# Resource Bounded Unprovability of Computational Lower Bounds (Part 1)

Tatsuaki Okamoto* Ryo Kashima**

* NTT Laboratories, Nippon Telegraph and Telephone Corporation
1-1 Hikarino-oka, Yokosuka-shi, Kanagawa, 239-0847 Japan
** Dept. of Mathematical and Computing Sciences, Tokyo Institute of Technology
1-12-1 O-okayama Meguro-ku, Tokyo, 152-8552 Japan

January 7, 2005

**Abstract.** This paper introduces new notions of asymptotic proofs, PT(polynomial-time)-extensions, PTM(polynomial-time Turing machine)-$\omega$-consistency, etc. on formal theories of arithmetic including PA (Peano Arithmetic). An asymptotic proof is a set of infinitely many formal proofs, which is introduced to define and characterize a property, PTM-$\omega$-consistency, of a formal theory. Informally speaking, PTM-$\omega$-consistency is a *polynomial-time bounded* version (in asymptotic proofs) of $\omega$-consistency, and characterized in two manners: (1) (in the light of the *extension of PTM to TM*) the resource *unbounded* version of PTM-$\omega$-consistency is equivalent to $\omega$-consistency, and (2) (in the light of *asymptotic proofs by PTM*) a PTM-$\omega$-*inconsistent* theory includes an axiom that only a super-polynomial-time Turing machine can prove asymptotically over PA, under some assumptions. This paper shows that *P≠NP (more generally, any super-polynomial-time lower bound in PSPACE) is unprovable in a PTM-$\omega$-consistent theory $T$*, where $T$ is a consistent PT-extension of PA (although this paper does not show that P≠NP is unprovable in PA, since PA has not been proven to be PTM-$\omega$-consistent). This result implies that to prove P≠NP by any technique requires a PTM-$\omega$-*inconsistent* theory, which should include an axiom that only a super-polynomial-time machine can prove asymptotically over PA (or implies a super-polynomial-time computational upper bound) under some assumptions. This result is a kind of generalization of the result of "Natural Proofs" by Razborov and Rudich [21], who showed that to prove "P≠NP" by a class of techniques called "Natural Proofs" implies a super-polynomial-time (e.g., sub-exponential-time) algorithm that can break a typical cryptographic primitive, a pseudo-random generator. Our result also implies that any relativizable proof of P≠NP requires the *resource unbounded version* of PTM-$\omega$-*inconsistent* theory, $\omega$-*inconsistent* theory, which suggests another negative result by Baker, Gill and Solovay [1] that no relativizable proof can prove "P≠NP" in PA, which is a $\omega$-consistent theory. Therefore, our result gives a unified view to the existing two major negative results on proving P≠NP, Natural Proofs and relativizable proofs, through the two manners of characterization of PTM-$\omega$-consistency. We also show that the PTM-$\omega$-consistency of $T$ cannot be proven in any PTM-$\omega$-consistent theory $S$, where $S$ is a consistent PT-extension of $T$. That is, to prove the independence of P vs NP from $T$ by proving the PTM-$\omega$-consistency of $T$ requires a PTM-$\omega$-*inconsistent* theory, or implies a super-polynomial-time computational upper bound under some assumptions. This seems to be related to the results of Ben-David and Halevi [4] and Kurz, O'Donnell and Royer [17], who showed that to prove the independence of P vs NP from PA using any currently known mathematical paradigm implies an extremely-close-to-polynomial-time (but still super-polynomial-time) algorithm that can solve NP-complete problems. Based on this result, we show that *the security of any computational cryptographic scheme is unprovable* in the setting where adversaries and provers are modeled as polynomial-time Turing machines and only a PTM-$\omega$-consistent theory is allowed to prove the security.

**Key Words:** computational complexity, computational lower bound, P vs NP, natural proofs, relativizable proofs, cryptography, unprovability, undecidability, mathematical logic, proof theory, incompleteness theorem

# Table of Contents

# 1 Introduction

## 1.1 Background

It looks very mysterious that proving computational lower bounds is extremely difficult, although many people believe that there exist various natural intractable problems that have no efficient algorithms that can solve them. A classical technique, diagonalization, can separate some computational classes like P $\neq$ EXP, but it fails to separate computational classes between P and PSPACE, which covers almost all practically interesting computational problems. Actually, we have very few results on the lower bounds of computational natural problems between P and PSPACE. The best known result of computational lower bounds (in standard computation models such as Turing machines and Boolean circuits) of a computational natural problem is about $5n$ in circuit complexity [16], where $n$ is problem size. Therefore, surprisingly, it is still very hard for us to prove even the $6n$ lower bound of TQBF, a PSPACE complete problem, which is considered to be much more intractable than NP complete problems.

Considering this situation, it seems natural to think that there is some substantial reason why proving computational lower bounds is so difficult. The ultimate answer to this question would be to show that such computational lower bounds are impossible to prove, e.g., showing its independence from a formal proof system like Peano Arithmetic (a formal system for number theory) and ZFC (a formal system for set theory).

This paper gives a new type of impossibility result, *resource bounded* impossibility, in the proof of computational lower bounds.

## 1.2 Our Results

Let theory $T$, on which we are assumed to try to prove P$\neq$NP, be a consistent PT-extension of PA, throughout this paper (and hereafter in this section), where theory $T$ is called PT-extension if there exists a polynomial-time algorithm that, given $n \in \mathbb{N}$, decides whether $n$ is the Gödel number of an axiom of $T$ (Section 2.3).

This paper shows the following results.

**New Notions** We introduce notions of asymptotic proofs, polynomial-time proofs, polynomial-time decisions, PT-extensions, PTM-$\omega$-consistency etc. on formal theories of arithmetic including PA (Peano Arithmetic).

– Asymptotic Proofs (Section 2.8): $Q_1\mathbf{x}_1 \cdots Q_k\ \mathbf{x}_k\ \varphi(\mathbf{x}_1,\ldots,\mathbf{x}_k)$ has an asymptotic proof over $T$ if

$$Q_1 x_1 \in \mathbb{N}\ \cdots\ Q_k x_k \in \mathbb{N} \quad T \vdash \varphi(\mathbf{x}_1,\ldots,\mathbf{x}_k),$$

where a boldfaced symbol (e.g, $\mathbf{x}$) denotes a variable in theory $T$ or numeral (e.g., $\mathbf{x}$ is the numeral of $x \in \mathbb{N}$), and $Q_i$ ($i \in \{1,\ldots,k\}$) denotes an unbounded quantifier.

– Polynomial-time proofs (Section 2.7):

$$\mathrm{PTM}_e(x) \vdash_T \varphi(\mathbf{x})$$

denotes that a PTM (polynomial-time Turing machine) coded by $e \in \mathbb{N}$, given $x \in \mathbb{N}$ and the Gödel number of the expression of $\{\varphi(\mathbf{a}) \mid a \in \mathbb{N}\}$ (constant in $|x|$), produces a proof (tree) of formula $\varphi(\mathbf{x})$ in theory $T$.

– Let theory $S$ be a PT-extension of theory $T$. PTM-$\omega$-consistency (Definition 62): Theory $S$ is PTM-$\omega$-consistent for $\Delta_1$-formula $\varphi(\mathbf{e}^*, \mathbf{x})$ over theory $T$, if the following condition holds.

$$\begin{aligned} & \forall e \in \mathbb{N} \ \ \exists e^* \in \mathbb{N} \ \ \exists \ell \in \mathbb{N} \ \ \forall n \geq \ell \ \ \forall c \in \mathbb{N} \quad \mathrm{PTM}_e(n) \ \not\vdash_T \ \exists \mathbf{x} \ (\mathbf{n} \leq \mathbf{x} < \mathbf{n} + |\mathbf{n}|^{\mathbf{c}}) \ \ \varphi(\mathbf{e}^*, \mathbf{x}) \\ \Rightarrow \quad & \forall e \in \mathbb{N} \ \ \exists e^* \in \mathbb{N} \ \ \exists \ell \in \mathbb{N} \ \ \forall n \geq \ell \quad \mathrm{PTM}_e(n) \ \not\vdash_S \ \exists \mathbf{x} \geq \mathbf{n} \ \ \varphi(\mathbf{e}^*, \mathbf{x}), \end{aligned}$$

where $|\mathbf{n}|$ denotes the numeral of $|n|$ (see Section 2.1).
Theory $T$ is PTM-$\omega$-consistent for $\varphi(\mathbf{e}^*, \mathbf{x})$, if $T$ is PTM-$\omega$-consistent for $\varphi(\mathbf{e}^*, \mathbf{x})$ over $T$.

**Formalization of P≠NP** We formalize P≠NP as follows (Definition 53):

$$\overline{\mathrm{P}{\neq}\mathrm{NP}} \ \ \equiv \ \ \forall \mathbf{e} \ \ \forall \mathbf{n} \ \ \exists \mathbf{x} \geq \mathbf{n} \quad \neg \mathrm{DecSAT}(\mathbf{e}, \mathbf{x}), \tag{1}$$

where $\mathrm{DecSAT}(\mathbf{e}, \mathbf{x})$ is a formula in PA which informally means that a PTM coded by $e$ correctly decides the satisfiability or unsatisfiability of a 3CNF coded by $x$.

**Unprovability of $\overline{\mathbf{P{\neq}NP}}$ in a PTM-$\boldsymbol{\omega}$-consistent theory** $\overline{\mathrm{P}{\neq}\mathrm{NP}}$ cannot be proven in $T$ that is PTM-$\omega$-consistent for any $\Delta_2^P$ formula (Theorem 67):

$$T \not\vdash \ \overline{\mathrm{P} \neq \mathrm{NP}}.$$

**Unprovability of PTM-$\boldsymbol{\omega}$-consistency in a PTM-$\boldsymbol{\omega}$-consistent theory** Let theory $S$ be a consistent PT-extension of theory $T$, and $S$ be PTM-$\omega$-consistent for any $\Delta_2^P$-formula. Then, PTM-$\omega$-consistency of $T$ for a $\Delta_2^P$-formula cannot be proven in $S$. (Theorem 73)

Thus, the independence of P vs NP from $T$ by proving PTM-$\omega$-consistency of $T$ for a $\Delta_2^P$-formula (i.e., through Theorem 67) cannot be proven in $S$.

In fact, the existence of PTM-$\omega$-consistent theory $T$ for a $\Delta_2^P$-formula has not been proven, and the independence of P vs NP from PA has not been proven.

**Unprovability of the Security of Computational Cryptography** The one-wayness of any function family is *unprovable* in the setting where an adversary and a prover are modeled to be polynomial-time Turing machines, and the security proof should be made in a PTM-$\omega$-consistent theory $T$ for $\Delta_2^P$ (Theorem 81). In other words, the security of any computational cryptographic scheme is unprovable under this setting.

### 1.3 An Implication of Our Results

To interpret our results, let assume the following hypotheses:

– (Hypothesis 1) $\mathfrak{N} \models \overline{\mathrm{P}{\neq}\mathrm{NP}}$, where $\mathfrak{N}$ is the standard model of natural numbers (i.e., P≠NP is true.)

– (Hypothesis 2) PA is PTM-$\omega$-consistent for $\Delta_2^P$.

We then have the following consequence from our results.

– *P vs NP is independent from PA.*
  This is because $\overline{\mathrm{P}{\neq}\mathrm{NP}}$ is consistent with PA from Hypothesis 1, and $\neg\overline{\mathrm{P}{\neq}\mathrm{NP}}$ is consistent with PA, since PA $\nvdash \overline{\mathrm{P}{\neq}\mathrm{NP}}$ (from Theorem 67 and Hypothesis 2).
– *Hypothesis 2 cannot be proven in a PTM-$\omega$-consistent theory* $T$, where $T$ is a consistent PT-extension of PA.
  That is, even if P vs NP is independent from PA, the independence (by proving Hypothesis 2) cannot be proven in a PTM-$\omega$-consistent theory $T$.

### 1.4 Characterization of PTM-$\omega$-consistency

Informally speaking, PTM-$\omega$-consistency is a *polynomial-time bounded* version (in asymptotic proofs) of $\omega$-consistency, and characterized in two manners:

1. (Characterization in the light of the *extension of PTM to TM*) The resource *unbounded* version of PTM-$\omega$-consistency is equivalent to $\omega$-consistency.
2. (Characterization in the light of *asymptotic proofs by PTM*) A PTM-$\omega$-*inconsistent* theory includes an axiom that only a super-polynomial-time Turing machine can prove asymptotically over PA, under some assumptions.

First, PTM-$\omega$-consistency can be extended to a resource *unbounded* TM (Turing machine) version of PTM-$\omega$-consistency, TM-$\omega$-consistency, which is equivalent to $\omega$-consistency (Remark 3 of Definition 62 in Section 7.1).

Second, PTM-$\omega$-consistent theory $T$ is a formal theory, but is characterized by asymptotic proofs of PTM provers over $T$. A proof in a formal theory itself is a finite length proof and has no asymptotic property as well as no implication of prover's computational capability. However, PTM-$\omega$-consistency is defined through asymptotic proofs of PTM provers, and an axiom in a PTM-$\omega$-consistent theory may be characterized by asymptotic proofs of a PTM prover.

For example, a PTM-$\omega$-*inconsistent* theory $T$, which is a consistent PT-extension of PA, should include an axiom outside PA that only a super-polynomial-time Turing machine can prove asymptotically over PA, assuming that PA is PTM-$\omega$-consistent and deduction in $T$ can be made asymptotically by PTM (Remark 5 of Definition 62 in Section 7.1). Let $T$ be a theory in which an axiom, $X$, outside PA is added to PA. Although $X$ cannot be proven in PA, it can be *asymptotically* proven over PA if it is true, since any true $\Delta_1$-sentence can be proven in PA. Therefore, a resource *unbounded* Turing machine can always produce an asymptotic proof of $X$ over PA, but a resource *bounded* (e.g., polynomial-time) Turing machine may produce no asymptotic proof of $X$ over PA. Hence, axiom $X$ (and theory $T$) can be characterized by the computational complexity of a prover for producing an asymptotic proof of $X$. If $T$ is PTM-$\omega$-*inconsistent*, the computational complexity of a prover for producing an asymptotic proof of $X$ should be super-polynomial-time, under the above-mentioned assumption (Remark 5 of Definition 62). Thus, PTM-$\omega$-consistency bridges a formal proof and prover's (asymptotic) computational capability through asymptotic proofs.

In accordance with the two manners of characterization of PTM-$\omega$-consistency, our main result that $\overline{\mathrm{P}{\neq}\mathrm{NP}}$ cannot be proven in a PTM-$\omega$-consistent theory (Theorem 67) suggests two avenues towards negative results:

- To prove $\overline{\mathrm{P}\neq\mathrm{NP}}$ requires a PTM-$\omega$-*inconsistent* theory, which should include an axiom that only a super-polynomial-time machine can prove asymptotically over PA (or implies a super-polynomial-time computational upper bound), under the assumption. This is a kind of generalization of the result of "Natural Proofs" by Razborov and Rudich [21]. See Section 1.5.
- To prove $\overline{\mathrm{P}\neq\mathrm{NP}}$ by a relativizable proof, i.e., to prove $\overline{\mathrm{P}^A \neq\mathrm{NP}^A}$ with oracle $A$ requires a PTM$^A$-$\omega$-*inconsistent* theory (Proposition 68). Therefore, if there exists a relativizable proof of $\overline{\mathrm{P}\neq\mathrm{NP}}$, which implies a proof of $\overline{\mathrm{P}^A \neq\mathrm{NP}^A}$ for any oracle $A$, it will require an $\omega$-*inconsistent* theory, since a PTM$^A$-$\omega$-*inconsistent* theory with any oracle $A$ is equivalent to a $\omega$-*inconsistent* theory. This suggests the result that no relativizable proof can prove "P$\neq$NP" in PA (or any $\omega$-consistent theory), which was shown by Baker, Gill and Solovay [1]. See the remark of Theorem 67.

Therefore, our result, Theorem 67 (and its generalization, Proposition 68), gives a unified view to the existing two major negative results on proving P$\neq$NP, Natural Proofs and relativizable proofs, through the above-mentioned two manners of characterization of PTM-$\omega$-consistency.

PTM-$\omega$-consistency has also the following properties:

- PTM-$\omega$-consistency and $\omega$-consistency do not imply each other. (Remark 2 of Definition 62)
- Although the PTM-$\omega$-consistency of PA seems to be as natural as the $\omega$-consistency of PA, no PTM-$\omega$-consistent theory $T$, which is a consistent PT-extension of PA, can prove the PTM-$\omega$-consistency of PA. (Theorem 73 and Remark 4 of Definition 62)

### 1.5 Related Works

**Self-defeating results** Our result is considered to be a kind of generalization of or a close relation to the previously known self-defeating results as follows:

- Our result that PTM-$\omega$-consistent theory cannot prove $\overline{\mathrm{P}\neq\mathrm{NP}}$ (Theorem 67) implies a *self-defeating* property such that to prove a super-polynomial-time lower bound like P$\neq$NP requires a PTM-$\omega$-*inconsistent* theory, which should include an axiom that only a super-polynomial-time machine can prove asymptotically over PA (or implies a super-polynomial-time computational upper bound) under the assumption described in the previous section.
  "Natural Proofs" by Razborov and Rudich [21] showed that to prove a computational lower bound (e.g., a super-polynomial-time lower bound like P$\neq$NP) by a class of techniques called "Natural Proofs" implies a comparable level of computational upper bound (e.g., a super-polynomial-time algorithm to break a typical cryptographic primitive, a pseudo-random generator). In other words, to prove P$\neq$NP by a "Natural Proof" requires an additional axiom $X$ that implies a super-polynomial-time (e.g., sub-exponential-time) algorithm to break a pseudo-random generator and that can be proven asymptotically only by a super-polynomial-time machine, since no polynomial-time machine is considered to be able to *asymptotically* prove an upper bound property of a super-polynomial-time machine. Therefore, to prove P$\neq$NP by a specific type of proof called "Natural Proof" requires a specific type of PTM-$\omega$-*inconsistent* theory, which is PA + $X$. That is, the negative result regarding "Natural Proofs" is considered to be a special case of our result, Theorem 67.
- Our results imply another *self-defeating* property such that PTM-$\omega$-consistent theory $S$ over $T$ cannot prove the independence of P vs NP from $T$ by proving PTM-$\omega$-consistency of $T$ for a $\Delta_2^P$-formula (Theorem 73). In other words, to prove the independence of P vs NP from $T$ through Theorem 67 (i.e., to prove $T \nvdash \overline{\mathrm{P}\neq\mathrm{NP}}$ by proving PTM-$\omega$-consistency of $T$ and to

prove $T \nvdash \neg\overline{\mathrm{P}{\neq}\mathrm{NP}}$ by some way) requires PTM-$\omega$-*inconsistent* theory over $T$, or implies a super-polynomial-time upper bound under the above-mentioned assumption.

Ben-David and Halevi [4] and Kurz, O'Donnell and Royer [17] showed that to prove the independence of P vs NP from PA using any currently known mathematical paradigm implies a comparable level of computational upper bound, an extremely-close-to-polynomial time algorithm to solve NP-complete problems. In other words, to prove the independence of P vs NP from PA using any currently known mathematical paradigm requires an additional axiom $Y$ that implies an extremely-close-to-polynomial time (but still super-polynomial-time) algorithm to solve NP-complete problems and that can be proven asymptotically only by a super-polynomial-time machine. Therefore, to prove the independence of P vs NP from PA by a specific type of proof using currently known mathematical paradigms requires a specific type of PTM-$\omega$-*inconsistent* theory, which is PA + $Y$. That is, the negative result by Ben-David et.al. is considered to be a special case of our result, Theorem 73, provided that Hypothesis 1 in Section 1.3 is true and PA $\nvdash \overline{\mathrm{P}{\neq}\mathrm{NP}}$ implies Hypothesis 2.

**Relativizable proofs** Our result that PTM-$\omega$-consistent theory cannot prove $\overline{\mathrm{P}{\neq}\mathrm{NP}}$ (Theorem 67) suggests the result by Baker, Gill and Solovay [1], who showed that there is no relativizable proof of "P≠NP", and the result by Hartmanis and Hopcroft [13, 14], who showed that for any reasonable theory $T$ we can effectively construct a TM $M$ such that relative to oracle $L(M)$, "P≠NP" cannot be proven in $T$. (See the remark of Theorem 67.)

Our result might be related to the result by da Costa and Doria [6], but the relationship between their result and ours is unclear for us.

**Mathematical logic approaches** The results of this paper are constructed on the theory and techniques of mathematical logic, especially proof theory. Several mathematical logic approaches to solve the P vs NP problem have been investigated such as bounded arithmetic [5, 18], propositional proof length [3, 18, 20] and descriptive complexity [8].

Bounded arithmetic characterizes an analogous notion of PH (polynomial hierarchy of computational complexity), which is a hierarchy of weak arithmetic theories, so-called bounded arithmetic classes, wherein only bounded quantifiers are allowed. The target of the bounded arithmetic approach is to separate one class from another in bounded arithmetic, which may imply a separation of one class from another in PH (i.e., typically P≠NP).

The proof length of propositional logic can characterize the NP vs co-NP problem, since TAUT, the set of propositional tautologies, is co-NP complete. Therefore, the main target of this approach is to prove NP≠co-NP by showing a super-polynomial length lower bound of a formal propositional proof of TAUT. In this approach, the lower bounds of the proof lengths and limitation of provability of some specific propositional proof systems (e.g., resolution, Frege system and extended Frege system) have been investigated.

The descriptive complexity characterizes NP by a class of problems definable by existential second order formulas and P by a class of problems definable in first order logic with an operator. The target of this approach is to separate P and NP using these logical characterizations.

This paper characterizes the concepts of P and P≠NP etc., by formulas in Peano Arithmetic (PA). A novel viewpoint of our approach is to introduce the concept of an asymptotic proof produced by a polynomial-time Turing machine as a prover, to characterize a property of a formal theory, PTM-$\omega$-consistency, by using this concept, and to show that no PTM-$\omega$-consistent theory can prove a super-polynomial-time computational lower bound such as P≠NP.

To the best of our knowledge, no existing approach has studied computational lower bounds from such a viewpoint. [1]

**Proof systems** In order to define the PTM-$\omega$-consistency, this paper introduces a new concept of proof systems, *asymptotic proofs* and *polynomial-time proofs* where the computational complexity of (prover's) proving a set of statements asymptotically is bounded by polynomial-time. In the conventional proof theory, the properties and capability of a proof system (e.g., consistency, completeness, incompleteness etc.) are of prime interest, but the required properties and capability of the prover are not considered (i.e., no explicit restriction nor condition is placed on the prover).

Note that the bounded arithmetic approach seems to follow this conventional paradigm and bounds the capability of the proof system (axioms and rule of inferences) to meet the capability of resource bounded computational classes. That is, the prover is still thought to exceed the scope of the approach.

In this paper, the computational complexity of a prover is investigated through the concept of an asymptotic proof system. An asymptotic proof is a set of an infinite number of formal proofs, and a resource bounded (e.g., polynomial-time bounded or exponential-time bounded etc.) prover asymptotically produces an asymptotic proof of a set of infinitely many formal statements.

This paper then introduces a new concept, PTM-$\omega$-consistency, which is a property of a conventional proof system, but is defined and characterized by the concept of asymptotic proofs with a polynomial-time bounded prover. PTM-$\omega$-consistency plays a key role in our results (for example, see Section 1.4).

**Undecidability** Although the computational complexity theory is a resource bounded version of the recursion theory, to the best of our knowledge, little research has been made on resource bounded undecidability of formal statements.

This paper introduces a resource bounded (asymptotic) decision system, which corresponds to a resource bounded (asymptotic) proof system, and presents the incompleteness theorems (Sections 4 and 5). Using the incompleteness theorem of resource bounded (asymptotic) decision systems yields the resource bounded unprovability of $\overline{\mathrm{P}\neq\mathrm{NP}}$ (Section 7).

### 1.6 Key Ideas of Our Results

In order to obtain our main result (Theorem 67: $\overline{\mathrm{P}\neq\mathrm{NP}}$ cannot be proven in a PTM-$\omega$-consistent theory), this paper introduces the concept of *polynomial-time decision systems* (Section 4). In a proof system, we usually consider only one side, a proof of a true statement. In a decision system, however, we have to consider two sides, CA (correctly accept: accept of a true statement) and CR (correctly reject: reject of a false statement). CD (correctly decide) means CA or CR.

The key idea to prove Theorem 67 is a *polynomial-time decision version of incompleteness theorems*. Informally speaking, we introduce a special sentence, $\rho_e^A(\mathbf{x})$, (an analogue of the so-called Gödel sentence) like "this statement, $\rho_e^A(\mathbf{x})$, cannot be correctly accepted by a polynomial-time Turing machine (PTM) encoded by $e$." (Hereafter, "a PTM encoded by $e$" is called "PTM $e$") If $\rho_e^A(\mathbf{x})$ can be correctly accepted by PTM $e$, it contradicts the definition of $\rho_e^A(\mathbf{x})$. It follows

[1] A prover is modeled as a Turing machine in the interactive proof system theory, and the computational complexity of a prover has been investigated [12, 11]. However, no proof system with a polynomial-time Turing machine prover that produces an asymptotic proof of a computational lower bound has been studied.

that $\rho_e^A(\mathbf{x})$ cannot be correctly accepted by PTM $e$. We also define another sentence, $\rho_e^R(\mathbf{x})$, which cannot be correctly rejected by PTM $e$. (First incompleteness theorems of polynomial-time decisions: Theorems 39 and 40). Based on these theorems, we show that, for any formula set $\{\psi(\mathbf{x}) \mid x \in \mathbb{N}\}$ (e.g., formula set on the satisfiability of 3CNF), for any PTM $e$, there exists another PTM $e^*$ such that PTM $e$, on input $x \in \mathbb{N}$, cannot asymptotically prove that PTM $e^*$ cannot correctly decide $\psi(\mathbf{x})$ (Second incompleteness theorem of polynomial-time decisions: Theorem 45). By using Theorem 45, we show that no PTM can prove $\overline{\mathrm{P}\neq\mathrm{NP}}$ asymptotically (Lemma 64).

This paper then introduces the PTM-*$\omega$-consistency* of $T$, which is a PTM version of $\omega$-consistency and plays a key role in our result (for its semantics and rationale, see Section 1.4 and the remarks of Definition 62). Combining Lemma 65 and PTM-$\omega$-consistency of $T$, we can show that $\overline{\mathrm{P}\neq\mathrm{NP}}$ cannot be proven in PTM-$\omega$-consistent theory $T$ (Theorem 67).

This paper also introduces the notion of *polynomial-time proof systems*, and obtains a *polynomial-time proof version of incompleteness theorems* (Sections 2 and 3). Informally speaking, we introduce a special sentence, $\rho_{e,T}$, like "this statement, $\rho_{e,T}$, cannot be proven by a polynomial-time Turing machine (PTM) $e$ in theory $T$." If $\rho_{e,T}$ can be proven by PTM $e$ in $T$, it contradicts the definition of $\rho_{e,T}$, assuming that $T$ is consistent. It follows that $\rho_{e,T}$ cannot be proven by PTM $e$ in $T$, although another PTM can prove it (First incompleteness theorem of polynomial-time proofs: Theorem 20). Based on this theorem, we show that, for any formula set $\{\psi(\mathbf{x}) \mid x \in \mathbb{N}\}$ for any PTM $e$, there exists another PTM $e^*$ such that PTM $e$, on input $x \in \mathbb{N}$, cannot asymptotically prove that PTM $e^*$ cannot prove $\psi(\mathbf{x})$ (Second incompleteness theorem of polynomial-time proofs: Theorem 21).

By using Theorem 21 and PTM-$\omega$-consistency, we show that the PTM-$\omega$-consistency of $T$ cannot be proven in a PTM-$\omega$-consistent theory $S$, where $S$ is a consistent PT-extension of $T$ (Theorem 73). (In fact, we have not shown the existence of a consistent and PTM-$\omega$-consistent PT-extension of PA; therefore, we have not shown the unprovability of $\overline{\mathrm{P}\neq\mathrm{NP}}$ in PA.)

Finally, based on Theorem 67, the unprovability of the security of the computational cryptography is obtained (Theorems 81 and 82) in a setting that provers as well as adversaries are modeled as PTMs and only PTM-$\omega$-consistent theory is allowed to prove the security.

## 2 Polynomial-Time Proofs

This paper follows the standard notions and definitions of computational complexity theory (e.g., definitions of P and NP) and mathematical logic (e.g., definition of a formal proof in Peano Arithmetic). See [23] for such standard notions and definitions of computational complexity theory and see [2, 7, 22] for the standard notions and definitions of mathematical logic.

The central interest of this paper is the difficulty of proving the lower bound of computational problems by resource bounded Turing machines. For this purpose, first, we need to formalize the notion of a formal proof produced by a resource bounded Turing machine. This section introduces our formalization of a proof produced by a polynomial-time Turing machine (polynomial-time proof: PTP) in a theory that is an extension of Peano Arithmetic (hereafter Peano Arithmetic is abbreviated to PA).

**Remark:**
This paper is based on the standard notion of formal proofs in first order logic [2, 7, 22]. There are, however, many possible ways of formalizing such formal proofs, especially with regard to the style of formalizing the deduction system; alternatives to the selection of logical axioms and rules of inference. There are two typical styles: one is the Hilbert-style, which has several logical axioms and a few rules of inference, and the other is the Gentzen-style, which has just one logical axiom

and several rules of inference. However, the results in this paper are not affected by the way of formalizing the deduction system, and almost all descriptions in this paper are independent of the style of formal deduction system adopted. When we need to make an explicit description on a specific deduction system, this paper adopts the Hilbert style, which has two rules of inference; Modus Ponens and Generalization rules.

### 2.1 Notations

Let $\mathbb{N}$ be the set of natural numbers including 0.

When $w$ is a bit string, $|w|$ denotes the bit length of $w$.

When $w \in \mathbb{N}$, $[w]$ denotes the binary representation of $w$, i.e., bit string $w_{k-1}w_{k-2}\cdots w_0$ with $w = w_{k-1}2^{k-1}+w_{k-2}2^{k-2}+\cdots w_0$, $k = \lfloor \log_2 w \rfloor + 1$ $(w > 0)$, and $w_i \in \{0,1\}$ for $i = 0,1,2,\ldots,k-1$. When $w = 0$, $[w], i.e., [0]$, denotes the binary representation, 0. When $w \in \mathbb{N}$, $|w|$ denotes the bit length of $[w]$.

PA has a constant symbol, $\mathbf{0}$, intended to denote the number 0, and has three function symbols, $\mathbf{S}, +, \cdot$, where $\mathbf{S}$ is a one-place function symbol intended to denote the successor function $S : \mathbb{N} \to \mathbb{N}$, i.e., the function for which $S(n) = n+1$, and symbols $+$ and $\cdot$ are two-place function symbols of addition and multiplication, respectively. PA also has symbols of predicate logic such as logical symbols ($\neg$, $\wedge$, $\vee$, $\to$, $\forall$, $\exists$, etc.), relation symbols ($=$, $<$, etc.), and variable symbols ($\mathbf{x}, \mathbf{y}, \mathbf{z}$, etc.).

The numerals of PA are denoted by boldfaced number symbols such as $\mathbf{1}$, $\mathbf{2}$, $\mathbf{3}$, ..., for $\mathbf{S0}$, $\mathbf{SS0}$, $\mathbf{SSS0}$, ... . Boldfaced alphabet symbols such as $\mathbf{x}$, $\mathbf{y}$, $\mathbf{x}_1$, $\mathbf{x}_i$, etc., are also used for variables in theory $T$.

Throughout this paper, we assume that the numeral, $\overbrace{\mathbf{SS}\cdots\mathbf{S}}^{n \text{ times}}\mathbf{0}$, of natural number $n$ is expressed by the following binary form in a theory including PA:

$$\mathbf{n}_0 + \mathbf{n}_1 \cdot \mathbf{SS0} + \cdots + \mathbf{n}_{k-1} \cdot \overbrace{\mathbf{SS0} \cdot \mathbf{SS0} \cdots \mathbf{SS0}}^{k-1 \text{ times}},$$

where $n = n_0 + n_1 \cdot 2 + \cdots n_{k-1} \cdot 2^{k-1}$, $n_i \in \{0,1\}$, $\mathbf{n}_i = \mathbf{0}$ if $n_i = 0$ and $\mathbf{n}_i = \mathbf{1} (= \mathbf{S0})$ if $n_i = 1$ $(i = 0,1,\ldots,k-1)$. Here we denote this expression of the numeral of natural number $n$ by $\mathbf{S}^n\mathbf{0}$ or just $\mathbf{n}$. Similarly, if alphabet $a$ denotes a natural number, $\mathbf{a}$ denotes $\mathbf{S}^a\mathbf{0}$.

We will now introduce two additional function symbols in PA. (A function symbol, $f$, of a primitive recursive function is considered to be implicitly included in PA, i.e., $f$ can be identified with a formula, $\rho_f$, in PA, since $f$ is representable by a $\Delta_1$formula, $\rho_f$, in PA and PA $\vdash$ $\forall \mathbf{x}_1 \cdots \forall \mathbf{x}_k \exists! \mathbf{y}\ \rho_f(\mathbf{x}_1, \ldots, \mathbf{x}_k, \mathbf{y})$. See Subsection 2.4.)

Here it is worth noting that these function symbols, which correspond to primitive recursive functions, are introduced for improving the readability of formulas, not for increasing proving ability. Therefore, in this paper we assume that no Gödel number for a function symbol of a primitive recursive function (except $\mathbf{S}$, $+$ and $\cdot$ ) is provided (see the next section for Gödel numbers). The Gödel number of a formula including such a function symbol is calculated on the formula without using the function symbol, i.e., the formula in which only function symbols in PA are employed. This assumption is applied for any theory $T$ which is a PT-extension of PA throughout this paper. Hence the Gödel number of a formula in $T$ is uniquely defined even if some function symbols of primitive recursive functions are employed in the formula.

If $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ are numerals in PA, $\mathbf{x} \dot{-} \mathbf{y}$ denotes a two-place function: $(\mathbf{x}, \mathbf{y}) \mapsto \mathbf{z}$, such that $\mathbf{z} = \mathbf{S}^{\max\{x-y,0\}}\mathbf{0}$, $\mathbf{x} = \mathbf{S}^x\mathbf{0}$, $\mathbf{y} = \mathbf{S}^y\mathbf{0}$, $x \in \mathbb{N}$ and $y \in \mathbb{N}$.

If $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ are numerals in PA, $\mathbf{x}^{\mathbf{y}}$ denotes a two-place function: $(\mathbf{x}, \mathbf{y}) \mapsto \mathbf{z}$, such that $\mathbf{z} = \mathbf{S}^{x^y}\mathbf{0}$, $\mathbf{x} = \mathbf{S}^{x}\mathbf{0}$, $\mathbf{y} = \mathbf{S}^{y}\mathbf{0}$, $x \in \mathbb{N}$ and $y \in \mathbb{N}$.

By using these function symbols (notations), the notation of a numeral, $\mathbf{n}$, is defined by

$$\mathbf{n}_0 + \mathbf{n}_1 \cdot \mathbf{2}^{\mathbf{1}} + \cdots + \mathbf{n}_{k-1} \cdot \mathbf{2}^{\mathbf{k} \dot{-} \mathbf{1}}.$$

When $\mathbf{n}$ is a numeral, $|\mathbf{n}|$ denotes the numeral of $|n|$. The function symbol, $|\cdot|$, is justified by the first claim in the proof of Theorem 11.

Some other notations are:

- $\psi \leftrightarrow \varphi$ denotes
$$(\psi \rightarrow \varphi) \wedge (\psi \leftarrow \varphi),$$
- $\exists ! \mathbf{y}\ \varphi(\mathbf{y})$ denotes
$$\exists \mathbf{y}\ \varphi(\mathbf{y})\ \wedge\ \forall \mathbf{y}_1 \forall \mathbf{y}_2 (\varphi(\mathbf{y}_1) \wedge \varphi(\mathbf{y}_2) \rightarrow \mathbf{y}_1 = \mathbf{y}_2),$$
which means $\mathbf{y}$ *uniquely* exists to satisfy $\varphi(\mathbf{y})$.
- $\forall \mathbf{x} \geq \mathbf{n}\ \varphi(\mathbf{x})$ denotes
$$\forall \mathbf{x}\ (\mathbf{x} \geq \mathbf{n}\ \rightarrow\ \varphi(\mathbf{x})).$$
- $\exists \mathbf{x} \geq \mathbf{n}\ \varphi(\mathbf{x})$ denotes
$$\exists \mathbf{x}\ (\mathbf{x} \geq \mathbf{n}\ \wedge\ \varphi(\mathbf{x})).$$
- Some basic notations in proof theory [2]:
$$T \vdash \varphi,$$
which informally denotes "the truth of formula $\varphi$ is provable in theory $T$".
$$\mathrm{Pr}_T(\ulcorner \varphi \urcorner),$$
which denotes a formula in $T$, which informally means "there exists a proof for the truth of formula $\varphi$ in theory $T$". Here $\ulcorner \varphi \urcorner$ denotes $\mathbf{S}^{\#\varphi}\mathbf{0}$.
- $T$ is *inconsistent* if there exists a formula $\varphi$ in $T$ such that $T \vdash \varphi$ and $T \vdash \neg\varphi$, which is also denoted by $T \vdash \perp$. $T$ is *consistent* if there exists no such formula $\varphi$ in $T$. Here, $\perp\ \equiv\ \neg \forall \mathbf{x}(\mathbf{x} = \mathbf{x})$.
- $T$ is $\omega$*-inconsistent* if there exists a formula $\varphi(\mathbf{x})$ in $T$ such that
$$T \vdash \exists \mathbf{x}\ \varphi(\mathbf{x}), \quad \text{and}$$
$$\forall a \in \mathbb{N}\ \ T \vdash \neg\varphi(\mathbf{a}).$$
$T$ is $\omega$*-consistent* if there exists no such formula $\varphi(\mathbf{x})$ in $T$. (If $T$ is $\omega$-consistent, $T$ is also consistent. The reverse is not always true.)
- $\mathfrak{N}$ is the standard model of natural numbers. When $\varphi$ is a formula in PA,
$$\mathfrak{N} \models \varphi$$
denotes that $\varphi$ is true in $\mathfrak{N}$.

### 2.2 Gödel numbers

There are many ways of defining the Gödel numbers, and the way introduced in this section differs from those described in Gödel's original paper and textbooks (e.g., [7]), since in this paper we require a polynomial time algorithm to make unique encoding and decoding. We basically follow the approach introduced by [5]. (We can also adopt a coding method employed in actual current computer systems.)

Let $\#\varphi$ be a Gödel number of $\varphi$. First, we define Gödel numbers of basic symbols in $L$ as follows: (for example) $\#\forall$ is 0, $\#($ is 1, $\#\mathbf{0}$ is 2, $\#)$ is 3, $\#\mathbf{S}$ is 4, $\#\neg$ is 5, $\# <$ is 6, $\# \rightarrow$ is 7, $\#+$ is 8, $\# =$ is 9, $\#\cdot$ is 10, $\#,$ is 11, $\#\mathbf{a}_1$ is 20, $\#\mathbf{x}_1$ is 22, $\#\mathbf{a}_2$ is 24, $\#\mathbf{x}_2$ is 26, etc.

We then use the following method to obtain the Gödel number of a sequence of natural numbers, $a_1, a_2, \ldots, a_k$ [5]:

1. Represent $a_i$ by the binary representation with the least significant bit on the right, as is traditional. Then, $a_1, a_2, \ldots, a_k$ can be represented by the sequence of three symbols '0', '1' and ','.
2. Reverse the order of the sequence of '0', '1' and ',', and replace '0' by '10', '1' by '11' and ',' by '01'. We then obtain a sequence of '0' and '1'.
3. The natural number whose binary representation is this sequence is the Gödel number of the number sequence, $a_1, a_2, \ldots, a_k$. It is denoted by $\langle a_1, a_2, \ldots, a_k \rangle$.

