that the following holds.*

- *If $\varphi \in \mathsf{SAT}$, then $\mathbf{Pr}[\ \mathcal{M}(\varphi, m) \in \oplus\mathsf{SAT}\ ] \geqslant 1 - 2^{-m}$.*
- *If $\varphi \notin \mathsf{SAT}$, then $\mathbf{Pr}[\ \mathcal{M}(\varphi, m) \in \oplus\mathsf{SAT}\ ] = 0$.*

Moreover, the output $\mathcal{M}(\varphi, m)$ uses $O(mn^2)$ variables, where n is the number of variables in φ.[*]

[*] Abusing the notation, $O(mn^2)$ denotes $\leqslant cmn^2$, for some constant c.

Proof. On input φ with n variables, the algorithm $\mathcal{M}$ first runs the reduction in Theorem 8.1 on φ for $8mn$ times to obtain formulas $\psi_1, \ldots, \psi_{8mn}$. Then, it outputs $\sim (\sim \psi_1 \sqcap \cdots \sqcap \sim \psi_{8mn})$.[†] Obviously, $\mathcal{M}$ runs in polynomial time. Note also that the output formula uses $8mn(n+1)+1 = O(mn^2)$ variables.

Recall that on input φ with n variables, the reduction in Theorem 8.1 outputs a formula ψ such that the following holds.

- If $\varphi \in \mathsf{SAT}$, then $\mathbf{Pr}[\ \psi \in \mathsf{USAT}\ ] \geqslant 1/(8n)$.
- If $\varphi \notin \mathsf{SAT}$, then $\mathbf{Pr}[\ \psi \in \mathsf{SAT}\ ] = 0$.

Note the following.

- If $\psi \notin \oplus\mathsf{SAT}$, then $\psi \notin \mathsf{USAT}$. Thus, $\mathbf{Pr}[\ \psi \notin \oplus\mathsf{SAT}\ ] \leqslant \mathbf{Pr}[\ \psi \notin \mathsf{USAT}\ ]$.
- $\bigsqcup_{i=1}^{8mn} \psi_i \in \oplus\mathsf{SAT}$ if and only if one of $\psi_i \in \oplus\mathsf{SAT}$.

Thus, on input φ, the output $\bigsqcup_{i=1}^{8mn} \psi_i$ satisfies the following.

- If $\varphi \notin \mathsf{SAT}$, then none of the ψ_i is satisfiable. Thus, $\bigsqcup_{i=1}^{8mn} \psi_i \notin \oplus\mathsf{SAT}$. Therefore,

$$\mathbf{Pr}\Big[\ \bigsqcup_{i=1}^{8mn} \psi_i \ \in\ \oplus\mathsf{SAT}\ \Big] \ =\ 0$$

- If $\varphi \in \mathsf{SAT}$, the following holds.

$$\mathbf{Pr}\Big[\ \bigsqcup_{i=1}^{8mn} \psi_i \notin \oplus\mathsf{SAT}\ \Big] \ =\ \prod_{i=1}^{8mn} \mathbf{Pr}[\ \psi_i \notin \oplus\mathsf{SAT}\ ] \ \leqslant\ \Big(1 - \frac{1}{8n}\Big)^{8mn} \ \leqslant\ (1/e)^m \ \leqslant\ (1/2)^m$$

  Therefore, $\mathbf{Pr}[\ \bigsqcup_{i=1}^{8mn} \psi_i \in \oplus\mathsf{SAT}\ ] \geqslant 1 - (1/2)^m$.

This completes the proof of Lemma 8.2. ∎

Lemma 8.3 (Reduction from $\overline{\mathsf{SAT}}$ to $\oplus\mathsf{SAT}$) *There is a polynomial time PTM $\mathcal{M}$ that on input formula φ and a positive integer m (in unary), outputs a formula, denoted by $\mathcal{M}(\varphi, m)$, such that the following holds.*

- *If $\varphi \in \overline{\mathsf{SAT}}$, then $\mathbf{Pr}[\ \mathcal{M}(\varphi, m) \in \oplus\mathsf{SAT}\ ] = 1$.*
- *If $\varphi \notin \overline{\mathsf{SAT}}$, then $\mathbf{Pr}[\ \mathcal{M}(\varphi, m) \in \oplus\mathsf{SAT}\ ] \leqslant (1/2)^m$.*

Proof. The PTM $\mathcal{M}$ works as follows. On input φ and m, it runs the reduction in Lemma 8.2 to obtain a formula ψ, and then outputs $\sim \psi$.

