On independence of variants of the weak pigeonhole principle

Emil Jeřábek*
Institute of Mathematics, AS CR
Žitná 25
115 67 Praha 1
Czech Republic
jerabek@math.cas.cz

March 16, 2007

Abstract

The principle $sPHP^a_b(PV(\alpha))$ states that no oracle circuit can compute a surjection of a onto b. We show that $sPHP^{\varrho(a)}_{P(a)}(PV(\alpha))$ is independent of $PV_1(\alpha)+sPHP^{\pi(a)}_{\Pi(a)}(PV(\alpha))$ for various choices of the parameters π, Π, ϱ, P. We also improve the known separation of $iWPHP(PV)$ from $S^1_2 + sWPHP(PV)$ under cryptographic assumptions.

Keywords: bounded arithmetic, pigeonhole principle, KPT witnessing, Boolean circuit.

1 Introduction

Many variants of the weak pigeonhole principle for polynomial-time functions have been used in the literature on bounded arithmetic: on the one hand, we have the injective version $iPHP^a_b(PV)$, which prohibits the existence of injective functions from a to $b < a$, and the surjective (or dual) version $sPHP^a_b(PV)$, which prohibits surjective functions from a onto $b > a$. On the other hand, we may vary the parameters a and b; for example, we may consider $sPHP^a_{a\#a}(PV)$, $sPHP^a_{a^2}(PV)$, $sPHP^a_{2a}(PV)$, or $sPHP^{a|a|}_{a(|a|+1)}(PV)$. For many choices of the parameters of $sWPHP(PV)$, the resulting principles are equivalent over S^1_2, by a well-known construction which goes back to Paris and Wilkie [14]. In the case of $iWPHP$, similar equivalences are provable even in the weaker theory PV_1.

We will show that in the case of $sWPHP$, we cannot weaken the base theory to PV_1 if we relativize the principles: $sPHP^a_{a\#a}(PV(\alpha))$, $sPHP^a_{a^2}(PV(\alpha))$, $sPHP^a_{2a}(PV(\alpha))$, etc., are not equivalent over $PV_1(\alpha)$. In fact, we will find a general condition on PV-functions π, Π, ϱ, and P which guarantees that $PV_1(\alpha)+\forall a\, sPHP^{\pi(a)}_{\Pi(a)}(PV(\alpha))$ does not prove $\forall a\, sPHP^{\varrho(a)}_{P(a)}(PV(\alpha))$, and we will show that the condition is almost optimal.

*The research was done while the author was visiting the Department of Computer Science of the University of Toronto. Supported by NSERC Discovery grant, grant IAA1019401 of GA AV ČR, and grant 1M0545 of MŠMT ČR.

Another interesting problem is the relation of the (unrelativized) surjective and injective versions of $sWPHP$. By a result of Krajíček and Pudlák [10], later improved by Thapen [16], the theory $S^1_2 + sWPHP(PV)$ does not prove $iWPHP(PV)$, assuming the security of the RSA cryptosystem against randomized polynomial-time attacks. We will show that the same conclusion holds under a weaker assumption, viz. there is no randomized poly-time algorithm for factoring of integers; moreover, we exhibit a simple number-theoretic statement about quadratic residues which separates $S^1_2 + iWPHP(PV)$ from $S^1_2 + sWPHP(PV)$ under this assumption. We also point out that $S^2_2(\alpha) + iWPHP(PV(\alpha))$ does not prove $sWPHP(PV(\alpha))$.

2 Preliminaries

We briefly summarize basic definitions and facts about bounded arithmetic to fix the notation. More details can be found in Krajíček [9] or Buss [3].

Buss' theories are formulated in the language $L = \langle 0, S, +, \times, \leq, \#, |x|, \lfloor \frac{x}{2} \rfloor \rangle$. The intended meaning of the symbols are the usual arithmetical operations on non-negative integers, and $|x| = \lceil \log_2(x+1) \rceil$, $x \# y = 2^{|x|\,|y|}$. *Bounded quantifiers* are introduced by

$$\exists x \leq t\, \varphi :\Leftrightarrow \exists x\, (x \leq t \wedge \varphi),$$
$$\forall x \leq t\, \varphi :\Leftrightarrow \forall x\, (x \leq t \rightarrow \varphi),$$

where t is a term without an occurrence of the variable x. Such a quantifier is *sharply bounded*, if the head function symbol of t is $|\cdot|$. A formula φ is bounded (sharply bounded) if all quantifiers in φ are bounded (sharply bounded). A formula is Σ^b_1 if it is constructed from sharply bounded formulas by means of $\wedge$, $\vee$, sharply bounded, and existential bounded quantifiers. In general, Σ^b_i-formulas consist of i alternating blocks of bounded quantifiers followed by a sharply bounded formula, where the first block is existential, and we ignore sharply bounded quantifiers which are allowed to appear anywhere in the quantifier prefix. The theory S^i_2 is axiomatized by a finite set of open axioms denoted by $BASIC$, which state elementary properties of the symbols of L, and the schema of *polynomial induction*

$$(PIND) \qquad \varphi(0) \wedge \forall x \leq a\, (\varphi(\lfloor \tfrac{x}{2} \rfloor) \rightarrow \varphi(x)) \rightarrow \varphi(a)$$

for Σ^b_i-formulas φ. Alternatively, S^i_2 can be axiomatized by $BASIC$ and the *length induction* schema

$$(LIND) \qquad \varphi(0) \wedge \forall x < |a|\, (\varphi(x) \rightarrow \varphi(x+1)) \rightarrow \varphi(|a|)$$

for Σ^b_i-formulas.

The theory T^i_2 is axiomatized over $BASIC$ by the *induction* schema

$$(IND) \qquad \varphi(0) \wedge \forall x < a\, (\varphi(x) \rightarrow \varphi(x+1)) \rightarrow \varphi(a)$$

for Σ^b_i-formulas φ.

PV is an equational theory introduced by Cook [5]. Its language contains function symbols for all polynomial-time algorithms, introduced inductively using limited recursion on notation (cf. Cobham [4]). It is axiomatized by defining equations of its function symbols, and a derivation rule similar to *PIND*. PV_1, also known as QPV, $T^0_2(\Box^p_1)$, or $\forall\Sigma^b_1(S^1_2)$, is a first-order variant of PV. It can be axiomatized by equations provable in PV together with the axioms $0 \neq 1$ and $\lfloor \frac{x}{2} \rfloor = 0 \to x = 0 \lor x = 1$, and it proves the *PIND* and *IND* schemata for sharply bounded formulas. We will also use the symbol PV to denote the set of function symbols of PV.

The theory $S^1_2(PV)$, which is axiomatized by PV_1 and $\Sigma^b_1(PV)$-*PIND*, is an extension of S^1_2 by definitions; for this reason, we will usually work in the more convenient language of PV, and identify S^1_2 with $S^1_2(PV)$. By Buss' theorem, S^1_2 is Σ^b_1-conservative over PV_1.