For example, $\langle 3, 4, 5 \rangle$ is a natural number, whose binary representation is 11101101101011011111, because $3, 4, 5$ is binary-represented along with commas by $11, 100, 101$ and is encoded to a binary sequence, 11101101101011011111.

When $\varphi$ is an expression in language $L$, it is a sequence of symbols, $s_0 s_1 \cdots s_k$, of $L$. We then define the Gödel number, $\#\varphi$, of $\varphi$ as follows:

$$\#\varphi \equiv \langle \#s_0, \#s_1, \ldots, \#s_k \rangle.$$

For example, when $\varphi$ is $\neg(\forall \mathbf{x}_1(\mathbf{x}_1 < \mathbf{S0}))$,

$$\begin{aligned}\#\varphi &\equiv \langle \#\neg, \#(, \#\forall, \#\mathbf{x}_1, \#(, \#\mathbf{x}_1, \# <, \#\mathbf{S}, \#\mathbf{0}, \#), \#) \rangle \\ &\equiv \langle 5, 1, 0, 13, 1, 13, 6, 4, 2, 3, 3 \rangle.\end{aligned}$$

Remember here that numeral $\mathbf{n}$ $(\equiv \mathbf{S}^n\mathbf{0})$ denotes the binary form, i.e.,

$$\mathbf{n}_0 + \mathbf{n}_1 \cdot \mathbf{SS0} + \cdots + \mathbf{n}_{k-1} \cdot \overbrace{\mathbf{SS0} \cdot \mathbf{SS0} \cdots \mathbf{SS0}}^{k-1 \text{ times}}.$$

Hence, $|\#\mathbf{n}|$ (i.e., $|\#\mathbf{S}^n\mathbf{0}|$) is of the order of $\log_2^2 n$.

Here also remember that we provide no Gödel numbers of additionally introduced function symbols of primitive recursive functions such as $\mathbf{2^n}$. That is, the Gödel number of a formula including such a function symbol is calculated on the formula with only function symbols in PA. For example,

$$\#\mathbf{2^n} \equiv \# \overbrace{\mathbf{SS0} \cdot \mathbf{SS0} \cdots \mathbf{SS0}}^{n \text{ times}} \equiv \langle \#\mathbf{S}, \#\mathbf{S}, \#\mathbf{0}, \#\cdot, \ldots, \#\mathbf{0} \rangle.$$

Therefore, $|\#\mathbf{2^n}| = O(n)$.[2]

[2] If we have the Gödel number of the function symbol EXP such as $\mathrm{EXP}(\mathbf{x}, \mathbf{y}) = \mathbf{x^y}$, then $|\#\mathbf{2^n}| = |\#\mathrm{EXP}(\mathbf{2}, \mathbf{n})| = |\#\mathrm{EXP}(\mathbf{SS0}, \mathbf{n}_0 + \mathbf{n}_1 \cdot \mathbf{SS0} + \cdots + \mathbf{n}_{k-1} \cdot \mathbf{SS0} \cdots \mathbf{SS0})| = |\langle \#\mathrm{EXP}, \#(, \#\mathbf{S}, \ldots, \#\mathbf{0} \rangle| = O(\log_2 n)$, since $k = O(\log_2 n)$.

We then introduce a concatenation operation $||$ of two Gödel numbers, $\#\varphi$ and $\#\psi$, where $\#\varphi \equiv \langle \#s_0, \#s_1, \ldots, \#s_k \rangle$ and $\#\psi \equiv \langle \#t_0, \#t_1, \ldots, \#t_l \rangle$. $\#\varphi || \#\psi$ is defined by

$$\langle \#s_0, \#s_1, \ldots, \#s_k, \#t_0, \#t_1, \ldots, \#t_l \rangle.$$

### 2.3 Polynomial-Time Extension of PA

Let formula $\mathbf{Axiom}_T(\mathbf{n})$ be true if and only if $n$ is the Gödel number of an axiom of $T$. If the truth of $\mathbf{Axiom}_T(\mathbf{n})$ can be correctly decided by a polynomial-time algorithm in $|n|$, on input $n$, we say that $T$ is polynomial-time axiomizable. If $T$ is an extension of $T_0$ and polynomial-time axiomizable, then we say that $T$ is a polynomial-time (PT) extension of $T_0$.

Using the notations introduced in Section 4.1, a polynomial-time axiomizable theory, $T$, is defined as follows: Let $\mathcal{AX} \equiv \{\mathbf{Axiom}_T(\mathbf{n}) \mid n \in \mathbb{N}\}$ and $\mathrm{Size}_{\mathcal{AX}}(n) = |n|$. There exists $e \in \mathbb{N}$ such that for all $n \in \mathbb{N}$

$$\mathrm{PTM}_e^{\mathcal{AX}}(n) \triangleright\ \mathbf{Axiom}_T(\mathbf{n}) \quad \vee \quad \mathrm{PTM}_e^{\mathcal{AX}}(n) \triangleright\ \neg\mathbf{Axiom}_T(\mathbf{n}).$$

### 2.4 Representability Theorem in Mathematical Logic

This section introduces the representability theorem in the conventional mathematical logic [2, 7, 22]. This theorem plays an important role in many situations as well as in constructing the polynomial-time version of the representability theorem (Theorem 11), which is essential to formalize the execution of PTM in PA.

In this paper, we use the standard notions and notations of mathematical logic, such as $T \vdash \varphi$ (informally, a sentence $\varphi$ is provable in theory $T$), with no introduction (see [2, 7, 22]).

**Definition 1.** *1. Let $R$ be a $k$-ary relation on $\mathbb{N}$; i.e., $R \subseteq \mathbb{N}^k$. A formula $\rho_R(x_1, \ldots, x_k)$ (in which only $x_1, \ldots, x_k$ occur free) will be said to represent a relation $R$ in theory $T$ if and only if for every $a_1, \ldots, a_k$ in $\mathbb{N}^k$*

$$\begin{aligned} (a_1, \ldots, a_k) \in R &\quad\Rightarrow\quad T \vdash \rho_R(\mathbf{a}_1, \ldots, \mathbf{a}_k), \\ (a_1, \ldots, a_k) \notin R &\quad\Rightarrow\quad T \vdash \neg\rho_R(\mathbf{a}_1, \ldots, \mathbf{a}_k). \end{aligned}$$

*A relation $R$ is said to be representable in $T$ if and only if there exists some formula $\rho_R$ that represents $R$ in $T$.*

*2. Let $f$ be a $k$-place function on the natural numbers. A formula $\rho_f(x_1, \ldots, x_k, y)$ (in which only $x_1, \ldots, x_k, y$ occur free) will be said to functionally represent $f$ in theory $T$ if and only if for every $a_1, \ldots, a_k$ in $\mathbb{N}^k$*

$$T \vdash \forall y(\rho_f(\mathbf{a}_1, \ldots, \mathbf{a}_k, \mathbf{y}) \leftrightarrow \mathbf{y} = \mathbf{S}^{f(a_1, \ldots, a_k)}\mathbf{0}).$$

*A function $f$ is said to be functionally representable in $T$ if and only if there exists some formula $\rho$ that functionally represents $f$ in $T$.*

**Proposition 2.** *(Representability Theorem) For any primitive recursive relation on $\mathbb{N}^k$, $R$, and any primitive recursive function on $\mathbb{N}^k$, $f$, there exist formulas, $\rho_R(\mathbf{x}_1, \ldots, \mathbf{x}_k)$ and $\rho_f(\mathbf{x}_1, \ldots, \mathbf{x}_k, \mathbf{y})$, such that:*

- *$\rho_R(\mathbf{x}_1, \ldots, \mathbf{x}_k)$ represents $R$, and $\rho_f(\mathbf{x}_1, \ldots, \mathbf{x}_k, \mathbf{y})$ functionally represents $f$ in PA.*

– $\rho_R(\mathbf{x}_1, \ldots, \mathbf{x}_k)$ *and* $\rho_f(\mathbf{x}_1, \ldots, \mathbf{x}_k, \mathbf{y})$ *are* $\Delta_1$ *in PA.*
–
$$\mathrm{PA} \vdash \forall \mathbf{x}_1 \cdots \forall \mathbf{x}_k \exists ! \mathbf{y}\ \rho_f(\mathbf{x}_1, \ldots, \mathbf{x}_k, \mathbf{y}).$$

**Proposition 3.** *Let only* $\mathbf{x}_1, \ldots, \mathbf{x}_k, \mathbf{y}$ *occur free in formula* $\varphi(\mathbf{x}_1, \ldots, \mathbf{x}_k, \mathbf{y})$*. If* $T \vdash \forall \mathbf{x}_1 \cdots \forall \mathbf{x}_k \exists ! \mathbf{y}\ \varphi(\mathbf{x}_1, \ldots, \mathbf{x}_k, \mathbf{y})$*, then theory* $T'$

$$\equiv\ T \cup \{\forall \mathbf{x}_1 \cdots \forall \mathbf{x}_k \forall \mathbf{y}\ (\varphi(\mathbf{x}_1, \ldots, \mathbf{x}_k, \mathbf{y}) \leftrightarrow f(\mathbf{x}_1 \cdots \mathbf{x}_k) = \mathbf{y})\}$$

*is a conservative extension of* $T$*.*

From Proposition 3, we can identify theory $T'$, which has function symbol $f$, with theory $T$, in the light of provability and representability. In other words, we can consider that function symbol $f$ ( and the corresponding axiom, $\forall \mathbf{x}_1 \cdots \forall \mathbf{x}_k \forall \mathbf{y}$ ( $\varphi(\mathbf{x}_1, \ldots, \mathbf{x}_k, \mathbf{y}) \leftrightarrow f(\mathbf{x}_1 \cdots \mathbf{x}_k) = \mathbf{y}$ ) ) is implicitly included in theory $T$. Therefore, from Propositions 2 and 3, we can consider that a function symbol of any primitive recursive function is implicitly included in PA. Later in this paper, we will introduce several primitive recursive function symbols in theory $T$ which is a PT-extension of PA.

**Proposition 4.** *Let* $g$ *be an* $n$*-place function, let* $h_1, \ldots, h_n$ *be* $m$*-place functions, and let* $f$ *be defined by*

$$v = f(x_1, \ldots, x_m) \equiv g(h_1(x_1, \ldots, x_m), \ldots, (x_1, \ldots, x_m)).$$

*Let formulas,* $\psi$ *and* $\theta_1, \ldots, \theta_n$*, functionally represent* $g$ *and* $h_1, \ldots, h_n$*, and formula* $\rho_f$ *be defined as follows:*

$$\begin{aligned}&\rho_f((\mathbf{x}_1, \ldots, \mathbf{x}_m, \mathbf{v}) \equiv \\ &\qquad \exists \mathbf{y}_1 \ldots \exists \mathbf{y}_k\ \ (\ \theta_1(\mathbf{x}_1, \ldots, \mathbf{x}_m, \mathbf{y}_1) \wedge \ldots \wedge \theta_k(\mathbf{x}_1, \ldots, \mathbf{x}_m, \mathbf{y}_k) \wedge \psi(\mathbf{y}_1, \ldots, \mathbf{y}_k, \mathbf{v})\ ).\end{aligned}$$

*Then,* $\rho_f$ *functionally represents* $f$*.*

### 2.5 Turing Machines

A Turing machine (TM) is represented by $(Q, \Sigma, \Gamma, \delta, q_0, q_{accept}, q_{reject})$, where $Q$ is a set of states, $\Sigma = \{0, 1\}$ is the input alphabet, $\Gamma$ is the tape alphabet with blank symbol $\sqcup$ and $\{0, 1\}$, $\delta : Q \times \Gamma \rightarrow Q \times \Gamma \times \{\mathrm{L}, \mathrm{R}\}$ is transition function, $q_0$ is the start state, $q_{accept}$ is the accept state, and $q_{reject}$ is the reject state [23].

The computation process of a Turing machine can be represented by the sequence of *configurations*, $C_0, C_1, \ldots, C_k$. Each configuration $C_i$ consists of three items, the current state, $q_i \in Q$, the current tape contents, and the current tape head location. It is convenient to represent a configuration by triple $(u, q, v)$, where the current state is $q$, the current tape contents is $uv$ and the current head location is the leftmost bit of $v$, where $uv$ denotes the concatenation of bit strings $u$ and $v$. When a configuration $C_i$ is $(ua, q, bv)$ $(a, b \in \{0, 1\})$, transition function $\delta$ yields configuration $C_{i+1}$ such that

$$\begin{aligned} C_{i+1} &= (u, q', acv) \ \text{ if } \delta(q, b) = (q', c, \mathrm{L}), \\ C_{i+1} &= (uac, q', v) \ \text{ if } \delta(q, b) = (q', c, \mathrm{R}). \end{aligned}$$

We can also define a Turing machine whose output is not just accept/reject, but a finite sequence of $\Sigma$. Here, $q_{halt}$ is used in place of $q_{accept}$ and $q_{reject}$. The output value is the tape contents in state $q_{halt}$.

Let $e_M$ be a natural number whose binary representation, $[e_M]$, is a part of the input to a universal Turing machine U, and denotes the description of Turing machine $M$. In other words, U can simulate $M$ by reading $[e_M]$. Let $w$ be a natural number and $[w]$ be the input to $M$. We then use $\mathrm{U}(e_M, w)$ $(= M(w))$ to denote a natural number whose binary representation, $[\mathrm{U}(e_M, w)]$, is the output of U with input $[e_M]$ and $[w]$. So, we abuse notation U for a function over natural numbers, which is defined by universal Turing machine U.

### 2.6 Polynomial-Time Turing Machines

Let $M$ be a *polynomial-time Turing machine* (PTM). W.l.o.g., we assume $[e_M]$ consists of a pair of bit strings, $(t, [c])$: $t$ is a description of a Turing machine that allows a universal Turing machine to simulate $M$, and $c$ is a constant natural number such that $M$'s running time is bounded by $\mathrm{Size}(w)^c$. Here $w \in \mathbb{N}$, $[w] \in W$ ($W$: set of input strings to $M$) is an input to $M$.

Size$(\cdot)$ is a function,

$$\mathrm{Size} : \mathbb{N} \longrightarrow \mathbb{N}, \qquad \mathrm{Size} : w \mapsto \mathrm{Size}(w),$$

which determines the size (bit length) of input $[w] \in W$ such that, for positive constants $c_1$ and $c_2$, for all $[w] \in W$, $|[w]|^{c_1} \leq \mathrm{Size}(w)| \leq |[w]|^{c_2}$, and Size$(\cdot)$ is a polynomial-time (in $|w|$) computable function. The size function, Size$(\cdot)$, is uniquely determined by each class of problems such as 3SAT and Hamiltonean circuit. If Size$(\cdot)$ is not explicitly defined, $\mathrm{Size}(a) \equiv |a|$. For example, if the underlying class of problems is 3SAT, $[w]$ is a binary-code description of a 3CNF formula, and Size$(w)$ is the number of variables of the 3CNF formula. Then, we may use $\mathrm{Size}_{3\mathrm{SAT}}(w)$ to explicitly represent this function for a specific problem, 3SAT. If $[w]$ is not a (syntactically) valid value that describes a 3CNF formula, the function value of $\mathrm{Size}_{3\mathrm{SAT}}(w)$ is defined to be $|w|$, and a PTM specific to 3SAT, which reads such an invalid input value, immediately moves to the reject state (or outputs "invalid input" etc). (In Section 6, function $\mathrm{Size}_{\mathcal{SAT}}$ is more simply defined by $\mathrm{Size}_{\mathcal{SAT}}(w) \equiv |w|$ for all $w \in \mathbb{N}$.)

It is easy to convert any Turing machine described by $t$ into a Turing machine described by $(t, [c])$, by just adding a running step number counter (with a specific tape for counting). Note that the counter does not count the running steps for counting. $M$ accepts $[w]$ if it accepts $[w]$ within $\mathrm{Size}_W(w)^c$ steps, and it rejects $[w]$ otherwise. Given $(t, [c])$, universal Turing machine U simulates PTM $M$ by description $t$ and counts the running step number of the simulated machine up to $\mathrm{Size}_W(w)^c$ and halts the machine when the number exceeds $\mathrm{Size}_W(w)^c$. Such a special universal Turing machine for PTMs, which only accepts the form of $(t, [c])$ as input $[e_M]$, is denoted by $\mathrm{U}_{\mathrm{PTM}}$ in this paper. We assume a single (fixed) $\mathrm{U}_{\mathrm{PTM}}$ for PTMs. Then, natural number $e_M$ implies a unique PTM $M$. It is clear that any PTM $M$ can be simulated by $\mathrm{U}_{\mathrm{PTM}}$ with input $[e_M]$ with form of $(t, [c])$. That is, $M(w)$ is exactly simulated by $\mathrm{U}_{\mathrm{PTM}}(e_M, w)$.

Here, w.l.o.g., we assume $\mathrm{U}_{\mathrm{PTM}}$ can syntactically check whether bit string $t$ is a syntactically correct description of a Turing machine for $\mathrm{U}_{\mathrm{PTM}}$. Such syntactic rules of describing a Turing machine for $\mathrm{U}_{\mathrm{PTM}}$ can be clearly specified. $\mathrm{U}_{\mathrm{PTM}}$ can effectively check whether $t$ is a syntactically correct description or not, in a manner similar to that used by computer language compilers. $\mathrm{U}_{\mathrm{PTM}}$ can also effectively check whether the format of $[e_M] = (t, [c])$ is syntactically valid or not. If $\mathrm{U}_{\mathrm{PTM}}$ recognizes $[e_M]$ to be syntactically incorrect (e.g., the part of $[c]$ is not syntactically recognized), $\mathrm{U}_{\mathrm{PTM}}$ outputs a special string denoting "syntactically invalid code". Here it is essential that $\mathrm{U}_{\mathrm{PTM}}$ be able to correctly simulate a PTM if $(t, [c])$ is valid, and such a valid string $[e_M]$, which is syntactically recognized valid by $\mathrm{U}_{\mathrm{PTM}}$, always exists for any PTM $M$. Note: it is not essential how well $\mathrm{U}_{\mathrm{PTM}}$ can find an invalid string. If $\mathrm{U}_{\mathrm{PTM}}$ incorrectly recognizes an invalid string as a valid one, and executes the input, then $\mathrm{U}_{\mathrm{PTM}}$ may run abnormally (e.g., runs in an infinite loop or immediately halts). If it immediately halts (i.e., in a halting state), it is the output of the execution. If it runs in an infinite loop, the step counter of $\mathrm{U}_{\mathrm{PTM}}$ executes independently and halts when the number of steps exceeds $\mathrm{Size}_W(w)^c$.

We then use $\mathrm{U}_{\mathrm{PTM}}(e_M, w)$ to denote a natural number whose binary representation, $[\mathrm{U}_{\mathrm{PTM}}(e_M, w)]$, is the output of $\mathrm{U}_{\mathrm{PTM}}$ with input $[e_M]$ and $[w]$. Therefore, similarly to U, we also abuse the notation of $\mathrm{U}_{\mathrm{PTM}}$ for a function over natural numbers: $(e_M, w) \mapsto \mathrm{U}_{\mathrm{PTM}}(e_M, w)$. Clearly, it is a totally recursive (for input $(e_M, w)$) and polynomial-time (in $\mathrm{Size}_W(w)$) function.

When the input to PTM $M$ is a tuple of natural numbers, $(w_1, w_2, \ldots, w_k) \in W$, we denote $\mathrm{U}_{\mathrm{PTM}}(e_M, (w_1, w_2, \ldots, w_k))$ as its output natural number. Here, we can consider $\mathrm{U}_{\mathrm{PTM}}$ as a totally recursive function over $(k+1)$-tuple natural numbers $(e_M, w_1, w_2, \ldots, w_k)$ and a polynomial-time (in $\mathrm{Size}_W(w_1, w_2, \ldots, w_k)$) function.

We then introduce a classical result on the relationship between the time complexity of a Turing machine and Boolean circuit complexity (Theorem 9.25 in [23]).

**Proposition 5.** *Let $t : \mathbb{N} \rightarrow \mathbb{N}$ be a function, where $t(n) \geq n$, and $X = \bigcup_{n \in \mathbb{N}} X_n$, where $X_n \equiv \{x_n \mid n \in \mathbb{N}\}$ is a set of problems $x_n$ with $\mathrm{Size}_{X_n}(x_n) = n$. If all problems in $X_n$ can be computed/decided by a Turing machine within time $t(n)$, then they can be computed/decided by a Boolean circuit with size $O(t^2(n))$.*

This proposition implies that the functionality of a polynomial-time Turing machine can be realized by a polynomial size (uniform) Boolean circuit. This property is used in the proof of Theorem 11.

### 2.7 Polynomial-Time Proofs

A formal proof, $\pi$, of a formula, $\varphi$, is expressed in tree form, called a proof tree, as follows: A proof tree consists of nodes and directed branches. When node $a$ is connected with node $b$ through a branch directed from $b$ to $a$ (i.e., $b \rightarrow a$), $a$ is called a child of $b$ and $b$ is called a parent of $a$: we denote the relation as $a[b]$. If $b$ and $c$ are parents of $a$, the relation is denoted by $a[b, c]$. If $a[b[c, [d, ], ]]$, then $a$ is called a descendent of $c$ and $d$, and $c$ and $d$ are called ancestors of $a$. A node with no child node is called a root, and a node with no parent node is called a leaf. A proof tree has only one root node. (Thus, the image of a proof tree is similar to an actual tree: the root is located at the bottom of a tree and the leaves are at upper on branches.) Node $x$ has form $< x_0, x_1 >$, where $x_0$ is a formula and $x_1$ is a rule of inference of the predicate logic in theory $T$. If $a, b, c$ are nodes of a proof tree of $\pi \equiv a[b, c]$, and $a \equiv < a_0, a_1 >$, $b \equiv < b_0, b_1 >$ and $c \equiv < c_0, c_1 >$, then $\pi$ means that formula $a_0$ is deduced from formulas $b_0$ and $c_0$ through a rule of inference, $a_1$. If no rule of inference is used for the deduction, the part of $a_1$ is empty. If node $a \equiv < a_0, a_1 >$ is a leaf, then $a_0$ is an axiom of the underlying theory $T$ of the proof tree, and $a_1$ is empty. If node $r \equiv < r_0, r_1 >$ is the root of a proof tree of formula $\varphi$, $r_0 = \varphi$.

If theory $T$ is a polynomial-time extension of PA, the Gödel numbers of all axioms and rules of inference in $T$ are polynomial-time (in the size of axioms) decidable. Hence it is clear that the validity of proof tree $\pi$ can be verified within polynomial-time in the size of axioms and the number of nodes of $\pi$. At the end of this subsection, we will show a more precise description of the polynomial-time algorithm to verify the validity of proof tree $\pi$.

Let $\Phi \equiv \{\varphi(\mathbf{a}) \mid a \in \mathbb{N}\}$ be a set of an infinite number of formulas in $T$. The size function, $\mathrm{Size}_\Phi(\cdot)$, over natural numbers $\{a \in \mathbb{N}\}$, is uniquely defined in each $\Phi$. If $\mathrm{Size}_\Phi(\cdot)$ is not explicitly defined, $\mathrm{Size}_\Phi(a) \equiv |a|$. Let $\#\Phi$ be the Gödel number of the (finite-size) expression of the symbol sequence, $\{\varphi(\mathbf{a}) \mid a \in \mathbb{N}\}$ (or the Gödel number of any description of $\Phi$). Note that the size of $\#\Phi$ is finite i.e., a constant in $|a|$.

If $\mathrm{U_{PTM}}(e, (p, \#\Phi, a)) = \#\pi$ and $\pi$ is a valid proof tree of $\varphi(\mathbf{a}) \in \Phi$ in theory $T$, we denote

$$\mathrm{PTM}_e(a) \vdash_T \varphi(\mathbf{a}).$$

Here $p$ denotes a natural number (e.g., 0), which indicates that the output target of $\mathrm{U_{PTM}}(e, \cdot)$ is a proof of the formula's truth.

If a natural number (e.g., 1), $d$, is input to $\mathrm{U_{PTM}}(e, \cdot)$ in place of $p$, it indicates that the output target of $\mathrm{U_{PTM}}(e, \cdot)$ is a decision (accept or reject) of the formula's truth. That is, "$\mathrm{U_{PTM}}(e, (d, \#\Phi, a))$ accepts" implies that $\mathrm{U_{PTM}}(e, \cdot)$, given $(d, \#\Phi, a)$, decides that formula $\varphi(\mathbf{a})$ is true. (See Section 4.1.)

In other words,

$$\begin{aligned} &\mathrm{PTM}_e(a) \vdash_T \varphi(\mathbf{a}) \\ &\quad \Leftrightarrow \quad \mathrm{U_{PTM}}(e, (p, \#\Phi, a)) = \#\pi \ \wedge \ \mathrm{U_{PTM}}(v_T, (\#\varphi(\mathbf{a}), \#\pi)) \text{ accepts}, \end{aligned}$$

where $\mathrm{U_{PTM}}(v_T, (\#\varphi(\mathbf{a}), \#\pi))$accepts, if and only if PTM $\mathrm{U_{PTM}}(v_T, \cdot)$ accepts its input $(\#\varphi(\mathbf{a}), \#\pi)$ as $\pi$ is a valid proof tree of $\varphi(\mathbf{a}) \in \Phi$ in theory $T$. Here, $|\#\pi|$ is clearly polynomially (in $\mathrm{Size}_\Phi(a)$) bounded, since $\#\pi$ is the output of $\mathrm{U_{PTM}}(e, (p, \#\Phi, a))$. In addition, we use the notation

$$\mathrm{PTM}_e(a) \not\vdash_T \varphi(\mathbf{a})$$

if and only if $\neg($ $\mathrm{PTM}_e(a) \vdash_T \varphi(\mathbf{a})$ $)$.

We now describe PTM $\mathrm{U_{PTM}}(v_T, \cdot)$ more precisely.

1. (Input to $\mathrm{U_{PTM}}(v_T, \cdot)$) $(\#\varphi(\mathbf{a}), \#\pi)$, where $\varphi(\mathbf{a})$ is a formula and $\pi$ is a proof tree.
2. $(\#\varphi(\mathbf{a}), \#\pi)$ is interpreted as the Gödel numbers of $\varphi(\mathbf{a})$ and $\pi$ in the manner described in Subsection 2.2.
3. Check the validity of the syntactic form of $\pi$.
4. Search all nodes of $\pi$, and, for each node, decide whether the node is leaf, root or other (say "middle nodes").
5. Repeat the following procedure for all leaf nodes, $a^{(i)}$ $(i = 1, \ldots, \ell)$:
   Pick up $a^{(i)}$, and check whether $a_0^{(i)}$ is an axiom of theory $T$, where $a^{(i)} =< a_0^{(i)}, a_1^{(i)} >$ and $a_1^{(i)}$ is empty string.
6. Repeat the following procedure for all "middle nodes" and the root note, $b^{(i)}$ $(i = 1, \ldots, m)$:
   Pick up $b^{(i)}$ along with its parent nodes (say $c^{(i,j)}$ $(j = 1, \ldots, p_i)$), and check whether $b_0^{(i)}$ is deduced from $c_0^{(i,j)}$ $(j = 1, \ldots, p_i)$, by using a rule of inference $b_1^{(i)}$ (or by using no rule of inference when $b_1^{(i)}$ is empty), where $b^{(i)} =< b_0^{(i)}, b_1^{(i)} >$, and $c^{(i,j)} =< c_0^{(i,j)}, c_1^{(i,j)} >$.
7. Let $r =< r_0, r_1 >$ be the root node. Check whether $r_0 = \varphi(\mathbf{a})$.
8. If all of the above-mentioned checks are passed correctly, the machine accepts the input. Otherwise rejects.

A series of formal proofs produced by a PTM is called a series of "polynomial-time proofs". Here, each polynomial-time proof is a formal proof, $\pi$, of each formula $\varphi(\mathbf{a})$ in theory $T$ (i.e., "a polynomial-time proof" does not mean a set of formal proofs. We will introduce a notion of a set of formal proofs in the next section).

In addition, we introduce the following notation:

$$\begin{aligned} &\mathrm{TM}_e(a) \vdash_T \varphi(\mathbf{a}) \\ &\quad \Leftrightarrow \quad \mathrm{U}(e, (p, \#\Phi, a)) = \#\pi \ \wedge \ \mathrm{U_{PTM}}(v_T, (\#\varphi(\mathbf{a}), \#\pi)) \text{ accepts}. \end{aligned}$$

### 2.8 Asymptotic Proofs

This section introduces a notion called "asymptotic proof".

**Definition 6.** *Let $T$ be a theory, $\Phi \equiv \{\varphi(\mathbf{a}_1, \ldots, \mathbf{a}_k) \mid (a_1, \ldots, a_k) \in \mathbb{N}^k\}$ be a set of (infinite number of) formulas, $\varphi(\mathbf{a}_1, \ldots, \mathbf{a}_k)$ in $T$, and*

$$Q_1 x_1 \in \mathbb{N} \ \cdots \ Q_k x_k \in \mathbb{N} \quad T \vdash \varphi(\mathbf{x}_1, \ldots, \mathbf{x}_k),$$

*where $Q_1, \ldots, Q_k$ are unbounded quantifiers (including partially bounded ones like $\exists x \geq n$ ).*

*Then, a set of an infinite number of formal proofs, $\Pi$, in $T$, is called an "asymptotic proof" of $Q_1 \mathbf{x}_1 \ \cdots Q_k \ \mathbf{x}_k \ \varphi(\mathbf{x}_1, \ldots, \mathbf{x}_k)$, over $T$ if*

$$\begin{aligned} & Q_1 a_1 \in \mathbb{N} \quad \cdots Q_k a_k \in \mathbb{N} \\ & \qquad ( \ \pi(\mathbf{a}_1, \ldots, \mathbf{a}_k) \in \Pi \ \wedge \ \mathrm{U}_{\mathrm{PTM}}(v_T, (\#\varphi(\mathbf{a}_1, \ldots, \mathbf{a}_k), \#\pi(\mathbf{a}_1, \ldots, \mathbf{a}_k))) \mathit{accepts} \ ). \end{aligned}$$

The descriptive size of an asymptotic proof, $\Pi$, can be infinite. Therefore, such an asymptotic proof, $\Pi$, cannot be formulated as a conventional formal proof in $T$, which should be finite-length.

The following lemma demonstrates the difference in provability between formal proofs and asymptotic proofs.

**Lemma 7.** *Let $T$ be a primitive recursive extension of PA and consistent.*

*There exists an asymptotic proof of the consistency of $T$ over PA.*

*On the other hand, there exists no formal proof of the consistency of $T$ in $T$.*

*Proof.* Let $\mathbf{Prov}_T$ be a relation over $(n, m) \in \mathbb{N}^2$ such that $(n, m) \in \mathbf{Prov}_T$ if and only if $n$ is the Gödel number of a formula (say $\psi$) and $m$ is the Gödel number of the proof of $\psi$ in $T$.

Then, $T$ is consistent if and only if

$$\forall m \in \mathbb{N} \quad (n^*, m) \notin \mathbf{Prov}_T,$$

where $n^*$ is the Gödel number of $\perp$ ($\perp$ is $\varphi \wedge \neg\varphi$ for a formula $\varphi$).

Since $T$ is a primitive recursive extension of PA, $\mathbf{Prov}_T$ is a primitive recursive relation. Then, from Proposition 2, there exists a $\Delta_1$-formula, $\mathrm{Prov}_T(\mathbf{x}, \mathbf{y})$, that represents relation $\mathbf{Prov}_T$ in PA.

Therefore, there exists an *asymptotic* proof of the consistency of $T$ over PA as follows:

$$\forall x \in \mathbb{N} \quad \mathrm{PA} \vdash \ \neg\mathrm{Prov}_T(\ulcorner \perp \urcorner, \mathbf{x}) \tag{2}$$

if and only if $T$ is consistent.

On the other hand, even if $T$ is consistent,

$$T \nvdash \ \forall \mathbf{x} \ \neg\mathrm{Prov}_T(\ulcorner \perp \urcorner, \mathbf{x})$$

by the second Gödel incompleteness theorem.

$\dashv$

We now consider the computational complexity of producing an asymptotic proof. Section 2.7 introduced the concept of a polynomial-time proof, that is a proof produced by a PTM. Then, we have a combined concept, an *asymptotic proof produced by a PTM* as follows:

**Definition 8.** *If an asymptotic proof of* $Q_1\mathbf{x}_1 \cdots Q_k\ \mathbf{x}_k\ \varphi(\mathbf{x}_1,\ldots,\mathbf{x}_k)$ *is produced by a PTM, i.e.,*

$$\exists e \in \mathbb{N}\ Q_1 x_1 \in \mathbb{N}\ \cdots\ Q_k x_k \in \mathbb{N} \quad \mathrm{PTM}_e(x_1,\ldots,x_k) \vdash_T \varphi(\mathbf{x}_1,\ldots,\mathbf{x}_k),$$

*then we say "a PTM asymptotically produces a proof of* $Q_1\mathbf{x}_1 \cdots Q_k\ \mathbf{x}_k\ \varphi(\mathbf{x}_1,\ldots,\mathbf{x}_k)$ *over* $T$*," or "*$Q_1\mathbf{x}_1 \cdots Q_k\ \mathbf{x}_k\ \varphi(\mathbf{x}_1,\ldots,\mathbf{x}_k)$ *has a polynomial-time proof (is polynomial-time provable) over* $T$*."*

*Similarly, if an asymptotic proof of* $Q_1\mathbf{x}_1 \cdots Q_k\ \mathbf{x}_k\ \varphi(\mathbf{x}_1,\ldots,\mathbf{x}_k)$ *is produced by a machine in computational class* $\mathcal{C}$*, then we say "a machine in* $\mathcal{C}$ *asymptotically produces a proof of* $Q_1\mathbf{x}_1 \cdots Q_k\ \mathbf{x}_k\ \varphi(\mathbf{x}_1,\ldots,\mathbf{x}_k)$ *over* $T$*," or "*$Q_1\mathbf{x}_1 \cdots Q_k\ \mathbf{x}_k\ \varphi(\mathbf{x}_1,\ldots,\mathbf{x}_k)$ *has a* $\mathcal{C}$ *proof (is* $\mathcal{C}$ *provable) over* $T$*."*

A variant of Lemma 7 demonstrates an example of asymptotic proofs produced by a PTM.

If $T$ is a "PT-extension" of PA, then $\mathrm{Prov}_T(\ulcorner\perp\urcorner,\mathbf{x})$ in Lemma 7 can be equivalent to $\mathrm{PTM\text{-}Acpt}(\mathbf{v}_T,\ulcorner\perp\urcorner,\mathbf{x})$ (see Section 2.10 for the notation of $\mathrm{PTM\text{-}Acpt}(\mathbf{v}_T,\cdot,\cdot)$). Next, we obtain the following lemma by Theorem 11.

**Lemma 9.** *Let* $T$ *be a consistent PT-extension of PA.*

*There exists an polynomial-time proof of the consistency of* $T$ *over PA. That is,*

$$\exists e \in \mathbb{N}\ \ \forall x \in \mathbb{N} \quad \mathrm{PTM}_e(x) \vdash_{\mathrm{PA}} \neg\mathrm{PTM\text{-}Acpt}(\mathbf{v}_T,\ulcorner\perp\urcorner,\mathbf{x}). \tag{3}$$

### 2.9 Representability Theorem of Polynomial-Time Proofs

**Definition 10.** *1. Let* $R$ *be a* $k$*-ary relation on* $\mathbb{N}$*, i.e.,* $R \subseteq \mathbb{N}^k$*. A formula* $\rho_R(\mathbf{x}_1,\ldots,\mathbf{x_k})$ *(in which only* $\mathbf{x}_1,\ldots,\mathbf{x}_k$ *occur free) will be said to polynomial-time represent relation* $R$ *in theory* $T$ *if and only if there exists* $e_R \in \mathbb{N}$ *such that for every* $a_1,\ldots,a_k$ *in* $\mathbb{N}^k$*,*

$$(a_1,\ldots,a_k) \in R \ \ \Rightarrow\ \ \mathrm{PTM}_{e_R}(a_1,\ldots,a_k) \vdash_T \rho_R(\mathbf{a}_1,\ldots,\mathbf{a}_k),$$
$$(a_1,\ldots,a_k) \notin R \ \ \Rightarrow\ \ \mathrm{PTM}_{e_R}(a_1,\ldots,a_k) \vdash_T \neg\rho_R(\mathbf{a}_1,\ldots,\mathbf{a}_k).$$

*2. Let* $f$ *be a* $k$*-place function on natural numbers* $a_1,\ldots,a_k$*. A formula* $\rho_f(\mathbf{x}_1,\ldots,\mathbf{x}_k,\mathbf{y})$ *(in which only* $\mathbf{x}_1,\ldots,\mathbf{x}_k,\mathbf{y}$ *occur free) will be said to functionally polynomial-time represent* $f$ *in theory* $T$ *if and only if there exists* $e_f$ *such that for every* $a_1,\ldots,a_k$ *in* $\mathbb{N}^k$

$$\mathrm{PTM}_{e_f}(a_1,\ldots,a_k) \vdash_T \forall\mathbf{y}(\rho_f(\mathbf{a}_1,\ldots,\mathbf{a}_k,\mathbf{y}) \leftrightarrow \mathbf{y} = \mathbf{S}^{f(a_1,\ldots,a_k)}\mathbf{0}).$$

**Theorem 11.** *(Polynomial-Time Representability Theorem) For any polynomial-time computable relation on* $\mathbb{N}^k$*,* $R$*, and any polynomial-time computable function on* $\mathbb{N}^k$*,* $f$*, there exist formulas,* $\rho_R(\mathbf{x}_1,\ldots,\mathbf{x}_k)$ *and* $\rho_f(\mathbf{x}_1,\ldots,\mathbf{x}_k,\mathbf{y})$*, such that:*

- $\rho_R(\mathbf{x}_1,\ldots,\mathbf{x}_k)$ *polynomial-time represents* $R$*, and* $\rho_f(\mathbf{x}_1,\ldots,\mathbf{x}_k,\mathbf{y})$ *functionally polynomial-time represents* $f$ *in PA.*
- $\rho_R(\mathbf{x}_1,\ldots,\mathbf{x}_k)$ *and* $\rho_f(\mathbf{x}_1,\ldots,\mathbf{x}_k,\mathbf{y})$ *are* $\Delta_1$ *in PA.*
-
$$\mathrm{PA} \vdash \forall\mathbf{x}_1\cdots\forall\mathbf{x}_k\exists!\mathbf{y}\ \rho_f(\mathbf{x}_1,\ldots,\mathbf{x}_k,\mathbf{y}).$$

*Proof.* For simplicity of description, we consider the case of relation $R$ with only one free variable $x$. It is straightforward to extend this result to the cases with multiple free variables and functional representability.

First, we will introduce two function symbols, $|\cdot|$ and $\mathrm{Bit}(\cdot,\cdot)$, which are intended to denote the length of the binary representation of a numeral and the $i$-th rightmost numeral ($\mathbf{0}$ or $\mathbf{1}$) of the binary representation of a numeral, respectively.