If $\varphi \in \overline{\mathsf{SAT}}$, then $\mathbf{Pr}[\ \psi \notin \oplus\mathsf{SAT}\ ] = 1$, and hence, $\mathbf{Pr}[\ \sim \psi \in \oplus\mathsf{SAT}\ ] = 1$.
If $\varphi \notin \overline{\mathsf{SAT}}$, then $\mathbf{Pr}[\ \sim \psi \in \oplus\mathsf{SAT}\ ] = \mathbf{Pr}[\ \psi \notin \oplus\mathsf{SAT}\ ] \leqslant (1/2)^m$. ∎

Combining Lemmas 8.2 and 8.3 and Cook-Levin reduction, we have the following.

Theorem 8.4 (Reductions from languages in $\mathbf{NP} \cup \mathbf{coNP}$ to $\oplus\mathsf{SAT}$) *For every language $L \in \mathbf{NP} \cup \mathbf{coNP}$, there is a polynomial time PTM $\mathcal{M}$ that on input word w and a number m (in unary), outputs a formula $\mathcal{M}(w, m)$ such that the following holds.*

- *If $w \in L$, then $\mathbf{Pr}[\ \mathcal{M}(w, m) \in \oplus\mathsf{SAT}\ ] \geqslant 1 - (1/2)^m$.*
- *If $w \notin L$, then $\mathbf{Pr}[\ \mathcal{M}(w, m) \in \oplus\mathsf{SAT}\ ] \leqslant (1/2)^m$.*

[†] Note that $\sim (\sim \psi_1 \sqcap \cdots \sqcap \sim \psi_{8mn})$ is equivalent to $\psi_1 \sqcup \cdots \sqcup \psi_{8mn}$.

4 Probabilistic reductions from languages in PH to $\oplus$SAT

In this section we will show how to extend Theorem 8.4 to all languages in **PH**. We need some terminology and notations. We write $\bar{x}$, $\bar{y}$ or $\bar{z}$ to denote a sequence of variables, and the length is denoted by $|\bar{x}|$, $|\bar{y}|$ or $|\bar{z}|$, respectively.

Recall that a QBF is formula of the form: $Q_1\bar{z}_1 \cdots Q_k\bar{z}_k \quad \phi$ where each $Q_i \in \{\forall, \exists\}$ and $Q_i \neq Q_{i+1}$, each $\bar{z}_i$ is a vector of variables and ϕ is a formula that uses variables $\bar{z}_1, \ldots, \bar{z}_k$. Note that all variables used in ψ are "quantified."

QBF with free variables. A *QBF with free variables* is a QBF formula that has variables that are not quantified, i.e., of the form:

$$\varphi \stackrel{\text{def}}{=} Q_1\bar{z}_1 \cdots Q_k\bar{z}_k \quad \phi$$

where ϕ uses some variables $\bar{y}$ that are "free," i.e., not quantified by any quantifiers, in addition to the variables $\bar{z}_1, \ldots, \bar{z}_k$. In this case, we write $\varphi(\bar{y})$ to indicate that $\bar{y}$ are free. For example, in the formula $\forall x \exists z(x \vee y \vee z)$, variables x, z are quantified, but variable y is free.

We usually denote an assignment that assigns variables in $\bar{y}$ as a string $\bar{a} \in \{0,1\}^n$ with the same length as $\bar{y}$. For a QBF $\varphi(\bar{y})$ with free variable $\bar{y}$ and $\bar{a}$ be an assignment on $\bar{y}$, we write $\varphi(\bar{a})$ to denote the QBF (without free variables) obtained by substituting every variable in $\bar{y}$ according to $\bar{a}$.

In the following the term "QBF" means a QBF which may or may not contain free variables. A *k-QBF* is a QBF in which there are k alternating quantifiers, i.e., $Q_1\bar{z}_1 \cdots Q_k\bar{z}_k \ \psi$, where each $Q_i \neq Q_{i+1}$.

The operations $\sim$, $\sqcap$ and $\sqcup$ with formulas with "free" variables. In the following we will deal with boolean formulas φ with "free" variables. Intuitively, free variables in a boolean formula are variables that cannot be renamed. We write $\varphi(\bar{y})$ to indicate that $\bar{y}$ are the free variables in φ.