For any function f, the surjective (also called dual) and injective variants of the pigeonhole principle are defined as

$$\begin{aligned} sPHP^a_b(f) &:\Leftrightarrow \exists v < b\, \forall u < a\, f(u) \neq v, \\ iPHP^a_b(f) &:\Leftrightarrow \exists u < a\, f(u) \geq b \lor \exists u_0 < a \exists u_1 < u_0\, f(u_0) = f(u_1). \end{aligned}$$

We also define

$$\begin{aligned} sPHP^a_b(PV) &:\Leftrightarrow \forall C\, sPHP^a_b(\mathrm{eval}(C, \cdot)), \\ iPHP^a_b(PV) &:\Leftrightarrow \forall C\, iPHP^a_b(\mathrm{eval}(C, \cdot)), \end{aligned}$$

where $\mathrm{eval}(C, x)$ is a PV-function which evaluates a circuit C on an input x. This is equivalent to taking $sPHP^a_b(f)$ (or $iPHP^a_b(f)$) for all PV-functions f (with parameters), but our formulation is a single formula rather than an infinite schema. We define $sWPHP(PV)$ and $iWPHP(PV)$ as $\forall a > 0\, sPHP^a_{2a}(PV)$ and $\forall a > 0\, iPHP^{2a}_a(PV)$ (respectively), but we will only use this notation in contexts where Theorem 3.1 is applicable, so that the exact value of the bounds does not matter.

All these theories can be *relativized* in a straightforward way. We include a new predicate α in the language, and define $\Sigma^b_i(\alpha)$ as before, but allowing α to be used in atomic formulas. The theory $S^i_2(\alpha)$ consists of *BASIC* and $\Sigma^b_i(\alpha)$-*PIND* (i.e., there are no axioms involving α apart from the induction axioms), and similarly $T^i_2(\alpha) = BASIC + \Sigma^b_i(\alpha)\text{-}IND$. In the case of $PV(\alpha)$ and $PV_1(\alpha)$, we allow the characteristic function of α to appear in functions constructed by limited recursion on notation, so that function symbols of $PV(\alpha)$ correspond to polynomial-time oracle algorithms.

We define

$$\begin{aligned} sPHP^a_b(PV(\alpha)) &:\Leftrightarrow \forall C\, sPHP^a_b(\mathrm{eval}^\alpha(C, \cdot)), \\ iPHP^a_b(PV(\alpha)) &:\Leftrightarrow \forall C\, iPHP^a_b(\mathrm{eval}^\alpha(C, \cdot)), \end{aligned}$$

where the function $\mathrm{eval}^\alpha(C, x)$ evaluates an *oracle circuit* C on input x, supplying values of α for the oracle. Both $sWPHP(PV(\alpha))$ and $iWPHP(PV(\alpha))$ are provable in $T^2_2(\alpha)$ [14, 9, 12], but neither is provable in $S^2_2(\alpha)$ [15].

We will generally reserve the variables C, D, … for oracle circuits, and we will write $C(u)$ instead of $\mathrm{eval}^\alpha(C, u)$. When we wish to emphasize the dependence on the oracle, we write it as a superscript C^α. We identify a number a with the integer interval $[0, a)$. The notation

$$C\colon a \to b$$

means that the oracle circuit C computes a function from a to b, i.e., $\forall u < a\, C(u) < b$. We write

$$C\colon a \twoheadrightarrow b,$$
$$C\colon a \hookrightarrow b,$$

if the function is surjective or injective (respectively). Thus, we can express $sPHP^a_b(PV(\alpha))$ and $iPHP^a_b(PV(\alpha))$ concisely as

$$sPHP^a_b(PV(\alpha)) \Leftrightarrow \neg\exists C\,(C\colon a \twoheadrightarrow b),$$
$$iPHP^a_b(PV(\alpha)) \Leftrightarrow \neg\exists C\,(C\colon a \hookrightarrow b).$$

We will need several results on provability in bounded arithmetic, most importantly witnessing theorems. The first one is *Parikh's theorem.*

Theorem 2.1 (Parikh [13]) *Let T be an extension of BASIC axiomatized by bounded formulas. If*

$$T \vdash \forall x\, \exists y\, \varphi(x, y),$$

where φ is a bounded formula, then there exists a term t such that

$$T \vdash \forall x\, \exists y \le t(x)\, \varphi(x, y).$$

The next one is *Buss' witnessing theorem.*

Theorem 2.2 (Buss [2]) *Let $i > 0$, and $\varphi \in \Sigma^b_i(\alpha)$ such that $S^i_2(\alpha) \vdash \forall x\, \exists y\, \varphi(x, y)$. Then there exists an $FP^{\Sigma^b_{i-1}(\alpha)}$-algorithm f^α such that $\langle \mathbb{N}, A\rangle \vDash \varphi(x, f(x))$ for every x and every oracle A.*

Wilkie's witnessing theorem is a variant of Buss' theorem for $sWPHP(PV(\alpha))$. (The result was never published by A. Wilkie, see Krajíček [9] or Thapen [17] for a proof.) We only formulate it in the unrelativized version for simplicity.

Theorem 2.3 *If $S^1_2 + sWPHP(PV) \vdash \forall x\, \exists y\, \varphi(x, y)$, where φ is a Σ^b_1-formula, then there exists a probabilistic polynomial-time algorithm which, for every x, computes with high probability a y such that $\mathbb{N} \vDash \varphi(x, y)$.*

As noted in [7], Thapen's proof of Wilkie's witnessing theorem also implies a description of $\forall\Sigma^b_1(\alpha)$-theorems of $S^1_2(\alpha) + sWPHP(PV(\alpha))$. We need to introduce another variant of *PHP* to state it:

$$rPHP^a_b(f, g) \Leftrightarrow \exists v < b\, (g(v) \ge a \vee f(g(v)) \ne v).$$

(Here r stands for "retraction pair", borrowed from category theory terminology.)

Theorem 2.4 ([17, 7]) *$\forall\Sigma^b_1(\alpha)$-consequences of $S^1_2(\alpha) + sWPHP(PV(\alpha))$ are axiomatized by $PV_1(\alpha) + rWPHP(PV(\alpha))$.*

Finally, we will make use of the *KPT witnessing theorem.* In its basic form, it is a variant of Herbrand's theorem.

Theorem 2.5 (Krajíček, Pudlák, Takeuti [11]) *Assume that*

$$PV_1(\alpha) \vdash \forall x\, \exists y\, \forall z\, \varphi(x, y, z),$$

where φ is an existential formula. There exist $PV(\alpha)$-functions $f_0, \ldots, f_k$ such that

$$PV_1(\alpha) \vdash \varphi(x, f_0(x), z_0) \vee \varphi(x, f_1(x, z_0), z_1) \vee \cdots \vee \varphi(x, f_k(x, z_0, \ldots, z_{k-1}), z_k).$$

The KPT witnessing theorem also has a computational interpretation. We consider an interactive protocol with two players, a student and a teacher (who both have access to the oracle α). They are given a number x, and the student tries to find y such that $\forall z\, \varphi(x, y, z)$. The teacher either accepts the student's answer, or responds with a counterexample: z such that $\neg\varphi(x, y, z)$. The KPT witnessing theorem can be restated as follows: if $PV(\alpha)$ proves $\exists y\, \forall z\, \varphi(x, y, z)$, there exists a constant k, and a polynomial-time strategy for the student, which makes any correct (computationally unlimited) teacher accept after at most k rounds. Here, polynomial-time means time polynomial in the length of x, and of the previous responses of the teacher; however, if the quantifier $\forall z$ is bounded, the time is polynomial in $|x|$ alone.

3 Variants of $sWPHP$

As we already mentioned in the introduction, variants of $sWPHP(PV(\alpha))$ are equivalent in $S^1_2(\alpha)$ for a wide range of the parameters, and the corresponding variants of $iWPHP(PV(\alpha))$ and $rWPHP(PV(\alpha))$ are equivalent over $PV_1(\alpha)$. The result is implicit in [14], but for completeness and illustrative purposes, we include a proof.