*Claim.*

$$\mathrm{PA} \vdash \forall \mathbf{x} > \mathbf{0}\ \exists ! \mathbf{n}\ \mathbf{2}^{\mathbf{n} \dot{-} \mathbf{1}} \leq \mathbf{x} < \mathbf{2}^{\mathbf{n}}.$$

*Proof.* We will use the induction axiom in PA. We can prove the following by using the axioms of PA easily (e.g., by proving $\mathbf{1} + \mathbf{1} = \mathbf{2} = \mathbf{2}^{\mathbf{1}}$):

$$\mathrm{PA} \vdash\ (\mathbf{2}^{\mathbf{0}} \leq \mathbf{1} < \mathbf{2}^{\mathbf{1}}).$$

In addition, we can also prove the following by using the axioms of PA (e.g., by using the axiom, $\forall \mathbf{x} \forall \mathbf{y} (\mathbf{x} + \mathbf{S}\mathbf{y}) = \mathbf{S}(\mathbf{x} + \mathbf{y})$ etc.):

$$\begin{aligned} &\mathrm{PA} \vdash\ \forall \mathbf{x}\ (\ \exists ! \mathbf{n}\ (\mathbf{2}^{\mathbf{n} \dot{-} \mathbf{1}} \leq \mathbf{x} < \mathbf{2}^{\mathbf{n}} \dot{-} \mathbf{1}) \rightarrow \exists ! \mathbf{n}\ (\mathbf{2}^{\mathbf{n} \dot{-} \mathbf{1}} \leq \mathbf{x} + \mathbf{1} < \mathbf{2}^{\mathbf{n}})\ ), \\ &\mathrm{PA} \vdash\ \forall \mathbf{x}\ (\exists ! \mathbf{n}\ (\mathbf{x} = \mathbf{2}^{\mathbf{n}} \dot{-} \mathbf{1}) \rightarrow \mathbf{2}^{\mathbf{n}} = \mathbf{x} + \mathbf{1} \\ &\qquad \rightarrow \mathbf{2}^{\mathbf{n}} \leq \mathbf{x} + \mathbf{1} < \mathbf{2}^{\mathbf{n} + \mathbf{1})}) \rightarrow \exists ! \mathbf{n}'\ (\mathbf{2}^{\mathbf{n}' \dot{-} \mathbf{1}} \leq \mathbf{x} + \mathbf{1} < \mathbf{2}^{\mathbf{n}'})). \end{aligned}$$

Combining the above results, we obtain

$$\mathrm{PA} \vdash\ \ (\mathbf{2}^{\mathbf{0}} \leq \mathbf{1} < \mathbf{2}^{\mathbf{1}})\ \wedge\ \forall \mathbf{x}\ ((\exists ! \mathbf{n}\ \mathbf{2}^{\mathbf{n} \dot{-} \mathbf{1}} \leq \mathbf{x} < \mathbf{2}^{\mathbf{n}}) \rightarrow (\exists ! \mathbf{n}\ \mathbf{2}^{\mathbf{n} \dot{-} \mathbf{1}} \leq \mathbf{x} + \mathbf{1} < \mathbf{2}^{\mathbf{n}})).$$

The induction axiom of PA implies

$$\begin{aligned} \mathrm{PA} \vdash\ \ &(\mathbf{2}^{\mathbf{0}} \leq \mathbf{1} < \mathbf{2}^{\mathbf{1}})\ \wedge\ \forall \mathbf{x}\ (\exists ! \mathbf{n}\ \mathbf{2}^{\mathbf{n} \dot{-} \mathbf{1}} \leq \mathbf{x} < \mathbf{2}^{\mathbf{n}} \rightarrow \exists ! \mathbf{n}\ \mathbf{2}^{\mathbf{n} \dot{-} \mathbf{1}} \leq \mathbf{x} + \mathbf{1} < \mathbf{2}^{\mathbf{n}}) \\ &\rightarrow \forall \mathbf{x} > \mathbf{0}\ \exists ! \mathbf{n}\ \ \mathbf{2}^{\mathbf{n} \dot{-} \mathbf{1}} \leq \mathbf{x} < \mathbf{2}^{\mathbf{n}}. \end{aligned}$$

Hence we obtain finally

$$\mathrm{PA} \vdash \forall \mathbf{x} > \mathbf{0}\ \exists ! \mathbf{n}\ \ \mathbf{2}^{\mathbf{n} \dot{-} \mathbf{1}} \leq \mathbf{x} < \mathbf{2}^{\mathbf{n}}.$$

$\dashv$

Following the claim above, we will introduce a function symbol, $|\cdot|$, in PA, which is intended to denote the binary expression length of numeral $\mathbf{x}$, such that

$$\mathrm{PA} \vdash \forall \mathbf{x} > \mathbf{0}\ \forall \mathbf{n}\ (\mathbf{2}^{\mathbf{n} \dot{-} \mathbf{1}} \leq \mathbf{x} < \mathbf{2}^{\mathbf{n}} \leftrightarrow \mathbf{n} = |\mathbf{x}|). \tag{4}$$

*Claim.*

$$\mathrm{PA} \vdash \forall \mathbf{x} > \mathbf{0}\ \forall \mathbf{n}\ (\mathbf{2}^{\mathbf{n} \dot{-} \mathbf{1}} \leq \mathbf{x} < \mathbf{2}^{\mathbf{n}} \rightarrow ((\mathbf{2}^{\mathbf{n} \dot{-} \mathbf{1}} \leq \mathbf{x} < \mathbf{2}^{\mathbf{n} \dot{-} \mathbf{1}} + \mathbf{2}^{\mathbf{n} \dot{-} \mathbf{2}}) \vee (\mathbf{2}^{\mathbf{n} \dot{-} \mathbf{1}} + \mathbf{2}^{\mathbf{n} \dot{-} \mathbf{2}} \leq \mathbf{x} < \mathbf{2}^{\mathbf{n}}))$$

We omit the proof since it is similarly obtained.

*Claim.*

$$\begin{aligned} \mathrm{PA} \vdash\ \ &\forall \mathbf{x} > \mathbf{1}\ \ \forall \mathbf{i} < |\mathbf{x}|\ \ \exists ! \mathbf{x}_i < \mathbf{2}\ \ \exists ! \mathbf{y} < \mathbf{2}^{\mathbf{i}}\ \ \exists ! \mathbf{z} < \mathbf{2}^{|\mathbf{x}| \dot{-} \mathbf{i} \dot{-} \mathbf{1}} \\ &(\mathbf{x} = \mathbf{y} + \mathbf{x}_i \cdot \mathbf{2}^{\mathbf{i}} + \mathbf{z} \cdot \mathbf{2}^{\mathbf{i} + \mathbf{1}}). \end{aligned}$$

This claim can be proven by applying the previous claims repeatedly.

Based on the claim above, we will introduce a function symbol, $\mathrm{Bit}(\cdot)$, in PA, which is intended to denote the $i$-th rightmost value of the binary expression of a numeral

$$\begin{aligned} \mathrm{PA} \vdash\ \ &\forall \mathbf{x} > \mathbf{1}\ \ \forall \mathbf{i} < |\mathbf{x}|\ \ \forall \mathbf{x}_i < \mathbf{2}\ \ \exists ! \mathbf{y} < \mathbf{2}^{\mathbf{i}}\ \ \exists ! \mathbf{z} < \mathbf{2}^{|\mathbf{x}| \dot{-} \mathbf{i} \dot{-} \mathbf{1}} \\ &(\ \mathbf{x} = \mathbf{y} + \mathbf{x}_i \cdot \mathbf{2}^{\mathbf{i}} + \mathbf{z} \cdot \mathbf{2}^{\mathbf{i} + \mathbf{1}}\ \ \leftrightarrow\ \ \mathbf{x}_i = \mathrm{Bit}(\mathbf{x}, \mathbf{i})\ ). \end{aligned} \tag{5}$$

Hereafter, we will also denote the binary representation
of variable $\mathbf{x}$ by $[\mathbf{x}] = \mathrm{Bit}(\mathbf{x}, \mathbf{n} \dot{-} \mathbf{1})\ \mathrm{Bit}(\mathbf{x}, \mathbf{n} \dot{-} \mathbf{2}) \cdots \mathrm{Bit}(\mathbf{x}, \mathbf{0})$, where $\mathbf{n} = |\mathbf{x}|$.

In order to construct a formula, $\rho_R(\mathbf{x})$, in PA which polynomial-time represents relation $R$, we will employ the approach of constructing a family of polynomial size Boolean circuits that represents relation $R$, which is introduced in Proposition 5 (Theorem 9.25 in [23]).

Since $R$ is polynomial-time computable relation, there exists a PTM, $\mathrm{U}_{\mathrm{PTM}}(e_R, \cdot)$, that computes relation $R$ correctly. Then, $R$ can be decided by a family of Boolean circuits, $\{B_n \mid n \in \mathbb{N}\}$, that are polynomial size in $n$.

Before showing formula $\rho_R(\mathbf{x})$, we will show how to construct a family of Boolean circuits, $\{B_n \mid n \in \mathbb{N}\}$, based on the description of Theorem 9.25 in [23].

Let the size of input $x$ be $n$ bits, and the computation time be $t(n) = n^c$ steps ($c$ is a constant determined by each PTM). The circuit is constructed by $(n^c)^2 k$ nodes. The value of each node is F (false/0/off) or T (true/1/on), and each value is denoted by $light[i, j, s]$ ($0 \leq i < n^c$, $0 \leq j < n^c$, $0 \leq s < k$).

$light[i, j, s] = \mathrm{T}$ ($light[i, j, s]$ is on) denotes the element of $cell[i, j]$ (i.e., in the $i$-th computation step and at the $j$-th leftmost tape square) is the $s$-th element, where there are $k(= 3+3\ell)$ elements, $\Gamma \cup \Gamma \times Q$, $\Gamma \equiv \{0, 1, \sqcup\}$ is the set of tape alphabets, and $Q \equiv \{q_0$ (initial state), $q_1, \ldots, q_{\ell-2}$ (reject state), $q_{\ell-1}$ (accept state) $\}$ is the set of states of the underlying PTM to decide $x \in R$. $light[i, j, s] = \mathrm{F}$ ($light[i, j, s]$ is off) denotes that the element of $cell[i, j]$ is not the $s$-th element. The set of the elements is $\{(0), (1), (\sqcup), (q_0, 0), (q_0, 1), (q_0, \sqcup), \ldots, (q_i, 0), (q_i, 1), (q_i, \sqcup), \ldots, (q_\ell, 0), (q_{\ell-1}, 1), (q_{\ell-1}, \sqcup)\}$. So, for each $cell[i, j]$, only one $light[i, j, s]$ (i.e., only one $s$) is T (true/1/on) and the others are F (false/0/off). Each node is connected to $3k$ nodes through $\wedge$ and $\vee$ gates. More precisely, for all $i (1 \leq i < n^c)$, for all $j (0 \leq j < n^c)$, for all $s (0 \leq s < k)$,

$$light[i, j, s] = \bigvee_{(a,b,c) \in A_s} (light[i-1, j-1, a] \wedge light[i-1, j, b] \wedge light[i-1, j+1, c],$$

where subset $A_s \equiv \{(a_0, b_0, c_0), \ldots, (a_t, b_t, c_t)\}$ ($t < k^3$) is uniquely determined for each $s$ based on the transition function $\delta$ of the underlying PTM to decide $x \in R$. For example,

- $A_1 \equiv \{(1, 1, 1), (1, 1, 0), (2, 3+3i-1, 1), \ldots\}$, where $\delta(q_i, 1) = (q_j, 0, \mathrm{L})$.
- $A_2 \equiv \{(1, 2, 1), (2, 3+3i-2, 1), \ldots\}$, where $\delta(q_i, 0) = (q_j, 1, \mathrm{R})$.
- $A_{3+3i-1} \equiv \{(0, 1, 3+3j-2), \ldots\}$, where $\delta(q_j, 0) = (q_i, 1, \mathrm{L})$.

The values of $light[0, j, s]$, for $0 \leq j < n^c$ and $0 \leq s < k$, are determined by the input $[x] = “x_{n-1} x_{n-2} \cdots x_0''$, i.e.,

$$\begin{cases} light[0,0,3] = 1 \text{ iff} x_{n-1} = 0, \\ light[0,0,4] = 1 \text{ iff} x_{n-1} = 1, \\ light[0,0,s] = 0 \text{ for all } s \text{ with } s \neq 3 \text{ and } s \neq 4. \end{cases}$$

$$\begin{cases} light[0,1,0] = 1 \text{ iff} x_{n-2} = 0, \\ light[0,1,1] = 1 \text{ iff} x_{n-2} = 1, \\ light[0,1,s] = 0 \text{ for all } s \text{ with } s \neq 0 \text{ and } s \neq 1. \end{cases}$$

$$\cdots$$

$$\begin{cases} light[0,n-1,0] = 1 \text{ iff} x_0 = 0, \\ light[0,n-1,1] = 1 \text{ iff} x_0 = 1, \\ light[0,n-1,s] = 0 \text{ for all } s \text{ with } s \neq 0 \text{ and } s \neq 1. \end{cases}$$

$$\begin{cases} light[0,j,2] = 1 \text{ for all } j \geq n. \\ light[0,j,s] = 0 \text{ for all } j \geq n \text{ and for all } s \text{ with } s \neq 2. \end{cases}$$

The input $[w]$ ($n^c k$ bit string) to Boolean circuit $B_n$ is

$$[w] = \text{"}light[0,1,1], light[0,1,2], \ldots, light[0, n^c-1, k-1]'',$$

The output of the circuit is the value of node $light[n^c-1,1,k-6]$, $light[n^c-1,1,k-5]$, $light[n^c-1,1,k-4]$ (reject) or $light[n^c-1,1,k-3]$, $light[n^c-1,1,k-2]$, $light[n^c-1,1,k-1]$ (accept).

We now show formula $\rho_R(\mathbf{x})$ in PA based on the above construction of Boolean circuit $B_n$.

First we define three formulas: $\mathrm{ISET}(\mathbf{x},\mathbf{y})$ in which only $\mathbf{x}$, $\mathbf{y}$ occurs free, $\mathrm{TRANS}(\mathbf{y})$ in which only $\mathbf{y}$ occurs free, and $\mathrm{EVAL}(\mathbf{y})$ in which only $\mathbf{y}$ occurs free. Formula $\mathrm{ISET}(\mathbf{x},\mathbf{y})$ denotes that the information of $x$ is transformed/copied to the value of a part of $y$, formula $\mathrm{TRANS}(\mathbf{y})$ denotes that the transition history of computing $R(x)$ is mapped to the value of the other part of $y$, and $\mathrm{EVAL}(\mathbf{y})$ is true if and only if the evaluation result of $R(x)$ is true.

$$\begin{aligned}
&\mathrm{ISET}(\mathbf{x},\mathbf{y}) \;\equiv \\
&\quad (\ (\mathrm{Bit}(\mathbf{x},\mathbf{0})=\mathbf{0} \;\rightarrow\; (\mathrm{Bit}(\mathbf{y},\mathbf{3})=\mathbf{1} \;\wedge\; \forall \mathbf{s}(\mathbf{0}<\mathbf{s}<\mathbf{3} \;\vee\; \mathbf{3}<\mathbf{s}<\mathbf{k})\ \mathrm{Bit}(\mathbf{y},\mathbf{s})=\mathbf{0}))\ \wedge \\
&\qquad (\mathrm{Bit}(\mathbf{x},\mathbf{0})=\mathbf{1} \rightarrow\; (\mathrm{Bit}(\mathbf{y},\mathbf{4})=\mathbf{1} \;\wedge\; \forall \mathbf{s}(\mathbf{0}<\mathbf{s}<\mathbf{4} \;\vee\; \mathbf{4}<\mathbf{s}<\mathbf{k})\ \mathrm{Bit}(\mathbf{y},\mathbf{s})=\mathbf{0}))\ ) \\
&\quad \wedge \\
&\quad (\ (\forall \mathbf{j}(\mathbf{0}<\mathbf{j}<\mathbf{n}) \;\wedge\; \mathrm{Bit}(\mathbf{x},\mathbf{j})=\mathbf{0} \;\rightarrow\; \mathrm{Bit}(\mathbf{y},\mathbf{j}\cdot\mathbf{k})=\mathbf{1} \;\wedge\; \forall \mathbf{s}(\mathbf{0}<\mathbf{s}<\mathbf{k})\ \mathrm{Bit}(\mathbf{y},\mathbf{j}\cdot\mathbf{k}+\mathbf{s})=\mathbf{0})\ \wedge \\
&\qquad (\forall \mathbf{j}<\mathbf{n} \;\wedge\; \mathrm{Bit}(\mathbf{x},\mathbf{j})=\mathbf{1} \;\rightarrow\; \mathrm{Bit}(\mathbf{y},\mathbf{j}\cdot\mathbf{k}+\mathbf{1})=\mathbf{1} \;\wedge\; \forall \mathbf{s}(\mathbf{s}=\mathbf{0} \;\vee\; \mathbf{1}<\mathbf{s}<\mathbf{k})\ \mathrm{Bit}(\mathbf{y},\mathbf{j}\cdot\mathbf{k}+\mathbf{s})=\mathbf{0})\ ) \\
&\quad \wedge \\
&\quad (\ \forall \mathbf{j}\ (\mathbf{n}<\mathbf{j}<\mathbf{n^c})\ (\mathrm{Bit}(\mathbf{y},\mathbf{j}\cdot\mathbf{k}+\mathbf{2})=\mathbf{1} \;\wedge\; \forall \mathbf{s}\ (\mathbf{s}<\mathbf{2} \;\wedge\; \mathbf{2}<\mathbf{s}<\mathbf{k})\ \mathrm{Bit}(\mathbf{y},\mathbf{j}\cdot\mathbf{k}+\mathbf{s})=\mathbf{0})\ ).
\end{aligned}$$

$$\begin{aligned}
&\mathrm{TRANS}(\mathbf{y}) \;\equiv \\
&\quad \forall \mathbf{i}(\mathbf{0}<\mathbf{i}<\mathbf{n^c})\ \ \forall \mathbf{j}<\mathbf{n^c}\ \ \forall \mathbf{s}<\mathbf{k} \\
&\qquad (\ (\exists \mathbf{a}<\mathbf{k}\ \ \exists \mathbf{b}<\mathbf{k}\ \ \exists \mathbf{c}<\mathbf{k}\ \ (\eta(\mathbf{y},\mathbf{i}\dot{-}\mathbf{1},\mathbf{j},\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{s}) \;\rightarrow\; \mathrm{Bit}(\mathbf{y},\mathbf{i}\cdot\mathbf{n^c}\cdot\mathbf{k}+\mathbf{j}\cdot\mathbf{k}+\mathbf{s})=\mathbf{1}) \\
&\qquad \wedge \\
&\qquad (\forall \mathbf{a}<\mathbf{k}\ \ \forall \mathbf{b}<\mathbf{k}\ \ \forall \mathbf{c}<\mathbf{k}\ \ \neg\eta(\mathbf{y},\mathbf{i}\dot{-}\mathbf{1},\mathbf{j},\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{s}) \;\rightarrow\; \mathrm{Bit}(\mathbf{y},\mathbf{i}\cdot\mathbf{n^c}\cdot\mathbf{k}+\mathbf{j}\cdot\mathbf{k}+\mathbf{s})=\mathbf{0})\ ).
\end{aligned}$$

Here, formula $\eta(\cdot)$ is uniquely fixed for each $s$ based on the transition function $\delta$ of the underlying PTM to decide $x \in R$, and corresponds to subset $A_s$ ($0 \le s < k$) in the above-mentioned Boolean circuit $B_n$. In more detail, $\eta(\cdot)$ is formulated as follows:

$$\begin{aligned}
\eta(\mathbf{y},\mathbf{i}\dot{-}\mathbf{1},\mathbf{j},\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{s}) \;\equiv\; & \eta_0(\mathbf{a},\mathbf{s},\mathrm{Bit}(\mathbf{y},(\mathbf{i}\dot{-}\mathbf{1})\cdot\mathbf{n^c}\cdot\mathbf{k}+(\mathbf{j}\dot{-}\mathbf{1})\cdot\mathbf{k}+\mathbf{a}))\ \wedge \\
& \eta_1(\mathbf{b},\mathbf{s},\mathrm{Bit}(\mathbf{y},(\mathbf{i}\dot{-}\mathbf{1})\cdot\mathbf{n^c}\cdot\mathbf{k}+\mathbf{j}\cdot\mathbf{k}+\mathbf{b}))\ \wedge \\
& \eta_2(\mathbf{c},\mathbf{s},\mathrm{Bit}(\mathbf{y},(\mathbf{i}\dot{-}\mathbf{1})\cdot\mathbf{n^c}\cdot\mathbf{k}+(\mathbf{j}+\mathbf{1})\cdot\mathbf{k}+\mathbf{c})).
\end{aligned}$$

Remark: If $\mathbf{j}=\mathbf{0}$ (or $\mathbf{j}=\mathbf{n^c}\dot{-}\mathbf{1}$), then $\mathbf{a}$ (or $\mathbf{c}$) is ignored.

$$\mathrm{EVAL}(\mathbf{y}) \;\equiv\; \exists \mathbf{s}\ (\mathbf{k}\dot{-}\mathbf{4}<\mathbf{s}<\mathbf{k})\quad \mathrm{Bit}(\mathbf{y},(\mathbf{n^c}\dot{-}\mathbf{1})\cdot\mathbf{n^c}\cdot\mathbf{k}+\mathbf{s})=\mathbf{1}.$$

Finally

$$\rho_R(\mathbf{x}) \;\equiv\; \exists!\mathbf{y}<\mathbf{2^{n^{2c}k}}\ (\mathbf{n}=|\mathbf{x}|\ \wedge \mathrm{ISET}(\mathbf{x},\mathbf{y})\ \wedge \mathrm{TRANS}(\mathbf{y})\ \wedge \mathrm{EVAL}(\mathbf{y})).$$

By the above-mentioned formula, $\rho_R(\mathbf{x})$, for any input $x \in \mathbb{N}$ ($[\mathbf{x}] = \mathbf{x}_{n-1} \cdots \mathbf{x}_0$), numeral $\mathbf{y}$ ($[\mathbf{y}] = \mathbf{y}_{(n^c)^2k-1} \cdots \mathbf{y}_0$), is uniquely determined, and the truth or falsity of $\rho_R(\mathbf{x})$ is also determined by the truth or falsity of term $\mathrm{EVAL}(\mathbf{y})$.

It is clear from Proposition 5 that formula $\rho_R(\mathbf{x})$ represents $R$, since each atomic formula of $\rho_R(\mathbf{x})$ represents the corresponding atomic execution of $\mathrm{U_{PTM}}(e_R, x)$, and such atomic execution is primitive recursive. That is, for all $x \in \mathbb{N}$

$$\mathrm{U_{PTM}}(e_R, x)\text{accepts} \;\Rightarrow\; \mathrm{PA} \vdash \; \exists ! \mathbf{y} < \mathbf{2^{n^{2c}k}} \; (\mathbf{n} = |\mathbf{x}| \; \wedge \mathrm{ISET}(\mathbf{x}, \mathbf{y}) \; \wedge \mathrm{TRANS}(\mathbf{y}) \; \wedge \mathrm{EVAL}(\mathbf{y})).$$

$$\mathrm{U_{PTM}}(e_R, x)\text{rejects} \;\Rightarrow\; \mathrm{PA} \vdash \; \exists ! \mathbf{y} < \mathbf{2^{n^{2c}k}} \; (\mathbf{n} = |\mathbf{x}| \; \wedge \mathrm{ISET}(\mathbf{x}, \mathbf{y}) \; \wedge \mathrm{TRANS}(\mathbf{y}) \; \wedge \neg\mathrm{EVAL}(\mathbf{y})).$$

Since $\mathbf{a}$ is represented by the binary form of numerals, and the proof tree of the formula,

$$\mathbf{2^{n \dot{-} 1}} \leq \mathbf{a} < \mathbf{2^n} \leftrightarrow \mathbf{n} = |\mathbf{a}|,$$

can be constructed in $O(|a|)$, there exists $e_L \in \mathbb{N}$ such that for every $a$ in $\mathbb{N}$,

$$\mathrm{PTM}_{e_L}(a) \vdash_{\mathrm{PA}} \forall \mathbf{n} \; (\mathbf{2^{n \dot{-} 1}} \leq \mathbf{a} < \mathbf{2^n} \leftrightarrow \mathbf{n} = |\mathbf{a}|).$$

Similarly, $\mathrm{Bit}(\cdot)$ can be also functionally polynomial-time represented.

Given $a \in \mathbb{N}$, in order to evaluate formula $\rho_R(\mathbf{a})$, we need to evaluate function $|\cdot|$ once, polynomially many repetitions of function Bit, and polynomially many repetitions of formula $\eta$, where the size of formula $\eta$ is constant in $|a|$ and $\eta$ is $\Delta_1$-formula (since all quantifiers in $\eta$ are bounded).

Therefore, in total, formula $\rho_R(\mathbf{a})$ can be functionally polynomial-time represented.

We now introduce formula $\widetilde{\rho_R}(\mathbf{x}, \mathbf{y})$ that is defined by

$$\mathbf{n} = |\mathbf{x}| \; \wedge \mathrm{ISET}(\mathbf{x}, \mathbf{y}) \; \wedge \mathrm{TRANS}(\mathbf{y}).$$

Here, $\widetilde{\rho_R}(\mathbf{x}, \mathbf{y})$ represents a polynomial-time function, which, given $x$, computes $y$. (Usually, a part of execution history, $y$, is output in a polynomial-time function.)

Then,

$$\rho_R(\mathbf{x}) = \exists ! \mathbf{y} < \mathbf{2^{n^{2c}k}} \; ( \; \widetilde{\rho_R}(\mathbf{x}, \mathbf{y}) \; \wedge \; \mathrm{EVAL}(\mathbf{y}) \; ).$$

By repeatedly proving an atomic formula on a pair of $kn^c$ bit parts (pair of laws) of the binary expression of $\mathbf{y}$, we obtain

$$\mathrm{PA} \vdash \; \forall \mathbf{x} \; \exists ! \mathbf{y} < \mathbf{2^{n^{2c}k}} \; (\mathbf{n} = |\mathbf{x}| \; \wedge \mathrm{ISET}(\mathbf{x}, \mathbf{y}) \; \wedge \mathrm{TRANS}(\mathbf{y}) \; ),$$

where $|\mathbf{x}|$ and $\mathrm{Bit}(\mathbf{x}, \mathbf{i})$ for $\mathbf{x} < \mathbf{2}$ are defined additionally. That is, we obtain

$$\mathrm{PA} \vdash \forall \mathbf{x} \; \exists ! \mathbf{y} < \mathbf{2^{n^{2c}k}} \; \widetilde{\rho_R}(\mathbf{x}, \mathbf{y}).$$

$\dashv$

### 2.10 Formalization of Polynomial-Time Proofs

Let $\Phi$ be a set of an infinite number of formulas, $\{\varphi(\mathbf{a}) \mid a \in \mathbb{N}\}$.

Let formula

$$\text{PTM-Out}(\mathbf{e}, \ulcorner \Phi \urcorner, \mathbf{a}, \mathbf{b})$$

polynomial-time represent

$$\mathrm{U_{PTM}}(e,(p,\#\Phi,a)) = b$$

over natural numbers, $(e,\#\Phi,a,b)$, and formula

$$\text{PTM-Acpt}(\mathbf{v}_T, \ulcorner\varphi(\mathbf{a})\urcorner, \mathbf{b})$$

polynomial-time represent

$$\mathrm{U_{PTM}}(v_T,(\#\varphi(a),b))\text{accepts}$$

over natural numbers, $(v_T,\#\varphi(a),b)$. (For the definition of $v_T$, see Section 2.7.) Here, these formulas are constructed by following (the multiple-variable version of) the method of constructing a formula that was shown in the proof of the polynomial-time representability theorem (Theorem 11).

Then,

$$\mathrm{Pr}_T[\varphi(\mathbf{a})](\mathbf{e},\ulcorner\Phi\urcorner,\mathbf{a}) \quad\equiv\quad \exists\mathbf{b} < 2^{\mathrm{Size}(\mathbf{a})^{\mathbf{c}}} \ (\,\text{PTM-Out}(\mathbf{e},\ulcorner\Phi\urcorner,\mathbf{a},\mathbf{b}) \ \wedge\ \text{PTM-Acpt}(\mathbf{v}_T,\ulcorner\varphi(\mathbf{a})\urcorner,\mathbf{b})\,),$$

where $c$ is uniquely determined by $e$ (i.e., there is a primitive recursive function $f$ such that $c = f(e)$.)

Clearly, $\mathrm{Pr}_T[\varphi(\mathbf{a})](\mathbf{e},\ulcorner\Phi\urcorner,\mathbf{a})$ represents the relation

$$\mathrm{PTM}_e(a) \vdash_T \varphi(\mathbf{a})$$

over natural numbers, $(e,\#\Phi,a) \in \mathbb{N}^4$ (for the definition, see Section 2.7). Then, for any $(e,\#\Phi,a) \in \mathbb{N}^3$,

$$\begin{array}{lcl} \mathrm{PTM}_e(a) \vdash_T \varphi(\mathbf{a}) & \Rightarrow & \mathrm{PA} \vdash\ \mathrm{Pr}_T[\varphi(\mathbf{a})](e,\ulcorner\Phi\urcorner,\mathbf{a}), \\ \mathrm{PTM}_e(a) \not\vdash_T \varphi(\mathbf{a}) & \Rightarrow & \mathrm{PA} \vdash\ \neg\mathrm{Pr}_T[\varphi(\mathbf{a})](e,\ulcorner\Phi\urcorner,\mathbf{a}). \end{array}$$

Here, note that the above-mentioned relation over $(e,\#\Phi,a)$ is polynomial-time decidable in $a$ with a fixed value of $(e,\#\Phi)$, but that the asymptotic computational complexity of this relation in $(e,\#\Phi)$ is not explicitly specified. However, the way of constructing a formula shown in the proof of Theorem 11 can be applied to any primitive recursive relation.

Here it is worth noting that, although formula $\mathrm{Pr}_T[\varphi(\mathbf{z})](\mathbf{x},\mathbf{y},\mathbf{z})$, with free variables $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$, is specified by the construction shown in the proof of Theorem 11, there still exists ambiguity with regard to details of the formula. However, notation $\mathrm{Pr}_T[\varphi(\mathbf{z})](\mathbf{x},\mathbf{y},\mathbf{z})$ means a fixed formula selected from among the possible formulas. The difference of a formula selected from them does not affect the results in this paper. It is important to note that the fixed formula of $\mathrm{Pr}_T[\varphi(\mathbf{z})](\mathbf{x},\mathbf{y},\mathbf{z})$ is assumed throughout this paper.

Informally, formula (sentence) $\mathrm{Pr}_T[\varphi(\mathbf{a})](\mathbf{e},\ulcorner\Phi\urcorner,\mathbf{a})$ is true if and only if $\mathrm{U_{PTM}}(e,\cdot)$, on input $(p,\#\Phi,a) \in \mathbb{N}^3$, outputs a proof tree of formula $\varphi(\mathbf{a}) \in \Phi$ in theory $T$. Here, note that $[\ \cdot\ ]$ does not mean a variable part of the formula, but just implies the target for $\mathrm{U_{PTM}}(e,(p,\#\Phi,a))$ to prove, while $(\ \cdot,\ \cdot,\ \cdot\ )$ means a variable part of the formula. Therefore, the part of $\mathrm{Pr}_T$ in formula $\mathrm{Pr}_T[\varphi(\mathbf{a})](\mathbf{e},\ulcorner\Phi\urcorner,\mathbf{a})$ identifies the form of the formula (like $\rho$ in $\rho(\mathbf{e},\ulcorner\Phi\urcorner,\mathbf{a})$). The part of $[\varphi(\mathbf{a})]$ in the formula is perfectly redundant and is not necessary to identify the formula, but helps readers in understanding the meaning of the formula. (Note that $\mathrm{Pr}_T[X](\cdot,\cdot,\cdot)$ is a single formula, regardless of $X$.)

## 3 Incompleteness Theorems of Polynomial-Time Proofs

This section shows the *polynomial-time proof* version of the (second) Gödel incompleteness theorem. First, we introduce the Gödel sentences of polynomial-time proofs, and the first incompleteness theorem of polynomial-time proofs. We then present the second incompleteness theorem of polynomial-time proofs, based on the the first incompleteness theorem and the derivability conditions of polynomial-time proofs.

### 3.1 Derivability Conditions of Polynomial-Time Proofs

This section introduces several properties, the *derivability conditions* of polynomial-time proofs. (They correspond to the derivability conditions regarding conventional incompleteness theorems.) These properties are used to prove the (first and second) incompleteness theorems of polynomial-time proofs in this paper.

**Lemma 12.** *(D.1 of PTPs)* *Let $\Phi \equiv \{\varphi(\mathbf{a}) \mid a \in \mathbb{N}\}$ be a set of an infinite number of formulas. Suppose that $T$ is a PT-extension of PA. Then the following holds for all $e, T$:*

*For any $e \in \mathbb{N}$ and any $a \in \mathbb{N}$ there exists $e^* \in \mathbb{N}$ such that*

$$\begin{aligned} & \mathrm{PTM}_e(a) \vdash_T \varphi(\mathbf{a}) \\ \Rightarrow\ & \mathrm{PTM}_{e^*}(a) \vdash_{\mathrm{PA}} \mathrm{Pr}_T[\varphi(\mathbf{a})](\mathbf{e}, \ulcorner\Phi\urcorner, \mathbf{a}). \end{aligned}$$

*Proof.* Since

$$\mathrm{PTM}_e(a) \vdash_T \varphi(\mathbf{a})$$

is a polynomial-time relation computed by $\mathrm{U_{PTM}}(e, (\#\Phi, \cdot))$ and $\mathrm{U_{PTM}}(v_T, (\cdot, \cdot))$, given $a \in \mathbb{N}$, such that

$$\mathrm{U_{PTM}}(e, (\#\Phi, a)) = \#\pi \ \wedge\ \mathrm{U_{PTM}}(v_T, (\#\varphi(\mathbf{a}), \#\pi)) \text{ accepts.}$$

Therefore, this result is obtained immediately from Theorem 11.

$\dashv$

**Lemma 13.** *(D.2 of PTPs)* *Let $\Phi \equiv \{\varphi(\mathbf{a}) \mid a \in \mathbb{N}\}$, $\Omega \equiv \{\varphi(\mathbf{a}) \rightarrow \psi(\mathbf{a}) \mid a \in \mathbb{N}\}$, and $\Psi \equiv \{\psi(\mathbf{a}) \mid a \in \mathbb{N}\}$. Suppose that $T$ is a PT-extension of PA.*

*For all $e_1 \in \mathbb{N}$ and for all $e_2 \in \mathbb{N}$, there exists $e_3 \in \mathbb{N}$ such that*

$$\begin{aligned} \mathrm{PA} \vdash\quad \forall \mathbf{x}\ (\quad & \mathrm{Pr}_T[\varphi(\mathbf{x})](\mathbf{e}_1, \ulcorner\Phi\urcorner, \mathbf{x})\ \wedge \mathrm{Pr}_T[\varphi(\mathbf{x}) \rightarrow \psi(\mathbf{x})](\mathbf{e}_2, \ulcorner\Omega\urcorner, \mathbf{x}) \\ & \rightarrow\ \mathrm{Pr}_T[\psi(\mathbf{x})](\mathbf{e}_3, \ulcorner\Psi\urcorner, \mathbf{x})\quad ). \end{aligned}$$

*Proof.* First, we introduce a two-place polynomial-time function, $h$, over $\mathbb{N}^2$ such that

$$h(s,t) \equiv \begin{cases} \#\pi & \text{if there exist proof trees } \pi_1 \text{ and } \pi_2 \text{ in } T \\ & \text{such that } s = \#\pi_1,\ t = \#\pi_2. \text{ Here} \\ & \pi \equiv <\psi, \text{Modus Ponens}> [\pi_1, \pi_2]. \\ 0 & \text{otherwise.} \end{cases}$$

(Given $s \in \mathbb{N}$, it is polynomial-time (in $|s|$) computable to check whether $u$ is the Gödel number of a proof tree in $T$ in a syntactic sense as a symbol sequence.)

PTM $\mathrm{U_{PTM}}(e_3, \cdot)$ is constructed by using two PTMs, $\mathrm{U_{PTM}}(e_1, (p, \#\Phi, \cdot))$ and $\mathrm{U_{PTM}}(e_2, (p, \#\Omega, \cdot))$, and function $h$ as follows:

1. (Input: ) $(p, \#\Psi, x) \in \mathbb{N}^3$.
2. (Output: ) Gödel number of a proof tree of $\psi(\mathbf{x})$ or 0.
3. Run the following computation

$$\mathrm{U_{PTM}}(e_1, (p, \#\Phi, x)) = s,$$

$$\mathrm{U_{PTM}}(e_2, (p, \#\Omega, x)) = t.$$

4. Compute $h(s,t)$ and output the result.

Since function $h$ is primitive recursive, there exists a $\Delta_1$-formula, $\mu(\mathbf{s},\mathbf{t},\mathbf{u})$, in PA which represents function $h$ such that (from Proposition 2)

$$\mathrm{PA} \vdash\ \forall \mathbf{s}\, \forall \mathbf{t}\, \exists ! \mathbf{u} \quad \mu(\mathbf{s},\mathbf{t},\mathbf{u}).$$

We now introduce function symbol $h$ in PA, then

$$\mathrm{PA} \vdash\ \forall \mathbf{s}\, \forall \mathbf{t}\, \forall \mathbf{u} \quad (\mu(\mathbf{s},\mathbf{t},\mathbf{u})\ \leftrightarrow\ \mathbf{u} = h(\mathbf{s},\mathbf{t})\ ).$$

That is,

$$\mathrm{PA} \vdash\ \forall \mathbf{s}\, \forall \mathbf{t}\, \exists ! \mathbf{u}\ \ \mathbf{u} = h(\mathbf{s},\mathbf{t}).$$

Therefore, for all $e_1 \in \mathbb{N}$ and for all $e_2 \in \mathbb{N}$,

$$\begin{aligned}\mathrm{PA} \vdash\ \forall \mathbf{x}\, \forall \mathbf{s}\, \forall \mathbf{t}\, \exists ! \mathbf{u}\ \ (\ \ & \text{PTM-Out}(\mathbf{e}_1, \ulcorner\Phi\urcorner, \mathbf{x}, \mathbf{s})\ \wedge\ \text{PTM-Out}(\mathbf{e}_2, \ulcorner\Omega\urcorner, \mathbf{x}, \mathbf{t})\ ) \\ & \rightarrow\ (\ \text{PTM-Out}(\mathbf{e}_1, \ulcorner\Phi\urcorner, \mathbf{x}, \mathbf{s})\ \wedge\ \text{PTM-Out}(\mathbf{e}_2, \ulcorner\Omega\urcorner, \mathbf{x}, \mathbf{t})\ \ \wedge\ \ \mathbf{u} = h(\mathbf{s},\mathbf{t})\ ).\end{aligned}$$

See Section 2.10 for the definition of notation PTM-Out$(\cdot)$.