- $\sim \varphi(\bar{y})$ is defined as before and the resulting formula $\sim(\varphi(\bar{y}))$ also have free variables $\bar{y}$.
- For $\varphi(\bar{y})$ and $\phi(\bar{y})$, we rename the variables so that $\bar{y}$ are the only common variables in φ and ϕ and define $\varphi(\bar{y}) \sqcap \phi(\bar{y}) \stackrel{\text{def}}{=} \varphi(\bar{y}) \wedge \phi(\bar{y})$ with free variables $\bar{y}$.
- For $\varphi(\bar{y})$ and $\phi(\bar{y})$, we define $\varphi(\bar{y}) \sqcup \phi(\bar{y}) \stackrel{\text{def}}{=} \sim(\sim \varphi(\bar{y}) \sqcap \sim \phi(\bar{y}))$ with free variables $\bar{y}$.

Lemma 8.5 (Reductions from Σ_k-SAT and Π_k-SAT to $\oplus$SAT) *For every $k \geqslant 1$, there is a probabilistic polynomial time algorithm $\mathcal{M}$ that on input a k-QBF $\varphi(\bar{y})$ and a positive integer m (in unary), outputs a formula $\psi(\bar{y})$ such that*

$$\mathbf{Pr}[\ \psi(\bar{y}) \text{ is "correct"}\ ] \quad \geqslant \quad 1 - (1/2)^m$$

Here we define a formula $\psi(\bar{y})$ to be "correct" when $\varphi(\bar{a})$ is a true QBF if and only if $\psi(\bar{a}) \in \oplus$SAT, for every assignment $\bar{a}$ on $\bar{y}$.

Proof. The proof is by induction on k. The base case $k = 1$ is similar to Lemmas 8.2 and 8.3. On input 1-QBF $\varphi(\bar{y})$ and integer m, the algorithm $\mathcal{M}$ works as follows.

- If $\varphi(\bar{y})$ is of the form $\exists \bar{x}\ \psi(\bar{x}, \bar{y})$, where $\bar{x}$ contains n variables, do the following.

  For each $i = 1, \ldots, 8mn$, construct formula $\alpha_i(\bar{y})$ as follows.

  - Randomly choose $k \in \{2, \ldots, n+1\}$.
  - Randomly choose a hash function $h \in \mathcal{H}_{n,k}$, where $\mathcal{H}_{n,k}$ is pair-wise independent.
  - Let $\alpha_i(\bar{y})$ denote the formula $\psi(\bar{x}, \bar{y}) \wedge (h(\bar{x}) = 0)$.

  Then, output the formula $\psi(\bar{y})$ where $\psi(\bar{y})$ is the formula $\bigsqcup_{i=1}^{8mn} \alpha_i(\bar{y})$.

- If $\varphi(\bar{y})$ is of the form $\forall \bar{x}\ \psi(\bar{x}, \bar{y})$, where $\bar{x}$ contains n variables, do the following

  For each $i = 1, \ldots, 8mn$, construct formula $\alpha_i(\bar{y})$ as follows.

  - Randomly choose $k \in \{2, \ldots, n+1\}$.
  - Randomly choose a hash function $h \in \mathcal{H}_{n,k}$, where $\mathcal{H}_{n,k}$ is pair-wise independent.
  - Let $\alpha_i(\bar{y})$ denote the formula $\neg\psi(\bar{x}, \bar{y}) \wedge (h(\bar{x}) = 0)$.

  Then, output the formula $\psi(\bar{y})$, where $\psi(\bar{y})$ is the formula $\sim \bigsqcup_{i=1}^{8mn} \alpha_i(\bar{y})$.

The proof that $\mathbf{Pr}[\ \psi(\bar{y}) \text{ is correct }] \geqslant 1 - (1/2)^m$ is similar to Lemmas 8.2 and 8.3.

For the induction hypothesis, we assume Lemma 8.5 holds for k, i.e., there is a probabilistic algorithm $\mathcal{M}_0$ that on input a k-QBF $\varphi(\bar{y})$ and a positive integer m (in unary), outputs a formula $\psi(\bar{y})$ such that $\mathbf{Pr}[\ \psi(\bar{y}) \text{ is correct }] \geqslant 1 - (1/2)^m$.

For the induction step, on input $(k+1)$-QBF $\varphi(\bar{y})$ and m, the algorithm $\mathcal{M}$ works as follows.

- $\varphi(\bar{y})$ is of the form $\exists \bar{x}\ \phi(\bar{x}, \bar{y})$, where $\bar{x}$ contains n variables.