Theorem 3.1

(i) The following are equivalent over $S^1_2(\alpha)$.

- (α) $\forall a\, \exists b\, sPHP^a_b(PV(\alpha))$
- (β) $\forall a > 0\, sPHP^a_{2a}(PV(\alpha))$
- (γ) $\forall a > 0\, \forall b\, sPHP^{a|b|}_{a(|b|+1)}(PV(\alpha))$

(ii) The following are equivalent over $PV_1(\alpha)$.

- (α) $\forall a\, \exists b\, iPHP^b_a(PV(\alpha))$
- (β) $\forall a > 0\, iPHP^{2a}_a(PV(\alpha))$
- (γ) $\forall a > 0\, \forall b\, iPHP^{a(|b|+1)}_{a|b|}(PV(\alpha))$

(iii) The following are equivalent over $PV_1(\alpha)$.

(α) $\forall a \, \exists b \, rPHP^a_b(PV(\alpha))$

(β) $\forall a > 0 \, rPHP^a_{2a}(PV(\alpha))$

(γ) $\forall a > 0 \, \forall b \, rPHP^{a|b|}_{a(|b|+1)}(PV(\alpha))$

Proof: Let us start with (i). The implication $(\gamma) \to (\alpha)$ is trivial. $(\beta) \to (\gamma)$: assume that there exists an oracle circuit C such that $C\colon a|b| \twoheadrightarrow a(|b|+1)$. We define a new circuit $C'\colon a(2|b|-1) \to 2a|b|$ by

$$C'(u) = \begin{cases} C(u), & u < a|b|, \\ u + a, & \text{otherwise.} \end{cases}$$

Notice that $C'\colon a(|b|+i) \twoheadrightarrow a(|b|+i+1)$ for every $i < |b|$. Let C^i be the composition of i copies of C'. For any $v < 2a|b|$, we can prove

$$\exists u < a(2|b|-i) \, C^i(u) = v$$

by $\Sigma^b_1(\alpha)$-*LIND* on $i \le |b|$, thus $C^{|b|}\colon a|b| \twoheadrightarrow 2a|b|$.

$(\alpha) \to (\beta)$: let $C\colon a \twoheadrightarrow 2a$, and fix any $b > 0$. Define $C'\colon a2^{|b|-1} \twoheadrightarrow a2^{|b|}$ by

$$C'(u) = C(u \bmod a) + 2a\lfloor u/a \rfloor,$$

and let C^i be the composition of i copies of C'. We have $C'\colon a2^i \twoheadrightarrow a2^{i+1}$ for every $i < |b|$. For any fixed $v < a2^{|b|}$, we prove

$$\exists u < a2^{|b|-i} \, C^i(u) = v$$

by $\Sigma^b_1(\alpha)$-*LIND* on $i \le |b|$. Thus $C^{|b|}\colon a \twoheadrightarrow a2^{|b|} > b$, and $D\colon a \twoheadrightarrow b$, where $D(u) = \min(b-1, C^{|b|}(u))$.

The proofs of (ii) and (iii) are quite similar. We will only show the proof of $(\beta) \to (\gamma)$ in (ii), to see why the complexity of the induction formula drops. Let $C\colon a(|b|+1) \hookrightarrow a|b|$. Define $C'\colon 2a|b| \hookrightarrow a(2|b|-1)$ by

$$C'(u) = \begin{cases} C(u), & u < a(|b|+1), \\ u - a, & \text{otherwise,} \end{cases}$$

and for every $i \le |b|$, let C^i be the composition of i copies of C'. We have $C'\colon a(|b|+i+1) \hookrightarrow a(|b|+i)$ for every $i < |b|$. For any $u < u' < 2a|b|$, we prove

$$C^i(u) \ne C^i(u')$$

by induction on $i \le |b|$, thus $C^{|b|}\colon 2a|b| \hookrightarrow a|b|$. $\square$

Corollary 3.2 *Let $\pi(a)$, $\Pi(a)$ be PV-functions such that PV_1 proves that π is unbounded, and*

$$\frac{\Pi(a)}{\pi(a)} \geq 1 + \frac{1}{|a|^c}$$

for some constant c. Then

$$\begin{aligned}
S^1_2(\alpha) + \forall a\, sPHP^{\pi(a)}_{\Pi(a)}(PV(\alpha)) &= S^1_2(\alpha) + sWPHP(PV(\alpha)),\\
PV_1(\alpha) + \forall a\, iPHP^{\Pi(a)}_{\pi(a)}(PV(\alpha)) &= PV_1(\alpha) + iWPHP(PV(\alpha)),\\
PV_1(\alpha) + \forall a\, rPHP^{\pi(a)}_{\Pi(a)}(PV(\alpha)) &= PV_1(\alpha) + rWPHP(PV(\alpha)),
\end{aligned}$$

and moreover, the theories

$$\begin{aligned}
&PV_1(\alpha) + \forall a\, sPHP^{\pi(a)}_{\Pi(a)}(PV(\alpha))\\
&PV_1(\alpha) + sWPHP(PV(\alpha))
\end{aligned}$$

prove the same $\forall\Sigma^b_1(\alpha)$-sentences.

Proof: Use Theorems 3.1 and 2.4. □

We are going to show that $PV_1(\alpha)$ does not prove the equivalence from Theorem 3.1 (i), or even its typical special cases. Before doing that, we remark that the *proof* of 3.1 (i) we have given requires $S^1_2(\alpha)$. We used $\Sigma^b_1(\alpha)$-induction to show that the composition of a sequence of surjective circuits (or iteration of a single circuit) is itself surjective. Let us define the *Surjective Circuit Iteration Principle* (*SCIP*) to be the following statement: if $a = \{a_i;\, i \leq k\}$ is a sequence of numbers, and C is an oracle circuit such that $C\colon a_i \twoheadrightarrow a_{i+1}$ for every $i < k$, then the k-fold iterate of C is a surjection of a_0 to a_k.

Theorem 3.3 $PV_1(\alpha) + SCIP = S^1_2(\alpha)$.

Proof: Work in $PV_1(\alpha) + SCIP$, and assume

$$\varphi(0) \wedge \forall i < |x|\, (\varphi(i) \to \varphi(i+1)),$$

where φ is a strict $\Sigma^b_1(\alpha)$-formula, we will show $\varphi(|x|)$. There exists a number $b \geq 2$, and a Boolean oracle circuit D such that

$$\varphi(i) \leftrightarrow \exists u < b\, D(u,i) = 1$$

for every $i \leq |x|$. For any $i \leq |x|$, we identify b^i with the set of sequences of length i with elements from b. We put

$$a_i = \frac{b^{i+1} - 1}{b - 1} + 1$$

for every $i \leq |x|$, and we identify a_i with the disjoint union

$$\{*\} \cup \bigcup_{j \leq i} b^j,$$

where $*$ is a new element. (All these identifications can be realized by polynomial-time functions in a straightforward way.) We define a circuit $C\colon a_{|x|} \to a_{|x|-1}$ by

$$C(\langle u_j;\, j \le i\rangle) = \begin{cases} \langle u_j;\, j < i\rangle, & \text{if } D(u_i, i) = 1 \text{ or } i > 0 \wedge D(u_{i-1}, i-1) = 0, \\ *, & \text{otherwise,} \end{cases}$$

$$C(\langle\rangle) = C(*) = *.$$

We claim $C\colon a_{k+1} \twoheadrightarrow a_k$ for every $k < |x|$. Consider any $u = \langle u_j;\, j < i\rangle \in a_k$. If $i = 0$, we can pick $u_0 < b$ such that $D(u_0, 0) = 1$, as $\varphi(0)$. If $i > 0$, and $D(u_{i-1}, i-1) = 0$, we take any $u_i < b$. If $i > 0$, and $D(u_{i-1}, i-1) = 1$, then $\varphi(i-1)$, thus $\varphi(i)$, and we may choose $u_i < b$ such that $D(u_i, i) = 1$. In each case, $u' = \langle u_j;\, j \le i\rangle \in a_{k+1}$, and $C(u') = u$.