Then, by the construction of $\mathrm{U_{PTM}}(e_3, \cdot)$, for all $e_1 \in \mathbb{N}$ and for all $e_2 \in \mathbb{N}$, there exists $e_3 \in \mathbb{N}$ such that

$$\begin{aligned}\mathrm{PA} \vdash\ \ & \forall \mathbf{x}\, \forall \mathbf{u} \\ & (\ \exists ! \mathbf{s}\ \exists ! \mathbf{t}\ (\text{PTM-Out}(\mathbf{e}_1, \ulcorner\Phi\urcorner, \mathbf{x}, \mathbf{s})\ \wedge\ \text{PTM-Out}(\mathbf{e}_2, \ulcorner\Omega\urcorner, \mathbf{x}, \mathbf{t})\ \wedge\ \mathbf{u} = h(\mathbf{s},\mathbf{t})\ )\ ) \\ & \quad \rightarrow\ \ \text{PTM-Out}(\mathbf{e}_3, \ulcorner\Psi\urcorner, \mathbf{x}, \mathbf{u})\ ).\end{aligned}$$

Therefore, for all $e_1 \in \mathbb{N}$ and for all $e_2 \in \mathbb{N}$, there exists $e_3 \in \mathbb{N}$ such that

$$\begin{aligned}\mathrm{PA} \vdash\ \forall \mathbf{x}\, \forall \mathbf{s}\, \forall \mathbf{t}\, \exists ! \mathbf{u}\ (\ \ & \text{PTM-Out}(\mathbf{e}_1, \ulcorner\Phi\urcorner, \mathbf{x}, \mathbf{s})\ \wedge\ \text{PTM-Out}(\mathbf{e}_2, \ulcorner\Omega\urcorner, \mathbf{x}, \mathbf{t}) \\ & \rightarrow\ \ \text{PTM-Out}(\mathbf{e}_3, \ulcorner\Psi\urcorner, \mathbf{x}, \mathbf{u}).\end{aligned} \tag{6}$$

On the other hand, a polynomial-time computation (relation) of $\mathrm{U_{PTM}}(v_T, (\#\psi, u))$ over $(v_T, \#\psi, u)$ is composed of two computation parts as follows:

1. If $u = 0$, reject. Otherwise, check whether there exists a proof tree $\pi$, in which $u = \#\pi$, and whether the inference of the root node of $\pi$ is correct. If both of them are valid, go to next step. (For example, if $\pi = < r_0, r_1 > [\pi_1, \pi_2]$, then check whether inference from $(\pi_1, \pi_2)$ to $r_0$ by the rule of inference $r_1$ is correct. If $u = h(s,t)$ and $h(s,t) \neq 0$, then $s = \#\pi_1$, $t = \#\pi_2$, $r_0 = \psi(\mathbf{x})$ and $r_1$ is Modus Ponens. Hence, if the inference is correct, $\pi_2$ should be $\pi_1 \rightarrow \psi$. )
2. Let $\pi_1$ and $\pi_2$ be parent nodes of the root node of $\pi$. Check whether $\pi_1$ and $\pi_2$ are valid proof trees in $T$.

Since the first computation part is primitive recursive, we can construct formula $\nu(\mathbf{u})$ to represent the first computation (relation) part of $\mathrm{U_{PTM}}(v_T, (\#\psi, u))$, over $u \in \mathbb{N}$. Then from the definition of $h$, we obtain

$$\mathrm{PA} \vdash \ \forall \mathbf{s}\ \forall \mathbf{t}\ \exists ! \mathbf{u}\ \ (\ \nu(\mathbf{u})\ \wedge\ \mathbf{u} = h(\mathbf{s}, \mathbf{t})\ ).$$

Then,

$$\begin{aligned}\mathrm{PA} \vdash \quad \forall \mathbf{x}\ \forall \mathbf{s}\ \forall \mathbf{t}\ \exists ! \mathbf{u}\ (\quad & (\text{PTM-Acpt}(\mathbf{v}_T, \#\varphi(\mathbf{x}), \mathbf{s})\ \wedge\ \text{PTM-Acpt}(\mathbf{v}_T, \#\varphi(\mathbf{x}) \rightarrow \psi(\mathbf{x}), \mathbf{t})) \\ & \rightarrow\ \ (\ \text{PTM-Acpt}(\mathbf{v}_T, \#\varphi(\mathbf{x}), \mathbf{s})\ \wedge\ \text{PTM-Acpt}(\mathbf{v}_T, \#\varphi(\mathbf{x}) \rightarrow \psi(\mathbf{x}), \mathbf{t}) \\ & \quad \wedge\ (\ \nu(\mathbf{u})\ \wedge\ \mathbf{u} = h(\mathbf{s}, \mathbf{t})\ )\ ).\end{aligned}$$

See Section 2.10 for the definition of notation, PTM-Acpt$(\cdot)$.

Here,

$$\begin{aligned}\mathrm{PA} \vdash \quad & \forall \mathbf{x}\ \forall \mathbf{s}\ \forall \mathbf{t}\ \exists ! \mathbf{u} \\ & (\ \text{PTM-Acpt}(\mathbf{v}_T, \#\varphi(\mathbf{x}), \mathbf{s})\ \wedge\ \ \text{PTM-Acpt}(\mathbf{v}_T, \#\varphi(\mathbf{x}) \rightarrow \psi(\mathbf{x}), \mathbf{t})\ \ \wedge\ \ (\ \nu(\mathbf{u})\ \wedge\ \mathbf{u} = h(\mathbf{s}, \mathbf{t})\ )\ ) \\ & \ \rightarrow \quad \text{PTM-Acpt}(\mathbf{v}_T, \#\psi(\mathbf{x}), \mathbf{u})\ ).\end{aligned}$$

Therefore,

$$\begin{aligned}\mathrm{PA} \vdash \ \ \forall \mathbf{x}\ \forall \mathbf{s}\ \forall \mathbf{t}\ \exists ! \mathbf{u}\ \ (\ \ & \text{PTM-Acpt}(\mathbf{v}_T, \#\varphi(\mathbf{x}), \mathbf{s})\ \wedge\ \text{PTM-Acpt}(\mathbf{v}_T, \#\varphi(\mathbf{x}) \rightarrow \psi(\mathbf{x}), \mathbf{t}) \\ & \rightarrow \text{PTM-Acpt}(\mathbf{v}_T, \#\psi(\mathbf{x}), \mathbf{u})\ ).\end{aligned} \tag{7}$$

Hence, combining Eqs. (6) and (7), for all $e_1 \in \mathbb{N}$ and for all $e_2 \in \mathbb{N}$, there exists $e_3 \in \mathbb{N}$ such that

$$\begin{aligned}\mathrm{PA} \vdash \quad & \forall \mathbf{x}\ \forall \mathbf{s}\ \forall \mathbf{t}\ \exists ! \mathbf{u} \\ & (\ (\ \text{PTM-Out}(\mathrm{e}_1, \ulcorner \Phi \urcorner, \mathbf{x}, \mathbf{s})\ \wedge\ \text{PTM-Acpt}(\mathbf{v}_T, \#\varphi(\mathbf{x}), \mathbf{s})\ ) \\ & \wedge\ \ (\ \text{PTM-Out}(\mathrm{e}_2, \ulcorner \Omega \urcorner, \mathbf{x}, \mathbf{t})\ \wedge\ \text{PTM-Acpt}(\mathbf{v}_T, \#\varphi(\mathbf{x}) \rightarrow \psi(\mathbf{x}), \mathbf{t})\ ) \\ & \rightarrow \quad \text{PTM-Out}(\mathrm{e}_3, \ulcorner \Psi \urcorner, \mathbf{x}, \mathbf{u})\ \ \wedge\ \ \text{PTM-Acpt}(\mathbf{v}_T, \#\psi(\mathbf{x}), \mathbf{u})\ ).\end{aligned}$$

Therefore, finally we obtain that for all $e_1 \in \mathbb{N}$ and for all $e_2 \in \mathbb{N}$, there exists $e_3 \in \mathbb{N}$ such that

$$\mathrm{PA} \vdash \ \ \forall \mathbf{x}\ \ (\ \mathrm{Pr}_T[\varphi(\mathbf{x})](\mathrm{e}_1, \ulcorner \Phi \urcorner, \mathbf{x})\ \ \wedge \mathrm{Pr}_T[\varphi(\mathbf{x}) \rightarrow \psi(\mathbf{x})](\mathrm{e}_2, \ulcorner \Omega \urcorner, \mathbf{x}) \quad \rightarrow \quad \mathrm{Pr}_T[\psi(\mathbf{x})](\mathrm{e}_3, \ulcorner \Psi \urcorner, \mathbf{x})\ ).$$

⊣

**Corollary 14.** *Let* $\Phi \equiv \{\varphi(\mathbf{a}) \mid a \in \mathbb{N}\}$ *and* $\Psi \equiv \{\psi(\mathbf{a}) \mid a \in \mathbb{N}\}$. *Suppose that* $T$ *is a consistent PT-extension of PA. We assume*

$$T \vdash \ \forall \mathbf{x}\ (\varphi(\mathbf{x}) \rightarrow \psi(\mathbf{x})).$$

*Then, for all* $e_1 \in \mathbb{N}$ *there exists* $e_2 \in \mathbb{N}$ *such that*

$$\mathrm{PA} \vdash \quad \forall \mathbf{x}\ (\ \mathrm{Pr}_T[\varphi(\mathbf{x})](\mathrm{e}_1, \ulcorner \Phi \urcorner, \mathbf{x})\ \ \rightarrow\ \ \mathrm{Pr}_T[\psi(\mathbf{x})](\mathrm{e}_2, \ulcorner \Psi \urcorner, \mathbf{x})\ ).$$

*Proof.* From the first derivability condition (D.1) of a traditional proof theory [2] and the assumption of this lemma, we obtain

$$\mathrm{PA} \vdash \ \mathrm{Pr}_T(\ulcorner \forall \mathbf{x}\ (\varphi(\mathbf{x}) \rightarrow \psi(\mathbf{x})) \urcorner).$$

Then, PTM $\mathrm{U_{PTM}}(e_2, \cdot)$ is constructed by using PTM $\mathrm{U_{PTM}}(e_1, (p, \#\Phi, \cdot))$ as follows:

1. (Input :) $(p, \#\Psi, x)$
2. (Output: ) Gödel number of a proof tree of $\psi(\mathbf{x})$ or 0.
3. Run the following computation

$$\mathrm{U}_{\mathrm{PTM}}(e_1, (p, \#\Phi, x)) = z, \quad \mathrm{U}_{\mathrm{PTM}}(v_T, (\#\varphi(\mathbf{x}), z)).$$

4. Compute the proof (say $\pi_2$) of $\forall \mathbf{y}\ (\varphi(\mathbf{y}) \to \psi(\mathbf{y}))$, since there exists a proof for the predicate from the assumption. (The computation time is finite i.e., constant in $\mathrm{Size}_\Phi(x)$ and $\mathrm{Size}_\Psi(x)$.)
5. Check whether $\mathrm{U}_{\mathrm{PTM}}(v_T, (\#\varphi(\mathbf{x}), z))$ accepts or rejects. If it rejects, output 0 and halt. If it accepts, then combine $\pi_1$ ($z = \#\pi_1$) and $\pi_2$ and make a new proof tree, $\pi_3$, for $\psi(\mathbf{x})$, as follows:

$$\pi_3 \ \equiv\ <\psi(\mathbf{x}), \text{Modus Ponens}> [\pi_1, <\varphi(\mathbf{x}) \to \psi(\mathbf{x}), \text{Modus Ponens}> [\pi_2, \text{Axiom X}]],$$

where Axiom X is a logical axiom, "$\forall \mathbf{y}\ (\varphi(\mathbf{y}) \to \psi(\mathbf{y})) \ \to\ (\varphi(\mathbf{x}) \to \psi(\mathbf{x}))$".
6. Output $\pi_3$ for the proof tree of formula $\psi(\mathbf{x})$.

The other part of the proof can be completed in an analogous manner to that in Lemma 13 except for the constructions of functions $h$ and $g$ to meet the above-mentioned construction of $\mathrm{U}_{\mathrm{PTM}}(e_2, \cdot))$ in this proof.

$\dashv$

**Corollary 15.** *Let $\Phi \equiv \{\varphi(\mathbf{a}) \mid a \in \mathbb{N}\}$, $\Psi \equiv \{\psi(\mathbf{a}) \mid a \in \mathbb{N}\}$, and $\Omega \equiv \{\varphi(\mathbf{a}) \wedge \psi(\mathbf{a}) \mid a \in \mathbb{N}\}$. Suppose that $T$ is a PT-extension of PA. For all $e_1 \in \mathbb{N}$ and all $e_2 \in \mathbb{N}$, there exists $e_3 \in \mathbb{N}$ such that*

$$\begin{aligned} T \vdash \quad \forall \mathbf{x} \ \ ( \ \mathrm{Pr}_T[\varphi(\mathbf{x})](\mathrm{e}_1, \ulcorner\Phi\urcorner, \mathbf{x}) \ \wedge \mathrm{Pr}_T[\psi(\mathbf{x})](\mathrm{e}_2, \ulcorner\Psi\urcorner, \mathbf{x}) \\ \to \ \mathrm{Pr}_T[\varphi(\mathbf{x}) \wedge \psi(\mathbf{x})](\mathrm{e}_3, \ulcorner\Omega\urcorner, \mathbf{x}) \ ). \end{aligned}$$

*Proof.* By using the following logical axiom of first order logic:

$$\varphi \to (\psi \to (\varphi \wedge \psi)),$$

and the derivability condition D.1. of the standard proof theory [2], we can obtain

$$T \vdash \ \mathrm{Pr}_T(\ulcorner \forall \mathbf{y}\ (\varphi(\mathbf{y}) \to (\psi(\mathbf{y}) \to (\varphi(\mathbf{y}) \wedge \psi(\mathbf{y}))))\urcorner).$$

By applying Corollary 14, we obtain

$$\begin{aligned} T \vdash \quad \forall \mathbf{x}( \ ( \ & \mathrm{Pr}_T[\varphi(\mathbf{x})](\mathrm{e}_1, \ulcorner\Phi\urcorner, \mathbf{x}) \ \wedge \mathrm{Pr}_T[\psi(\mathbf{x})](\mathrm{e}_2, \ulcorner\Psi\urcorner, \mathbf{x})) \\ \to \ & (\mathrm{Pr}_T[(\psi(\mathbf{x}) \to (\varphi(\mathbf{x}) \wedge \psi(\mathbf{x}))](\mathrm{e}'_3, \ulcorner\Omega'\urcorner, \mathbf{x}) \ \wedge \mathrm{Pr}_T[\psi(\mathbf{x})](\mathrm{e}_2, \ulcorner\Psi\urcorner, \mathbf{x})) \\ \to \ & \mathrm{Pr}_T[\varphi(\mathbf{x}) \wedge \psi(\mathbf{x})](\mathrm{e}_3, \ulcorner\Omega\urcorner, \mathbf{x}) \ ). \end{aligned}$$

$\dashv$

**Lemma 16.** *(D.3 of PTPs) Let $R$ be a polynomial-time relation over $\mathbb{N}$. Let formula $\rho_R(x)$ (in which only $x$ occurs free) polynomial-time represent relation $R$ in theory $T$, and the concrete form of formula $\rho_R(\mathbf{x})$ follow the construction given in the proof of Theorem 11. Let $\mathcal{R} \equiv \{\rho_R(\mathbf{a}) \mid a \in \mathbb{N}\}$. Suppose that $T$ is a consistent PT-extension of PA. Then, there exists $e \in \mathbb{N}$ such that*

$$\mathrm{PA} \vdash \ \forall \mathbf{x} \ ( \ \rho_R(\mathbf{x}) \ \to \ \mathrm{Pr}_T[\rho_R(\mathbf{x})](\mathrm{e}, \ulcorner\mathcal{R}\urcorner, \mathbf{x}) \ ).$$

*Proof.* Here we will follow the notations employed in the proof of Theorem 11.

Formula $\rho_R(\mathbf{x})$ has two atomic functions, $|\cdot|$ and $\mathrm{Bit}(\cdot)$, and three atomic formulas $\eta_0(\cdot)$, $\eta_1(\cdot)$ and $\eta_2(\cdot)$. Since $\rho_R(\mathbf{x})$ is a $\Delta_1$-formula, these atomic functions and formulas are composed by a finite number of logical symbols, $\wedge$, $\vee$, $\rightarrow$, $\neg$, and bounded quantifiers. Here bounded quantifiers can be replaced by a finite number of $\wedge$ and $\vee$.

Hence, by applying Lemma 13 and Corollaries 14 and 15, formula $\mathrm{Pr}_T[\rho_R(\mathbf{x})](\mathbf{e}, \ulcorner\mathcal{R}\urcorner, \mathbf{x})$ can be deduced from a logical composition of the corresponding atomic formulas,

$$\begin{aligned}
&\mathrm{Pr}_T[\mathbf{w} = |\mathbf{z}|](\mathrm{e}_3, \ulcorner\mathcal{L}\urcorner, (\mathbf{z}, \mathbf{w})), \quad \mathrm{Pr}_T[\mathbf{w} = \mathrm{Bit}(\mathbf{z})](\mathrm{e}_4, \ulcorner\mathcal{B}\urcorner, (\mathbf{z}, \mathbf{w})),\\
&\mathrm{Pr}_T[\eta_i(\mathbf{z}, \mathbf{w}, \mathbf{v})](\mathrm{e}_i, \ulcorner\mathcal{E}_i\urcorner, (\mathbf{z}, \mathbf{w}, \mathbf{v})),\\
&\mathrm{Pr}_T[\neg(\mathbf{w} = |\mathbf{z}|)](\mathrm{e}_3^*, \ulcorner\mathcal{L}^*\urcorner, (\mathbf{z}, \mathbf{w})), \quad \mathrm{Pr}_T[\neg(\mathbf{w} = \mathrm{Bit}(\mathbf{z}))](\mathrm{e}_4^*, \ulcorner\mathcal{B}\urcorner, (\mathbf{z}, \mathbf{w})),\\
&\mathrm{Pr}_T[\neg\eta_i(\mathbf{z}, \mathbf{w}, \mathbf{v})](\mathrm{e}_i^*, \ulcorner{\mathcal{E}_i}^*\urcorner, (\mathbf{z}, \mathbf{w}, \mathbf{v})),
\end{aligned}$$

where $i = 0, 1, 2$.

Therefore, to prove this Lemma it is sufficient to prove the following atomic formulas:

$$\begin{aligned}
&\exists \mathrm{e} \in \mathbb{N} \quad \mathrm{PA} \vdash\ \forall \mathbf{z}\ \forall \mathbf{w}\ (\ (\mathbf{w} = |\mathbf{z}|)\ \rightarrow\ \mathrm{Pr}_T[\mathbf{w} = |\mathbf{z}|](\mathrm{e}, \ulcorner\mathcal{L}\urcorner, (\mathbf{z}, \mathbf{w}))\ ),\\
&\exists \mathrm{e} \in \mathbb{N} \quad \mathrm{PA} \vdash\ \forall \mathbf{z}\ \forall \mathbf{w}\ (\ (\mathbf{w} = \mathrm{Bit}(\mathbf{z}))\ \rightarrow\ \mathrm{Pr}_T[\mathbf{w} = \mathrm{Bit}(\mathbf{z})](\mathrm{e}, \ulcorner\mathcal{B}\urcorner, (\mathbf{z}, \mathbf{w}))\ ),\\
&\exists \mathrm{e} \in \mathbb{N} \quad \mathrm{PA} \vdash\ \forall \mathbf{z}\ \forall \mathbf{w}\ \forall \mathbf{v}\ (\ \eta_i(\mathbf{z})\ \rightarrow\ \mathrm{Pr}_T[\eta_i(\mathbf{z}, \mathbf{w}, \mathbf{v})](\mathrm{e}, \ulcorner\mathcal{E}_i\urcorner, (\mathbf{z}, \mathbf{w}, \mathbf{v}))\ ),\\
&\exists \mathrm{e} \in \mathbb{N} \quad \mathrm{PA} \vdash\ \forall \mathbf{z}\ \forall \mathbf{w}\ (\ \neg(\mathbf{w} = |\mathbf{z}|)\ \rightarrow\ \mathrm{Pr}_T[\neg(\mathbf{w} = |\mathbf{z}|](\mathrm{e}, \ulcorner\mathcal{L}^*\urcorner, (\mathbf{z}, \mathbf{w}))\ ),\\
&\exists \mathrm{e} \in \mathbb{N} \quad \mathrm{PA} \vdash\ \forall \mathbf{z}\ \forall \mathbf{w}\ (\ \neg(\mathbf{w} = \mathrm{Bit}(\mathbf{z}))\ \rightarrow\ \mathrm{Pr}_T[\neg(\mathbf{w} = \mathrm{Bit}(\mathbf{z}))](\mathrm{e}, \ulcorner\mathcal{B}^*\urcorner, (\mathbf{z}, \mathbf{w}))\ ),\\
&\exists \mathrm{e} \in \mathbb{N} \quad \mathrm{PA} \vdash\ \forall \mathbf{z}\ \forall \mathbf{w}\ \forall \mathbf{v}\ (\ \neg\eta_i(\mathbf{z}, \mathbf{w}, \mathbf{v})\ \rightarrow\ \mathrm{Pr}_T[\neg\eta_i(\mathbf{z}, \mathbf{w}, \mathbf{v})](\mathrm{e}, \ulcorner{\mathcal{E}_i}^*\urcorner, (\mathbf{z}, \mathbf{w}, \mathbf{v}))\ ),
\end{aligned}$$

where $i = 0, 1, 2$.

We will then show a construction of $\mathrm{U_{PTM}}(e, \cdot)$ that outputs a proof tree of each atomic formula.

First, $\mathrm{U_{PTM}}(e, \cdot)$ for atomic function $|\cdot|$ is as follows:

1. (Input:) $(p, \#\mathcal{L}, z, w)$, where $\mathcal{L} \equiv \{\mathbf{a} = |\mathbf{b}| \mid a \in \mathbb{N}, b \in \mathbb{N}\}$
2. (Output:) $\#\pi_L$ or 0, where $\pi_L$ is a proof tree of formula $\mathbf{w} = |\mathbf{z}|$ in PA, and 0 means "Fail". Note that $\mathbf{z}$ and $\mathbf{w}$ are given in binary form such as
$$\begin{aligned}
&\mathbf{z}_0 + \mathbf{z}_1 \cdot \mathbf{2} + \cdots + \mathbf{z}_{w-1} \cdot \mathbf{2}^{\mathbf{w} \dot{-} \mathbf{1}}\\
&\quad (\text{more precisely, } \mathbf{z}_0 + \mathbf{z}_1 \cdot \mathbf{SS0} + \cdots + \mathbf{z}_{n-1} \cdot \mathbf{SS0} \cdot \mathbf{SS0} \cdots \mathbf{SS0}),\\
&\mathbf{w}_0 + \mathbf{w}_1 \cdot \mathbf{2} + \cdots + \mathbf{w}_{\mathbf{t}-1} \cdot \mathbf{2}^{\mathbf{t} \dot{-} \mathbf{1}}.
\end{aligned}$$
3. Check whether $\mathbf{2}^{\mathbf{w} \dot{-} \mathbf{1}} \leq \mathbf{z} < \mathbf{2}^{\mathbf{w}}$ or not. If it is false, output 0. Otherwise, go to next step.
4. Make a proof tree, $\pi_L$, of $\mathbf{2}^{\mathbf{w} \dot{-} \mathbf{1}} \leq \mathbf{z} < \mathbf{2}^{\mathbf{w}}$ by showing $\mathbf{z}' = \mathbf{z}_0 + \mathbf{z}_1 \cdot \mathbf{2} + \cdots + \mathbf{z}_{w-2} \cdot \mathbf{2}^{\mathbf{w} \dot{-} \mathbf{2}}$ and $\mathbf{z}'' = \overline{\mathbf{z}_0} + \overline{\mathbf{z}_1} \cdot \mathbf{2} + \cdots + \overline{\mathbf{z}_{w-2}} \cdot \mathbf{2}^{\mathbf{w} \dot{-} \mathbf{2}}$, along with the proof tree of $\mathbf{z} = \mathbf{2}^{\mathbf{w} \dot{-} \mathbf{1}} + \mathbf{z}'$ and $\mathbf{z} + \mathbf{z}'' = \mathbf{2}^{\mathbf{w}}$, where $\overline{\mathbf{z}_i}$ denotes the complement of $\mathbf{z}_i$ (e.g., if $\mathbf{z}_i = \mathbf{0}$, $\overline{\mathbf{z}_i} = \mathbf{1}$).
(Note that $\mathbf{w}$ in the above equations is expressed in binary form.)

PTM-Out$(\mathrm{e}, \ulcorner\mathcal{L}\urcorner, \mathbf{z}, \mathbf{w}, \mathbf{y})$ represents the above-mentioned computation (function) of $\mathrm{U_{PTM}}(e,$ $(p,\ \#\mathcal{L},\ z, w))$ to output $y$ such that $y = \#\pi_L$ or $y = 0$. (For PTM-Out, see Section 2.10). From the definition of $|\cdot|$,

$$\mathrm{PA} \vdash\ \forall \mathbf{z} \forall \mathbf{w}\ (\mathbf{w} = |\mathbf{z}| \leftrightarrow (\mathbf{2}^{\mathbf{w} \dot{-} \mathbf{1}} \leq \mathbf{z} < \mathbf{2}^{\mathbf{w}})).$$

In addition, from the construction of $\mathrm{U_{PTM}}(e, (p, \#\mathcal{L}, z, w))$,

$$\mathrm{PA} \vdash\ \forall \mathbf{z} \forall \mathbf{w}\ (\mathbf{2}^{\mathbf{w} \dot{-} \mathbf{1}} \leq \mathbf{z} < \mathbf{2}^{\mathbf{w}})\ \rightarrow\ \text{PTM-Out}(\mathrm{e}, \ulcorner\mathcal{L}\urcorner, \mathbf{z}, \mathbf{w}, \ulcorner\pi_L\urcorner).$$

Since PTM-Acpt$(\mathbf{v}_T, \ulcorner \mathbf{z} = |\mathbf{w}| \urcorner, \ulcorner \pi_L \urcorner)$ represents computation $\mathrm{U_{PTM}}(v_T, (\#\mathbf{z} = |\mathbf{w}|, \#\pi_L))$ (see Section 2.10),

$$\mathrm{PA} \vdash \ \forall \mathbf{z}\ \forall \mathbf{w}\ \exists! \mathbf{y}\ \ (\mathbf{w} = |\mathbf{z}|\ \wedge\ \text{PTM-Out}(\mathbf{e}, \ulcorner \mathcal{L} \urcorner, \mathbf{z}, \mathbf{w}, \mathbf{y})) \rightarrow\ \text{PTM-Acpt}(\mathbf{v}_T, \ulcorner \mathbf{w} = |\mathbf{z}| \urcorner, \mathbf{y})\ ).$$

Namely,

$$\mathrm{PA} \vdash\ \forall \mathbf{z}\ \forall \mathbf{w}\ (\ (\mathbf{w} = |\mathbf{z}|)\ \rightarrow\ \mathrm{Pr}_T[\mathbf{w} = |\mathbf{z}|](\mathbf{e}, \ulcorner \mathcal{L} \urcorner, (\mathbf{z}, \mathbf{w}))\ ).$$

We can also prove similar results on $\neg(\mathbf{w} = |\mathbf{z}|)$, $(\mathbf{w} = \mathrm{Bit}(\mathbf{z}))$, and $\neg(\mathbf{w} = \mathrm{Bit}(\mathbf{z}))$ in a manner similar to that on $(\mathbf{w} = |\mathbf{z}|)$.

We will now prove the results on formulas $\eta_i$ $(i = 0, 1, 2)$.

Since the values of variables of formula $\eta_i$ $(i = 0, 1, 2)$ are bounded by constant $\mathbf{k}$ and the number of variables is also bounded by 3, all possible evaluation values of $\eta_i$ $(i = 0, 1, 2)$ with possible values of variable are bounded by a constant. This means that a proof of each possibility of formula $\eta_i$ $(i = 0, 1, 2)$ can be created ahead of time and stored by PTM $\mathrm{U_{PTM}}(e, \cdot)$. So, the role of PTM $\mathrm{U_{PTM}}(e, \cdot)$ is just pattern matching against the value of the input variables.

Given an input value, $\mathrm{U_{PTM}}(e, \cdot)$ outputs the Gödel number of a proof tree of formula $\eta_i$ $(i = 0, 1, 2)$ as follows:

1. (Input:) $(p, \ulcorner \mathcal{E}_i \urcorner, z, w, v)$, where $\mathcal{E}_i \equiv \{\eta_i(\mathbf{a}, \mathbf{b}, \mathbf{c})\ \mid\ a \in \mathbb{N}, b \in \mathbb{N}, c \in \mathbb{N}\}$
2. (Output:) $\#\pi_{E,i}(\mathbf{z}, \mathbf{w}, \mathbf{v})$ or 0, where $\pi_{E,i}(\mathbf{z}, \mathbf{w}, \mathbf{v})$ is a proof tree of formula $\eta_i(\mathbf{z}, \mathbf{w}, \mathbf{v})$ in theory PA.
3. (Preprocessing Phase before getting Input) List up all input values of $(z, w, v)$ for which $\eta_i(\mathbf{z}, \mathbf{w}, \mathbf{v})$ is true (say the list "TList"). Make (the Gödel number of) a proof tree, $\pi_{E,i}(\mathbf{z}, \mathbf{w}, \mathbf{v})$, of $\eta_i(\mathbf{z}, \mathbf{w}, \mathbf{v})$ for all values of $(z, w, v) \in$ TList. Make a list of (the Gödel number of) the proof trees along with TList, which is retrieved by entry $(z, w, v)$ (say PList; $\{((z, w, v), \#\pi_{E,i}(\mathbf{z}, \mathbf{w}, \mathbf{v}))\ |$ $(z, w, v) \in$ TList$\}$ ). Note that the size of PList is finite and constant in the size of input $\mathbf{x}$ to $\rho_R(\cdot)$.
4. Given input $(z, w, v)$, search PList by the input. If entry $(z, w, v)$ is found in PList, output the corresponding $\#\pi_{E,i}(\mathbf{z}, \mathbf{w}, \mathbf{v})$. Otherwise, output 0.

Let PTM-Out$(\mathbf{e}, \ulcorner \mathcal{E}_i \urcorner, \mathbf{z}, \mathbf{w}, \mathbf{v}, \mathbf{y})$ be a formula to represent the computation, $\mathrm{U_{PTM}}(e, (p, \ulcorner \mathcal{E}_i \urcorner,$ $z,\ w, v)) = y$, where $y = \pi_{E,i}(\mathbf{z}, \mathbf{w}, \mathbf{v})$ or $y = 0$.

Since the computation is just pattern matching, the formula should be effectively equivalent to the following form:

$$\begin{array}{ll}
\forall \mathbf{z}\ \forall \mathbf{w}\ \forall \mathbf{v}\ \exists! \mathbf{y} & (\ ((\mathbf{z}, \mathbf{w}, \mathbf{v}) = (\mathbf{z}_0, \mathbf{w}_0, \mathbf{v}_0)\ \rightarrow\ \mathbf{y} = \pi_0) \\
& \wedge\ ((\mathbf{z}, \mathbf{w}, \mathbf{v}) = (\mathbf{z}_1, \mathbf{w}_1, \mathbf{v}_1)\ \rightarrow\ \mathbf{y} = \pi_1) \\
& \quad \cdots \\
& \quad \cdots \\
& \wedge\ ((\mathbf{z}, \mathbf{w}, \mathbf{v}) = (\mathbf{z}_K, \mathbf{w}_K, \mathbf{v}_K)\ \rightarrow\ \mathbf{y} = \pi_K), \\
& \wedge\ ((\mathbf{z}, \mathbf{w}, \mathbf{v}) \neq (\mathbf{z}_0, \mathbf{w}_0, \mathbf{v}_0)\ \wedge\ (\mathbf{z}, \mathbf{w}, \mathbf{v}) \neq (\mathbf{z}_1, \mathbf{w}_1, \mathbf{v}_1)\ \cdots \\
& \qquad \wedge\ (\mathbf{z}, \mathbf{w}, \mathbf{v}) \neq (\mathbf{z}_K, \mathbf{w}_K, \mathbf{v}_K)\ \rightarrow\ \mathbf{y} = \mathbf{0})\ ),
\end{array}$$

where TList $\equiv \{(z_0, w_0, v_0), (z_1, w_1, v_1), \ldots, (z_K, w_K, v_K)\}$.

From the construction,

$$\begin{array}{ll}
\mathrm{PA} \vdash & \forall \mathbf{z}\ \forall \mathbf{w}\ \forall \mathbf{v}\ (\ \eta_i(\mathbf{z}, \mathbf{w}, \mathbf{v}) \\
& \leftrightarrow\ \ (\ (\mathbf{z}, \mathbf{w}, \mathbf{v}) = (\mathbf{z}_0, \mathbf{w}_0, \mathbf{v}_0)\ \vee\ (\mathbf{z}, \mathbf{w}, \mathbf{v}) = (\mathbf{z}_1, \mathbf{w}_1, \mathbf{v}_1)\ \cdots\ \vee\ (\mathbf{z}, \mathbf{w}, \mathbf{v}) = (\mathbf{z}_K, \mathbf{w}_K, \mathbf{v}_K)\ )\ ).
\end{array}$$

For all $i$ $(0 \leq i \leq K)$,

$$\text{PA} \vdash \quad \forall \mathbf{z}\, \forall \mathbf{w}\, \forall \mathbf{v}\ (\ (\mathbf{z},\mathbf{w},\mathbf{v}) = (\mathbf{z}_i,\mathbf{w}_i,\mathbf{v}_i) \ \rightarrow \ \text{PTM-Out}(\mathbf{e}, \ulcorner \mathcal{E}_i \urcorner, \mathbf{z}, \mathbf{w}, \mathbf{v}, \ulcorner \pi_i \urcorner)\ ).$$

For all $i$ $(0 \leq i \leq K)$,

$$\text{PA} \vdash \quad \forall \mathbf{z}\, \forall \mathbf{w}\, \forall \mathbf{v}\ (\ (\mathbf{z},\mathbf{w},\mathbf{v}) = (\mathbf{z}_i,\mathbf{w}_i,\mathbf{v}_i) \ \rightarrow \ \text{PTM-Acpt}(\mathbf{v}_T, \ulcorner \eta_i(\mathbf{z},\mathbf{w},\mathbf{v}) \urcorner, \ulcorner \pi_i \urcorner)\ ).$$

Hence,

$$\text{PA} \vdash \forall \mathbf{z}\, \forall \mathbf{w}\, \forall \mathbf{v}\, \exists! \mathbf{y}\ (\ \eta_i(\mathbf{z},\mathbf{w},\mathbf{v}) \ \rightarrow \ \text{PTM-Out}(\mathbf{e}, \ulcorner \mathcal{E}_i \urcorner, \mathbf{z}, \mathbf{w}, \mathbf{v}, \mathbf{y}) \wedge \text{PTM-Acpt}(\mathbf{v}_T, \ulcorner \eta_i(\mathbf{z},\mathbf{w},\mathbf{v}) \urcorner, \mathbf{y})\ ).$$

Namely,

$$T \vdash \ \forall \mathbf{z}\, \forall \mathbf{w}\ (\ \eta_i(\mathbf{z},\mathbf{w},\mathbf{v}) \ \rightarrow \ \text{Pr}_T[\eta_i(\mathbf{z},\mathbf{w},\mathbf{v})](\mathbf{e}, \ulcorner \mathcal{E}_i \urcorner, \mathbf{z}, \mathbf{w}, \mathbf{v})\ ).$$

We can also prove a similar result on $\neg\eta_i(\mathbf{z},\mathbf{w},\mathbf{v})$ in a manner similar to that on $\eta_i(\mathbf{z},\mathbf{w},\mathbf{v})$.

$\dashv$

### 3.2 Recursion Theorem of Polynomial-Time Proofs

**Proposition 17.** *(Recursion Theorem) Let* $\text{U}(t,(\cdot,\cdot))$ *be a Turing machine that computes a two-place function:* $\mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$*. There exists a Turing machine* $\text{U}(k,\cdot)$ *(i.e., there exists* $k \in \mathbb{N}$*) that computes a function:* $\mathbb{N} \rightarrow \mathbb{N}$*, where for every* $w \in \mathbb{N}$*,*

$$\text{U}(k,w) = \text{U}(t,(k,w)).$$

(Note: for notation $\text{U}(\cdot,\cdot)$, see Section 2.5.)

For the proof of this proposition, see [23](Section 6.1). The point is that we can construct a Turing machine $\text{U}_{\text{PTM}}(k,\cdot)$ that can read its own code, $k$. Note that the computational complexity of reading its own code is constant in (independent from) input size, $|w|$. $\text{U}(k,\cdot)$, on input $w$, first reads $k$, and then simulates $\text{U}(t,(\cdot,\cdot))$ on input $(k,w)$.

By using this proposition, we can obtain the PTM version of the recursion theorem.

**Lemma 18.** *(PTM and formula version of Recursion Theorem) Given* $t \in \mathbb{N}$*, let formula* $\xi_t(\mathbf{k},\mathbf{w})$*, in which only* $\mathbf{k}$ *and* $\mathbf{w}$ *occur free, polynomial-time represent function* $\text{U}_{\text{PTM}}(t,(k,w))$ *on* $(k,w) \in \mathbb{N}^2$*. Then, for any* $t \in \mathbb{N}$*, there exists* $k \in \mathbb{N}$ *and formula* $\rho_k$ *such that formula* $\rho_k(\mathbf{w})$*, in which only* $\mathbf{w}$ *occurs free, polynomial-time represents function* $\text{U}_{\text{PTM}}(k,w)$ *on* $w \in \mathbb{N}$*, and*

$$\text{PA} \vdash \forall \mathbf{w} \quad (\rho_k(\mathbf{w}) \ \leftrightarrow \ \xi_t(\mathbf{k},\mathbf{w}))$$

*Proof.* From the recursion theorem (Proposition 17), for any $t \in \mathbb{N}$, there exists $k \in \mathbb{N}$ such that for any $w \in \mathbb{N}$,

$$\text{U}_{\text{PTM}}(k,w) = \text{U}_{\text{PTM}}(t,(k,w)).$$

Here, PTM $\text{U}_{\text{PTM}}(k,\cdot)$ runs as follows:

1. (Input:) $w \in \mathbb{N}$
2. (Output:) accept/reject
3. First, read its own code, $k \in \mathbb{N}$ via the recursion theorem (Proposition 17).
4. Simulate PTM $\text{U}_{\text{PTM}}(t,(\cdot,\cdot))$ on input $(k,w)$.
5. Accept if and only if $\text{U}_{\text{PTM}}(t,(k,w))$ accepts.

Therefore, the difference between $\mathrm{U}_{\mathrm{PTM}}(k, w)$ and $\mathrm{U}_{\mathrm{PTM}}(t,(k, w))$ is the step in which $\mathrm{U}_{\mathrm{PTM}}(k, w)$ reads its own code, $k$, while $\mathrm{U}_{\mathrm{PTM}}(t,(k, w))$ obtains $k$ as an input.

Let $\rho_k(\mathbf{w})$ polynomial-time represent $\mathrm{U}_{\mathrm{PTM}}(k, w)$. Let $\theta(\mathbf{k})$ represent the computation of the step in which $\mathrm{U}_{\mathrm{PTM}}(k, w)$ reads its own code, $k$.

Since $\mathrm{U}_{\mathrm{PTM}}(k, w)$ can always reads its own code, $k$, clearly from Proposition 2 there exists $k \in \mathbb{N}$ such that

$$\mathrm{PA} \vdash \ \theta(\mathbf{k}).$$

Then, there exists $k \in \mathbb{N}$ such that

$$\mathrm{PA} \vdash \ \forall \mathbf{w} \ ( \ \theta(\mathbf{k}) \wedge \xi_t(\mathbf{k}, \mathbf{w}) \leftrightarrow \ \xi_t(\mathbf{k}, \mathbf{w}) \ ).$$

Since formula $(\theta(\mathbf{k}) \wedge \xi_t(\mathbf{k}, \mathbf{w}))$ polynomial-time represents $\mathrm{U}_{\mathrm{PTM}}(k, w)$ from Proposition 4, we identify it by $\rho_k(\mathbf{w})$. Then, there exists $k \in \mathbb{N}$ and $\rho_k(\mathbf{w})$ such that

$$\mathrm{PA} \vdash \ \forall \mathbf{w} \ ( \ \rho_k(\mathbf{w}) \ \leftrightarrow \ \xi_t(\mathbf{k}, \mathbf{w}) \ ),$$

$\dashv$

### 3.3 Gödel Sentences of Polynomial-Time Proofs

**Lemma 19.** *Let $T$ be a consistent PT-extension of PA. Then, for any $e \in \mathbb{N}$, there exists a set of formulas, $\mathcal{G} \equiv \{\rho_{e,T}(\mathbf{a}) \mid a \in \mathbb{N}\}$, such that*

$$\mathrm{PA} \vdash \forall \mathbf{x} \ \ (\rho_{e,T}(\mathbf{x}) \ \leftrightarrow \ \neg \mathrm{Pr}_T[\rho_{e,T}(\mathbf{x})](\mathbf{e}, \ulcorner \mathcal{G} \urcorner, \mathbf{x}) \ ).$$

*For all $x$, $\rho_{e,T}(\mathbf{x})$ is called a "Gödel sentence" with respect to* PTM.