  For each $i = 1, \ldots, 8mn$, construct a formula $\alpha_i(\bar{y})$ as follows.

  - Let $\beta_i(\bar{x}, \bar{y})$ be the output of $\mathcal{M}_0$ on input $\phi(\bar{x}, \bar{y})$ and $(m+1)$.
  - Randomly choose $k \in \{2, \ldots, n+1\}$.
  - Randomly choose a hash function $h \in \mathcal{H}_{n,k}$, where $\mathcal{H}_{n,k}$ is pair-wise independent.
  - Let $\alpha_i(\bar{y})$ denote the formula $\beta_i(\bar{x}, \bar{y}) \wedge (h(\bar{x}) = 0)$.

  Then, output the formula $\psi(\bar{y})$ where $\psi(\bar{y}) \stackrel{\mathsf{def}}{=} \bigsqcup_{i=1}^{8mn} \alpha_i(\bar{y})$.

- $\varphi(\bar{y})$ is of the form $\forall \bar{x}\ \psi(\bar{x}, \bar{y})$, where $\bar{x}$ contains n variables.

  For each $i = 1, \ldots, 8mn$, construct a formula α_i, as follows.

  - Let $\beta_i(\bar{x}, \bar{y})$ be the output of $\mathcal{M}_0$ on input $\neg\psi(\bar{x}, \bar{y})$ and $(m+1)$.
  - Randomly choose $k \in \{2, \ldots, n+1\}$.
  - Randomly choose a hash function $h \in \mathcal{H}_{n,k}$, where $\mathcal{H}_{n,k}$ is pair-wise independent.
  - Let $\alpha_i(\bar{y})$ be the formula $\beta_i(\bar{x}, \bar{y}) \wedge (h(\bar{x}) = 0)$.

  Then, output the formula $\psi(\bar{y})$ where $\psi(\bar{y}) \stackrel{\mathsf{def}}{=} \sim \bigsqcup_{i=1}^{8mn} \alpha_i(\bar{y})$.

We now calculate the probability of the event that $\psi(\bar{y})$ is correct.

We first consider the case that $\varphi(\bar{y})$ is of the form $\exists \bar{x} \phi(\bar{x}, \bar{y})$. By the induction hypothesis, $\mathbf{Pr}[\ \beta_i(\bar{x}, \bar{y}) \text{ is correct }] \geqslant 1 - (1/2)^{m+1}$, for each $i = 1, \ldots, 8mn$. Note that $\beta_i(\bar{x}, \bar{y})$ is correct, if for every assignment $\bar{a}$ and $\bar{b}$ on $\bar{x}$ and $\bar{y}$, respectively, $\beta_i(\bar{a}, \bar{b}) \in \oplus\mathsf{SAT}$ if and only if $\phi(\bar{a}, \bar{b})$ is a true QBF.

Assume that $\beta_i(\bar{x}, \bar{y})$ is correct. Let $\bar{b} : \bar{y} \to \{0, 1\}$ be such that $\varphi(\bar{b})$ is true QBF. Thus, for every assignment $\bar{a} : \bar{x} \to \{0, 1\}$, if $\varphi(\bar{a}, \bar{b})$ is true QBF, $\beta_i(\bar{a}, \bar{b}) \in \oplus\mathsf{SAT}$. Otherwise, $\beta_i(\bar{a}, \bar{b}) \notin \oplus\mathsf{SAT}$. So, we only need to consider all those assignments $\bar{a}$ such that $\phi_i(\bar{a}, \bar{b})$ is true, which by

the induction hypothesis, is equivalent to saying that $\beta_i(\bar{a}, \bar{b}) \in \oplus\mathsf{SAT}$. By applying the same technique as in Lemma 8.11 on the set of $\bar{a}$ such that $\beta_i(\bar{a}, \bar{b}) \in \oplus\mathsf{SAT}$, we randomly "choose" the hash function h such that there is unique assignment $\bar{a}$ such that $h(\bar{a}) = 0$, and the probability that we choose such h is $\geqslant 3/(16n)$. Thus, we have:

$$\mathbf{Pr}[\ \beta_i(\bar{x}, \bar{y}) \wedge h(\bar{x}) = 0 \text{ is correct } \mid \ \beta_i(\bar{x}, \bar{y}) \text{ is correct }] \ \geqslant \ \frac{3}{16n}$$

Thus,

$$\begin{aligned}\mathbf{Pr}[\ \psi_i(\bar{x}, \bar{y}) \text{ is correct }] \ = \ \mathbf{Pr}[\ \beta_i(\bar{x}, \bar{y}) \wedge h(\bar{x}) = 0 \text{ is correct }] \ &\geqslant \ \frac{3}{16n}\Big(1 - (1/2)^{m+1}\Big) \\ &\geqslant \ \frac{1}{8n}\end{aligned}$$

where in the last inequality we assume that $m \geqslant 1$.