For every $i \le |x|$, let C^i be the composition of i copies of C. *SCIP* implies that $C^{|x|}\colon a_{|x|} \twoheadrightarrow a_0$, thus there exists $u \in a_{|x|}$ such that $C^{|x|}(u) = \langle\rangle$. As $C^{|x|}(*) = *$, we have $u = \langle u_j;\, j < k\rangle$ for some $k \le |x|$. We have

$$C^i(u) = \langle u_j;\, j < k - i\rangle$$

by induction on $i \le |x|$, thus $k = |x|$. Then we can prove

$$\forall j \le i\, D(u_j, j) = 1$$

by induction on $i \le |x|$, which implies $\varphi(|x|)$. □

Of course, Theorem 3.3 does not exclude the possibility that we can prove the equivalence of variants of $sWPHP(PV(\alpha))$ in $PV_1(\alpha)$ in a completely different way. We will need a more complicated argument to rule it out. This is the main result of this paper.

Theorem 3.4 *Let π, Π, ϱ, P be PV-functions such that $\Pi(x) > \pi(x)$ and $P(x) > \varrho(x)$ for every x. Assume that*

$$PV_1(\alpha) + \forall b\, sPHP^{\pi(b)}_{\Pi(b)}(PV(\alpha)) \vdash \forall a\, sPHP^{\varrho(a)}_{P(a)}(PV(\alpha)).$$

Then there exists a constant $c > 1$ such that the following holds. For every $a \ge c$ such that $\varrho(a) \ge |a|^c$, there exists b such that

$$|b| \le |a|^c \wedge \pi(b) \ge \frac{\varrho(a)}{|a|^c} \wedge \left(\left(\frac{P(a)}{\varrho(a)}\right)^c \ge \frac{\Pi(b)}{\pi(b)} \vee \frac{P(a)}{\varrho(a)} - 1 \ge \left(\frac{\Pi(b)}{\pi(b)} - 1\right)^c\right).$$

Proof: For convenience, we will consider α as a function rather than predicate (we could encode it by its bit graph). By assumption, $PV_1(\alpha)$ proves

$$\exists b, C\, \forall v < \Pi(b)\, \exists u < \pi(b)\, C(u) = v \vee \exists y < P(a)\, \forall x < \varrho(a)\, \alpha(x) \ne y.$$

By Parikh's theorem and prenexing, there exists a constant k such that $PV_1(\alpha)$ proves the $\exists\forall\exists$-formula

$$\exists b, C \le 2^{|a|^k}\, \exists y < P(a)\, \forall x < \varrho(a)\, \forall v < \Pi(b)\, \exists u < \pi(b)\, (C(u) = v \vee \alpha(x) \ne y).$$

By the KPT witnessing theorem, there exists a constant c', and an interactive protocol between a student (S) and teacher (T) with the following properties. Given a number a, S tries to find either (1) a $y \in P(a) \smallsetminus \mathrm{rng}(\alpha \restriction \varrho(a))$, or (2) a number b, and an oracle circuit C such that $C\colon \pi(b) \twoheadrightarrow \Pi(b)$. T provides counterexamples to S's suggestions: $x < \varrho(a)$ such that $\alpha(x) = y$ in case (1), and $v \in \Pi(b) \smallsetminus \mathrm{rng}(C \restriction \pi(b))$ in case (2). Both S and T have oracle access to α; S works deterministically in time $|a|^{c'}$, and T is computationally unlimited. S succeeds after at most c' rounds for every honest T. Notice that S's final answer must be of type (1), as in fact no surjection from $\pi(b)$ to $\Pi(b)$ exists. We may assume that S verifies the correctness of T's counterexamples in case (1) by an oracle call to α.

Fix a such that $\varrho(a) > 4|a|^{c'}$, and put $s := \lfloor \varrho(a)/(2|a|^{c'}) \rfloor$. We modify S as follows: whenever she would ask T for a counterexample to $C\colon \pi(b) \twoheadrightarrow \Pi(b)$ such that $\pi(b) \leq s$, she computes the answer herself by evaluating C on all possible inputs. (When $\pi(b) > s$, the algorithm is unchanged.) S no longer works in polynomial time, but the number of oracle calls is bounded by $\varrho(a)/2$: at most $s|C|$ calls (i.e., $|C|$ for each input) are used to evaluate C^α, and the sum of the sizes of all C is bounded by $|a|^{c'}$, the original running time. We consider the following randomized procedure:

- choose a uniformly random injection $\alpha\colon \varrho(a) \hookrightarrow P(a)$
- choose a sufficiently large pool of random bits w
- run the interactive protocol described above, using the following answers for T:
  - in case (1), reply with the (unique) $x < \varrho(a)$ such that $\alpha(x) = y$, or with $*$ if $y \notin \mathrm{rng}(\alpha)$
  - in case (2), use fresh random bits from w to select uniformly a $v < \Pi(b)$

Notice that T may break the protocol: in case (2), there is no guarantee that $v \notin \mathrm{rng}(C^\alpha)$. However, we have a simple upper bound on the probability of such event.

Claim 1 *For any fixed α, and random w, all answers of T are correct with probability at least $(1-p)^{c'}$, where*

$$p = \max\left\{ \frac{\pi(b)}{\Pi(b)};\ |b| \leq |a|^{c'}, \pi(b) \geq s \right\}.$$

Proof: For a given $C\colon \pi(b) \to \Pi(b)$, the probability of $v \in \mathrm{rng}(C^\alpha)$ is at most $\pi(b)/\Pi(b) \leq p$. T answers at most c' questions in total, and the answers of type (2) are independent. $\square$ (Claim 1)

The next task is to bound the probability that S fails to find a $y \in P(a) \smallsetminus \mathrm{rng}(\alpha)$, i.e., that T does not answer $*$ to any question. Fix w, and for random α, let E_i denote the event "T did not answer $*$ to any of the first i questions of type (1)", where $i \leq c'$. Clearly, S fails iff E_c. We have $\Pr_\alpha(E_0) = 1$, and E_{i+1} implies E_i, thus

$$\Pr_\alpha(E_{c'}) = \Pr_\alpha(E_1 \mid E_0) \Pr_\alpha(E_2 \mid E_1) \cdots \Pr_\alpha(E_{c'} \mid E_{c'-1}).$$