*Proof.* Given $e \in \mathbb{N}$ and theory $T$, PTM $\mathrm{U}_{\mathrm{PTM}}(t,(k, x))$ in Lemma 18 is specialized to this lemma as follows:

1. (Input:) $(k, x) \in \mathbb{N}^2$
2. (Output:) accept/reject
3. Construct formula $\rho_k(\mathbf{x})$ that polynomial-time represents the computation of $\mathrm{U}_{\mathrm{PTM}}(k, x)$ via the polynomial-time representability theorem (Theorem 11). Let $\mathcal{G}_k \equiv \{\rho_k(\mathbf{a}) \mid a \in \mathbb{N}\}$.
4. Construct PTM $\mathrm{U}_{\mathrm{PTM}}(e,(p, \#\mathcal{G}_k, x))$ to produce a proof of formula $\rho_k(\mathbf{x})$. Then, check whether it outputs a valid proof tree of the input by using $\mathrm{U}_{\mathrm{PTM}}(v_T, \cdot)$. That is, verify whether the following holds or not:
$$\mathrm{PTM}_e(x) \vdash_T \rho_k(\mathbf{x}),$$
i.e.,, check
$$\mathrm{U}_{\mathrm{PTM}}(e,(p, \#\mathcal{G}_k, x))=y \ \wedge \ \mathrm{U}_{\mathrm{PTM}}(v_T,(\#\rho_k(\mathbf{x}), y)) \text{ accepts},$$
5. Accept if and only if the above-mentioned relation "does not" hold.

It is clear from the definition of formula $\mathrm{Pr}_T[\cdot](\cdot,\cdot,\cdot)$ in Section 2.10 that $\neg \mathrm{Pr}_T[\rho_k(\mathbf{x})](\mathbf{e}, \ulcorner \mathcal{G}_k \urcorner, \mathbf{x})$ represents the above-mentioned relation that $\mathrm{U}_{\mathrm{PTM}}(t,(k, x))$ accepts.

Therefore, from Lemma 18, for any $t \in \mathbb{N}$ (i.e., for any $e \in \mathbb{N}$), there exists $k \in \mathbb{N}$ and formula $\rho_k$ such that

$$\mathrm{PA} \vdash \forall \mathbf{x} \ \ (\rho_k(\mathbf{x}) \ \leftrightarrow \ \neg \mathrm{Pr}_T[\rho_k(\mathbf{x})](\mathbf{e}, \ulcorner \mathcal{G}_k \urcorner, \mathbf{x}) \ ).$$

We rename $\rho_k$ as $\rho_{e,T}$, which is a special symbol for a "Gödel sentence" with respect to PTM, in this paper. (We also rename $\mathcal{G}_k$ as $\mathcal{G}$.)

$\dashv$

### 3.4 The First Incompleteness Theorem of Polynomial-Time Proofs

**Theorem 20.** *Let $T$ be a consistent PT-extension of PA. Let $\rho_{e,T}(\mathbf{a})$ be a Gödel sentence with respect to PTM, where $a \in \mathbb{N}$.*

*For all $e \in \mathbb{N}$ and all $x \in \mathbb{N}$,*

$$\mathrm{PTM}_e(x) \nvdash_T \ \rho_{e,T}(\mathbf{x}).$$

*Proof.* Assuming

$$\exists e \in \mathbb{N} \ \ \exists x \in \mathbb{N} \quad \mathrm{PTM}_e(x) \vdash_T \ \rho_{e,T}(\mathbf{x}), \tag{8}$$

then

$$\exists e \in \mathbb{N} \ \ \exists x \in \mathbb{N} \quad \mathrm{PA} \vdash \ \mathrm{Pr}_T[\rho_{e,T}(\mathbf{x})](\mathbf{e}, \ulcorner \mathcal{G} \urcorner, \mathbf{x}), \tag{9}$$

from Lemma 12.

On the other hand, Eq. (8) implies

$$\exists e \in \mathbb{N} \ \ \exists x \in \mathbb{N} \quad T \vdash \ \rho_{e,T}(\mathbf{x}).$$

According to the property of the Gödel sentence with respect to PTM (Lemma 19), for any $e \in \mathbb{N}$

$$\mathrm{PA} \vdash \forall \mathbf{x} \ \ (\rho_{e,T}(\mathbf{x}) \ \leftrightarrow \ \neg \mathrm{Pr}_T[\rho_{e,T}(\mathbf{x})](\mathbf{e}, \ulcorner \mathcal{G} \urcorner, \mathbf{x}) \ ).$$

Therefore,

$$\exists e \in \mathbb{N} \ \ \exists x \in \mathbb{N} \quad T \vdash \ \neg \mathrm{Pr}_T[\rho_{e,T}(\mathbf{x})](\mathbf{e}, \ulcorner \mathcal{G} \urcorner, \mathbf{x}). \tag{10}$$

Since $T$ is a consistent PT-extension of PA, Eq. (9) contradicts Eq. (10).

Thus,

$$\mathrm{PTM}_e(x) \nvdash_T \ \rho_{e,T}(\mathbf{x}).$$

$\dashv$

### 3.5 The Second Incompleteness Theorem of Polynomial-Time Proofs

**Theorem 21.** *Let $T$ be a consistent PT-extension of PA. For any $e \in \mathbb{N}$ and any set of formulas $\Psi \equiv \{\psi(\mathbf{a}) \mid a \in \mathbb{N}\}$, there exists $e^* \in \mathbb{N}$ such that for any $x \in \mathbb{N}$*

$$\mathrm{PTM}_e(x) \nvdash_T \ \neg \mathrm{Pr}_T[\psi(\mathbf{x})](\mathbf{e}^*, \ulcorner \Psi \urcorner, \mathbf{x}). \tag{11}$$

*Proof.* Let $\mathcal{G} \equiv \{\rho_{e,T}(\mathbf{a}) \mid a \in \mathbb{N}\}$ be a set of Gödel sentences with respect to PTM. Let $\mathcal{G}^+ \equiv \{\mathrm{Pr}_T[\rho_{e,T}(\mathbf{a})](\mathbf{e}, \ulcorner \mathcal{G} \urcorner, \mathbf{a}) \mid a \in \mathbb{N}\}$, $\mathcal{G}^{++} \equiv \{\neg \rho_{e,T}(\mathbf{a}) \mid a \in \mathbb{N}\}$, and $\mathcal{G}^{+++} \equiv \{\rho_{e,T}(\mathbf{a}) \wedge \neg \rho_{e,T}(\mathbf{a}) \mid a \in \mathbb{N}\}$.

For any $e \in \mathbb{N}$, there exist $e^+ \in \mathbb{N}$, $e^{++} \in \mathbb{N}$ and $e^{+++} \in \mathbb{N}$ such that

$$\begin{aligned}
\mathrm{PA} \vdash \ \forall \mathbf{x} \ \ ( \quad & \mathrm{Pr}_T[\rho_{e,T}(\mathbf{x})](\mathbf{e}, \ulcorner \mathcal{G} \urcorner, \mathbf{x}) \\
\rightarrow \quad & \mathrm{Pr}_T[\mathrm{Pr}_T[\rho_{e,T}(\mathbf{x})](\mathbf{e}, \ulcorner \mathcal{G} \urcorner, \mathbf{x})](\mathbf{e}^+, \ulcorner \mathcal{G}^+ \urcorner, \mathbf{x}) \quad \text{(by Lemma 16)} \\
\rightarrow \quad & \mathrm{Pr}_T[\neg \rho_{e,T}(\mathbf{x})](\mathbf{e}^{++}, \ulcorner \mathcal{G}^{++} \urcorner, \mathbf{x}) \quad \text{(by Lemma 19 and Corollary 14)} \\
\rightarrow \quad & \mathrm{Pr}_T[\rho_{e,T}(\mathbf{x})](\mathbf{e}, \ulcorner \mathcal{G} \urcorner, \mathbf{x}) \wedge \ \mathrm{Pr}_T[\neg \rho_{e,T}(\mathbf{x})](\mathbf{e}^{++}, \ulcorner \mathcal{G}^{++} \urcorner, \mathbf{x}) \\
\rightarrow \quad & \mathrm{Pr}_T[\rho_{e,T}(\mathbf{x}) \wedge \neg \rho_{e,T}(\mathbf{x})](\mathbf{e}^{+++}, \ulcorner \mathcal{G}^{+++} \urcorner, \mathbf{x}) \ \ ). \quad \text{(by Corollary 15)}
\end{aligned} \tag{12}$$

For any formula family $\Psi \equiv \{\psi(\mathbf{a}) \mid a \in \mathbb{N}\}$,

$$\mathrm{PA} \vdash \ \forall \mathbf{x} \ ( \ \rho_{e,T}(\mathbf{x}) \wedge \neg \rho_{e,T}(\mathbf{x}) \ \rightarrow \ \psi(\mathbf{x}) \ ). \tag{13}$$

Hence by Corollary 14, for any $e^{+++} \in \mathbb{N}$, there exists $e^* \in \mathbb{N}$ such that

$$\mathrm{PA} \vdash \ \forall \mathbf{x} \ ( \ \mathrm{Pr}_T[\rho_{e,T}(\mathbf{x}) \wedge \neg\rho_{e,T}(\mathbf{x})](\mathbf{e}^{+++}, \ulcorner\mathcal{G}^{+++}\urcorner, \mathbf{x}) \ \rightarrow \ \mathrm{Pr}_T[\psi(\mathbf{x})](\mathbf{e}^*, \ulcorner\Psi\urcorner, \mathbf{x}) \ ). \tag{14}$$

Therefore, for any $e \in \mathbb{N}$, there exists $e^* \in \mathbb{N}$ such that

$$\mathrm{PA} \vdash \ \forall \mathbf{x} \ \ ( \ \mathrm{Pr}_T[\rho_{e,T}(\mathbf{x})](\mathbf{e}, \ulcorner\mathcal{G}\urcorner, \mathbf{x}) \rightarrow \ \mathrm{Pr}_T[\psi(\mathbf{x})](\mathbf{e}^*, \ulcorner\Psi\urcorner, \mathbf{x}) \ ).$$

That is, for any $e \in \mathbb{N}$, there exists $e^* \in \mathbb{N}$ such that

$$\mathrm{PA} \vdash \ \forall \mathbf{x} \ \ ( \ \neg\mathrm{Pr}_T[\psi(\mathbf{x})](\mathbf{e}^*, \ulcorner\Psi\urcorner, \mathbf{x}) \ \ \rightarrow \ \neg\mathrm{Pr}_T[\rho_{e,T}(\mathbf{x})](\mathbf{e}, \ulcorner\mathcal{G}\urcorner, \mathbf{x})).$$

Since $\rho_{e,T}(\mathbf{x})$ is a "Gödel sentence" with respect to PTM, from Lemma 19,

$$\mathrm{PA} \vdash \forall \mathbf{x} \ \ (\rho_{e,T}(\mathbf{x}) \ \leftrightarrow \ \neg\mathrm{Pr}_T[\rho_{e,T}(\mathbf{x})](\mathbf{e}, \ulcorner\mathcal{G}\urcorner, \mathbf{x}) \ ).$$

Hence, for any $e \in \mathbb{N}$, there exists $e^* \in \mathbb{N}$ such that

$$\mathrm{PA} \vdash \ \forall \mathbf{x} \ \ ( \ \neg\mathrm{Pr}_T[\psi(\mathbf{x})](\mathbf{e}^*, \ulcorner\Psi\urcorner, \mathbf{x}) \ \rightarrow \ \rho_{e,T}(\mathbf{x}) \ ). \tag{15}$$

We now assume that there exist $e \in \mathbb{N}$ and a formula set $\Psi$ such that

$$\forall e^* \in \mathbb{N} \ \ \exists x \in \mathbb{N} \quad \mathrm{PTM}_e(x) \vdash_T \ \neg\mathrm{Pr}_T[\psi(\mathbf{x})](\mathbf{e}^*, \ulcorner\Psi\urcorner, \mathbf{x}). \tag{16}$$

Then, PTM $\mathrm{U_{PTM}}(e', \cdot)$ is constructed using PTM $\mathrm{U_{PTM}}(e, \cdot)$ as follows:

- (Input: ) $(p, \#\mathcal{G}', a) \in \mathbb{N}^3$, where $\mathcal{G}' \equiv \{\rho_{e',T}(\mathbf{a}) \mid a \in \mathbb{N}\}$.
- (Output: ) Gödel number of a proof tree of $\rho_{e',T}(\mathbf{a})$ or 0.
- First, read its own code, $e' \in \mathbb{N}$, via the recursion theorem (Proposition 17).
- Syntactically check whether the input has the form of $(p, \#\mathcal{G}', a)$ and $\#\mathcal{G}' = \#\{\rho_{e',T}(\mathbf{a}) \mid a \in \mathbb{N}\}$. If it is not correct, output 0. Otherwise, go to the next step.
- Find a proof, $\pi$, of formula

  $$\forall \mathbf{x} \ \ ( \ \neg\mathrm{Pr}_T[\psi(\mathbf{x})](\mathbf{e}^{*\prime}, \ulcorner\Psi\urcorner, \mathbf{x}) \ \rightarrow \ \rho_{e',T}(\mathbf{x}) \ ),$$

  where there exists $e^{*\prime} \in \mathbb{N}$ such that a proof of the formula exists, according to Eq. (15). Here, the size of $\pi$ is constant in $|a|$.
- Simulate $\mathrm{U_{PTM}}(e, (p, \#\Phi[e^{*\prime}], a))$, and check whether its output is the Gödel number of a valid proof tree of $\neg\mathrm{Pr}_T[\psi(\mathbf{a})](\mathbf{e}^{*\prime}, \ulcorner\Psi\urcorner, \mathbf{a})$ by using $\mathrm{U_{PTM}}(v_T, \cdot)$, where $\Phi[e^{*\prime}] \equiv \{\neg\mathrm{Pr}_T[\psi(\mathbf{a})](\mathbf{e}^{*\prime}, \ulcorner\Psi\urcorner, \mathbf{a}) \mid a \in \mathbb{N}\}$.
- If it is not a valid proof tree, then output 0.
- If it is a valid proof tree (say $\theta$), using proofs, $\theta$ and $\pi$, construct the following proof tree of $\rho_{e',T}(\mathbf{a})$:

  $$< \rho_{e',T}(\mathbf{a}), \text{Modus Ponens} > \ [ \ \theta, \ < \neg\mathrm{Pr}_T[\psi(\mathbf{a})](\mathbf{e}^{*\prime}, \ulcorner\Psi\urcorner, \mathbf{a}) \ \rightarrow \ \rho_{e',T}(\mathbf{a}), \text{Modus Ponens} >$$

  $$[ \ \pi, \forall \mathbf{x} \ \ ( \ \neg\mathrm{Pr}_T[\psi(\mathbf{x})](\mathbf{e}^{*\prime}, \ulcorner\Psi\urcorner, \mathbf{x}) \ \rightarrow \ \rho_{e',T}(\mathbf{x}) \ ) \ \ \rightarrow \ \ ( \ \neg\mathrm{Pr}_T[\psi(\mathbf{a})](\mathbf{e}^{*\prime}, \ulcorner\Psi\urcorner, \mathbf{a}) \ \rightarrow \ \rho_{e',T}(\mathbf{a}) \ ) \, ] \, ].$$

  Output the Gödel number of the proof tree.

The running time of $\mathrm{U_{PTM}}(e', \cdot)$ is that of $\mathrm{U_{PTM}}(e, \cdot)$ plus polynomial-time in $|a|$.

Since we assume that Eq. (16) holds, $\mathrm{U_{PTM}}(e', (p, \#\mathcal{G}', a))$ outputs the Gödel number of a valid proof tree of $\rho_{e',T}(\mathbf{a})$. Thus,

$$\exists x \in \mathbb{N} \quad \mathrm{PTM}_{e'}(x) \vdash_T \ \rho_{e',T}(\mathbf{x}).$$

This contradicts Theorem 20. Therefore, Eq. (16) does not hold. That is, for any $e \in \mathbb{N}$ and any $\Psi$, there exists $e^* \in \mathbb{N}$ such that for any $x \in \mathbb{N}$

$$\mathrm{PTM}_e(x) \nvdash_T \ \neg\mathrm{Pr}_T[\psi(\mathbf{x})](\mathbf{e}^*, \ulcorner\Psi\urcorner, \mathbf{x}).$$

⊣

## 4 Polynomial-Time Decisions

In order to prove the (resource bounded) unprovability of $\overline{\mathrm{P}\neq\mathrm{NP}}$, this section introduces our formalization of a decision made by a polynomial-time Turing machine (polynomial-time decision: PTD).

### 4.1 Polynomial-Time Decisions

This section introduces the formalization of a decision made by a polynomial-time Turing machine (polynomial-time decision: PTD).

Let $\Phi \equiv \{\varphi(\mathbf{a}) \mid a \in \mathbb{N}\}$ be a set of an infinite number of formulas in PA. If $\mathrm{U_{PTM}}(e,(d,\#\Phi,a))$ accepts and $\mathfrak{N} \models \varphi(\mathbf{a})$ (i.e., $\varphi(\mathbf{a})$ is true in the standard model of natural numbers), we then denote

$$\mathrm{PTM}_e^{\Phi}(a) \triangleright \varphi(\mathbf{a}).$$

(This can be interpreted as "$\mathrm{U_{PTM}}(e,\cdot)$ *correctly accepts* $\varphi(\mathbf{a})$.) Here $d$ denotes a natural number (e.g., 1), which indicates $\mathrm{U_{PTM}}(e,\cdot)$ that the output target is a decision on the formula's truth. In other words,

$$\begin{aligned}&\mathrm{PTM}_e^{\Phi}(a) \triangleright \varphi(\mathbf{a})\\ &\quad\Leftrightarrow\quad \mathrm{U_{PTM}}(e,(d,\#\Phi,a))\text{accepts} \ \wedge \ \mathfrak{N} \models \varphi(\mathbf{a}).\end{aligned}$$

If $\mathrm{U_{PTM}}(e,(d,\#\Phi,a))$ rejects and $\mathfrak{N} \models \neg\varphi(\mathbf{a})$ (i.e., $\varphi(\mathbf{a})$ is false in the standard model of natural numbers), we then denote $\mathrm{PTM}_e^{\Phi}(a) \triangleright \neg\varphi(\mathbf{a})$. (This can be interpreted as "$\mathrm{U_{PTM}}(e,\cdot)$ *correctly rejects* $\varphi(\mathbf{a})$.) In other words,

$$\begin{aligned}&\mathrm{PTM}_e^{\Phi}(a) \triangleright \neg\varphi(\mathbf{a})\\ &\quad\Leftrightarrow\quad \mathrm{U_{PTM}}(e,(d,\#\Phi,a))\text{rejects} \ \wedge \ \mathfrak{N} \models \neg\varphi(\mathbf{a}).\end{aligned}$$

Here note that $\mathrm{PTM}_e^{\Phi}(a) \triangleright \neg\varphi(\mathbf{a})$ is different from $\mathrm{PTM}_e^{\Omega}(a) \triangleright \neg\varphi(\mathbf{a})$, where $\Omega \equiv \{\neg\varphi(\mathbf{a}) \mid a \in \mathbb{N}\}$. [3]

In addition, we use the notation $\mathrm{PTM}_e^{\Phi}(a) \not\triangleright \varphi(\mathbf{a})$ if and only if $\neg($ $\mathrm{PTM}_e^{\Phi}(a) \triangleright \varphi(\mathbf{a})$ $)$.

[3] In the notation of polynomial-time proofs,

$$\mathrm{PTM}_e(a) \vdash \varphi(\mathbf{a}),$$

we omit $\Phi \equiv \{\varphi(\mathbf{a}) \mid a \in \mathbb{N}\}$ in a place of $\mathrm{PTM}_e(a) \vdash \varphi(\mathbf{a})$ (e.g., the upper right position of $\mathrm{PTM}_e$), since $\Phi$ is uniquely determined by the object of the proof, $\varphi(\mathbf{a})$. However, in polynomial-time decisions, we have two different types of decisions as follows, as described above:

$$\begin{aligned}&\mathrm{PTM}_e^{\Phi}(a) \triangleright \varphi(\mathbf{a}),\\ &\mathrm{PTM}_e^{\Omega}(a) \triangleright \varphi(\mathbf{a}),\end{aligned}$$

where $\Omega \equiv \{\neg\varphi(\mathbf{a}) \mid a \in \mathbb{N}\}$. In the former notation, $\varphi(\mathbf{a})$ is *correctly accepted*, while, in the latter notation, $\neg\varphi(\mathbf{a})$ is *correctly rejected*. Therefore, in the notation of polynomial-time decisions,

$$\mathrm{PTM}_e^{\Phi}(a) \triangleright \varphi(\mathbf{a}),$$

we *cannot* omit $\Phi$ in the upper right position of $\mathrm{PTM}_e$.

We now introduce a relaxed notion of $\mathrm{PTM}_e^{\Phi}(a) \triangleright \varphi(\mathbf{a})$. We denote

$$\mathrm{PTM}_e^{\Phi}(a) \triangleright_v \varphi(\mathbf{a})$$

if and only if

$$\mathrm{U}_{\mathrm{PTM}}(e, (d, \#\Phi, a))\text{accepts} \ \wedge \ \mathrm{U}(v, (d, \#\Phi, a))\text{accepts}.$$

**Lemma 22.** *Let $\Phi \equiv \{\varphi(\mathbf{a}) \mid a \in \mathbb{N}\}$ be a set of an infinite number of $\Delta_1$-formulas in PA, and $\varphi$ represent a primitive recursive relation $R_\Phi$, where there exists a Turing machine $\mathrm{U}(v_\varphi, \cdot)$ such that, for every $a \in \mathbb{N}$,*

$$a \in R_\Phi \quad \Leftrightarrow \quad \mathrm{U}(v_\varphi, (d, \#\Phi, a))\textit{accepts}.$$

*Then,*

$$\mathrm{PTM}_e^{\Phi}(a) \triangleright_{v_\varphi} \varphi(\mathbf{a}) \quad \Leftrightarrow \quad \mathrm{PTM}_e^{\Phi}(a) \triangleright \varphi(\mathbf{a}),$$

*and*

$$\mathrm{PTM}_e^{\Phi}(a) \triangleright_{v_\varphi} \neg\varphi(\mathbf{a}) \quad \Leftrightarrow \quad \mathrm{PTM}_e^{\Phi}(a) \triangleright \neg\varphi(\mathbf{a}).$$

*Proof.* Since $R_\Phi$ is a relation that formula $\varphi$ represents, for every $a \in \mathbb{N}$

$$\mathrm{U}(v_\varphi, (d, \#\Phi, a))\text{accepts} \quad \Rightarrow \quad a \in R_\Phi \quad \Rightarrow \quad \mathrm{PA} \vdash \varphi(\mathbf{a}),$$

$$\mathrm{U}(v_\varphi, (d, \#\Phi, a))\text{rejects} \quad \Rightarrow \quad a \notin R_\Phi \quad \Rightarrow \quad \mathrm{PA} \vdash \neg\varphi(\mathbf{a}).$$

Since $\mathfrak{N}$ is a model of PA, from the soundness of PA, for every $a \in \mathbb{N}$

$$\mathrm{U}(v_\varphi, (d, \#\Phi, a))\text{accepts} \quad \Rightarrow \quad \mathfrak{N} \models \varphi(\mathbf{a}),$$

$$\mathrm{U}(v_\varphi, (d, \#\Phi, a))\text{rejects} \quad \Rightarrow \quad \mathfrak{N} \models \neg\varphi(\mathbf{a}).$$

$\dashv$

### 4.2 Formalization of Polynomial-Time Decisions

A formula to represent the relation on polynomial-time decisions, $\mathrm{PTM}_e^{\Phi}(a) \triangleright \varphi(\mathbf{a})$, is obtained, in a manner similar to that shown in Section 2.10.

Let $\Phi \equiv \{\varphi(\mathbf{a}) \mid a \in \mathbb{N}\}$ be a set of an infinite number of $\Delta_1$-formulas in PA. Let formula

$$\text{PTM-Acc}(\mathbf{e}, \ulcorner\Phi\urcorner, \mathbf{a})$$

polynomial-time represent

$$\mathrm{U}_{\mathrm{PTM}}(e, (d, \#\Phi, a))\text{accepts}$$

over natural numbers, $(e, \#\Phi, a)$.

Let formula

$$\mathrm{Acc}(\mathbf{e}, \ulcorner\Phi\urcorner, \mathbf{a})$$

represent

$$\mathrm{U}(e, (d, \#\Phi, a))\text{accepts}$$

over natural numbers, $(e, \#\Phi, a)$. Here, if $a = (a_1, \ldots, a_k)$, we then denote

$$\mathrm{Acc}(\mathbf{e}, \ulcorner\Phi\urcorner, \mathbf{a}_1, \ldots, \mathbf{a}_k)$$

to represent

$$\mathrm{U}(e,(d,\#\Phi,(a_1,\ldots,a_k)))\text{accepts}.$$

The method of constructing these formulas is the same as that described in Section 2.10.

We then define the following formulas: for all $a \in \mathbb{N}$,

$$\begin{aligned}
\mathrm{CA}_v[\varphi(\mathbf{a})](\mathbf{e},\ulcorner\Phi\urcorner,\mathbf{a}) &\equiv \text{PTM-Acc}(\mathbf{e},\ulcorner\Phi\urcorner,\mathbf{a}) \ \wedge\ \mathrm{Acc}(\mathbf{v},\ulcorner\Phi\urcorner,\mathbf{a}),\\
\mathrm{CR}_v[\varphi(\mathbf{a})](\mathbf{e},\ulcorner\Phi\urcorner,\mathbf{a}) &\equiv \neg\text{PTM-Acc}(\mathbf{e},\ulcorner\Phi\urcorner,\mathbf{a}) \ \wedge\ \neg\mathrm{Acc}(\mathbf{v},\ulcorner\Phi\urcorner,\mathbf{a}),\\
\mathrm{CD}_v[\varphi(\mathbf{a})](\mathbf{e},\ulcorner\Phi\urcorner,\mathbf{a}) &\equiv \mathrm{CA}_v[\varphi(\mathbf{a})](\mathbf{e},\ulcorner\Phi\urcorner,\mathbf{a}) \ \vee\ \mathrm{CR}_v[\varphi(\mathbf{a})](\mathbf{e},\ulcorner\Phi\urcorner,\mathbf{a}),
\end{aligned}$$

(Here, CA, CR, and CD stand for 'correctly accept', 'correctly reject', and 'correctly decide', respectively.)

We also define the following formulas: for all $a \in \mathbb{N}$,

$$\begin{aligned}
\mathrm{CA}[\varphi(\mathbf{a})](\mathbf{e},\ulcorner\Phi\urcorner,\mathbf{a}) &\equiv \text{PTM-Acc}(\mathbf{e},\ulcorner\Phi\urcorner,\mathbf{a}) \ \wedge\ \varphi(\mathbf{a}),\\
\mathrm{CR}[\varphi(\mathbf{a})](\mathbf{e},\ulcorner\Phi\urcorner,\mathbf{a}) &\equiv \neg\text{PTM-Acc}(\mathbf{e},\ulcorner\Phi\urcorner,\mathbf{a}) \ \wedge\ \neg\varphi(\mathbf{a}),\\
\mathrm{CD}[\varphi(\mathbf{a})](\mathbf{e},\ulcorner\Phi\urcorner,\mathbf{a}) &\equiv \mathrm{CA}[\varphi(\mathbf{a})](\mathbf{e},\ulcorner\Phi\urcorner,\mathbf{a}) \ \vee\ \mathrm{CR}[\varphi(\mathbf{a})](\mathbf{e},\ulcorner\Phi\urcorner,\mathbf{a}).
\end{aligned}$$

**Lemma 23.** *Let $\Omega \equiv \{\omega(\mathbf{a}) \mid a \in \mathbb{N}\}$ be a set of an infinite number of $\Delta_1$-formulas in PA.*

*Then, there exists a primitive recursive function $f$ such that*

$$\begin{aligned}
&\exists y < f(\#\Omega, \mathrm{Size}_\Omega(a)) \quad \mathrm{U}_{\mathrm{PTM}}(v_{\mathrm{PA}},(\#\omega(\mathbf{a}),y))\textit{accepts} \quad\Leftrightarrow\quad \mathrm{PA}\vdash\ \omega(\mathbf{a}), \quad \textit{and}\\
&\exists z < f(\#\Omega, \mathrm{Size}_\Omega(a)) \quad \mathrm{U}_{\mathrm{PTM}}(v_{\mathrm{PA}},(\#\neg\omega(\mathbf{a}),z))\textit{accepts} \quad\Leftrightarrow\quad \mathrm{PA}\vdash\ \neg\omega(\mathbf{a}).
\end{aligned}$$

*Proof.* Since $\omega(\mathbf{a})$ is a $\Delta_1$-formula, there exists a TM $\mathrm{U}(e_0,\cdot)$ such that

$$\forall a \in \mathbb{N}\ \ (\ \mathrm{PA}\vdash\ \omega(\mathbf{a})\ \Rightarrow\ \mathrm{TM}^{\Omega}_{e_0}(a)\vdash_{\mathrm{PA}}\ \omega(\mathbf{a})\ \ \vee\ \ \mathrm{PA}\vdash\ \neg\omega(\mathbf{a})\ \Rightarrow\ \mathrm{TM}^{\Omega'}_{e_0}(a)\vdash_{\mathrm{PA}}\ \neg\omega(\mathbf{a})\ ),$$

where $\Omega' \equiv \{\neg\omega(\mathbf{a}) \mid a \in \mathbb{N}\}$, and $\mathrm{Size}_\Omega(a) = \mathrm{Size}_{\Omega'}(a)$ for all $a \in \mathbb{N}$.

Therefore, there exists another TM $\mathrm{U}(e_1,\cdot)$ such that

$$\forall a \in \mathbb{N}\ \ \mathrm{U}(e_1,(\#\Omega,a)) = |\pi|\ \wedge\ (\ \mathrm{U}(e_0,(p,\#\Omega,a)) = \#\pi\ \vee \mathrm{U}(e_0,(p,\#\Omega',a)) = \#\pi.$$

(That is, $\pi$ is a proof tree of $\omega(\mathbf{a})$ or $\neg\omega(\mathbf{a})$, generated by $\mathrm{U}(e_0,\cdot)$.)

Hence, there exists a TM $\mathrm{U}(e_2,\cdot)$ such that

$$\begin{aligned}
\mathrm{U}(e_2,(\#\Omega,n))\ &= \max\{\ \ |\pi(a)|\ \ |\\
&a \in \mathbb{N}\ \ \wedge\ \ n = \mathrm{Size}_\Omega(a)\ \ \wedge\ \ (\mathrm{U}(e_0,(p,\#\Omega,a)) = \#\pi\ \ \vee\ \ \mathrm{U}(e_0,(p,\#\Omega',a)) = \#\pi\ )\ \ \}.
\end{aligned}$$

(That is, $\mathrm{U}(e_2,(\#\Omega,n))$ computes the maximum length of proofs that $\mathrm{U}(e_0,\cdot)$ outputs where the input size is $n$.)

Thus, there exists the above-mentioned primitive recursive function $f$ that is computed by TM $\mathrm{U}(e_2,\cdot)$.

⊣

**Definition 24.** *Let $\Omega \equiv \{\omega(\mathbf{a}) \mid a \in \mathbb{N}\}$ and $\Phi \equiv \{\varphi(\mathbf{a}) \mid a \in \mathbb{N}\}$ be sets of an infinite number of $\Delta_1$-formulas in PA.*

*Let $\mathrm{U}(v^A_\Omega,\cdot)$ be a TM as follows:*

- *(Input: )* $(d, \#\Phi, a)$
- *(Output: )* *accept or reject*
- *Let* $\Phi_1 \equiv \{\varphi_1(\mathbf{a}) \mid a \in \mathbb{N}\}$ *and* $\Phi_2 \equiv \{\varphi_2(\mathbf{a}) \mid a \in \mathbb{N}\}$. *If* $\varphi(\mathbf{a}) \equiv \varphi_1(\mathbf{a}) \wedge \varphi_2(\mathbf{a})$, *then let* $\Phi \equiv \{\varphi_1(\mathbf{a}) \wedge \varphi_2(\mathbf{a}) \mid a \in \mathbb{N}^2\}$, *where* $\mathrm{Size}_\Phi(a) = \mathrm{Size}_{\Phi_1}(a) + \mathrm{Size}_{\Phi_2}(a)$. *Then, simulate* $\mathrm{U}(v_\Omega^A, (d, \#\Phi_1, a))$ *and* $\mathrm{U}(v_\Omega^A, (d, \#\Phi_2, a))$. *(Here, whether* $\varphi(\mathbf{a})$ *is the form of* $\varphi_1(\mathbf{a}) \wedge \varphi_2(\mathbf{a})$ *is syntactically checked by some rule, and is uniquely decided. For example, search a formula from left to right and syntactically check the form based on the leftmost* $\wedge$*, and if it is not the form then move to the right direction to find another* $\wedge$*, etc.)*
  *Accept if and only if both of them accept.*
- *If theorem in PA,*
$$\mathrm{PA} \vdash \ \forall \mathbf{x}(\ \psi(\mathbf{x}) \ \rightarrow \ \varphi(\mathbf{x})\ ),$$
  *is installed in* $\mathrm{U}(v_\Omega^A, \cdot)$, *then simulate* $\mathrm{U}(v_\Omega^A, (d, \#\Psi, a))$, *where* $\Psi \equiv \{\psi(\mathbf{a}) \mid a \in \mathbb{N}\}$ *and* $\mathrm{Size}_\Phi(a) = \mathrm{Size}_\Psi(a) + c$ *(c: constant).*
  *Accept if and only if* $\mathrm{U}(v_\Omega^A, (d, \#\Psi, a))$ *accepts.*
  *A finite number of the theorems explicitly shown in this paper are installed in* $\mathrm{U}(v_\Omega^A, \cdot)$.
- *Let* $\Psi \equiv \{\psi(\mathbf{a}) \mid a \in \mathbb{N}\}$ *be a set of an infinite number of* $\Delta_1$*-formulas in PA. If* $\Phi \equiv \{\mathrm{CA}_\Omega[\psi(\mathbf{a})](\mathbf{e}, \ulcorner\Psi\urcorner, \mathbf{a}) \mid a \in \mathbb{N}\}$, *then simulate* $\mathrm{U}_{\mathrm{PTM}}(e, (d, \#\Psi, a))$ *and* $\mathrm{U}(v_\Omega^A, (d, \#\Psi, a))$. *Here* $\mathrm{Size}_\Phi(a) = 2 \cdot \mathrm{Size}_\Psi(a)$.
  *Accept if and only if both of them accept.*
- *Unless the above-mentioned cases occur, check (by exhaustive search for* $y < f(\#\Omega, \mathrm{Size}_\Phi(a))$*) whether*
$$\exists y < f(\#\Omega, \mathrm{Size}_\Phi(a)) \quad \mathrm{U}_{\mathrm{PTM}}(v_{\mathrm{PA}}, (\#\varphi(\mathbf{a}), y)) \textit{accepts}, \tag{17}$$
  *where* $\mathrm{U}_{\mathrm{PTM}}(v_{\mathrm{PA}}, \cdot)$ *is defined in Section 2.7, and* $f$ *is a primitive recursive function defined in Lemma 23.*
  *Accept if and only if Eq. (17) holds.*

*Let* $\mathrm{U}(v_\Omega^R, \cdot)$ *be a TM as follows:*

- *(Input: )* $(d, \#\Phi, a)$
- *(Output: )* *accept or reject*
- *Let* $\Phi_1 \equiv \{\varphi_1(\mathbf{a}) \mid a \in \mathbb{N}\}$ *and* $\Phi_2 \equiv \{\varphi_2(\mathbf{a}) \mid a \in \mathbb{N}\}$. *If* $\varphi(\mathbf{a}) \equiv \varphi_1(\mathbf{a}) \vee \varphi_2(\mathbf{a})$, *then let* $\Phi \equiv \{\varphi_1(\mathbf{a}) \vee \varphi_2(\mathbf{a}) \mid a \in \mathbb{N}^2\}$, *where* $\mathrm{Size}_\Phi(a) = \mathrm{Size}_{\Phi_1}(a) + \mathrm{Size}_{\Phi_2}(a)$. *Then, simulate* $\mathrm{U}(v_\Omega^R, (d, \#\Phi_1, a))$ *and* $\mathrm{U}(v_\Omega^R, (d, \#\Phi_2, a))$. *(Here, whether* $\varphi(\mathbf{a})$ *is the form of* $\varphi_1(\mathbf{a}) \vee \varphi_2(\mathbf{a})$ *is syntactically checked by some rule, and is uniquely decided.)*
  *Reject if and only if both of them reject.*
- *If theorem in PA,*
$$\mathrm{PA} \vdash \ \forall \mathbf{x}(\ \neg\psi(\mathbf{x}) \ \rightarrow \ \neg\varphi(\mathbf{x})\ ),$$
  *is installed in* $\mathrm{U}(v_\Omega^R, \cdot)$, *then simulate* $\mathrm{U}(v_\Omega^R, (d, \#\Psi, a))$, *where* $\Psi \equiv \{\psi(\mathbf{a}) \mid a \in \mathbb{N}\}$ *and* $\mathrm{Size}_\Phi(a) = \mathrm{Size}_\Psi(a) + c$ *(c: constant).*
  *Reject if and only if* $\mathrm{U}(v_\Omega^R, (d, \#\Psi, a))$ *rejects.*
  *A finite number of the theorems explicitly shown in this paper are installed in* $\mathrm{U}(v_\Omega^R, \cdot)$.
- *Let* $\Psi \equiv \{\psi(\mathbf{a}) \mid a \in \mathbb{N}\}$ *be a set of an infinite number of* $\Delta_1$*-formulas in PA. If* $\Phi \equiv \{\neg\mathrm{CR}_\Omega[\psi(\mathbf{a})](\mathbf{e}, \ulcorner\Psi\urcorner, \mathbf{a}) \mid a \in \mathbb{N}\}$, *then simulate* $\mathrm{U}_{\mathrm{PTM}}(e, (d, \#\Psi, a))$ *and* $\mathrm{U}(v_\Omega^R, (d, \#\Psi, a))$. *Here* $\mathrm{Size}_\Phi(a) = 2 \cdot \mathrm{Size}_\Psi(a)$.
  *Reject if and only if both of them reject'.*

– *Unless the above-mentioned cases occur, check (by exhaustive search for* $y < f(\#\Omega, \mathrm{Size}_{\Phi}(a))$*) whether*

$$\exists y < f(\#\Omega, \mathrm{Size}_{\Phi}(a)) \quad \mathrm{U}_{\mathrm{PTM}}(v_{\mathrm{PA}}, (\#\neg\varphi(\mathbf{a}), y))\, accepts, \tag{18}$$