Note also that if $\bar{b} : \bar{y} \to \{0,1\}$ is an assignment such that $\varphi(\bar{b})$ is false QBF, then $\beta_i(\bar{a}, \bar{b}) \notin \oplus\mathsf{SAT}$, for every assignment $\bar{a}$ (since $\beta_i(\bar{x}, \bar{y})$ is a correct formula). Thus, for any choice of h, $\beta_i(\bar{x}, \bar{b}) \wedge h(\bar{x}) = 0 \notin \oplus\mathsf{SAT}$.

Finally, note that $\bigsqcup_{i=1}^{8mn} \alpha_i(\bar{y})$ is correct if and only if one of $\alpha_i(\bar{y})$ is correct. Therefore,

$$\begin{aligned}\mathbf{Pr}\Big[\ \bigsqcup_{i=1}^{8mn} \alpha_i(\bar{y}) \text{ is not correct }\Big] \ &= \ \mathbf{Pr}[\ \alpha_i(\bar{y}) \text{ is not correct, for each } i = 1, \ldots, 8mn\ ] \\ &\leqslant \ \Big(1 - 1/(8n)\Big)^{8mn} \ \leqslant \ (1/2)^m\end{aligned}$$

The proof for the case where $\varphi(\bar{y})$ is of the form $\forall \bar{x}\ \phi(\bar{x}, \bar{y})$ is similar. ∎

Combining Lemma 8.5 and the fact that Σ_k-SAT and Π_k-SAT are $\mathbf{\Sigma}_k^p$- and $\mathbf{\Pi}_k^p$-complete, for each $k \geqslant 1$, we have the following theorem.

Theorem 8.6 (Reductions from languages in PH to $\oplus\mathsf{SAT}$) *For every language $L \in \mathbf{PH}$, there is a probabilistic polynomial time algorithm $\mathcal{M}$ that on input w, outputs a formula ψ such that the following holds, where $n = |w|$.*

- *If $w \in L$, then $\mathbf{Pr}[\ \psi \in \oplus\mathsf{SAT}\ ] \geqslant 1 - (1/2)^n$.*
- *If $w \notin L$, then $\mathbf{Pr}[\ \psi \in \oplus\mathsf{SAT}\ ] \leqslant (1/2)^n$.*

Appendix

A Pair-wise independent collection of hash functions

Definition 8.7 For $n, k \geqslant 1$, let $\mathcal{H}_{n,k}$ be a collection of functions from $\{0,1\}^n$ to $\{0,1\}^k$. We say that $\mathcal{H}_{n,k}$ is *pair-wise independent*, if for every $x, x' \in \{0,1\}^n$ where $x \neq x'$ and for every $y, y' \in \{0,1\}^k$, the following holds.

$$\mathbf{Pr}_{h \in \mathcal{H}_{n,k}}[\ h(x) = y \ \wedge \ h(x') = y'\ ] \ = \ 2^{-2k}$$

In the following we show that $\mathcal{H}_{n,k}$ exists. First, we show that $\mathcal{H}_{n,n}$ exists. For every $n \geqslant 1$, for every $a, b \in \mathsf{GF}(2^n)$, define a function $h_{a,b}$ from $\{0,1\}^n$ to $\{0,1\}^n$ as follows.[‡]

$$h_{a,b}(x) \ \stackrel{\mathsf{def}}{=} \ xa + b$$

[‡] $\mathsf{GF}(2^n)$ denotes a finite field with 2^n elements, where each element can be encoded as a 0-1 string of length n.