Claim 2 *For any fixed w and $i < c'$,*

$$\Pr_\alpha(E_{i+1} \mid E_i) \geq \left(\frac{\varrho(a)}{P(a)}\right)^2.$$

Proof: For any α, there exists a set $Q_\alpha \subseteq \varrho(a)$ of size $\varrho(a)/2$ such that when we run our procedure on α, S only queries the oracle α for values on Q_α. Put $B = \{\alpha \restriction Q_\alpha; E_i(\alpha)\}$. If $\beta \in B$ and $\alpha \supseteq \beta$, then $E_i(\alpha)$ holds. Indeed, if α' is such that $\beta = \alpha' \restriction Q_{\alpha'}$ and $E_i(\alpha')$, then the procedure behaves identically on α and α' until the $(i+1)$th question of type (1): the computation of S is deterministic, answers of T of type (2) are predetermined as w is fixed, answers to oracle calls of S to α are determined by β from the definition of $Q_{\alpha'}$, and T's answers of type (1) are also fixed by β as all return an x such that $\alpha'(x) = y$, and we required S to query $\alpha'(x)$ after any such answer. Thus,

$$\Pr_\alpha(E_{i+1} \mid E_i) \geq \min_{\beta \in B} \Pr_\alpha(E_{i+1} \mid \alpha \supseteq \beta).$$

Fix $\beta \in B$, and let $y < P(a)$ be the $(i+1)$th question of type (1). If there exists $x \in \mathrm{dom}(\beta)$ such that $\beta(x) = y$, then clearly $\Pr_\alpha(E_{i+1} \mid \alpha \supseteq \beta) = 1$. Otherwise, choosing a random injection $\alpha \supseteq \beta$ is equivalent to choosing a random injection $\gamma\colon \varrho(a) \smallsetminus \mathrm{dom}(\beta) \hookrightarrow P(a) \smallsetminus \mathrm{rng}(\beta)$, hence

$$\begin{aligned}\Pr_\alpha(E_{i+1} \mid \alpha \supseteq \beta) = \Pr_\gamma(y \in \mathrm{rng}(\gamma)) &= \frac{\varrho(a) - |\mathrm{dom}(\beta)|}{P(a) - |\mathrm{rng}(\beta)|} \\ &= \frac{\varrho(a)/2}{P(a) - \varrho(a)/2} = \frac{\varrho^2(a)}{2P(a)\varrho(a) - \varrho^2(a)} \geq \left(\frac{\varrho(a)}{P(a)}\right)^2,\end{aligned}$$

as $P^2(a) - 2P(a)\varrho(a) + \varrho^2(a) = (P(a) - \varrho(a))^2 \geq 0$. □ (Claim 2)

By soundness of the protocol, S succeeds whenever all answers of T are correct, thus by Claim 1, there exists b such that $|b| \leq |a|^{c'}$, $\pi(b) \geq \lfloor \varrho(a)/(2|a|^{c'}) \rfloor \geq \varrho(a)/(4|a|^{c'})$, and S succeeds with probability at least

$$\left(1 - \frac{\pi(b)}{\Pi(b)}\right)^{c'}.$$

On the other hand, Claim 2 and the preceding discussion implies that S fails with probability at least

$$\left(\frac{\varrho(a)}{P(a)}\right)^{2c'},$$

thus

$$\left(\frac{\varrho(a)}{P(a)}\right)^d + \left(1 - \frac{\pi(b)}{\Pi(b)}\right)^d \leq 1,$$

where $d := 2c'$. Put $u = P(a)/\varrho(a)$ and $v = \Pi(b)/\pi(b)$. Notice that $|a|^d \geq 4|a|^{c'}$ for large enough a; the theorem thus follows from the following claim.

Claim 3 *For every constant $d \geq 1$, there exists a constant $c \geq d$ such that for every $u, v > 1$,*

$$u^{-d} + (1 - v^{-1})^d \leq 1 \to u^c \geq v \vee u - 1 \geq (v-1)^c.$$

Proof: As x^d is convex and $x^{1/d}$ is concave,

$$u^c \geq \frac{1}{1-(1-v^{-1})^d} \geq \frac{1}{1-(1-dv^{-1})} = d^{-1}v,$$
$$u \geq \frac{1}{(1-(1-v^{-1})^d)^{1/d}} \geq \frac{1}{1-d^{-1}(1-v^{-1})^d} \geq 1+d^{-1}(1-v^{-1})^d.$$

If $v \leq 3/2$, we have

$$u-1 \geq d^{-1}v^{-d}(v-1)^d \geq (v-1)^{2d+\log d}$$

from the second inequality, because $v^{-1} \geq 1/2 \geq v-1$. If $v \geq 3/2$, the second inequality gives $u \geq 1+d^{-1}3^{-d}$, thus $u^{d3^d} \geq 2$, and

$$u^{d(1+3^d \log d)} \geq du^d \geq v$$

from the first inequality. □ (Claim 3)

□

Remark 3.5 The disjuncts of the condition

$$(*) \qquad \left(\frac{P(a)}{\varrho(a)}\right)^c \geq \frac{\Pi(b)}{\pi(b)} \vee \frac{P(a)}{\varrho(a)} - 1 \geq \left(\frac{\Pi(b)}{\pi(b)} - 1\right)^c$$

are each relevant only for certain settings of the parameters, the cross-over point being when $\Pi(b)/\pi(b)$ is about $1+\Theta(1)$. More precisely, for every constant $0<\varepsilon<1$ and $c>1$, there is a constant $d>1$ such that

- if $\Pi(b)/\pi(b) \geq 1+\varepsilon$, then $(*)$ implies
$$\left(\frac{P(a)}{\varrho(a)}\right)^d \geq \frac{\Pi(b)}{\pi(b)},$$
- if $\Pi(b)/\pi(b) \leq 1+\varepsilon$, then $(*)$ implies
$$\frac{P(a)}{\varrho(a)} - 1 \geq \left(\frac{\Pi(b)}{\pi(b)} - 1\right)^d.$$

Corollary 3.6

(i) $PV_1(\alpha) + \forall a\, \exists b\, sPHP^a_b(PV(\alpha)) \nvdash \forall a\, sPHP^a_{f(a)}(PV(\alpha))$ *for any PV-function f,*

(ii) $PV_1(\alpha) + \forall a\, sPHP^a_{a\#a}(PV(\alpha)) \nvdash \forall a > 1\, sPHP^a_{a^2}(PV(\alpha))$,

(iii) $PV_1(\alpha) + \forall a > 1\, sPHP^a_{a^2}(PV(\alpha)) \nvdash \forall a > 0\, sPHP^a_{2a}(PV(\alpha))$,

(iv) $PV_1(\alpha) + \forall a > 0\, sPHP^a_{2a}(PV(\alpha)) \nvdash \forall a > 0\, sPHP^{a||a||}_{a(||a||+1)}(PV(\alpha))$,

(v) $PV_1(\alpha) + \forall a > 0\, sPHP^{a||a||}_{a(||a||+1)}(PV(\alpha)) \nvdash \forall a > 0\, sPHP^{a|a|}_{a(|a|+1)}(PV(\alpha))$.

(*vi*) $PV_1(\alpha) + \forall a > 0\, sPHP^{a|a|}_{a(|a|+1)}(PV(\alpha)) \nvdash \forall a\, sPHP^a_{a+1}(PV(\alpha))$.