*where* $\mathrm{U}_{\mathrm{PTM}}(v_{\mathrm{PA}}, \cdot)$ *is defined in Section 2.7, and* $f$ *is a primitive recursive function defined in Lemma 23.*
*Reject if and only if Eq. (18) holds.*

If $\mathrm{U}(v^A_\Omega, \cdot)$ and $\mathrm{U}(v^R_\Omega, \cdot)$ are TMs defined in Definition 24, we simply denote by

$$\begin{aligned}
\mathrm{CA}_\Omega[\varphi(\mathbf{a})](\mathbf{e}, \ulcorner\Phi\urcorner, \mathbf{a}) &\equiv \mathrm{CA}_{\mathbf{v}^A_\Omega}[\varphi(\mathbf{a})](\mathbf{e}, \ulcorner\Phi\urcorner, \mathbf{a}),\\
\mathrm{CR}_\Omega[\varphi(\mathbf{a})](\mathbf{e}, \ulcorner\Phi\urcorner, \mathbf{a}) &\equiv \mathrm{CR}_{\mathbf{v}^R_\Omega}[\varphi(\mathbf{a})](\mathbf{e}, \ulcorner\Phi\urcorner, \mathbf{a}),\\
\mathrm{CD}_\Omega[\varphi(\mathbf{a})](\mathbf{e}, \ulcorner\Phi\urcorner, \mathbf{a}) &\equiv \mathrm{CA}_{\mathbf{v}^A_\Omega}[\varphi(\mathbf{a})](\mathbf{e}, \ulcorner\Phi\urcorner, \mathbf{a}) \ \vee\ \mathrm{CR}_{\mathbf{v}^R_\Omega}[\varphi(\mathbf{a})](\mathbf{e}, \ulcorner\Phi\urcorner, \mathbf{a}),
\end{aligned}$$

**Definition 25.** *We say* $\mathrm{U}(v, \cdot)$ *"soundly accepts" if, for any* $\Phi \equiv \{\varphi(\mathbf{a}) \mid a \in \mathbb{N}\}$*, for all* $a \in \mathbb{N}$*,*

$$\mathrm{U}(v, (d, \#\Phi, a))\, accepts \quad \Rightarrow \quad \mathfrak{N} \models \varphi(a). \tag{19}$$

*We say* $\mathrm{U}(v, \cdot)$ *"soundly rejects" if, for any* $\Phi \equiv \{\varphi(\mathbf{a}) \mid a \in \mathbb{N}\}$*, for any* $a \in \mathbb{N}$*,*

$$\mathrm{U}(v, (d, \#\Phi, a))\, rejects \quad \Rightarrow \quad \mathfrak{N} \models \neg\varphi(a). \tag{20}$$

*The following lemma is obtained from Definitions 24 and 25, and Lemma 23.*

**Lemma 26.** *Let* $\mathrm{U}(v^A_\Phi, \cdot)$ *soundly accept, and* $\mathrm{U}(v^R_\Phi, \cdot)$ *soundly reject.*
*For all* $a \in \mathbb{N}$*,*

$$\mathrm{U}(v^A_\Phi, (d, \#\Phi, a))\, accepts \quad \Leftrightarrow \quad \mathfrak{N} \models \varphi(a). \tag{21}$$

$$\mathrm{U}(v^R_\Phi, (d, \#\Phi, a))\, rejects \quad \Leftrightarrow \quad \mathfrak{N} \models \neg\varphi(a). \tag{22}$$

$$\mathrm{PA} \vdash \ \forall \mathbf{x} \ \ (\ \mathrm{Acc}(\mathbf{v}^A_\Omega, \ulcorner\Phi\urcorner, (\mathbf{x}, \mathbf{x})) \ \leftrightarrow\ \mathrm{Acc}(\mathbf{v}^A_\Omega, \ulcorner\Phi_1\urcorner, \mathbf{x}) \ \wedge\ \mathrm{Acc}(\mathbf{v}^A_\Omega, \ulcorner\Phi_2\urcorner, \mathbf{x}), \tag{23}$$

*where* $\Phi \equiv \{\varphi_1(\mathbf{a}) \wedge \varphi_2(\mathbf{a}) \mid (a, a) \in \mathbb{N}^2\}$*.*

$$\mathrm{PA} \vdash \ \forall \mathbf{x} \ \ (\ \neg\mathrm{Acc}(\mathbf{v}^R_\Omega, \ulcorner\Phi'\urcorner, (\mathbf{x}, \mathbf{x})) \ \leftrightarrow\ \neg\mathrm{Acc}(\mathbf{v}^R_\Omega, \ulcorner\Phi_1\urcorner, \mathbf{x}) \ \wedge\ \neg\mathrm{Acc}(\mathbf{v}^R_\Omega, \ulcorner\Phi_2\urcorner, \mathbf{x}), \tag{24}$$

*where* $\Phi' \equiv \{\varphi_1(\mathbf{a}) \vee \varphi_2(\mathbf{a}) \mid (a, a) \in \mathbb{N}^2\}$*.*

$$\mathrm{PA} \vdash \quad \forall \mathbf{x}\ (\ \ \mathrm{CA}_\Omega[\psi(\mathbf{x})](\mathbf{e}, \ulcorner\Psi\urcorner, \mathbf{x}) \ \leftrightarrow \ \ \mathrm{Acc}(\mathbf{v}^A_\Omega, \ulcorner\mathcal{CA}[e, \Omega]\urcorner, \mathbf{x})\ ), \tag{25}$$

*where* $\Psi \equiv \{\psi(\mathbf{a}) \mid a \in \mathbb{N}\}$ *is a set of an infinite number of* $\Delta_1$*-formulas in PA, and* $\mathcal{CA}[e, \Omega] \equiv \{\mathrm{CA}_\Omega[\psi(\mathbf{a})](\mathbf{e}, \ulcorner\Psi\urcorner, \mathbf{a}) \mid a \in \mathbb{N}\}$*.*

$$\mathrm{PA} \vdash \quad \forall \mathbf{x}\ (\ \ \mathrm{CR}_\Omega[\psi(\mathbf{x})](\mathbf{e}, \ulcorner\Psi\urcorner, \mathbf{x}) \ \leftrightarrow \ \ \neg\mathrm{Acc}(\mathbf{v}^R_\Omega, \ulcorner\mathcal{CR}[e, \Omega]\urcorner, \mathbf{x})\ ), \tag{26}$$

*where* $\mathcal{CR}[e, \Omega] \equiv \{\neg\mathrm{CR}_\Omega[\psi(\mathbf{a})](\mathbf{e}, \ulcorner\Psi\urcorner, \mathbf{a}) \mid a \in \mathbb{N}\}$*.*

**Remark:** If

$$\exists x \in \mathbb{N} \quad \mathfrak{N} \models \ \neg\varphi(\mathbf{x}),$$

then

$$\mathrm{PA} \nvdash \ \forall \mathbf{x} \ \ (\ \mathrm{Acc}(\mathbf{v}^*_\Phi, \ulcorner\Phi\urcorner, \mathbf{x}) \ \rightarrow\ \varphi(\mathbf{x})\ ),$$

since if $\mathrm{PA} \vdash \forall \mathbf{x} \ \ (\ \mathrm{Pr}_{\mathrm{PA}}(\ulcorner\varphi(\mathbf{x})\urcorner) \ \rightarrow\ \varphi(\mathbf{x})\ )$, then $\exists x \in \mathbb{N} \quad \mathrm{PA} \vdash \neg\mathrm{Pr}_{\mathrm{PA}}(\ulcorner\varphi(\mathbf{x})\urcorner)$, which implies $\mathrm{PA} \vdash \ \mathrm{Con(PA)}$ and contradicts the second Gödel Incompleteness Theorem.

**Lemma 27.** *Let $\Phi \equiv \{\varphi(\mathbf{a}) \mid a \in \mathbb{N}\}$ be a set of an infinite number of $\Delta_1$-formulas in PA.*

*For all $e \in \mathbb{N}$, and for all $a \in \mathbb{N}$,*

$$\mathrm{PTM}^{\Phi}_{e}(a) \triangleright_{v^A_\Phi} \varphi(\mathbf{a}) \quad \Leftrightarrow \quad \mathrm{PTM}^{\Phi}_{e}(a) \triangleright \varphi(\mathbf{a}),$$

*For all $e \in \mathbb{N}$, and for all $a \in \mathbb{N}$,*

$$\mathrm{PTM}^{\Phi}_{e}(a) \triangleright_{v^R_\Phi} \neg\varphi(\mathbf{a}) \quad \Leftrightarrow \quad \mathrm{PTM}^{\Phi}_{e}(a) \triangleright \neg\varphi(\mathbf{a}).$$

*Proof.* For all $e \in \mathbb{N}$, and for all $a \in \mathbb{N}$,

$$\mathrm{PTM}^{\Phi}_{e}(a) \triangleright_{v^A_\Phi} \varphi(\mathbf{a}) \quad \Leftrightarrow \quad \mathrm{U}_{\mathrm{PTM}}(e, (d, \#\Phi, a))\text{accepts} \ \wedge \ \mathrm{U}(v^A_\Phi, (d, \#\Phi, a))\text{accepts}.$$

As shown in Eq. (21),

$$\mathrm{U}(v^A_\Phi, (d, \#\Phi, a))\text{accepts} \quad \Leftrightarrow \quad \mathfrak{N} \models \varphi(a).$$

Hence,

$$\begin{aligned} \mathrm{PTM}^{\Phi}_{e}(a) \triangleright_{v^A_\Phi} \varphi(\mathbf{a}) \quad &\Leftrightarrow \quad \mathrm{U}_{\mathrm{PTM}}(e, (d, \#\Phi, a))\text{accepts} \ \wedge \ \mathfrak{N} \models \varphi(a) \\ &\Leftrightarrow \quad \mathrm{PTM}^{\Phi}_{e}(a) \triangleright \varphi(\mathbf{a}). \end{aligned}$$

Similarly, from Eq. (22), we obtain that for all $e \in \mathbb{N}$, and for all $a \in \mathbb{N}$,

$$\mathrm{PTM}^{\Phi}_{e}(a) \triangleright_{v^R_\Phi} \neg\varphi(\mathbf{a}) \quad \Leftrightarrow \quad \mathrm{PTM}^{\Phi}_{e}(a) \triangleright \neg\varphi(\mathbf{a}).$$

$\dashv$

**Lemma 28.** *Let $\Omega \equiv \{\omega(\mathbf{a}) \mid a \in \mathbb{N}\}$ and $\Phi \equiv \{\varphi(\mathbf{a}) \mid a \in \mathbb{N}\}$ be sets of an infinite number of $\Delta_1$-formulas in PA.*

*Then for all $e \in \mathbb{N}$,*

$$\mathrm{PA} \vdash \ \forall \mathbf{x} \ \ ( \ \mathrm{CA}_\Omega[\varphi(\mathbf{x})](\mathbf{e}, \ulcorner\Phi\urcorner, \mathbf{x}) \ \rightarrow \ \mathrm{CA}_\Phi[\varphi(\mathbf{x})](\mathbf{e}, \ulcorner\Phi\urcorner, \mathbf{x}) \ ),$$

$$\mathrm{PA} \vdash \ \forall \mathbf{x} \ \ ( \ \mathrm{CR}_\Omega[\varphi(\mathbf{x})](\mathbf{e}, \ulcorner\Phi\urcorner, \mathbf{x}) \ \rightarrow \ \mathrm{CR}_\Phi[\varphi(\mathbf{x})](\mathbf{e}, \ulcorner\Phi\urcorner, \mathbf{x}) \ ).$$

*Proof.* For all $e \in \mathbb{N}$,

$$\begin{aligned} \mathrm{PA} \vdash \ \forall \mathbf{x} \ ( \quad & \mathrm{CA}_\Omega[\varphi(\mathbf{x})](\mathbf{e}, \ulcorner\Phi\urcorner, \mathbf{x}) \\ \leftrightarrow \quad & \mathrm{PTM\text{-}Acc}(\mathbf{e}, \ulcorner\Phi\urcorner, \mathbf{x}) \ \wedge \ \mathrm{Acc}(\mathbf{v}^A_\Omega, \ulcorner\Phi\urcorner, \mathbf{x}) \\ \rightarrow \quad & \mathrm{PTM\text{-}Acc}(\mathbf{e}, \ulcorner\Phi\urcorner, \mathbf{x}) \ \wedge \ \mathrm{Acc}(\mathbf{v}^A_\Phi, \ulcorner\Phi\urcorner, \mathbf{x}) \\ \leftrightarrow \quad & \mathrm{CA}_\Phi[\varphi(\mathbf{a})](\mathbf{e}, \ulcorner\Phi\urcorner, \mathbf{a}) \ ). \end{aligned}$$

For all $e \in \mathbb{N}$,

$$\begin{aligned} \mathrm{PA} \vdash \ \forall \mathbf{x} \ ( \quad & \mathrm{CR}_\Omega[\varphi(\mathbf{x})](\mathbf{e}, \ulcorner\Phi\urcorner, \mathbf{x}) \\ \leftrightarrow \quad & \neg\mathrm{PTM\text{-}Acc}(\mathbf{e}, \ulcorner\Phi\urcorner, \mathbf{x}) \ \wedge \ \neg\mathrm{Acc}(\mathbf{v}^R_\Omega, \ulcorner\Phi\urcorner, \mathbf{x}) \\ \rightarrow \quad & \neg\mathrm{PTM\text{-}Acc}(\mathbf{e}, \ulcorner\Phi\urcorner, \mathbf{x}) \ \wedge \ \neg\mathrm{Acc}(\mathbf{v}^R_\Phi, \ulcorner\Phi\urcorner, \mathbf{x}) \\ \leftrightarrow \quad & \mathrm{CR}_\Phi[\varphi(\mathbf{a})](\mathbf{e}, \ulcorner\Phi\urcorner, \mathbf{a}) \ ). \end{aligned}$$

$\dashv$

## 5 Incompleteness Theorems of Polynomial-Time Decisions

This section shows the *polynomial-time decision* version of the (second) Gödel incompleteness theorem. First, we introduce the Gödel sentences of polynomial-time decisions, and the first incompleteness theorems of polynomial-time decisions. We then present the second incompleteness theorem of polynomial-time decisions, based on the the first incompleteness theorems and the derivability conditions of polynomial-time decisions.

### 5.1 Derivability Conditions of Polynomial-Time Decisions

**Lemma 29.** *(D.1-CA)* *Let $\Phi \equiv \{\varphi(\mathbf{a}) \mid a \in \mathbb{N}\}$ be a set of an infinite number of $\Delta_1$-formulas in PA.*

*For any $e \in \mathbb{N}$, for any $v \in \mathbb{N}$, and for any $a \in \mathbb{N}$*

$$\mathrm{PTM}_e^{\Phi}(a) \rhd_v \varphi(\mathbf{a}) \quad \Rightarrow \quad \mathrm{PA} \vdash \mathrm{CA}_v[\varphi(\mathbf{a})](\mathbf{e}, \ulcorner\Phi\urcorner, \mathbf{a}).$$

*Proof.* For all $e \in \mathbb{N}$, and for all $a \in \mathbb{N}$,

$$\begin{aligned}
& \mathrm{PTM}_e^{\Phi}(a) \rhd_v \varphi(\mathbf{a}) \\
\Leftrightarrow\ & \mathrm{U}_{\mathrm{PTM}}(e, (d, \#\Phi, a))\mathrm{accepts}\ \wedge\ \mathrm{U}(v, (d, \#\Phi, a))\mathrm{accepts} \\
\Rightarrow\ & \mathrm{PA} \vdash\ \ \mathrm{PTM\text{-}Acc}(\mathbf{e}, \ulcorner\Phi\urcorner, \mathbf{a})\ \wedge\ \mathrm{Acc}(\mathbf{v}, \ulcorner\Phi\urcorner, \mathbf{a}) \quad \text{(from } \Sigma_1\text{-Completeness Theorem of PA)} \\
\Leftrightarrow\ & \mathrm{PA} \vdash\ \ \mathrm{CA}_v[\varphi(\mathbf{a})](\mathbf{e}, \ulcorner\Phi\urcorner, \mathbf{a}).
\end{aligned}$$

⊣

**Lemma 30.** *(D.1-CR)* *Let $\Phi \equiv \{\varphi(\mathbf{a}) \mid a \in \mathbb{N}\}$ be a set of an infinite number of $\Delta_1$-formulas in PA.*

*For any $e \in \mathbb{N}$, for any $v \in \mathbb{N}$, and for any $a \in \mathbb{N}$*

$$\mathrm{PTM}_e^{\Phi}(a) \rhd_v \neg\varphi(\mathbf{a}) \quad \Rightarrow \quad \mathrm{PA} \vdash \mathrm{CR}_v[\varphi(\mathbf{a})](\mathbf{e}, \ulcorner\Phi\urcorner, \mathbf{a}).$$

*Proof.* For all $e \in \mathbb{N}$, and for all $a \in \mathbb{N}$,

$$\begin{aligned}
& \mathrm{PTM}_e^{\Phi}(a) \rhd_v \neg\varphi(\mathbf{a}) \\
\Leftrightarrow\ & \mathrm{U}_{\mathrm{PTM}}(e, (d, \#\Phi, a))\mathrm{rejects}\ \wedge\ \mathrm{U}(v, (d, \#\Phi, a))\mathrm{rejects} \\
\Rightarrow\ & \mathrm{PA} \vdash\ \ \neg\mathrm{PTM\text{-}Acc}(\mathbf{e}, \ulcorner\Phi\urcorner, \mathbf{a})\ \wedge\ \neg\mathrm{Acc}(\mathbf{v}, \ulcorner\Phi\urcorner, \mathbf{a}) \quad \text{(from } \Sigma_1\text{-Completeness Theorem of PA)} \\
\Leftrightarrow\ & \mathrm{PA} \vdash\ \ \mathrm{CR}_v[\varphi(\mathbf{a})](\mathbf{e}, \ulcorner\Phi\urcorner, \mathbf{a}).
\end{aligned}$$

⊣

**Lemma 31.** *(D.2-CA)* *Let $\Omega \equiv \{\omega(\mathbf{a}) \mid a \in \mathbb{N}\}$, $\Phi \equiv \{\varphi(\mathbf{a}) \mid a \in \mathbb{N}\}$, $\Psi \equiv \{\psi(\mathbf{a}) \mid a \in \mathbb{N}\}$, and $\Gamma \equiv \{\varphi(\mathbf{a}) \wedge \psi(\mathbf{a}) \mid\ (a, a) \in \mathbb{N}^2\}$ be sets of an infinite number of $\Delta_1$-formulas in PA.*

*For all $e_1 \in \mathbb{N}$ and for all $e_2 \in \mathbb{N}$, there exists $e_3 \in \mathbb{N}$ such that*

$$\begin{aligned}
\mathrm{PA} \vdash\ \ \forall \mathbf{x}\ (\ \ & \mathrm{CA}_{\Omega}[\varphi(\mathbf{x})](\mathbf{e}_1, \ulcorner\Phi\urcorner, \mathbf{x})\ \wedge \mathrm{CA}_{\Omega}[\psi(\mathbf{x})](\mathbf{e}_2, \ulcorner\Psi\urcorner, \mathbf{x}) \\
& \rightarrow\ \mathrm{CA}_{\Omega}[\varphi(\mathbf{x})\ \wedge\ \psi(\mathbf{x})](\mathbf{e}_3, \ulcorner\Gamma\urcorner, \mathbf{x})\ \ ).
\end{aligned}$$

*Proof.* PTM $\mathrm{U}_{\mathrm{PTM}}(e_3, \cdot)$ is constructed by using two PTMs, $\mathrm{U}_{\mathrm{PTM}}(e_1, (d, \#\Phi, \cdot))$ and $\mathrm{U}_{\mathrm{PTM}}(e_2, (d, \#\Psi, \cdot))$ as follows:

1. (Input: ) $(d, \#\Gamma, (x, x)) \in \mathbb{N}^4$.
2. (Output: ) accept or reject
3. Run the following computation

$$\mathrm{U_{PTM}}(e_1, (d, \#\Phi, x)),$$

$$\mathrm{U_{PTM}}(e_2, (d, \#\Psi, x)).$$

4. If both of them accept, then accept. Otherwise reject.

From the construction of $\mathrm{U_{PTM}}(e_3, \cdot)$, clearly

$$\mathrm{PA} \vdash \ \forall \mathbf{x} \ ( \ \text{PTM-Acc}(\mathbf{e}_1, \ulcorner\Phi\urcorner, \mathbf{x}) \ \wedge \ \text{PTM-Acc}(\mathbf{e}_2, \ulcorner\Psi\urcorner, \mathbf{x}) \ \leftrightarrow \ \text{PTM-Acc}(\mathbf{e}_3, \ulcorner\Gamma\urcorner, \mathbf{x}) \ ).$$

As shown in Eq. (23),

$$\mathrm{PA} \vdash \ \forall \mathbf{x} \ ( \ \mathrm{Acc}(\mathbf{v}^A_\Omega, \ulcorner\Phi\urcorner, \mathbf{x}) \ \wedge \mathrm{Acc}(\mathbf{v}^A_\Omega, \ulcorner\Psi\urcorner, \mathbf{x}) \ \rightarrow \ \mathrm{Acc}(\mathbf{v}^A_\Omega, \ulcorner\Gamma\urcorner, \mathbf{x}) \ ). \tag{27}$$

Then, for all $e_1 \in \mathbb{N}$ and for all $e_2 \in \mathbb{N}$, there exists $e_3 \in \mathbb{N}$ such that

$$\begin{aligned}
\mathrm{PA} \vdash \ \forall \mathbf{x} \ ( \quad & \mathrm{CA}_\Omega[\varphi(\mathbf{x})](\mathbf{e}_1, \ulcorner\Phi\urcorner, \mathbf{x}) \ \wedge \mathrm{CA}_\Omega[\psi(\mathbf{x})](\mathbf{e}_2, \ulcorner\Psi\urcorner, \mathbf{x}) \\
\leftrightarrow \ & ( \ \text{PTM-Acc}(\mathbf{e}_1, \ulcorner\Phi\urcorner, \mathbf{x}) \ \wedge \ \mathrm{Acc}(\mathbf{v}^A_\Omega, \ulcorner\Phi\urcorner, \mathbf{x}) \ ) \\
& \quad \wedge \ ( \ \text{PTM-Acc}(\mathbf{e}_2, \ulcorner\Psi\urcorner, \mathbf{x}) \ \wedge \ \mathrm{Acc}(\mathbf{v}^A_\Omega, \ulcorner\Psi\urcorner, \mathbf{x}) \ ) \\
\rightarrow \ & ( \ \text{PTM-Acc}(\mathbf{e}_1, \ulcorner\Phi\urcorner, \mathbf{x}) \ \wedge \ \text{PTM-Acc}(\mathbf{e}_2, \ulcorner\Psi\urcorner, \mathbf{x}) \ ) \\
& \quad \wedge \ ( \ \mathrm{Acc}(\mathbf{v}^A_\Omega, \ulcorner\Phi\urcorner, \mathbf{x}) \ \wedge \ \mathrm{Acc}(\mathbf{v}^A_\Omega, \ulcorner\Psi\urcorner, \mathbf{x}) \ ) \\
\leftrightarrow \ & \ \text{PTM-Acc}(\mathbf{e}_3, \ulcorner\Gamma\urcorner, \mathbf{x}) \ \wedge \ \mathrm{Acc}(\mathbf{v}^A_\Omega, \ulcorner\Gamma\urcorner, \mathbf{x}) \\
\leftrightarrow \ & \ \mathrm{CA}_\Omega[\varphi(\mathbf{x}) \ \wedge \ \psi(\mathbf{x})](\mathbf{e}_3, \ulcorner\Gamma\urcorner, \mathbf{x}) \ \ ).
\end{aligned}$$

$\dashv$

**Lemma 32.** *(D.2-CR) Let $\Omega \equiv \{\omega(\mathbf{a}) \mid a \in \mathbb{N}\}$, $\Phi \equiv \{\varphi(\mathbf{a}) \mid a \in \mathbb{N}\}$, $\Psi \equiv \{\psi(\mathbf{a}) \mid a \in \mathbb{N}\}$, and $\Gamma \equiv \{\varphi(\mathbf{a}) \wedge \psi(\mathbf{a}) \mid \ (a, a) \in \mathbb{N}^2\}$ be sets of an infinite number of $\Delta_1$-formulas in PA.*
*For all $e_1 \in \mathbb{N}$ and for all $e_2 \in \mathbb{N}$, there exists $e_3 \in \mathbb{N}$ such that*

$$\begin{aligned}
\mathrm{PA} \vdash \ \forall \mathbf{x} \ ( \quad & \mathrm{CR}_\Omega[\varphi(\mathbf{x})](\mathbf{e}_1, \ulcorner\Phi\urcorner, \mathbf{x}) \ \wedge \mathrm{CR}_\Omega[\psi(\mathbf{x})](\mathbf{e}_2, \ulcorner\Psi\urcorner, \mathbf{x}) \\
\rightarrow \ & \mathrm{CR}_\Omega[\varphi(\mathbf{x}) \ \vee \ \psi(\mathbf{x})](\mathbf{e}_3, \ulcorner\Theta\urcorner, \mathbf{x}) \ \ ).
\end{aligned}$$

*Proof.* PTM $\mathrm{U_{PTM}}(e_3, \cdot)$ is constructed by using two PTMs, $\mathrm{U_{PTM}}(e_1, (d, \#\Phi, \cdot))$ and $\mathrm{U_{PTM}}(e_2, (d, \#\Psi, \cdot))$ as follows:

1. (Input: ) $(d, \#\Theta, (x, x)) \in \mathbb{N}^4$.
2. (Output: ) accept or reject
3. Run the following computation

$$\mathrm{U_{PTM}}(e_1, (d, \#\Phi, x)),$$

$$\mathrm{U_{PTM}}(e_2, (d, \#\Psi, x)).$$

4. If both of them reject, then reject. Otherwise accept.

From the construction of $\mathrm{U_{PTM}}(e_3, \cdot)$, clearly

$$\mathrm{PA} \vdash \ \forall \mathbf{x} \ ( \ \neg\text{PTM-Acc}(\mathbf{e}_1, \ulcorner\Phi\urcorner, \mathbf{x}) \ \wedge \ \neg\text{PTM-Acc}(\mathbf{e}_2, \ulcorner\Psi\urcorner, \mathbf{x}) \ \leftrightarrow \ \neg\text{PTM-Acc}(\mathbf{e}_3, \ulcorner\Theta\urcorner, \mathbf{x}) \ ).$$

As shown in Eq. (24),

$$\mathrm{PA} \vdash \ \forall \mathbf{x} \ ( \ \neg\mathrm{Acc}(\mathbf{v}^R_\Omega, \ulcorner\Phi\urcorner, \mathbf{x}) \ \wedge \neg\mathrm{Acc}(\mathbf{v}^R_\Omega, \ulcorner\Psi\urcorner, \mathbf{x}) \ \rightarrow \ \neg\mathrm{Acc}(\mathbf{v}^R_\Omega, \ulcorner\Theta\urcorner, \mathbf{x}) \ ). \tag{28}$$

Then, for all $e_1 \in \mathbb{N}$ and for all $e_2 \in \mathbb{N}$, there exists $e_3 \in \mathbb{N}$ such that

$$\begin{aligned}
\mathrm{PA} \vdash \ \forall \mathbf{x} \ ( \quad & \mathrm{CR}_\Omega[\varphi(\mathbf{x})](\mathbf{e}_1, \ulcorner\Phi\urcorner, \mathbf{x}) \ \wedge \mathrm{CR}_\Omega[\psi(\mathbf{x})](\mathbf{e}_2, \ulcorner\Psi\urcorner, \mathbf{x}) \\
\leftrightarrow \ & ( \ \neg\text{PTM-Acc}(\mathbf{e}_1, \ulcorner\Phi\urcorner, \mathbf{x}) \ \wedge \ \neg\mathrm{Acc}(\mathbf{v}^R_\Omega, \ulcorner\Phi\urcorner, \mathbf{x}) \ ) \\
& \quad \wedge \ ( \ \neg\text{PTM-Acc}(\mathbf{e}_2, \ulcorner\Psi\urcorner, \mathbf{x}) \ \wedge \ \neg\mathrm{Acc}(\mathbf{v}^R_\Omega, \ulcorner\Psi\urcorner, \mathbf{x}) \ ) \\
\rightarrow \ & ( \ \neg\text{PTM-Acc}(\mathbf{e}_1, \ulcorner\Phi\urcorner, \mathbf{x}) \ \wedge \ \neg\text{PTM-Acc}(\mathbf{e}_2, \ulcorner\Psi\urcorner, \mathbf{x}) \ ) \\
& \quad \wedge \ ( \ \neg\mathrm{Acc}(\mathbf{v}^R_\Omega, \ulcorner\Phi\urcorner, \mathbf{x}) \ \wedge \ \neg\mathrm{Acc}(\mathbf{v}^R_\Omega, \ulcorner\Psi\urcorner, \mathbf{x}) \ ) \\
\leftrightarrow \ & \neg\text{PTM-Acc}(\mathbf{e}_3, \ulcorner\Gamma\urcorner, \mathbf{x}) \ \wedge \ \neg\mathrm{Acc}(\mathbf{v}^R_\Omega, \ulcorner\Gamma\urcorner, \mathbf{x}) \\
\leftrightarrow \ & \mathrm{CR}_\Omega[\varphi(\mathbf{x}) \ \wedge \ \psi(\mathbf{x})](\mathbf{e}_3, \ulcorner\Gamma\urcorner, \mathbf{x}) \ ).
\end{aligned}$$

⊣

**Corollary 33.** *Let $\Omega \equiv \{\omega(\mathbf{a}) \mid a \in \mathbb{N}\}$, $\Phi \equiv \{\varphi(\mathbf{a}) \mid a \in \mathbb{N}\}$, and $\Psi \equiv \{\psi(\mathbf{a}) \mid a \in \mathbb{N}\}$ be sets of an infinite number of $\Delta_1$-formulas in PA.*

*We assume that*

$$\mathrm{PA} \vdash \ \forall \mathbf{x} \ (\varphi(\mathbf{x}) \rightarrow \psi(\mathbf{x}))$$

*is installed in $\mathrm{U}(v^A_\Omega, \cdot)$.*

*Then, for all $e_1 \in \mathbb{N}$ there exists $e_2 \in \mathbb{N}$ such that*

$$\mathrm{PA} \vdash \ \forall \mathbf{x} \ ( \ \mathrm{CA}_\Omega[\varphi(\mathbf{x})](\mathbf{e}_1, \ulcorner\Phi\urcorner, \mathbf{x}) \ \rightarrow \ \mathrm{CA}_\Omega[\psi(\mathbf{x})](\mathbf{e}_2, \ulcorner\Psi\urcorner, \mathbf{x}) \ ).$$

*Proof.* PTM $\mathrm{U_{PTM}}(e_2, \cdot)$ is constructed by using PTM $\mathrm{U_{PTM}}(e_1, (d, \#\Phi, \cdot))$ as follows:

1. (Input: ) $(d, \#\Psi, x)$
2. (Output: ) accept or reject
3. Run the following computation
$$\mathrm{U_{PTM}}(e_1, (d, \#\Phi, x)).$$
4. Accept if and only $\mathrm{U_{PTM}}(e_1, (d, \#\Phi, x))$ accepts.

From the construction of $\mathrm{U_{PTM}}(e_2, \cdot)$, clearly

$$\mathrm{PA} \vdash \ \forall \mathbf{x} \ ( \ \text{PTM-Acc}(\mathbf{e}_1, \ulcorner\Phi\urcorner, \mathbf{x}) \ \leftrightarrow \ \text{PTM-Acc}(\mathbf{e}_2, \ulcorner\Psi\urcorner, \mathbf{x}) \ ).$$

Since $\mathrm{PA} \vdash \ \forall \mathbf{x} \ (\varphi(\mathbf{x}) \rightarrow \psi(\mathbf{x}))$ is installed in $\mathrm{U}(v^A_\Omega, \cdot)$,

$$\mathrm{PA} \vdash \ \forall \mathbf{x} \ ( \ \mathrm{Acc}(\mathbf{v}^A_\Omega, \ulcorner\Phi\urcorner, \mathbf{x}) \ \rightarrow \ \mathrm{Acc}(\mathbf{v}^A_\Omega, \ulcorner\Psi\urcorner, \mathbf{x}) \ ).$$

Then, for all $e_1 \in \mathbb{N}$ there exists $e_2 \in \mathbb{N}$ such that

$$\begin{aligned}
\mathrm{PA} \vdash \ \forall \mathbf{x} \ ( \quad & \mathrm{CA}_\Omega[\varphi(\mathbf{x})](\mathbf{e}_1, \ulcorner\Phi\urcorner, \mathbf{x}) \\
\leftrightarrow \ & \text{PTM-Acc}(\mathbf{e}_1, \ulcorner\Phi\urcorner, \mathbf{x}) \ \wedge \ \mathrm{Acc}(\mathbf{v}^A_\Omega, \ulcorner\Phi\urcorner, \mathbf{x}) \\
\rightarrow \ & \text{PTM-Acc}(\mathbf{e}_2, \ulcorner\Psi\urcorner, \mathbf{x}) \ \wedge \ \mathrm{Acc}(\mathbf{v}^A_\Omega, \ulcorner\Psi\urcorner, \mathbf{x}) \\
\leftrightarrow \ & \mathrm{CA}_\Omega[\psi(\mathbf{x})](\mathbf{e}_2, \ulcorner\Psi\urcorner, \mathbf{x}) \ ).
\end{aligned}$$

$\dashv$

**Corollary 34.** *Let $\Omega \equiv \{\omega(\mathbf{a}) \mid a \in \mathbb{N}\}$, $\Phi \equiv \{\varphi(\mathbf{a}) \mid a \in \mathbb{N}\}$, and $\Psi \equiv \{\psi(\mathbf{a}) \mid a \in \mathbb{N}\}$ be sets of an infinite number of $\Delta_1$-formulas in PA.*

*We assume that*

$$\mathrm{PA} \vdash \ \forall \mathbf{x}\ (\neg\varphi(\mathbf{x}) \to \neg\psi(\mathbf{x}))$$

*is installed in $\mathrm{U}(v^R_\Omega, \cdot)$.*

*Then, for all $e_1 \in \mathbb{N}$ there exists $e_2 \in \mathbb{N}$ such that*

$$\mathrm{PA} \vdash \quad \forall \mathbf{x}\ (\ \mathrm{CR}_\Omega[\varphi(\mathbf{x})](\mathrm{e}_1, \ulcorner\Phi\urcorner, \mathbf{x}) \ \to \ \mathrm{CR}_\Omega[\psi(\mathbf{x})](\mathrm{e}_2, \ulcorner\Psi\urcorner, \mathbf{x})\ ).$$

*Proof.* PTM $\mathrm{U_{PTM}}(e_2, \cdot)$ is constructed by using PTM $\mathrm{U_{PTM}}(e_1, (d, \#\Phi, \cdot))$ as follows:

1. (Input: ) $(d, \#\Psi, x)$
2. (Output: ) accept or reject
3. Run the following computation
$$\mathrm{U_{PTM}}(e_1, (d, \#\Phi, x)).$$
4. Reject if and only $\mathrm{U_{PTM}}(e_1, (d, \#\Phi, x))$ rejects.

From the construction of $\mathrm{U_{PTM}}(e_2, \cdot)$, clearly

$$\mathrm{PA} \vdash \ \forall \mathbf{x}\ \ (\ \neg\text{PTM-Acc}(\mathrm{e}_1, \ulcorner\Phi\urcorner, \mathbf{x}) \ \leftrightarrow \ \neg\text{PTM-Acc}(\mathrm{e}_2, \ulcorner\Psi\urcorner, \mathbf{x})\ \ ).$$

Since $\mathrm{PA} \vdash \ \forall \mathbf{x}\ (\neg\varphi(\mathbf{x}) \to \neg\psi(\mathbf{x}))$ is installed in $\mathrm{U}(v^R_\Omega, \cdot)$,

$$\mathrm{PA} \vdash \ \forall \mathbf{x}\ (\ \neg\mathrm{Acc}(\mathbf{v}^R_\Omega, \ulcorner\Phi\urcorner, \mathbf{x}) \ \to \ \neg\mathrm{Acc}(\mathbf{v}^R_\Omega, \ulcorner\Psi\urcorner, \mathbf{x})\ ).$$

Then, for all $e_1 \in \mathbb{N}$ there exists $e_2 \in \mathbb{N}$ such that

$$\begin{aligned} \mathrm{PA} \vdash \quad \forall \mathbf{x} \quad ( \quad & \mathrm{CR}_\Omega[\varphi(\mathbf{x})](\mathrm{e}_1, \ulcorner\Phi\urcorner, \mathbf{x}) \\ \leftrightarrow\ & \neg\text{PTM-Acc}(\mathrm{e}_1, \ulcorner\Phi\urcorner, \mathbf{x}) \ \wedge\ \neg\mathrm{Acc}(\mathbf{v}^R_\Omega, \ulcorner\Phi\urcorner, \mathbf{x}) \\ \to\ & \neg\text{PTM-Acc}(\mathrm{e}_2, \ulcorner\Psi\urcorner, \mathbf{x}) \ \wedge\ \neg\mathrm{Acc}(\mathbf{v}^R_\Omega, \ulcorner\Psi\urcorner, \mathbf{x}) \\ \leftrightarrow\ & \mathrm{CR}_\Omega[\psi(\mathbf{x})](\mathrm{e}_2, \ulcorner\Psi\urcorner, \mathbf{x}) \ \ ). \end{aligned}$$

$\dashv$

**Lemma 35.** *(D.3-CA) Let $\Omega \equiv \{\omega(\mathbf{a}) \mid a \in \mathbb{N}\}$ and $\Phi \equiv \{\varphi(\mathbf{a}) \mid a \in \mathbb{N}\}$ be sets of an infinite number of $\Delta_1$-formulas in PA. Let $\mathcal{CA}[e, \Omega] \equiv \{\mathrm{CA}_\Omega[\varphi(\mathbf{a})](\mathrm{e}, \ulcorner\Phi\urcorner, \mathbf{a}) \mid a \in \mathbb{N}\}$.*

*For all $e_1 \in \mathbb{N}$, there exists $e_2 \in \mathbb{N}$ such that*

$$\mathrm{PA} \vdash \quad \forall \mathbf{x}\ (\ \mathrm{CA}_\Omega[\varphi(\mathbf{x})](\mathrm{e}_1, \ulcorner\Phi\urcorner, \mathbf{x}) \ \to \ \mathrm{CA}_\Omega[\mathrm{CA}_\Omega[\varphi(\mathbf{x})](\mathrm{e}_1, \ulcorner\Phi\urcorner, \mathbf{x})](\mathrm{e}_2, \ulcorner\mathcal{CA}[e_1, \Omega]\urcorner, \mathbf{x})\ ).$$

*Proof.* PTM $\mathrm{U_{PTM}}(e_2, \cdot)$ is constructed by using PTM $\mathrm{U_{PTM}}(e_1, (d, \#\Phi, \cdot))$ as follows:

1. (Input: ) $(d, \#\mathcal{CA}[e_1, \Omega], x) \in \mathbb{N}^3$
2. (Output: ) accept or reject
3. Run the following computation
$$\mathrm{U_{PTM}}(e_1, (d, \#\Phi, x)).$$
4. Accept if and only $\mathrm{U_{PTM}}(e_1, (d, \#\Phi, x))$ accepts.