Theorem 8.8 *The collection* $\mathcal{H}_{n,n} \stackrel{\mathsf{def}}{=} \{h_{a,b} : a, b \in \mathsf{GF}(2^n)\}$ *is pair-wise independent.*

We have another candidate for pair-wise independent collection. For every $n \geqslant 1$, for every $A \in \{0,1\}^{n\times n}$ and $b \in \{0,1\}^{n\times 1}$, define a function $h_{A,b}$ from $\{0,1\}^{n\times 1}$ to $\{0,1\}^{n\times 1}$ as follows.[§]

$$h_{A,b}(x) \stackrel{\mathsf{def}}{=} Ax + b$$

Theorem 8.9 *The collection* $\mathcal{H}_{n,n} \stackrel{\mathsf{def}}{=} \{h_{A,b} : A \in \{0,1\}^{n\times n}$ *and* $b \in \{0,1\}^{n\times 1}\}$ *is pair-wise independent.*

Remark 8.10 Note that the existence of $\mathcal{H}_{n,n}$ implies the existence of $\mathcal{H}_{n,k}$. If $n < k$, then we can use $\mathcal{H}_{k,k}$ and extend n bit inputs to k by padding with zeros. If $n > k$, then we can use $\mathcal{H}_{n,n}$ and reduce n bit outputs to k by truncating the last $(n-k)$ bits.

Lemma 8.11 (Valiant and Vazirani, 1986) *Let* $\mathcal{H}_{n,k}$ *be a pair-wise independent hash function collection. Let* $S \subseteq \{0,1\}^n$ *such that* $2^{k-2} \leqslant |S| \leqslant 2^{k-1}$. *Then, the following holds.*

$$\mathbf{Pr}_{h\in\mathcal{H}_{n,k}}[\text{ there is a unique } x \in S \text{ such that } h(x) = 0^k\ ] \ \geqslant\ \frac{3}{16}$$

Proof. Let N denote the number of x's such that $h(x) = 0$, where h is randomly chosen from $\mathcal{H}_{n,k}$ (with uniform distribution). We will calculate $\mathbf{Pr}[\ N = 1\ ]$. Note that:

$$\begin{aligned}\mathbf{Pr}[\ N = 1\ ] &= \mathbf{Pr}[\ N \geqslant 1\ ] \ - \ \mathbf{Pr}[\ N \geqslant 2\ ] \\ &= \mathbf{Pr}\Big[\ \bigcup_{x\in S} \mathcal{E}_x\ \Big] \ - \ \mathbf{Pr}\Big[\ \bigcup_{x,x'\in S \text{ and } x\neq x'} \mathcal{E}_x \cap \mathcal{E}_{x'}\ \Big]\end{aligned}$$

where $\mathcal{E}_x$ denotes the event that $h(x) = 0$. In the following, we let $p = 2^{-k}$.

Since $\mathcal{H}_{n,k}$ is pairwise independent, $\mathbf{Pr}[\ \mathcal{E}_x\ ] = p$ and $\mathbf{Pr}[\ \mathcal{E}_x \cap \mathcal{E}_{x'}\ ] = p^2$, whenever $x \neq x'$.

By the inclusion-exclusion principle, we have:

$$\mathbf{Pr}\Big[\ \bigcup_{x\in S} \mathcal{E}_x\ \Big] \ \geqslant\ \sum_{x\in S} \mathbf{Pr}[\ \mathcal{E}_x\ ] \ - \sum_{x,x'\in S \text{ and } x\neq x'} \mathbf{Pr}[\ \mathcal{E}_x \cap \mathcal{E}_{x'}\ ] \ =\ |S|p \ -\ \binom{|S|}{2} \cdot p^2$$

By union bound, we have:

$$\mathbf{Pr}\Big[\ \bigcup_{x,x'\in S \text{ and } x\neq x'} \mathcal{E}_x \cap \mathcal{E}_{x'}\ \Big] \ \leqslant\ \sum_{x,x'\in S \text{ and } x\neq x'} \mathbf{Pr}[\ \mathcal{E}_x \cap \mathcal{E}_{x'}\ ] \ \leqslant\ \binom{|S|}{2} \cdot p^2$$

Combining both, we have:

$$\mathbf{Pr}[\ N = 1\ ] \ =\ \mathbf{Pr}[\ N \geqslant 1\ ] - \mathbf{Pr}[\ N \geqslant 2\ ] \ \ \geqslant\ \ |S|p \ -\ |S|^2p^2$$

Since $1/4 \leqslant |S|p \leqslant 1/2$, a straightforward calculation shows that $|S|p - |S|^2p^2 \geqslant 3/16$. ∎

[§] $\{0,1\}^{n\times n}$ denotes the set of 0-1 matrices with n rows and n columns and $\{0,1\}^{n\times 1}$ denotes the set of 0-1 column vectors of n rows. Here the addition $+$ and multiplication $\cdot$ are defined over $\mathbb{Z}_2$.