Proof: We will show (*i*), the others are similar but easier. Let $f_k(a) = 2^{|a|^k}$. There exists a $k \geq 1$ such that PV_1 proves $f(a) \leq f_k(a)$, and obviously

$$PV_1(\alpha) + \forall a\, sPHP^a_{f_{k+1}(a)}(PV(\alpha)) \vdash \forall a\, \exists b\, sPHP^a_b(PV(\alpha)),$$

it thus suffices to prove

$$PV_1(\alpha) + \forall a\, sPHP^a_{f_{k+1}(a)}(PV(\alpha)) \nvdash sPHP^a_{f_k(a)}(PV(\alpha)).$$

Assume otherwise. By Theorem 3.4 and Remark 3.5, there exists a constant c such that for all sufficiently large a, there exists b such that $b \geq a/|a|^c$, and

$$\left(\frac{f_k(a)}{a}\right)^c \geq \frac{f_{k+1}(b)}{b}.$$

The function $f_{k+1}(b)/b$ is close enough to monotone to obtain

$$\left(\frac{f_k(a)}{a}\right)^c \geq \frac{f_{k+1}(a/|a|^{c+1})}{a/|a|^{c+1}}$$

for all $a \gg 0$. Taking logarithms, we have

$$c|a|^k - c|a| \geq (|a| - (c+1)||a||)^{k+1} - |a| + (c+1)||a||,$$

thus $|a|^{k+1} = O(|a|^k)$, a contradiction. $\square$

Example 3.7 If $\forall a\, sPHP^a_{P(a)}(PV(\alpha))$ is equivalent to $\forall a\, sPHP^a_{2a}(PV(\alpha))$ over $PV_1(\alpha)$, then

- $P(a) = a + \Omega(a)$,
- $\exists c\, \forall n\, \exists a\, (n \leq |a| \leq n^c \wedge P(a) \leq ca)$. $\square$

The reader may wish to see that the messiness of the condition in Theorem 3.4 is a property of nature, not just an accidental residue of our proof. We will indeed show that the condition is close to optimal; moreover, the only manipulations of circuits we need are the trivial constructions listed in the following lemma.

Lemma 3.8 $PV_1(\alpha)$ *proves:*

(*i*) *If* $C\colon \varrho \twoheadrightarrow P$, $\pi \geq \varrho$, *and* $0 < \Pi \leq P$, *there exists a* D *such that* $D\colon \pi \twoheadrightarrow \Pi$.

(*ii*) *If* $C_1\colon \varrho_1 \twoheadrightarrow P_1$ *and* $C_2\colon \varrho_2 \twoheadrightarrow P_2$, *there exist* D *and* E *such that* $D\colon \varrho_1 + \varrho_2 \twoheadrightarrow P_1 + P_2$ *and* $E\colon \varrho_1\varrho_2 \twoheadrightarrow P_1P_2$.

(*iii*) *If* $C_1\colon \varrho_1 \twoheadrightarrow \varrho_2$ *and* $C_2\colon \varrho_2 \twoheadrightarrow \varrho_3$, *there exists a* D *such that* $D\colon \varrho_1 \twoheadrightarrow \varrho_3$.

(*iv*) *If* $C\colon \varrho + |a| \twoheadrightarrow P + |a|$ *and* $P > 0$, *there exists a* D *such that* $D\colon \varrho \twoheadrightarrow P$.

Proof:
(i): put $D(u) = \min(C(\min(u, \varrho - 1)), \Pi - 1)$.
(ii): define $E(u) = P_1 C_2(\lfloor \frac{u}{\varrho_1} \rfloor) + C_1(u \bmod \varrho_1)$ and

$$D(u) = \begin{cases} C_1(u), & u < \varrho_1, \\ C_2(u - \varrho_1) + P_1, & u \geq \varrho_1. \end{cases}$$

(iii): take $D(u) = C_2(C_1(u))$.
(iv): $\Sigma_0^b(PV(\alpha))$-defined subsets of $k := [0, k)$, where $k = |a|$ for some a, can be encoded by numbers. If $X \subseteq k$ is encoded by a number x, we can define $\mathrm{card}(X)$ (i.e., the number of 1s in the binary expansion of x) by a PV-function, and we can prove basic properties of card, such as

$$\begin{gathered} \mathrm{card}(k) = k, \\ X \cap Y = \emptyset \Rightarrow \mathrm{card}(X \cup Y) = \mathrm{card}(X) + \mathrm{card}(Y), \\ \mathrm{card}(\mathrm{rng}(w)) \leq \mathrm{card}(\mathrm{dom}(w)), \end{gathered}$$

(where w is a function encoded by a number) by a straightforward application of $\Sigma_0^b(PV)$-*LIND*.

The sets

$$\begin{aligned} Y &= \{i < |a|;\, C(\varrho + i) \geq P\}, \\ X &= \{j < |a|;\, \exists i < |a|\, C(\varrho + i) = P + j\} \end{aligned}$$

are encoded by numbers, as they are $\Sigma_0^b(PV(\alpha))$-definable. As $i \mapsto C(\varrho + i) - P$ is a surjection of Y to X, we have $\mathrm{card}(X) \leq \mathrm{card}(Y)$, thus $\mathrm{card}(|a| \smallsetminus X) \geq \mathrm{card}(|a| \smallsetminus Y)$. For any encoded set $Z \subseteq k$, the mapping $c_Z(i) := \mathrm{card}(Z \cap i)$ provides a bijection of Z onto $|Z|$, which is definable by a PV-function, hence encoded by a number. Thus

$$w(j) := c^{-1}_{|a| \smallsetminus Y}\left(\min\left\{\mathrm{card}(|a| \smallsetminus Y) - 1, c_{|a| \smallsetminus X}(j)\right\}\right)$$

is an encoded surjection of $|a| \smallsetminus X$ onto $|a| \smallsetminus Y$. For any $u < \varrho$, we define

$$D(u) = \begin{cases} C(u), & C(u) < P, \\ C(\varrho + w(C(u) - P)), & C(u) \geq P, C(u) - P \notin X, \\ 0, & \text{otherwise.} \end{cases}$$

It is easy to see that $D\colon \varrho \twoheadrightarrow P$. $\square$

Theorem 3.9 *Let π, Π, ϱ, P be PV-functions, and $c > 0$ a constant such that PV_1 proves: $\Pi(b) > \pi(b)$, $P(a) > \varrho(a)$, and for every $a \geq c$ such that $\varrho(a) \geq |a|^c$, there exists b such that*

$$\pi(b) \geq \varrho(a) - |a|^c \wedge \left(\frac{P(a)}{\varrho(a)}\right)^c \geq \frac{\Pi(b)}{\pi(b)}.$$

Then $PV_1(\alpha) + \forall b\, sPHP^{\pi(b)}_{\Pi(b)}(PV(\alpha))$ proves $\forall a\, sPHP^{\varrho(a)}_{P(a)}(PV(\alpha))$.

Proof: Work in $PV_1(\alpha)$, and fix a. If $a < c$ or $\varrho(a) \leq |a|^c$, there exists x such that $\varrho(a) = |x|$, thus $sPHP^{\varrho(a)}_{P(a)}(PV(\alpha))$ holds by Lemma 3.8 (iv). Otherwise fix b as in the statement, put $\pi = \pi(b)$, $\Pi = \Pi(b)$, $\varrho = \varrho(a)$, $P = P(a)$, and assume that $C\colon \varrho \twoheadrightarrow P$, we will construct $D\colon \pi \twoheadrightarrow \Pi$. By Lemma 3.8 (iv) we may assume $\pi \geq \varrho$. We distinguish two cases.