From the construction of $\mathrm{U_{PTM}}(e_2, \cdot)$,

$$\mathrm{PA} \vdash \ \forall \mathbf{x} \ ( \ \mathrm{PTM\text{-}Acc}(\mathbf{e}_1, \ulcorner \Phi \urcorner, \mathbf{x}) \ \leftrightarrow \ \mathrm{PTM\text{-}Acc}(\mathbf{e}_2, \ulcorner \Psi \urcorner, \mathbf{x}) \ ).$$

As shown in Eq. (25),

$$\mathrm{PA} \vdash \ \forall \mathbf{x} \ ( \ \mathrm{CA}_\Omega[\varphi(\mathbf{x})](\mathbf{e}_1, \ulcorner \Phi \urcorner, \mathbf{x}) \ \rightarrow \ \mathrm{Acc}(\mathbf{v}^A_\Omega, \ulcorner \mathcal{CA}[e_1, \Omega] \urcorner, \mathbf{x}) \ ).$$

Then,

$$\begin{aligned} \mathrm{PA} \vdash \ \forall \mathbf{x} \ ( \ & \ \mathrm{CA}_\Omega[\varphi(\mathbf{x})](\mathbf{e}_1, \ulcorner \Phi \urcorner, \mathbf{x}) \\ & \leftrightarrow \mathrm{PTM\text{-}Acc}(\mathbf{e}_1, \ulcorner \Phi \urcorner, \mathbf{x}) \ \wedge \ \mathrm{Acc}(\mathbf{v}^A_\Omega, \ulcorner \Phi \urcorner, \mathbf{x}) \\ & \leftrightarrow \mathrm{PTM\text{-}Acc}(\mathbf{e}_1, \ulcorner \Phi \urcorner, \mathbf{x}) \ \wedge \ \mathrm{CA}_\Omega[\varphi(\mathbf{x})](\mathbf{e}_1, \ulcorner \Phi \urcorner, \mathbf{x}) \\ & \rightarrow \mathrm{PTM\text{-}Acc}(\mathbf{e}_2, \ulcorner \mathcal{CA}[e_1, \Omega] \urcorner, \mathbf{x}) \ \wedge \ \mathrm{Acc}(\mathbf{v}^A_\Omega, \ulcorner \mathcal{CA}[e_1, \Omega] \urcorner, \mathbf{x}) \\ & \rightarrow \mathrm{CA}_\Omega[\mathrm{CA}_\Omega[\varphi(\mathbf{x})](\mathbf{e}_1, \ulcorner \Phi \urcorner, \mathbf{x})](\mathbf{e}_2, \ulcorner \mathcal{CA}[e_1, \Omega] \urcorner, \mathbf{x}) \ \ ). \end{aligned}$$

$\dashv$

**Lemma 36.** *(D.3-CR) Let $\Omega \equiv \{\omega(\mathbf{a}) \mid a \in \mathbb{N}\}$ and $\Phi \equiv \{\varphi(\mathbf{a}) \mid a \in \mathbb{N}\}$ be sets of an infinite number of $\Delta_1$-formulas in PA. Let $\mathcal{CR}[e, \Omega] \equiv \{\neg \mathrm{CR}_\Omega[\varphi(\mathbf{a})](\mathbf{e}, \ulcorner \Phi \urcorner, \mathbf{a}) \mid a \in \mathbb{N}\}$.*

*For all $e_1 \in \mathbb{N}$, there exists $e_2 \in \mathbb{N}$ such that*

$$\mathrm{PA} \vdash \ \forall \mathbf{x} \ ( \ \mathrm{CR}_\Omega[\varphi(\mathbf{x})](\mathbf{e}_1, \ulcorner \Phi \urcorner, \mathbf{x}) \ \rightarrow \ \mathrm{CR}_\Omega[\neg \mathrm{CR}_\Omega[\varphi(\mathbf{x})](\mathbf{e}_1, \ulcorner \Phi \urcorner, \mathbf{x})](\mathbf{e}_2, \ulcorner \mathcal{CR}[e_1, \Omega] \urcorner, \mathbf{x}) \ ).$$

*Proof.* PTM $\mathrm{U_{PTM}}(e_2, \cdot)$ is constructed by using PTM $\mathrm{U_{PTM}}(e_1, (d, \#\Phi, \cdot))$ as follows:

1. (Input: ) $(d, \#\mathcal{CR}[e_1, \Omega], x) \in \mathbb{N}^3$
2. (Output: ) accept or reject
3. Run the following computation
$$\mathrm{U_{PTM}}(e_1, (d, \#\Phi, x)).$$
4. Reject if and only $\mathrm{U_{PTM}}(e_1, (d, \#\Phi, x))$ rejects.

From the construction of $\mathrm{U_{PTM}}(e_2, \cdot)$,

$$\mathrm{PA} \vdash \ \forall \mathbf{x} \ ( \neg \mathrm{PTM\text{-}Acc}(\mathbf{e}_1, \ulcorner \Phi \urcorner, \mathbf{x}) \ \leftrightarrow \ \neg \mathrm{PTM\text{-}Acc}(\mathbf{e}_2, \ulcorner \Psi \urcorner, \mathbf{x}) \ ).$$

As shown in Eq. (26),

$$\mathrm{PA} \vdash \ \forall \mathbf{x} \ ( \ \mathrm{CR}_\Omega[\varphi(\mathbf{x})](\mathbf{e}_1, \ulcorner \Phi \urcorner, \mathbf{x}) \ \rightarrow \ \neg \mathrm{Acc}(\mathbf{v}^R_\Omega, \ulcorner \mathcal{CR}[e_1, \Omega] \urcorner, \mathbf{x}) \ ).$$

Then,

$$\begin{aligned} \mathrm{PA} \vdash \ \forall \mathbf{x} \ ( \ & \ \mathrm{CR}_\Omega[\varphi(\mathbf{x})](\mathbf{e}_1, \ulcorner \Phi \urcorner, \mathbf{x}) \\ & \leftrightarrow \neg \mathrm{PTM\text{-}Acc}(\mathbf{e}_1, \ulcorner \Phi \urcorner, \mathbf{x}) \ \wedge \ \neg \mathrm{Acc}(\mathbf{v}^R_\Omega, \ulcorner \Phi \urcorner, \mathbf{x}) \\ & \leftrightarrow \neg \mathrm{PTM\text{-}Acc}(\mathbf{e}_1, \ulcorner \Phi \urcorner, \mathbf{x}) \ \wedge \ \mathrm{CR}_\Omega[\varphi(\mathbf{x})](\mathbf{e}_1, \ulcorner \Phi \urcorner, \mathbf{x}) \\ & \rightarrow \neg \mathrm{PTM\text{-}Acc}(\mathbf{e}_2, \ulcorner \mathcal{CR}[e_1, \Omega] \urcorner, \mathbf{x}) \ \wedge \ \neg \mathrm{Acc}(\mathbf{v}^R_\Omega, \ulcorner \mathcal{CR}[e_1, \Omega] \urcorner, \mathbf{x}) \\ & \rightarrow \mathrm{CR}_\Omega[\neg \mathrm{CR}_\Omega[\varphi(\mathbf{x})](\mathbf{e}_1, \ulcorner \Phi \urcorner, \mathbf{x})](\mathbf{e}_2, \ulcorner \mathcal{CR}[e_1, \Omega] \urcorner, \mathbf{x}) \ \ ). \end{aligned}$$

$\dashv$

## 5.2 Gödel Sentences of Polynomial-Time Decisions

**Lemma 37.** *For any $e \in \mathbb{N}$ and for any $v \in \mathbb{N}$, there exists a set of formulas, $\mathcal{G}^A \equiv \{\rho^A_{e,v}(\mathbf{a}) \mid a \in \mathbb{N}\}$, such that*

$$\mathrm{PA} \vdash \forall \mathbf{x} \ \ (\rho^A_{e,v}(\mathbf{x}) \ \leftrightarrow \ \neg \mathrm{CA}_v[\rho^A_{e,v}(\mathbf{x})](\mathbf{e}, \ulcorner \mathcal{G}^A \urcorner, \mathbf{x}) \ ).$$

*For all $x \in \mathbb{N}$, $\rho^A_{e,v}(\mathbf{x})$ is called a "Gödel sentence" with respect to CA.*

*Proof.* Let $e \in \mathbb{N}$ and $v \in \mathbb{N}$ be given.

Based on the recursion theorem (Proposition 17), TM $\mathrm{U}(k, \cdot)$ is constructed as follows:

1. (Input:) $(d, \#\mathbb{N}^3, (e, v, x)) \in \mathbb{N}^5$.,
2. (Output:) accept or reject
3. First, read its own code, $k \in \mathbb{N}$.
4. Construct a formula,
$$\rho^A_{e,v}(\mathbf{x}) \ \equiv \ \mathrm{Acc}(\mathbf{k}, \ulcorner \mathbb{N}^3 \urcorner, \mathbf{x}, \mathbf{e}, \mathbf{v}),$$
(i.e., $\rho^A_{e,v}(\mathbf{x})$ represents the relation that $\mathrm{U_{PTM}}(k, (d, \#\mathbb{N}^3, (e, v, x)))$ accepts). Let $\mathcal{G}^A \equiv \{\rho^A_{e,v}(\mathbf{a}) \mid a \in \mathbb{N}\}$.
5. Execute PTM $\mathrm{U_{PTM}}(e, (d, \#\mathcal{G}^A, x))$ to decide the truth of formula $\rho^A_{e,v}(\mathbf{x})$.
6. Execute TM $\mathrm{U}(v, (d, \#\mathcal{G}^A, x))$ to decide the truth of formula $\rho^A_{e,v}(\mathbf{x})$.
7. Reject if and only if both of $\mathrm{U_{PTM}}(e, (d, \#\mathcal{G}^A, x))$ and $\mathrm{U}(v, (d, \#\mathcal{G}^A, x))$ accepts.

Here, note that when only $\mathbf{x}$ occurs free in formula $\rho^A_{e,v}(\mathbf{x})$, $\#\rho^A_{e,v}(\mathbf{x})$ is a finite number. For $a \in \mathbb{N}$, $\rho^A_{e,v}(\mathbf{a})$ is equivalent to $\rho'_{e,v}(\mathbf{a}) \equiv (\rho^A_{e,v}(\mathbf{x}) \wedge \mathbf{x} = \mathbf{a})$, and $|\#\rho'_{e,v}(\mathbf{a})| = O(|a|)$.

Then, in a manner similar to Lemma 18, we obtain

$$\forall e \in \mathbb{N} \ \forall v \in \mathbb{N} \quad \mathrm{PA} \vdash \ \forall \mathbf{x} \ \ (\rho^A_{e,v}(\mathbf{x}) \ \leftrightarrow \ \neg \mathrm{CA}_v[\rho^A_{e,v}(\mathbf{x})](\mathbf{e}, \ulcorner \mathcal{G}^A \urcorner, \mathbf{x}) \ ).$$

⊣

**Lemma 38.** *For any $e \in \mathbb{N}$ and for any $v \in \mathbb{N}$, there exists a set of formulas, $\mathcal{G}^R \equiv \{\rho^R_{e,v}(\mathbf{a}) \mid a \in \mathbb{N}\}$, such that*

$$\mathrm{PA} \vdash \forall \mathbf{x} \ \ (\rho^R_{e,v}(\mathbf{x}) \ \leftrightarrow \ \mathrm{CR}_v[\rho^R_{e,v}(\mathbf{x})](\mathbf{e}, \ulcorner \mathcal{G}^R \urcorner, \mathbf{x}) \ ).$$

*For all $x \in \mathbb{N}$, $\rho^R_{e,v}(\mathbf{x})$ is called a "Gödel sentence" with respect to CR.*

*Proof.* Let $e \in \mathbb{N}$ and $v \in \mathbb{N}$ be given.

Based on the recursion theorem (Proposition 17), TM $\mathrm{U}(k, \cdot)$ is constructed as follows:

1. (Input:) $(d, \#\mathbb{N}^3, (e, v, x)) \in \mathbb{N}^5$.,
2. (Output:) accept or reject
3. First, read its own code, $k \in \mathbb{N}$.
4. Construct a formula,
$$\rho^R_{e,v}(\mathbf{x}) \ \equiv \ \mathrm{Acc}(\mathbf{k}, \ulcorner \mathbb{N}^3 \urcorner, \mathbf{x}, \mathbf{e}, \mathbf{v}),$$
(i.e., $\rho^R_{e,v}(\mathbf{x})$ represents the relation that $\mathrm{U_{PTM}}(k, (d, \#\mathbb{N}^3, (e, v, x)))$ accepts). Let $\mathcal{G}^R \equiv \{\rho^R_{e,v}(\mathbf{a}) \mid a \in \mathbb{N}\}$.
5. Execute PTM $\mathrm{U_{PTM}}(e, (d, \#\mathcal{G}^R, x))$ to decide on the truth of formula $\rho^R_{e,v}(\mathbf{x})$.
6. Execute TM $\mathrm{U}(v, (d, \#\mathcal{G}^R, x))$ to decide on the truth of formula $\rho^R_{e,v}(\mathbf{x})$.
7. Accept if and only if both of $\mathrm{U_{PTM}}(e, (d, \#\mathcal{G}^R, x))$ and $\mathrm{U}(v, (d, \#\mathcal{G}^R, x))$ reject.

Then, in a manner similar to Lemma 18, we obtain

$$\forall e \in \mathbb{N} \ \forall v \in \mathbb{N} \quad \mathrm{PA} \vdash \ \forall \mathbf{x} \ \ (\rho^R_{e,v}(\mathbf{x}) \ \leftrightarrow \ \mathrm{CR}_v[\rho^R_{e,v}(\mathbf{x})](\mathbf{e}, \ulcorner \mathcal{G}^R \urcorner, \mathbf{x}) \ ).$$

⊣

### 5.3 The First Incompleteness Theorems of Polynomial-Time Decisions

**Theorem 39.** *Let $\rho^A_{e,v}(\mathbf{a})$ be a Gödel sentence with respect to CA, where $a \in \mathbb{N}$. Let $\mathcal{G}^A \equiv \{\rho^A_{e,v}(\mathbf{a}) \mid a \in \mathbb{N}\}$. Let $\mathrm{U}(v,\cdot)$ soundly accept (see Definition 25).*

*For all $e \in \mathbb{N}$, and for all $x \in \mathbb{N}$,*

$$\mathrm{PTM}_e^{\mathcal{G}^A}(x) \not\triangleright_v \rho^A_{e,v}(\mathbf{x}).$$

*Proof.* Assume that there exist $e \in \mathbb{N}$, $v \in \mathbb{N}$ and $x \in \mathbb{N}$ such that

$$\mathrm{PTM}_e^{\mathcal{G}^A}(x) \triangleright_v \rho^A_{e,v}(\mathbf{x}). \tag{29}$$

From Lemma 29

$$\mathrm{PA} \vdash \mathrm{CA}_v[\rho^A_{e,v}(\mathbf{x})](\mathbf{e}, \ulcorner \mathcal{G}^A \urcorner, \mathbf{x}).$$

Since PA has model $\mathfrak{N}$,

$$\mathfrak{N} \models \mathrm{CA}_v[\rho^A_{e,v}(\mathbf{x})](\mathbf{e}, \ulcorner \mathcal{G}^A \urcorner, \mathbf{x}). \tag{30}$$

On the other hand, from assumption of Eq. (29), $\mathrm{U}(v,(d,\#\mathcal{G}^A,x))$ accepts. Since $\mathrm{U}(v,\cdot)$ soundly accepts,

$$\mathfrak{N} \models \rho^A_{e,v}(\mathbf{x}).$$

Applying Lemma 37 to the above equation,

$$\mathfrak{N} \models \neg\mathrm{CA}_v[\rho^A_{e,v}(\mathbf{x})](\mathbf{e}, \ulcorner \mathcal{G}^A \urcorner, \mathbf{x}).$$

This contradicts Eq. (30). Thus, for all $e \in \mathbb{N}$, and for all $x \in \mathbb{N}$,

$$\mathrm{PTM}_e^{\mathcal{G}^A}(x) \not\triangleright_v \rho^A_{e,v}(\mathbf{x}).$$

⊣

**Theorem 40.** *Let $\rho^R_{e,v}(\mathbf{a})$ be a Gödel sentence with respect to CR, where $a \in \mathbb{N}$. Let $\mathcal{G}^R \equiv \{\rho^R_{e,v}(\mathbf{a}) \mid a \in \mathbb{N}\}$. Let $\mathrm{U}(v,\cdot)$ soundly reject (see Definition 25).*

*For all $e \in \mathbb{N}$, and for all $x \in \mathbb{N}$,*

$$\mathrm{PTM}_e^{\mathcal{G}^R}(x) \not\triangleright_v \neg\rho^R_{e,v}(\mathbf{x}).$$

*Proof.* Assume that there exist $e \in \mathbb{N}$, $v \in \mathbb{N}$ and $x \in \mathbb{N}$ such that

$$\mathrm{PTM}_e^{\mathcal{G}^R}(x) \triangleright_v \neg\rho^R_{e,v}(\mathbf{x}). \tag{31}$$

From Lemma 30,

$$\mathrm{PA} \vdash \mathrm{CR}_v[\rho^R_{e,v}(\mathbf{x})](\mathbf{e}, \ulcorner \mathcal{G}^{\mathcal{R}} \urcorner, \mathbf{x}).$$

Since PA has model $\mathfrak{N}$,

$$\mathfrak{N} \models \mathrm{CR}_v[\rho^R_{e,v}(\mathbf{x})](\mathbf{e}, \ulcorner \mathcal{G}^{\mathcal{R}} \urcorner, \mathbf{x}). \tag{32}$$

On the other hand, from assumption of Eq. (31) $\mathrm{U}(v,(d,\#\mathcal{G}^R,x))$ rejects. Since $\mathrm{U}(v,\cdot)$ soundly rejects,

$$\mathfrak{N} \models \neg\rho^R_{e,v}(\mathbf{x}).$$

Applying Lemma 38 to the above equation,

$$\mathfrak{N} \models \neg\mathrm{CR}_v[\rho^R_{e,v}(\mathbf{x})](\mathbf{e}, \ulcorner\mathcal{G}^R\urcorner, \mathbf{x}).$$

This contradicts Eq. (32). Thus, for all $e \in \mathbb{N}$, and for all $x \in \mathbb{N}$,

$$\mathrm{PTM}^{\mathcal{G}^R}_e(x) \not\vdash_v \neg\rho^R_{e,v}(\mathbf{x}).$$

$\dashv$

The following Corollaries are immediately obtained from Theorems 39 and 40, since $\mathrm{U}(v^A_\Omega, \cdot)$ soundly accepts and $\mathrm{U}(v^R_\Omega, \cdot)$ soundly rejects, as shown in Lemma 26.

**Corollary 41.** *Let $\Omega \equiv \{\omega(\mathbf{a}) \mid a \in \mathbb{N}\}$, be a set of an infinite number of $\Delta_1$-formulas in PA. Let $\mathrm{U}(v^A_\Omega, \cdot)$ be a TM as defined in Definition 24. Let $\rho^A_{e,\Omega}(\mathbf{a}) \equiv \rho^A_{e,v^A_\Omega}(\mathbf{a})$ be a Gödel sentence with respect to CA, where $a \in \mathbb{N}$. Let $\mathcal{G}^A_\Omega \equiv \{\rho^A_{e,\Omega}(\mathbf{a}) \mid a \in \mathbb{N}\}$.*

*For all $e \in \mathbb{N}$, and for all $x \in \mathbb{N}$,*

$$\mathrm{PTM}^{\mathcal{G}^A_\Omega}_e(x) \not\vdash_{v^A_\Omega} \rho^A_{e,\Omega}(\mathbf{x}).$$

**Corollary 42.** *Let $\Omega \equiv \{\omega(\mathbf{a}) \mid a \in \mathbb{N}\}$, be a set of an infinite number of $\Delta_1$-formulas in PA. Let $\mathrm{U}(v^R_\Omega, \cdot)$ be a TM as defined in Definition 24. Let $\rho^R_{e,\Omega}(\mathbf{a}) \equiv \rho^R_{e,v^R_\Omega}(\mathbf{a})$ be a Gödel sentence with respect to CR, where $a \in \mathbb{N}$. Let $\mathcal{G}^R_\Omega \equiv \{\rho^R_{e,\Omega}(\mathbf{a}) \mid a \in \mathbb{N}\}$.*

*For all $e \in \mathbb{N}$, and for all $x \in \mathbb{N}$,*

$$\mathrm{PTM}^{\mathcal{G}^R_\Omega}_e(x) \not\vdash_{v^R_\Omega} \neg\rho^R_{e,\Omega}(\mathbf{x}).$$

### 5.4 The Second Incompleteness Theorem of Polynomial-Time Decisions

**Lemma 43.** *For $a \in \mathbb{N}$, let $\rho^A_{e,\Omega}(\mathbf{a}) \equiv \rho^A_{e,v^A_\Omega}(\mathbf{a})$ be a Gödel sentence with respect to CA, and $\rho^R_{e,\Omega}(\mathbf{a}) \equiv \rho^R_{e,v^R_\Omega}(\mathbf{a})$ be a Gödel sentence with respect to CR. (For the definition of $v^A_\Omega$ and $v^R_\Omega$, see Definition 24.)*

*Then, there exists a primitive recursive function $h$ such that for any $\Delta_1$-formula sets $\Psi \equiv \{\psi(\mathbf{a}) \mid a \in \mathbb{N}\}$ and $\Omega \equiv \{\omega(\mathbf{a}) \mid a \in \mathbb{N}\}$, and for any $e \in \mathbb{N}$,*

$$\mathrm{PA} \vdash \forall\mathbf{x}\ \ (\ \neg\mathrm{CA}_\Omega[\psi(\mathbf{x})](h(\mathbf{e}), \ulcorner\Psi\urcorner, \mathbf{x}) \quad \rightarrow \quad \rho^A_{e,\Omega}(\mathbf{x})\ ), \tag{33}$$

*and*

$$\mathrm{PA} \vdash \forall\mathbf{x}\ \ (\ \neg\mathrm{CR}_\Omega[\psi(\mathbf{x})](h(\mathbf{e}), \ulcorner\Psi\urcorner, \mathbf{x}) \quad \rightarrow \quad \neg\rho^R_{e,\Omega}(\mathbf{x})\ ). \tag{34}$$

*In other words, for any $e \in \mathbb{N}$, there exists $e^*$ such that*

$$\mathrm{PA} \vdash \forall\mathbf{x}\ \ (\ \neg\mathrm{CA}_\Omega[\psi(\mathbf{x})](\mathbf{e}^*, \ulcorner\Psi\urcorner, \mathbf{x}) \quad \rightarrow \quad \rho^A_{e,\Omega}(\mathbf{x})\ ), \tag{35}$$

$$\mathrm{PA} \vdash \forall\mathbf{x}\ \ (\ \neg\mathrm{CR}_\Omega[\psi(\mathbf{x})](\mathbf{e}^*, \ulcorner\Psi\urcorner, \mathbf{x}) \quad \rightarrow \quad \neg\rho^R_{e,\Omega}(\mathbf{x})\ ), \tag{36}$$

*Proof.* Let

$$\mathcal{G}^A_\Omega \equiv \{\rho^A_{e,\Omega}(\mathbf{a}) \mid a \in \mathbb{N}\}$$

be a set of Gödel sentences with respect to CA, and

$$\mathcal{G}^R_\Omega \equiv \{\rho^R_{e,\Omega}(\mathbf{a}) \mid a \in \mathbb{N}\}$$

be a set of Gödel sentences with respect to CR.

Let

$$\begin{aligned}
&\mathcal{G}^{A+} \equiv \{\mathrm{CA}_\Omega[\rho^A_{e,\Omega}(\mathbf{a})](\mathbf{e}, \ulcorner\mathcal{G}^A_\Omega\urcorner, \mathbf{a}) \mid a \in \mathbb{N}\},\\
&\mathcal{G}^{A++} \equiv \{\neg\rho^A_{e,\Omega}(\mathbf{a}) \mid a \in \mathbb{N}\},\\
&\mathcal{G}^{A+++} \equiv \{\rho^A_{e,\Omega}(\mathbf{a}) \wedge \neg\rho^A_{e,\Omega}(\mathbf{a}) \mid a \in \mathbb{N}\},\\
&\mathcal{G}^{R+} \equiv \{\neg\mathrm{CR}_\Omega[\rho^R_{e,\Omega}(\mathbf{a})](\mathbf{e}, \ulcorner\mathcal{G}^R_\Omega\urcorner, \mathbf{a}) \mid a \in \mathbb{N}\},\\
&\mathcal{G}^{R++} \equiv \{\neg\rho^R_{e,\Omega}(\mathbf{a}) \mid a \in \mathbb{N}\},\\
&\mathcal{G}^{R+++} \equiv \{\rho^R_{e,\Omega}(\mathbf{a}) \vee \neg\rho^R_{e,\Omega}(\mathbf{a}) \mid a \in \mathbb{N}\}.
\end{aligned}$$

For any $e \in \mathbb{N}$, there exist $e^+ \in \mathbb{N}$, $e^{++} \in \mathbb{N}$ and $e^{+++} \in \mathbb{N}$ such that

$$\begin{aligned}
\mathrm{PA} \vdash \forall\mathbf{x}\ (\quad & \mathrm{CA}_\Omega[\rho^A_{e,\Omega}(\mathbf{x})](\mathbf{e}, \ulcorner\mathcal{G}^A_\Omega\urcorner, \mathbf{x})\\
\rightarrow\ & \mathrm{CA}_\Omega[\mathrm{CA}_\Omega[\rho^A_{e,\Omega}(\mathbf{x})](\mathbf{e}, \ulcorner\mathcal{G}\urcorner, \mathbf{x})](\mathbf{e}^+, \ulcorner\mathcal{G}^{A+}\urcorner, \mathbf{x}) \quad \text{(by Lemma 35)}\\
\rightarrow\ & \mathrm{CA}_\Omega[\neg\rho^A_{e,\Omega}(\mathbf{x})](\mathbf{e}^{++}, \ulcorner\mathcal{G}^{A++}\urcorner, \mathbf{x}) \quad \text{(by Lemma 37 and Corollary 33 )}\\
\rightarrow\ & \mathrm{CA}_\Omega[\rho^A_{e,\Omega}(\mathbf{x})](\mathbf{e}, \ulcorner\mathcal{G}^A_\Omega\urcorner, \mathbf{x}) \wedge\ \mathrm{CA}_\Omega[\neg\rho^A_{e,\Omega}(\mathbf{x})](\mathbf{e}^{++}, \ulcorner\mathcal{G}^{A++}\urcorner, \mathbf{x})\\
\rightarrow\ & \mathrm{CA}_\Omega[\rho^A_{e,\Omega}(\mathbf{x})\ \wedge\ \neg\rho^A_{e,\Omega}(\mathbf{x})](\mathbf{e}^{+++}, \ulcorner\mathcal{G}^{A+++}\urcorner, \mathbf{x}), \quad \text{(by Lemma 31)}
\end{aligned} \tag{37}$$

and

$$\begin{aligned}
\mathrm{PA} \vdash \forall\mathbf{x}\ (\quad & \mathrm{CR}_\Omega[\rho^R_{e,\Omega}(\mathbf{x})](\mathbf{e}, \ulcorner\mathcal{G}^R_\Omega\urcorner, \mathbf{x})\\
\rightarrow\ & \mathrm{CR}_\Omega[\neg\mathrm{CR}_\Omega[\rho^R_{e,\Omega}(\mathbf{x})](\mathbf{e}, \ulcorner\mathcal{G}\urcorner, \mathbf{x})](\mathbf{e}^+, \ulcorner\mathcal{G}^{R+}\urcorner, \mathbf{x}) \quad \text{(by Lemma 36)}\\
\rightarrow\ & \mathrm{CR}_\Omega[\neg\rho^R_{e,\Omega}(\mathbf{x})](\mathbf{e}^{++}, \ulcorner\mathcal{G}^{R++}\urcorner, \mathbf{x}) \quad \text{(by Lemma 38 and Corollary 34 )}\\
\rightarrow\ & \mathrm{CR}_\Omega[\rho^R_{e,\Omega}(\mathbf{x})](\mathbf{e}, \ulcorner\mathcal{G}^R_\Omega\urcorner, \mathbf{x}) \wedge\ \mathrm{CR}_\Omega[\neg\rho^R_{e,\Omega}(\mathbf{x})](\mathbf{e}^{++}, \ulcorner\mathcal{G}^{\mathrm{CR}++}\urcorner, \mathbf{x})\\
\rightarrow\ & \mathrm{CR}_\Omega[\rho^R_{e,\Omega}(\mathbf{x})\ \vee\ \neg\rho^R_{e,\Omega}(\mathbf{x})](\mathbf{e}^{+++}, \ulcorner\mathcal{G}^{\mathrm{CR}+++}\urcorner, \mathbf{x})\ ). \quad \text{(by Lemma 32)}
\end{aligned} \tag{38}$$

For any formula set $\Psi \equiv \{\psi(\mathbf{a}) \mid a \in \mathbb{N}\}$,

$$\mathrm{PA} \vdash\ \forall\mathbf{x}\ (\ \rho^A_{e,\Omega}(\mathbf{x})\ \wedge\ \neg\rho^A_{e,\Omega}(\mathbf{x}) \rightarrow\ \psi(\mathbf{x})\ ), \tag{39}$$

$$\mathrm{PA} \vdash\ \forall\mathbf{x}\ (\ \neg(\ \rho^R_{e,\Omega}(\mathbf{x})\ \vee\ \neg\rho^R_{e,\Omega}(\mathbf{x})\ )\ \rightarrow\ \neg\psi(\mathbf{x})\ ). \tag{40}$$

Hence by Corollaries 33 and 34, for any $e^{+++} \in \mathbb{N}$, there exists $e^* \in \mathbb{N}$ such that

$$\begin{aligned}
\mathrm{PA} \vdash\ \forall\mathbf{x}\ (\quad & \mathrm{CA}_\Omega[\rho^A_{e,\Omega}(\mathbf{x})\ \wedge\ \neg\rho^A_{e,\Omega}(\mathbf{x})](\mathbf{e}^{+++}, \ulcorner\mathcal{G}^{A+++}\urcorner, \mathbf{x})\\
\rightarrow\ & \mathrm{CA}_\Omega[\psi(\mathbf{x})](\mathbf{e}^*, \ulcorner\Psi\urcorner, \mathbf{x})\ ),
\end{aligned} \tag{41}$$

$$\begin{aligned}
\mathrm{PA} \vdash\ \forall\mathbf{x}\ (\quad & \mathrm{CR}_\Omega[\rho^R_{e,\Omega}(\mathbf{x})\ \wedge\ \neg\rho^R_{e,\Omega}(\mathbf{x})](\mathbf{e}^{+++}, \ulcorner\mathcal{G}^{R+++}\urcorner, \mathbf{x})\\
\rightarrow\ & \mathrm{CR}_\Omega[\psi(\mathbf{x})](\mathbf{e}^*, \ulcorner\Psi\urcorner, \mathbf{x})\ ).
\end{aligned} \tag{42}$$

Therefore, for any $e \in \mathbb{N}$, there exists $e^* \in \mathbb{N}$ such that

$$\mathrm{PA} \vdash \ \forall \mathbf{x} \ \ ( \ \mathrm{CA}_{\Omega}[\rho^{A}_{e,\Omega}(\mathbf{x})](\mathrm{e}, \ulcorner \mathcal{G}^{A}_{\Omega} \urcorner, \mathbf{x}) \ \rightarrow \ \mathrm{CA}_{\Omega}[\psi(\mathbf{x})](\mathbf{e}^*, \ulcorner \Psi \urcorner, \mathbf{x}) \ ), \tag{43}$$

$$\mathrm{PA} \vdash \ \forall \mathbf{x} \ \ ( \ \mathrm{CR}_{\Omega}[\rho^{R}_{e,\Omega}(\mathbf{x})](\mathrm{e}, \ulcorner \mathcal{G}^{R}_{\Omega} \urcorner, \mathbf{x}) \ \rightarrow \ \mathrm{CR}_{\Omega}[\psi(\mathbf{x})](\mathbf{e}^*, \ulcorner \Psi \urcorner, \mathbf{x}) \ ). \tag{44}$$

That is,

$$\mathrm{PA} \vdash \ \forall \mathbf{x} \ \ ( \ \neg\mathrm{CA}_{\Omega}[\psi(\mathbf{x})](\mathbf{e}^*, \ulcorner \Psi \urcorner, \mathbf{x}) \ \ \rightarrow \ \neg\mathrm{CA}_{\Omega}[\rho^{A}_{e,\Omega}(\mathbf{x})](\mathrm{e}, \ulcorner \mathcal{G}^{A}_{\Omega} \urcorner, \mathbf{x}) \ ), \tag{45}$$

$$\mathrm{PA} \vdash \ \forall \mathbf{x} \ \ ( \ \neg\mathrm{CR}_{\Omega}[\psi(\mathbf{x})](\mathbf{e}^*, \ulcorner \Psi \urcorner, \mathbf{x}) \ \ \rightarrow \ \neg\mathrm{CR}_{\Omega}[\rho^{R}_{e,\Omega}(\mathbf{x})](\mathrm{e}, \ulcorner \mathcal{G}^{R}_{\Omega} \urcorner, \mathbf{x}) \ \ ). \tag{46}$$

In addition, from the property of Gödel sentences (Lemmas 37 and 38),

$$\mathrm{PA} \vdash \forall \mathbf{x} \ \ (\rho^{A}_{e,\Omega}(\mathbf{x}) \ \leftrightarrow \ \neg\mathrm{CA}_{\Omega}[\rho^{A}_{e,\Omega}(\mathbf{x})](\mathrm{e}, \ulcorner \mathcal{G}^{A}_{\Omega} \urcorner, \mathbf{x}) \ ),$$

$$\mathrm{PA} \vdash \forall \mathbf{x} \ \ (\neg\rho^{R}_{e,\Omega}(\mathbf{x}) \ \leftrightarrow \ \neg\mathrm{CR}_{\Omega}[\rho^{R}_{e,\Omega}(\mathbf{x})](\mathrm{e}, \ulcorner \mathcal{G}^{R}_{\Omega} \urcorner, \mathbf{x}) \ ).$$

Hence,

$$\mathrm{PA} \vdash \ \forall \mathbf{x} \ \ ( \ \neg\mathrm{CA}_{\Omega}[\psi(\mathbf{x})](\mathbf{e}^*, \ulcorner \Psi \urcorner, \mathbf{x}) \ \ \rightarrow \ \ \rho^{A}_{e,\Omega}(\mathbf{x}) \ ), \tag{47}$$

$$\mathrm{PA} \vdash \ \forall \mathbf{x} \ \ ( \ \neg\mathrm{CR}_{\Omega}[\psi(\mathbf{x})](\mathbf{e}^*, \ulcorner \Psi \urcorner, \mathbf{x}) \ \ \rightarrow \ \ \neg\rho^{R}_{e,\Omega}(\mathbf{x}) \ ). \tag{48}$$

Since $e^*$ is computed from $e$ in a manner similar to those used in the lemmas and corollaries in Section 5.1, there exists a primitive recursive function $h$ such that for any formula sets $\Psi$ and $\Omega$, and for any $e \in \mathbb{N}$,

$$\mathrm{PA} \vdash \ \forall \mathbf{x} \ \ ( \ \neg\mathrm{CA}_{\Omega}[\psi(\mathbf{x})](h(\mathbf{e}), \ulcorner \Psi \urcorner, \mathbf{x}) \ \ \rightarrow \ \ \rho^{A}_{e,\Omega}(\mathbf{x}) \ ),$$

$$\mathrm{PA} \vdash \ \forall \mathbf{x} \ \ ( \ \neg\mathrm{CR}_{\Omega}[\psi(\mathbf{x})](h(\mathbf{e}), \ulcorner \Psi \urcorner, \mathbf{x}) \ \ \rightarrow \ \ \neg\rho^{R}_{e,\Omega}(\mathbf{x}) \ ).$$

$\dashv$

**Lemma 44.** *Let $T$ be a consistent PT-extension of PA. Let assume that there exist $e \in \mathbb{N}$, $e^* \in \mathbb{N}$, $x \in \mathbb{N}$ and a $\Delta_1$-formula set $\Psi \equiv \{\psi(\mathbf{a}) \mid a \in \mathbb{N}\}$ such that*

$$\mathrm{PTM}_e(x) \vdash_T \ \neg\mathrm{CD}[\psi(\mathbf{x})](\mathbf{e}^*, \ulcorner \Psi \urcorner, \mathbf{x}). \tag{49}$$

*Then, there exists a primitive recursive function $s$ such that $\tilde{e} = s(e) \in \mathbb{N}$ and*

$$\mathfrak{N} \models \psi(\mathbf{x}) \quad \Rightarrow \quad \mathrm{PTM}^{\Theta^A[e^*]}_{\tilde{e}}(x) \triangleright_{v^A_{\Theta^A[e^*]}} \ \neg\mathrm{CA}_{\Psi}[\psi(\mathbf{x})](\mathbf{e}^*, \ulcorner \Psi \urcorner, \mathbf{x}), \tag{50}$$

$$\mathfrak{N} \models \neg\psi(\mathbf{x}) \quad \Rightarrow \quad \mathrm{PTM}^{\Theta^R[e^*]}_{\tilde{e}}(x) \triangleright_{v^R_{\Theta^R[e^*]}} \ \neg\mathrm{CR}_{\Psi}[\psi(\mathbf{x})](\mathbf{e}^*, \ulcorner \Psi \urcorner, \mathbf{x}), \tag{51}$$

*where $\Theta^A[e^*] \equiv \{\neg\mathrm{CA}_{\Psi}[\psi(\mathbf{a})](\mathbf{e}^*, \ulcorner \Psi \urcorner, \mathbf{a}) \mid a \in \mathbb{N}\}$, and $\Theta^R[e^*] \equiv \{\mathrm{CR}_{\Psi}[\psi(\mathbf{a})](\mathbf{e}^*, \ulcorner \Psi \urcorner, \mathbf{a}) \mid a \in \mathbb{N}\}$.*

*Proof.* PTM $\mathrm{U}_{\mathrm{PTM}}(\tilde{e}, \cdot)$ is constructed using PTM $\mathrm{U}_{\mathrm{PTM}}(e, \cdot)$ as follows:

- (Input: ) $(d, \#\Theta^A[e^*], x) \in \mathbb{N}^3$ or $(d, \#\Theta^R[e^*], x) \in \mathbb{N}^3$.
- (Output: ) accept or reject
- Simulate $\mathrm{U}_{\mathrm{PTM}}(e, (p, \#\Phi[e^*], x))$, and check whether its output is the Gödel number of a valid proof tree of $\neg\mathrm{CD}[\psi(\mathbf{x})](\mathbf{e}^*, \ulcorner \Psi \urcorner, \mathbf{x})$ by using $\mathrm{U}_{\mathrm{PTM}}(v_T, \cdot)$.
- Let input be $(d, \#\Theta^A[e^*], x)$. Then, accept if and only if it is a valid proof tree.

– Let input be $(d, \#\Theta^R[e^*], x)$. Then, reject if and only if it is a valid proof tree.

The running time of $\mathrm{U_{PTM}}(\tilde{e}, \cdot)$ is that of $\mathrm{U_{PTM}}(e, \cdot)$ plus polynomial-time in $|x|$. From the construction of $\mathrm{U_{PTM}}(\tilde{e}, \cdot)$, there exists a primitive recursive function $s$ such that $\tilde{e} = s(e)$.