Case 1: $P \geq 2\varrho$. We have

$$\frac{\pi}{\varrho} \leq 2\left\lfloor\frac{\pi}{\varrho}\right\rfloor, \qquad \left(\frac{P}{\varrho}\right)^c \leq 2^c\left\lfloor\frac{P}{\varrho}\right\rfloor^c,$$

thus

$$\Pi \leq 2^{c+1}\varrho\left\lfloor\frac{\pi}{\varrho}\right\rfloor\left\lfloor\frac{P}{\varrho}\right\rfloor^c \leq \varrho\left\lfloor\frac{\pi}{\varrho}\right\rfloor\left\lfloor\frac{P}{\varrho}\right\rfloor^{2c+1}.$$

By Lemma 3.8 (ii) and (i), we may construct surjections

$$\varrho\left\lfloor\frac{P}{\varrho}\right\rfloor^i \twoheadrightarrow P\left\lfloor\frac{P}{\varrho}\right\rfloor^i \geq \varrho\left\lfloor\frac{P}{\varrho}\right\rfloor^{i+1}$$

for every $i \leq 2c$, thus by Lemma 3.8 (iii), there exists a surjection

$$\varrho \twoheadrightarrow \varrho\left\lfloor\frac{P}{\varrho}\right\rfloor^{2c+1}.$$

By Lemma 3.8 (ii) and (i), there exists a surjection

$$\pi \geq \varrho\left\lfloor\frac{\pi}{\varrho}\right\rfloor \twoheadrightarrow \varrho\left\lfloor\frac{\pi}{\varrho}\right\rfloor\left\lfloor\frac{P}{\varrho}\right\rfloor^{2c+1} \geq \Pi.$$

Case 2: $P \leq 2\varrho$. We have

$$\left(\frac{P}{\varrho}\right)^c \leq 1 + 2^c\left(\frac{P}{\varrho} - 1\right),$$

thus

$$\Pi \leq \pi + 2^c\pi\left(\frac{P}{\varrho} - 1\right) \leq \pi + 2^{c+1}\left\lfloor\frac{\pi}{\varrho}\right\rfloor(P - \varrho).$$

By Lemma 3.8 (ii) there exists a surjection

$$\varrho\left\lfloor\frac{\pi}{\varrho}\right\rfloor \twoheadrightarrow P\left\lfloor\frac{\pi}{\varrho}\right\rfloor,$$

thus by Lemma 3.8 (ii) there exist surjections

$$\begin{aligned}\pi + i\left\lfloor\frac{\pi}{\varrho}\right\rfloor(P-\varrho) = \left(i\left\lfloor\frac{\pi}{\varrho}\right\rfloor(P-\varrho) + (\pi \bmod \varrho)\right) + \varrho\left\lfloor\frac{\pi}{\varrho}\right\rfloor \twoheadrightarrow\\ \twoheadrightarrow \left(i\left\lfloor\frac{\pi}{\varrho}\right\rfloor(P-\varrho) + (\pi \bmod \varrho)\right) + P\left\lfloor\frac{\pi}{\varrho}\right\rfloor = \pi + (i+1)\left\lfloor\frac{\pi}{\varrho}\right\rfloor(P-\varrho)\end{aligned}$$

for every $i < 2^{c+1}$, thus by Lemma 3.8 (iii) and (i), there exists a surjection

$$\pi \twoheadrightarrow \pi + 2^{c+1}\left\lfloor\frac{\pi}{\varrho}\right\rfloor(P-\varrho) \geq \Pi. \qquad \square$$

The main discrepancy between the bounds provided by Theorems 3.4 and 3.9 is in the second disjunct of $(*)$. It has the following explanation. In the proof of Theorem 3.9, we only use the assumption $sPHP^{\pi(b)}_{\Pi(b)}$ once: given a surjection $C\colon \varrho(a) \twoheadrightarrow P(a)$, we construct a single circuit which is a surjection $D\colon \pi(b) \twoheadrightarrow P(b)$. In terms of the proof of Theorem 3.4, we have a student-teacher protocol in which S is allowed only one query of type (2). We do not know which of the two bounds is closer to truth. On the one hand, it is hard to imagine why the restriction to one query of type (2) should be justified in general; on the other hand, we have no idea how to fruitfully use more queries.

Question 3.10 $PV_1(\alpha) + \forall a > 0\, sPHP^{a||a||}_{a(||a||+1)}(PV(\alpha)) \overset{?}{\vdash} \forall a > 0\, sPHP^{a||a||^2}_{a(||a||^2+1)}(PV(\alpha))$

4 Surjective vs. injective *WPHP*

We now turn to the question whether $S^1_2 + sWPHP(PV)$ proves $iWPHP(PV)$. The relativized case was solved by Thapen [17], who proved that $S^1_2(\alpha) + sWPHP(PV(\alpha)) \nvdash iWPHP(PV(\alpha))$. In the unrelativized case, similar results were obtained under cryptographic assumptions. Krajíček and Pudlák [10] have shown that $S^1_2 \nvdash iWPHP(PV)$, if the RSA cryptosystem is secure against polynomial-time attacks. Thapen [16] subsequently extended the result to $S^1_2 + sWPHP(PV) \nvdash iWPHP(PV)$, assuming security of RSA against randomized polynomial-time attacks. Cook and Thapen [6] derive other independences involving PV_1 on this basis.

It is well-known that RSA can be cracked if we can factorize integers efficiently. We will show that the weaker assumption of hardness of integer factoring for polynomial-time probabilistic algorithms is sufficient to obtain $S^1_2 + sWPHP(PV) \nvdash iWPHP(PV)$.

Let p be an odd prime, and a an integer. The *Legendre symbol* is defined by

$$\left(\frac{a}{p}\right) = \begin{cases} 0, & p \mid a, \\ 1, & a \text{ is a quadratic residue mod } p, \\ -1, & a \text{ is not a quadratic residue mod } p. \end{cases}$$

The following classical theorem was formalized in $I\Delta_0 + iWPHP(\Delta_0)$ by Berarducci and Intrigila; we observe that the proof goes through in $PV_1 + iWPHP(PV)$.

Theorem 4.1 ([1]) *$PV_1 + iWPHP(PV)$ proves multiplicativity of the Legendre symbol:*

$$p \text{ odd prime } \to \left(\frac{a}{p}\right)\left(\frac{b}{p}\right) = \left(\frac{ab}{p}\right).$$

Proof: Put $a_0 = a$, $a_1 = b$, $a_2 = ab$. It is easy to see that PV_1 proves the identity whenever $\left(\frac{a_i}{p}\right) = 0, 1$ for some i, thus assume that all a_i are quadratic non-residues. Then

$$f(i, u) := a_i u^2 \bmod p$$

defines an injection

$$f\colon 3 \times [1, (p-1)/2] \hookrightarrow [1, p-1],$$

contradicting $iPHP^{3n}_{2n}(PV)$. □

Theorem 4.2 *If $S^1_2 + sWPHP(PV)$ proves multiplicativity of the Legendre symbol, there exists a probabilistic polynomial-time algorithm for factoring of integers.*

Proof: The nontrivial part of the assumption can be written as a Σ^b_1-formula

$$p > 0 \to \exists u, v < p\, uv = p \vee \exists u < p\, (u^2 \equiv a \vee u^2 \equiv b \vee u^2 \equiv ab \pmod p).$$

By Wilkie's witnessing theorem [9], there exists a probabilistic poly-time algorithm $A(p, a, b)$ which computes either a nontrivial factor of p, or a square root of a, b, or ab modulo p with probability $1 - \varepsilon$.

Let n be an odd composite number which is not a prime power. Let G be the group of units in $\mathbb{Z}_n$, and H its subgroup of squares. As n has at least two distinct odd prime divisors, $[G : H] \geq 4$ by the Chinese remainder theorem. If we choose uniformly random $a, b \in G$, then ab is also a uniformly distributed element of G, thus a, b, and ab are quadratic non-residues modulo n with probability at least $1 - 3/4 = 1/4$.

It follows that the following algorithm finds a factor of n with probability at least $1/4 - \varepsilon$: if n is even, or a perfect power, construct a factor in an obvious way. Otherwise choose random $a, b \in [1, n - 1]$, compute $u := A(n, a, b)$, and check whether $\gcd(n, a)$, $\gcd(n, b)$, or u is a nontrivial divisor of n. □

Remark 4.3 With a bit of care, Theorem 4.2 can be formalized in $PV_1 + sWPHP(PV)$. That is, if $S^1_2 + sWPHP(PV)$ proves multiplicativity of the Legendre symbol, then $PV_1 + sWPHP(PV)$ proves that integers can be factored by a probabilistic poly-time algorithm.

Corollary 4.4 *Assume that integer factoring is impossible in probabilistic polynomial time.*

(*i*) $S^1_2 + sWPHP(PV) \nvdash iWPHP(PV)$,

(*ii*) *PV_1 does not prove the unique choice schema*

$$\forall i < |a|\, \exists! x < b\, \varphi(i, x) \to \exists w\, \forall i < |a|\, \varphi(i, (w)_i)$$

for a Σ^b_0-formula φ,

(*iii*) *PV_1 does not prove the Δ^b_1-comprehension schema*

$$\forall x\, (\varphi(x) \equiv \neg\psi(x)) \to \exists w\, \forall i < |a|\, (\varphi(i) \equiv (w)_i = 1),$$

where $\varphi, \psi \in \Sigma^b_1$.

Proof: (*i*) follows from Theorems 4.1 and 4.2, and (*i*) implies (*ii*) and (*iii*) by Cook and Thapen [6]. □

For the sake of completeness, we also mention the converse problem: whether the injective variant of the weak pigeonhole principle implies the surjective variant. First notice that $S^1_2(\alpha) + iWPHP(PV(\alpha)) \vdash sWPHP(PV(\alpha))$ is "almost true", due to Theorem 2.4.

Corollary 4.5
$S^1_2(\alpha) + sWPHP(PV(\alpha))$ is $\Sigma^b_1(\alpha)$-conservative over $S^1_2(\alpha) + iWPHP(PV(\alpha))$.

On the other hand, $iWPHP(PV(\alpha))$ is a $\Sigma^b_1(\alpha)$-statement, so we would not expect it to prove a genuinely $\Sigma^b_2(\alpha)$-principle such as $sWPHP(PV(\alpha))$. This intuition indeed proves correct, essentially due to Krajíček [8].

Definition 4.6 Let Γ be a set of $L(\alpha)$-formulas. $\mathrm{Th}_\Gamma(\mathbb{N})$ is the set of all $\varphi(\alpha) \in \Gamma$ such that the second-order formula $\forall\alpha\,\varphi(\alpha)$ holds in the standard model of arithmetic.

Theorem 4.7 $S^2_2(\alpha) + \mathrm{Th}_{\Pi^b_2(\alpha)}(\mathbb{N}) \nvdash sWPHP(PV(\alpha))$.

Proof: Krajíček [8] (cf. [9, Thm. 11.2.5]) has shown that $sWPHP(\alpha)$ is unprovable in $S^2_2(\alpha)$. The only property of $S^2_2(\alpha)$ used in his argument is Buss' witnessing theorem, and by a well-known observation, addition of true $\Pi^b_2(\alpha)$ axioms does not change witnessing of $\Sigma^b_2(\alpha)$-formulas. □

Corollary 4.8 $S^2_2(\alpha) + iWPHP(PV(\alpha)) \nvdash sWPHP(PV(\alpha))$.

5 Acknowledgement

I would like to thank Steve Cook, and the anonymous referees, for useful suggestions.

References

[1] Alessandro Berarducci and Benedetto Intrigila, *Combinatorial principles in elementary number theory*, Annals of Pure and Applied Logic 55 (1991), no. 1, pp. 35–50.

[2] Samuel R. Buss, *Bounded arithmetic*, Bibliopolis, Naples, 1986.

[3] ———, *First-order proof theory of arithmetic*, in: Handbook of Proof Theory (S. R. Buss, ed.), Studies in Logic and the Foundations of Mathematics vol. 137, Elsevier, Amsterdam, 1998, pp. 79–147.

[4] Alan Cobham, *The intrinsic computational difficulty of functions*, in: Proceedings of the 2nd International Congress of Logic, Methodology and Philosophy of Science (Y. Bar-Hillel, ed.), North–Holland, 1965, pp. 24–30.

[5] Stephen A. Cook, *Feasibly constructive proofs and the propositional calculus*, in: Proceedings of the 7th Annual ACM Symposium on Theory of Computing, 1975, pp. 83–97.

[6] Stephen A. Cook and Neil Thapen, *The strength of replacement in weak arithmetic*, ACM Transactions on Computational Logic 7 (2006), no. 4, pp. 749–764.

[7] Emil Jeřábek, *Dual weak pigeonhole principle, Boolean complexity, and derandomization*, Annals of Pure and Applied Logic 129 (2004), pp. 1–37.

[8] Jan Krajíček, *No counter-example interpretation and interactive computation*, in: Logic From Computer Science, Proceedings of a Workshop held November 13–17, 1989 in Berkeley (Y. N. Moschovakis, ed.), Mathematical Sciences Research Institute Publications vol. 21, Springer, 1992, pp. 287–293.

[9] ———, *Bounded arithmetic, propositional logic, and complexity theory*, Encyclopedia of Mathematics and Its Applications vol. 60, Cambridge University Press, 1995.

[10] Jan Krajíček and Pavel Pudlák, *Some consequences of cryptographical conjectures for S_2^1 and EF*, Information and Computation 140 (1998), no. 1, pp. 82–94.

[11] Jan Krajíček, Pavel Pudlák, and Gaisi Takeuti, *Bounded arithmetic and the polynomial hierarchy*, Annals of Pure and Applied Logic 52 (1991), no. 1–2, pp. 143–153.

[12] Alexis Maciel, Toniann Pitassi, and Alan R. Woods, *A new proof of the weak pigeonhole principle*, Journal of Computer and System Sciences 64 (2002), no. 4, pp. 843–872.

[13] Rohit Parikh, *Existence and feasibility in arithmetic*, Journal of Symbolic Logic 36 (1971), no. 3, pp. 494–508.

[14] Jeff B. Paris, Alex J. Wilkie, and Alan R. Woods, *Provability of the pigeonhole principle and the existence of infinitely many primes*, Journal of Symbolic Logic 53 (1988), no. 4, pp. 1235–1244.

[15] Søren M. Riis, *Making infinite structures finite in models of second order bounded arithmetic*, in: Arithmetic, Proof Theory, and Computational Complexity (P. Clote and J. Krajíček, eds.), Oxford Logic Guides vol. 23, Oxford University Press, 1993, pp. 289–319.

[16] Neil Thapen, *A model-theoretic characterization of the weak pigeonhole principle*, Annals of Pure and Applied Logic 118 (2002), no. 1–2, pp. 175–195.

[17] ———, *Structures interpretable in models of bounded arithmetic*, Annals of Pure and Applied Logic 136 (2005), no. 3, pp. 247–266.