Since $T$ is a consistent PT-extension of PA, if Eq. (49) holds,

$$\mathfrak{N} \models \neg\mathrm{CD}[\psi(\mathbf{x})](\mathbf{e}^*, \ulcorner\Psi\urcorner, \mathbf{x}).$$

Then,

$$\begin{array}{cl} & \mathfrak{N} \models \psi(\mathbf{x}) \\ \Rightarrow & \mathfrak{N} \models \psi(\mathbf{x}) \ \wedge \ \neg\mathrm{CD}[\psi(\mathbf{x})](\mathbf{e}^*, \ulcorner\Psi\urcorner, \mathbf{x}) \\ \Leftrightarrow & \mathfrak{N} \models \neg\text{PTM-Acc}(\mathbf{e}^*, \ulcorner\Psi\urcorner, \mathbf{x}) \ \wedge \ \psi(\mathbf{x}) \\ \Leftrightarrow & \mathfrak{N} \models \neg\text{PTM-Acc}(\mathbf{e}^*, \ulcorner\Psi\urcorner, \mathbf{x}) \ \wedge \ \mathfrak{N} \models \psi(\mathbf{x}) \\ \Leftrightarrow & \mathfrak{N} \models \neg\text{PTM-Acc}(\mathbf{e}^*, \ulcorner\Psi\urcorner, \mathbf{x}) \ \wedge \ \mathrm{PA} \vdash \psi(\mathbf{x}) \\ \Leftrightarrow & \mathfrak{N} \models \neg\text{PTM-Acc}(\mathbf{e}^*, \ulcorner\Psi\urcorner, \mathbf{x}) \ \wedge \ \mathrm{U_{PTM}}(v^A_{\Psi}, (d, \#\Psi, x)) \text{ (soundly) accepts} \\ \Leftrightarrow & \mathfrak{N} \models \neg\text{PTM-Acc}(\mathbf{e}^*, \ulcorner\Psi\urcorner, \mathbf{x}) \ \wedge \ \mathfrak{N} \models \mathrm{Acc}(\mathbf{v}^A_{\Psi}, \ulcorner\Psi\urcorner, \mathbf{x}) \\ \Leftrightarrow & \mathfrak{N} \models \neg\text{PTM-Acc}(\mathbf{e}^*, \ulcorner\Psi\urcorner, \mathbf{x}) \ \wedge \ \mathrm{Acc}(\mathbf{v}^A_{\Psi}, \ulcorner\Psi\urcorner, \mathbf{x}) \\ \Rightarrow & \mathfrak{N} \models \neg\mathrm{CA}_{\Psi}[\psi(\mathbf{x})](\mathbf{e}^*, \ulcorner\Psi\urcorner, \mathbf{x}) \end{array}$$

and

$$\begin{array}{cl} & \mathfrak{N} \models \neg\psi(\mathbf{x}) \\ \Rightarrow & \mathfrak{N} \models \neg\psi(\mathbf{x}) \ \wedge \ \neg\mathrm{CD}[\psi(\mathbf{x})](\mathbf{e}^*, \ulcorner\Psi\urcorner, \mathbf{x}) \\ \Leftrightarrow & \mathfrak{N} \models \text{PTM-Acc}(\mathbf{e}^*, \ulcorner\Psi\urcorner, \mathbf{x}) \ \wedge \ \neg\psi(\mathbf{x}) \ ) \\ \Leftrightarrow & \mathfrak{N} \models \text{PTM-Acc}(\mathbf{e}^*, \ulcorner\Psi\urcorner, \mathbf{x}) \ \wedge \ \mathfrak{N} \models \neg\psi(\mathbf{x}) \\ \Leftrightarrow & \mathfrak{N} \models \text{PTM-Acc}(\mathbf{e}^*, \ulcorner\Psi\urcorner, \mathbf{x}) \ \wedge \ \mathrm{PA} \vdash \neg\psi(\mathbf{x}) \\ \Leftrightarrow & \mathfrak{N} \models \text{PTM-Acc}(\mathbf{e}^*, \ulcorner\Psi\urcorner, \mathbf{x}) \ \wedge \ \mathrm{U_{PTM}}(v^R_{\Psi}, (d, \#\Psi, x)) \text{ (soundly) rejects} \\ \Leftrightarrow & \mathfrak{N} \models \text{PTM-Acc}(\mathbf{e}^*, \ulcorner\Psi\urcorner, \mathbf{x}) \ \wedge \ \mathfrak{N} \models \neg\mathrm{Acc}(\mathbf{v}^R_{\Psi}, \ulcorner\Psi\urcorner, \mathbf{x}) \\ \Leftrightarrow & \mathfrak{N} \models \text{PTM-Acc}(\mathbf{e}^*, \ulcorner\Psi\urcorner, \mathbf{x}) \ \wedge \ \neg\mathrm{Acc}(\mathbf{v}^R_{\Psi}, \ulcorner\Psi\urcorner, \mathbf{x}) \\ \Leftrightarrow & \mathfrak{N} \models \neg\mathrm{CR}_{\Psi}[\psi(\mathbf{x})](\mathbf{e}^*, \ulcorner\Psi\urcorner, \mathbf{x}). \end{array}$$

Therefore, if Eq. (49) holds,

$$\begin{array}{lcl} \mathfrak{N} \models \psi(\mathbf{x}) & \Rightarrow & \mathfrak{N} \models \neg\mathrm{CA}_{\Psi}[\psi(\mathbf{x})](\mathbf{e}^*, \ulcorner\Psi\urcorner, \mathbf{x}), \\ \mathfrak{N} \models \neg\psi(\mathbf{x}) & \Rightarrow & \mathfrak{N} \models \neg\mathrm{CR}_{\Psi}[\psi(\mathbf{x})](\mathbf{e}^*, \ulcorner\Psi\urcorner, \mathbf{x}). \end{array}$$

Since $\neg\mathrm{CA}_{\Psi}[\psi(\mathbf{x})](\mathbf{e}^*, \ulcorner\Psi\urcorner, \mathbf{x})$ and $\neg\mathrm{CR}_{\Psi}[\psi(\mathbf{x})](\mathbf{e}^*, \ulcorner\Psi\urcorner, \mathbf{x})$ are $\Delta_1$-formulas,

$$\begin{array}{lcl} \mathfrak{N} \models \psi(\mathbf{x}) & \Rightarrow & \mathrm{PA} \vdash \neg\mathrm{CA}_{\Psi}[\psi(\mathbf{x})](\mathbf{e}^*, \ulcorner\Psi\urcorner, \mathbf{x}), \\ \mathfrak{N} \models \neg\psi(\mathbf{x}) & \Rightarrow & \mathrm{PA} \vdash \neg\mathrm{CR}_{\Psi}[\psi(\mathbf{x})](\mathbf{e}^*, \ulcorner\Psi\urcorner, \mathbf{x}). \end{array}$$

Then, from the definition of $\mathrm{U}(v^A_{\Theta[e^*]}, \cdot)$ and $\mathrm{U}(v^R_{\Theta[e^*]}, \cdot)$ (see Definition 24), $\mathrm{U}(v^A_{\Theta[e^*]}, (d, \#\Theta^A[e^*], x))$ accepts if $\mathfrak{N} \models \psi(\mathbf{x})$, and $\mathrm{U}(v^A_{\Theta[e^*]}, (d, \#\Theta^R[e^*], x))$ rejects if $\mathfrak{N} \models \neg\psi(\mathbf{x})$.

On the other hand, from the construction of $\mathrm{U_{PTM}}(\tilde{e},\cdot)$, if Eq. (49) holds, $\mathrm{U_{PTM}}(\tilde{e},(d,\#\Theta^A[e^*],x))$ accepts, and $\mathrm{U_{PTM}}(\tilde{e},(d,\#\Theta^R[e^*],x))$ rejects. Thus,

$$\mathfrak{N} \models \psi(\mathbf{x}) \quad \Rightarrow \quad \mathrm{PTM}_{\tilde{e}}^{\Theta^A[e^*]}(x) \rhd_{v^A_{\Theta^A[e^*]}} \neg\mathrm{CA}_{\Psi}[\psi(\mathbf{x})](\mathbf{e}^*, \ulcorner\Psi\urcorner, \mathbf{x}),$$

$$\mathfrak{N} \models \neg\psi(\mathbf{x}) \quad \Rightarrow \quad \mathrm{PTM}_{\tilde{e}}^{\Theta^R[e^*]}(x) \rhd_{v^R_{\Theta^R[e^*]}} \neg\mathrm{CR}_{\Psi}[\psi(\mathbf{x})](\mathbf{e}^*, \ulcorner\Psi\urcorner, \mathbf{x}),$$

$\dashv$

**Theorem 45.** *Let $T$ be a consistent PT-extension of PA. For any set of $\Delta_1$-formulas $\Psi \equiv \{\psi(\mathbf{a}) \mid a \in \mathbb{N}\}$,*

$$\forall e \in \mathbb{N} \;\; \exists e^* \in \mathbb{N} \;\; \forall x \in \mathbb{N} \quad \mathrm{PTM}_e(x) \not\vdash_T \; \neg\mathrm{CD}[\psi(\mathbf{x})](\mathbf{e}^*, \ulcorner\Psi\urcorner, \mathbf{x}).$$

*Proof.* We assume that

there exist $e \in \mathbb{N}$ and a formula set $\Psi$ such that

$$\forall e^* \in \mathbb{N} \;\; \exists x \in \mathbb{N} \quad \mathrm{PTM}_e(x) \vdash_T \; \neg\mathrm{CD}[\psi(\mathbf{x})](\mathbf{e}^*, \ulcorner\Psi\urcorner, \mathbf{x}). \tag{52}$$

Then, from Lemma 44, we can construct $\mathrm{U_{PTM}}(\tilde{e},\cdot)$ using $\mathrm{U_{PTM}}(e,\cdot)$ such that $\tilde{e} = s(e)$ and $\mathrm{U_{PTM}}(\tilde{e},\cdot)$ satisfy Eqs. (50) and (51).

Then, there exists a primitive recursive function $t$ such that $e' = t(\tilde{e}) \in \mathbb{N}$ and PTM $\mathrm{U_{PTM}}(e',\cdot)$ is constructed using PTM $\mathrm{U_{PTM}}(\tilde{e},\cdot)$ as follows:

- (Input: ) $(d, \#\mathcal{G}^A_{\Theta^A[h(e')]}, a) \in \mathbb{N}^3$ or $(d, \#\mathcal{G}^R_{\Theta^R[h(e')]}, a) \in \mathbb{N}^3$ where $\mathcal{G}^A_{\Theta^A[h(e')]} \equiv \{\rho^A_{e',\Theta^A[h(e')]}(\mathbf{a}) \mid a \in \mathbb{N}\}$ and $\mathcal{G}^R_{\Theta^R[h(e')]} \equiv \{\rho^R_{e',\Theta^R[h(e')]}(\mathbf{a}) \mid a \in \mathbb{N}\}$. (For the definitions of $\Theta^A[\cdot]$ and $\Theta^R[\cdot]$, see Lemma 44.)
- (Output: ) accept or reject
- First, read its own code, $e' \in \mathbb{N}$ via the recursion theorem (Proposition 17).
- If input is $(d, \#\mathcal{G}^A_{\Theta^A[h(e')]}, a)$, then simulate $\mathrm{U_{PTM}}(\tilde{e},(d,\#\Theta^A[h(e')],a))$, and accept if and only if $\mathrm{U_{PTM}}(\tilde{e},(d,\#\Theta^A[h(e')],a))$ accepts.
- If input is $(d, \#\mathcal{G}^R_{\Theta^R[h(e')]}, a)$, then simulate $\mathrm{U_{PTM}}(\tilde{e},(d,\#\Theta^R[h(e')],a))$, and reject if and only if $\mathrm{U_{PTM}}(\tilde{e},(d,\#\Theta^R[h(e')],a))$ rejects.

The running time of $\mathrm{U_{PTM}}(e',\cdot)$ is that of $\mathrm{U_{PTM}}(e,\cdot)$ plus polynomial-time in $|a|$.

By substituting $\Theta[h(e')]$ for $\Omega$, in Eqs. (33) and (34), we obtain that for any formula set $\Psi \equiv \{\psi(\mathbf{a}) \mid a \in \mathbb{N}\}$ and for any $e' \in \mathbb{N}$,

$$\mathrm{PA} \vdash \;\forall \mathbf{x} \;\; (\; \neg\mathrm{CA}_{\Theta^A[h(e')]}[\psi(\mathbf{x})](h(\mathbf{e}'), \ulcorner\Psi\urcorner, \mathbf{x}) \quad \to \quad \rho^A_{e',\Theta^A[h(e')]}(\mathbf{x}) \;), \tag{53}$$

$$\mathrm{PA} \vdash \;\forall \mathbf{x} \;\; (\; \neg\mathrm{CR}_{\Theta^R[h(e')]}[\psi(\mathbf{x})](h(\mathbf{e}'), \ulcorner\Psi\urcorner, \mathbf{x}) \quad \to \quad \neg\rho^R_{e',\Theta^R[h(e')]}(\mathbf{x}) \;). \tag{54}$$

For all $e \in \mathbb{N}$ and all $a \in \mathbb{N}$,

$$\mathfrak{N} \models \; \psi(\mathbf{x}) \quad \Rightarrow \quad f(\#\Psi, \mathrm{Size}_{\Psi}(a)) < f(\#\Theta^A[e], \mathrm{Size}_{\Theta^A[e]}(a)),$$

$$\mathfrak{N} \models \; \neg\psi(\mathbf{x}) \quad \Rightarrow \quad f(\#\Psi, \mathrm{Size}_{\Psi}(a)) < f(\#\Theta^R[e], \mathrm{Size}_{\Theta^R[e]}(a))$$

(for the definition of function $f$, see Definition 24).

Therefore,

$$
\begin{aligned}
&\mathfrak{N} \models\ \psi(\mathbf{x}) \quad \Rightarrow \\
&\qquad \mathrm{PA} \vdash\ \forall \mathbf{x}\ \ (\ \neg \mathrm{CA}_{\Psi}[\psi(\mathbf{x})](h(\mathbf{e}'), \ulcorner \Psi \urcorner, \mathbf{x}) \quad \leftrightarrow \quad \neg \mathrm{CA}_{\Theta^A[h(e')]}[\psi(\mathbf{x})](h(\mathbf{e}'), \ulcorner \Psi \urcorner, \mathbf{x})\ ), \quad (55)\\
&\mathfrak{N} \models\ \neg\psi(\mathbf{x}) \quad \Rightarrow \\
&\qquad \mathrm{PA} \vdash\ \forall \mathbf{x}\ \ (\ \neg \mathrm{CR}_{\Psi}[\psi(\mathbf{x})](h(\mathbf{e}'), \ulcorner \Psi \urcorner, \mathbf{x}) \quad \leftrightarrow \quad \neg \mathrm{CR}_{\Theta^R[h(e')]}[\psi(\mathbf{x})](h(\mathbf{e}'), \ulcorner \Psi \urcorner, \mathbf{x})\ ). \quad (56)
\end{aligned}
$$

Hence, if Eq. (52) holds, then by applying Eqs. (50), (51), (53), (54), (55), and (56),

$$
\begin{aligned}
\mathfrak{N} \models\ \psi(\mathbf{x}) \quad &\Rightarrow \quad \mathrm{U}(v^A_{\Theta^A[h(e')]}, (d, \#\mathcal{G}^A_{\Theta^A[h(e')]}, a)) \text{ accepts},\\
\mathfrak{N} \models\ \neg\psi(\mathbf{x}) \quad &\Rightarrow \quad \mathrm{U}(v^R_{\Theta^R[h(e')]}, (d, \#\mathcal{G}^R_{\Theta^R[h(e')]}, a)) \text{ rejects}.
\end{aligned}
$$

On the other hand, if Eq. (52) holds, from the construction of $\mathrm{U}_{\mathrm{PTM}}(e', \cdot)$ and $\mathrm{U}_{\mathrm{PTM}}(\tilde{e}, \cdot)$,

$$
\begin{aligned}
&\mathrm{U}_{\mathrm{PTM}}(e', (d, \#\mathcal{G}^A_{\Theta^A[h(e')]}, a)) \text{ accepts},\\
&\mathrm{U}_{\mathrm{PTM}}(e', (d, \#\mathcal{G}^R_{\Theta^R[h(e')]}, a)) \text{ rejects}.
\end{aligned}
$$

Hence, if Eq. (52) holds, for a formula set $\Psi$,

$$
\begin{aligned}
\mathfrak{N} \models\ \psi(\mathbf{x}) \quad &\Rightarrow \quad \exists e' \in \mathbb{N}\ \ \exists x \in \mathbb{N} \quad \mathrm{PTM}_{e'}^{\mathcal{G}^A_{\Phi[h(e')]}}(x)\ \rhd_{v^A_{\Phi[h(e')]}}\ \ \rho^A_{e',\Phi[h(e')]}(\mathbf{x}),\\
\mathfrak{N} \models\ \neg\psi(\mathbf{x}) \quad &\Rightarrow \quad \exists e' \in \mathbb{N}\ \ \exists x \in \mathbb{N} \quad \mathrm{PTM}_{e'}^{\mathcal{G}^R_{\Phi[h(e')]}}(x)\ \rhd_{v^R_{\Phi[h(e')]}}\ \ \neg\rho^R_{e',\Phi[h(e')]}(\mathbf{x}).
\end{aligned}
$$

This contradicts Corollaries 41 and 42. Therefore, Eq. (52) does not hold for $e^* = h(e') = h(t(s(e)))$. That is, there exists a primitive recursive function $g$ such that for any $e \in \mathbb{N}$, for any $\Psi$, and for any $x \in \mathbb{N}$

$$
\mathrm{PTM}_e(x) \not\vdash_T\ \neg \mathrm{CD}[\psi(\mathbf{x})](g(\mathbf{e}), \ulcorner \Psi \urcorner, \mathbf{x}),
$$

where $g(e) = h(t(s(e)))$.

$\dashv$

**Corollary 46.** *Let $T$ be a consistent PT-extension of PA. There exists a primitive recursive function $g$ such that for any set of $\Delta_1$-formulas $\Psi \equiv \{\psi(\mathbf{a}) \mid a \in \mathbb{N}\}$,*

$$
\begin{aligned}
\mathfrak{N} \models \psi(\mathbf{x}) \quad &\Rightarrow \quad \forall e \in \mathbb{N}\ \ \forall x \in \mathbb{N} \quad \mathrm{PTM}_e(x) \not\vdash_T\ \neg \mathrm{CA}[\psi(\mathbf{x})](g(\mathbf{e}), \ulcorner \Psi \urcorner, \mathbf{x}),\\
\mathfrak{N} \models \neg\psi(\mathbf{x}) \quad &\Rightarrow \quad \forall e \in \mathbb{N}\ \ \forall x \in \mathbb{N} \quad \mathrm{PTM}_e(x) \not\vdash_T\ \neg \mathrm{CR}[\psi(\mathbf{x})](g(\mathbf{e}), \ulcorner \Psi \urcorner, \mathbf{x}).
\end{aligned}
$$

## 6 Formalization of P≠NP and a Super-Polynomial-Time Lower Bound

We now introduce the notations and definitions necessary to consider the P≠NP problem in this paper. We omit the fundamental concepts and definitions regarding P and NP (see [23] for them).

## 6.1 $\overline{\mathbf{P{\neq}NP}}$

**Definition 47.** *Let $R_{\text{3SAT}} \subset \mathbb{N}$ be a relation such that $x \in R_{\text{3SAT}}$ if and only if there exists a satisfiable 3CNF formula $\varphi$ and $x = \#\varphi$. Let* $\text{SAT}(\mathbf{x})$ *be a formula in* PA*and* $\text{SAT}(\mathbf{x})$ *represent relation $R_{\text{3SAT}}$ in* PA. *(see Section 2.4 for representability.) That is, for every $a \in \mathbb{N}$*

$$\begin{aligned} a \in R_{\text{3SAT}} \quad &\Rightarrow \quad \text{PA} \vdash \text{SAT}(\mathbf{a}), \\ a \notin R_{\text{3SAT}} \quad &\Rightarrow \quad \text{PA} \vdash \neg\text{SAT}(\mathbf{a}). \end{aligned}$$

*Let $\mathcal{SAT}$ be a set of formulas in* PA, $\{\text{SAT}(\mathbf{a}) \mid a \in \mathbb{N}\}$*, and co-$\mathcal{SAT}$ be a set of formulas in* PA, $\{\neg\text{SAT}(\mathbf{a}) \mid a \in \mathbb{N}\}$. *Let* $\text{Size}_{\mathcal{SAT}}(a)$ *and* $\text{Size}_{co\text{-}\mathcal{SAT}}(a)$ *be $|a|$. Let $\mathcal{DS}$ be $\mathcal{SAT} \cup co\text{-}\mathcal{SAT}$.*

**Definition 48.** *Let theory $T$ be a PT-extension of PA. For $e \in \mathbb{N}$, let* $\text{U}_{\text{PTM}}(e, (d, \#\mathcal{SAT}, \cdot))$*, on input $x \in \mathbb{N}$, output one bit decision, whether $x \in R_{\text{3SAT}}$ or $x \notin R_{\text{3SAT}}$; in other words,* $\text{SAT}(\mathbf{x})$ *is true or false.*

We then define a formula that characterizes the fact that a PTM, $\text{U}_{\text{PTM}}(e, (\cdot, \cdot))$, given $x \in \mathbb{N}$, can solve the problem of deciding the truth/falsity of formula $\text{SAT}(\mathbf{x})$.

**Definition 49.** *Let theory $T$ be a PT-extension of* PA.

$$\text{DecSAT}(\mathbf{e}, \mathbf{x})$$

*denotes a $\Delta_1$-formula in* PA, *which represents the following primitive recursive relation on $(e, x) \in \mathbb{N}^2$ such that*

$$\text{U}_{\text{PTM}}(e, (d, \#\mathcal{SAT}, x)) \text{ accepts} \;\Leftrightarrow\; x \in R_{\text{3SAT}}.$$

*More precisely, let*

$$\text{DecSAT}(\mathbf{e}, \mathbf{x}) \quad \equiv \quad \text{CD}[\text{SAT}(\mathbf{x})](\mathbf{e}, \ulcorner \mathcal{SAT} \urcorner, \mathbf{x})$$

*(For the definition of this notation, see Section 4.2). This primitive recursive relation on $(e, x)$ means whether the decision (on $x \in R_{\text{3SAT}}$) of PTM* $\text{U}_{\text{PTM}}(e, (d, \#\mathcal{SAT}, \cdot))$ *is correct or not.*

We now introduce the Cook-Levin Theorem [23], which characterizes the P vs NP problem by the satisfiability problem, 3SAT (an NP-complete problem).

**Proposition 50.** *(Cook-Levin Theorem)*

$$\exists e \in \mathbb{N} \;\; \exists n \in \mathbb{N} \;\; \forall x \geq n \quad (\; \text{U}_{\text{PTM}}(e, (d, \#\mathcal{SAT}, x)) \text{ accepts} \;\Leftrightarrow\; x \in R_{\text{3SAT}} \;)$$

*if and only if $\;P = NP$.*

**Lemma 51.** *Let theory $T$ be a consistent PT-extension of PA.*

$$\forall e \in \mathbb{N} \;\; \forall n \in N \;\; \exists x \geq n \quad T \vdash \neg\text{DecSAT}(\mathbf{e}, \mathbf{x}),$$

*if and only if $\;P{\neq}NP$.*

*Proof.* From the representability theorem (Proposition 2) regarding formula $\text{DecSAT}(\mathbf{e}, \mathbf{x})$, the statement of this lemma is equivalent to

$$\exists e \in \mathbb{N} \;\; \exists n \in N \;\; \forall x \geq n \quad T \vdash \text{DecSAT}(\mathbf{e}, \mathbf{x}),$$

if and only if P=NP.

Thus, we obtain the statement of this lemma from the definitions of formula $\text{DecSAT}(\mathbf{e}, \mathbf{x})$, and Proposition 50.

$\dashv$

**Lemma 52.** *Let theory $T$ be a PT-extension of* PA *and $\omega$-consistent.*

$$\forall e \in \mathbb{N} \ \ \forall n \in N \quad T \vdash \ \exists \mathbf{x} \geq \mathbf{n} \quad \neg \mathrm{DecSAT}(\mathbf{e}, \mathbf{x}),$$

*if and only if $P \neq NP$.*

*Proof.*
(If:)
When

$$\exists x \in \mathbb{N} \quad T \vdash \ \neg \mathrm{DecSAT}(\mathbf{e}, \mathbf{x}),$$

the following holds

$$T \vdash \ \exists \mathbf{x} \ \ \neg \mathrm{DecSAT}(\mathbf{e}, \mathbf{x}).$$

(Only if:)
The following claim is obtained from $\omega$-consistency.

*Claim.* Let theory $T$ be a PT-extension of PA and $\omega$-consistent.

$$\begin{aligned} &\exists e \in \mathbb{N} \ \ \exists n \in N \ \ \forall x \geq n \quad T \vdash \ \mathrm{DecSAT}(\mathbf{e}, \mathbf{x}). \\ \Rightarrow \quad &\exists e \in \mathbb{N} \ \ \exists n \in N \quad T \nvdash \ \exists \mathbf{x} \geq \mathbf{n} \quad \neg \mathrm{DecSAT}(\mathbf{e}, \mathbf{x}). \end{aligned}$$

From Definition 53, if P=NP,

$$\exists e \in \mathbb{N} \ \ \exists n \in N \ \ \forall x \geq n \quad T \vdash \ \mathrm{DecSAT}(\mathbf{e}, \mathbf{x}).$$

We then have the following equation from the above-mentioned claim,

$$\exists e \in \mathbb{N} \ \ \exists n \in N \quad T \nvdash \ \exists \mathbf{x} \geq \mathbf{n} \quad \neg \mathrm{DecSAT}(\mathbf{e}, \mathbf{x}).$$

Hence, if

$$\forall e \in \mathbb{N} \ \ \forall n \in N \quad T \vdash \ \exists \mathbf{x} \geq \mathbf{n} \quad \neg \mathrm{DecSAT}(\mathbf{e}, \mathbf{x}),$$

then P$\neq$NP.

$\dashv$

**Note:** This lemma implies

$$\begin{aligned} &\forall e \in \mathbb{N} \ \ \forall n \in N \quad T \vdash \ \exists \mathbf{x} \geq \mathbf{n} \quad \neg \mathrm{DecSAT}(\mathbf{e}, \mathbf{x}) \\ \Leftrightarrow \quad &\forall e \in \mathbb{N} \ \ \forall n \in N \ \ \exists x \geq n \quad T \vdash \ \neg \mathrm{DecSAT}(\mathbf{e}, \mathbf{x}). \end{aligned}$$

**Definition 53.** *Let $\overline{P \neq NP}$ be a formula (sentence) in* PA *such that*

$$\overline{P \neq NP} \quad \equiv \quad \forall \mathbf{e} \ \ \forall \mathbf{n} \ \ \exists \mathbf{x} \geq \mathbf{n} \quad \neg \mathrm{DecSAT}(\mathbf{e}, \mathbf{x}).$$

**Lemma 54.**

$$\mathfrak{N} \models \ \overline{P \neq NP},$$

*if and only if $P \neq NP$.*

*Proof.*

$$\begin{aligned}
& \mathrm{P}\neq\mathrm{NP} \\
\Leftrightarrow\ & \forall e \in \mathbb{N}\ \ \forall n \in N\ \ \exists x \geq n \quad \mathrm{PA} \vdash\ \neg\mathrm{DecSAT}(\mathbf{e},\mathbf{x}) \qquad \text{(from Lemma 51)} \\
\Leftrightarrow\ & \forall e \in \mathbb{N}\ \ \forall n \in N\ \ \exists x \geq n \quad \mathfrak{N} \models\ \neg\mathrm{DecSAT}(\mathbf{e},\mathbf{x}) \qquad \text{(since } \neg\mathrm{DecSAT}(\mathbf{e},\mathbf{x}) \text{ is } \Delta_1\text{-formula)} \\
\Leftrightarrow\ & \mathfrak{N} \models\ \overline{\mathrm{P}\neq\mathrm{NP}}.
\end{aligned}$$

⊣

**Lemma 55.** *Let theory $T$ be a PT-extension of* PA *and $\omega$-consistent. If*

$$T \vdash\ \overline{P\neq NP},$$

*then $P\neq NP$.*

*Proof.* If

$$T \vdash\ \forall \mathbf{e}\ \ \forall \mathbf{n}\ \ \exists \mathbf{x} \geq \mathbf{n} \quad \neg\mathrm{DecSAT}(\mathbf{e},\mathbf{x}),$$

then

$$\forall e \in \mathbb{N}\ \ \forall n \in N \quad T \vdash\ \exists \mathbf{x} \geq \mathbf{n} \quad \neg\mathrm{DecSAT}(\mathbf{e},\mathbf{x}).$$

We then obtain P≠NP by Lemma 52.

⊣

### 6.2 Formalization of a Super-Polynomial-Time Lower Bound

This section shows a formalization of a super-polynomial-time lower bound in PA in a manner similar to $\overline{\mathrm{P}\neq\mathrm{NP}}$.

**Definition 56.** *Let $L$ be a language (a set of binary strings) in PSPACE. Let $R_L \subset \mathbb{N}$ be a relation such that $x \in R_L$ if and only if $[x] \in L$. Let $\varphi_L(\mathbf{x})$ be a formula in* PA*and $\varphi_L(\mathbf{x})$ represent relation $R_L$ in* PA*. (see Section 2.4 for representability.) That is, for every $a \in \mathbb{N}$*

$$\begin{aligned}
a \in R_L &\quad\Rightarrow\quad \mathrm{PA} \vdash \varphi_L(\mathbf{a}), \\
a \notin R_L &\quad\Rightarrow\quad \mathrm{PA} \vdash \neg\varphi_L(\mathbf{a}).
\end{aligned}$$

*Let $\Phi_L$ be a set of formulas in* PA*, $\{\varphi_L(\mathbf{a}) \mid a \in \mathbb{N}\}$, and $\mathrm{Size}_{\Phi_L}(a)$ be $|a|$.*

**Definition 57.**

$$\forall e \in \mathbb{N}\ \ \forall n \in \mathbb{N}\ \ \exists x \geq n \quad \neg(\ \mathrm{U_{PTM}}(e,(d,\#\Phi_L,x)) \text{ accepts } \Leftrightarrow\ x \in R_L\ )$$

*if and only if $L$ has a super-polynomial-time computational lower bound.*

**Lemma 58.** *Let theory $T$ be a consistent PT-extension of PA.*

$$\forall e \in \mathbb{N}\ \ \forall n \in N\ \ \exists x \geq n \quad T \vdash\ \neg\mathrm{CD}[\varphi_L(\mathbf{x})](\mathbf{e},\ulcorner\Phi_L\urcorner,\mathbf{x}),$$

*if and only if $L$ has a super-polynomial-time computational lower bound.*

*Proof.* This is obtained from the representability theorem (Proposition 2) regarding formula $\mathrm{CD}[\varphi_L(\mathbf{x})](\mathbf{e}, \ulcorner \Phi_L \urcorner, \mathbf{x})$, and the definition of this formula notation, CD (see Section 4.2).

$\dashv$

The following lemmas can be proven in a manner similar to those used in Lemmas 52 and 55.

**Lemma 59.** *Let theory $T$ be a PT-extension of* PA *and $\omega$-consistent.*

$$\forall e \in \mathbb{N} \ \ \forall n \in N \quad T \vdash \ \exists \mathbf{x} \geq \mathbf{n} \quad \neg\mathrm{CD}[\varphi_L(\mathbf{x})](\mathbf{e}, \ulcorner \Phi_L \urcorner, \mathbf{x}),$$

*if and only if $L$ has a super-polynomial-time computational lower bound.*

**Lemma 60.** *Let theory $T$ be a PT-extension of PA and $\omega$-consistent. If*

$$T \vdash \ \forall \mathbf{e} \ \ \forall \mathbf{n} \ \ \exists \mathbf{x} \geq \mathbf{n} \quad \neg\mathrm{CD}[\varphi_L(\mathbf{x})](\mathbf{e}, \ulcorner \Phi_L \urcorner, \mathbf{x}),$$

*then $L$ has a super-polynomial-time computational lower bound.*

## 7 Unprovability of P$\neq$NP and Super-Polynomial-Time Lower Bounds

This section shows that there exists no formal proof of $\overline{\mathrm{P}\neq\mathrm{NP}}$ in $T$, if $T$ is a consistent PT-extension of PA and PTM-$\omega$-consistent for $\Delta_2^P$. This result is based on the second incompleteness theorem of polynomial-time decisions, Theorem 45.

### 7.1 PTM-$\omega$-Consistency

**Definition 61.** *Formula $\varphi(\mathbf{x})$ in PA is called $\Sigma_i^P$ $(i = 1, 2, \ldots)$ if there exists a formula $\psi(\mathbf{x})$ in PA such that*

$$\mathrm{PA} \vdash \ \forall \mathbf{x} \ (\varphi(\mathbf{x}) \ \leftrightarrow \ \psi(\mathbf{x}) \ ),$$

$$\psi(\mathbf{x}) \equiv \ \exists \mathbf{w}_1 < \mathbf{2}^{|\mathbf{x}|^{\mathbf{c}_1}} \ \forall \mathbf{w}_2 < \mathbf{2}^{|\mathbf{x}|^{\mathbf{c}_2}} \ \cdots \ Q_i \mathbf{w}_i < \mathbf{2}^{|\mathbf{x}|^{\mathbf{c}_i}} \ \psi_0(\mathbf{x}, \mathbf{w}_1, \ldots, \mathbf{w}_i),$$

*where $Q_i$ is $\forall$ or $\exists$, $\psi_0(\mathbf{x}, \mathbf{w}_1, \ldots, \mathbf{w}_i)$ is a formula that represents a polynomial-time relation over $(x, w_1, \ldots, w_i)$, $c_j$ $(0 \leq j \leq i)$ is a constant (in $|x|$).*

*Similarly, formula $\varphi(\mathbf{x})$ in PA is called $\Pi_i^P$ $(i = 1, 2, \ldots)$ if there exists a formula $\psi(\mathbf{x})$ in PA such that*

$$\mathrm{PA} \vdash \ \forall \mathbf{x} \ (\varphi(\mathbf{x}) \ \leftrightarrow \ \psi(\mathbf{x}) \ ),$$

$$\psi(\mathbf{x}) \equiv \ \forall \mathbf{w}_1 < \mathbf{2}^{|\mathbf{x}|^{\mathbf{c}_1}} \ \exists \mathbf{w}_2 < \mathbf{2}^{|\mathbf{x}|^{\mathbf{c}_2}} \ \cdots \ Q_i \mathbf{w}_i < \mathbf{2}^{|\mathbf{x}|^{\mathbf{c}_i}} \ \psi_0(\mathbf{x}, \mathbf{w}_1, \ldots, \mathbf{w}_i).$$

*Formula $\varphi(\mathbf{x})$ in PA is called $\Delta_i^P$ $(i = 1, 2, \ldots)$ if $\varphi(\mathbf{x})$ is $\Sigma_i^P$ and $\Pi_i^P$.*

*Formula $\varphi(\mathbf{x})$ in PA is called QBF if there exists a formula $\psi(\mathbf{x})$ in PA such that*

$$\mathrm{PA} \vdash \ \forall \mathbf{x} \ (\varphi(\mathbf{x}) \ \leftrightarrow \ \psi(\mathbf{x}) \ ),$$

$$\psi(\mathbf{x}) \equiv \ \forall \mathbf{w}_1 < \mathbf{2}^{|\mathbf{x}|^{\mathbf{c}_1}} \ \exists \mathbf{w}_2 < \mathbf{2}^{|\mathbf{x}|^{\mathbf{c}_2}} \ \cdots \ Q_i \mathbf{w}_i < \mathbf{2}^{|\mathbf{x}|^{\mathbf{c}_k}} \ \psi_0(\mathbf{x}, \mathbf{w}_1, \ldots, \mathbf{w}_k),$$

*where $k \equiv |x|^c$ for a constant $c$.*

**Definition 62.** *(*PTM-$\omega$*-consistency) Let theory $S$ be a PT-extension of theory $T$. $S$ is* PTM-$\omega$*-inconsistent for $\Delta_1$-formula $\varphi(\mathbf{e}^*,\mathbf{x})$ over $T$, if the following two conditions hold simultaneously.*

$$\forall e \in \mathbb{N}\ \exists e^* \in \mathbb{N}\ \exists \ell \in \mathbb{N}\ \forall n \geq \ell\ \forall c \in \mathbb{N} \quad \mathrm{PTM}_e(n)\ \not\vdash_T \exists \mathbf{x}\ (\mathbf{n} \leq \mathbf{x} < \mathbf{n} + |\mathbf{n}|^{\mathbf{c}})\ \ \varphi(\mathbf{e}^*,\mathbf{x}), \quad (57)$$

$$\exists e \in \mathbb{N}\ \forall e^* \in \mathbb{N}\ \forall \ell \in \mathbb{N}\ \exists n \geq \ell \quad \mathrm{PTM}_e(n)\ \vdash_S \exists \mathbf{x} \geq \mathbf{n}\ \ \varphi(\mathbf{e}^*,\mathbf{x}). \quad (58)$$

*Here,* $\mathrm{Size}_{\Phi[c]}(n) = |n|^{c+1}$*, and* $\Phi[c] \equiv \{\exists \mathbf{x}\ (\mathbf{a} \leq \mathbf{x} < \mathbf{n} + |\mathbf{n}|^{\mathbf{c}})\ \ \varphi(\mathbf{e}^*,\mathbf{x}) \mid a \in \mathbb{N}\}$*.*

*Theory $S$ is PTM-$\omega$-consistent for $\varphi(\mathbf{e}^*,\mathbf{x})$ over $T$, if theory $S$ is not PTM-$\omega$-inconsistent for $\varphi(\mathbf{e}^*,\mathbf{x})$ over $T$.*

*Theory $S$ is PTM-$\omega$-consistent for $\Sigma_i^P$ ($\Pi_i^P$, $\Delta_i^P$, resp.) over $T$, if $S$ is PTM-$\omega$-consistent for any $\Sigma_i^P$ ($\Pi_i^P$, $\Delta_i^P$, resp.) formula $\varphi(\mathbf{e}^*,\mathbf{x})$ over $T$.*

*Theory $T$ is PTM-$\omega$-consistent for $\varphi(\mathbf{e}^*,\mathbf{x})$ ($\Sigma_i^P$, $\Pi_i^P$, $\Delta_i^P$, resp.), if $T$ is PTM-$\omega$-consistent for $\varphi(\mathbf{e}^*,\mathbf{x})$ ($\Sigma_i^P$, $\Pi_i^P$, $\Delta_i^P$, resp.) over $T$.*

The following definition is equivalent to the above: Theory $S$ is PTM-$\omega$-consistent for $\varphi(\mathbf{e}^*,\mathbf{x})$ over $T$, if the following condition holds.

$$\begin{aligned} & \forall e \in \mathbb{N}\ \ \exists e^* \in \mathbb{N}\ \ \exists \ell \in \mathbb{N}\ \ \forall n \geq \ell\ \ \forall c \in \mathbb{N} \quad \mathrm{PTM}_e(n)\ \not\vdash_T\ \exists \mathbf{x}\ (\mathbf{n} \leq \mathbf{x} < \mathbf{n} + |\mathbf{n}|^{\mathbf{c}})\ \ \varphi(\mathbf{e}^*,\mathbf{x}) \\ \Rightarrow\ & \forall e \in \mathbb{N}\ \ \exists e^* \in \mathbb{N}\ \ \exists \ell \in \mathbb{N}\ \ \forall n \geq \ell \quad \mathrm{PTM}_e(n)\ \not\vdash_S\ \exists \mathbf{x} \geq \mathbf{n}\ \ \varphi(\mathbf{e}^*,\mathbf{x}). \end{aligned} \quad (59)$$

In the remarks below, we consider only the PTM-$\omega$-consistency of theory $T$, not the PTM-$\omega$-consistency of theory $S$ over $T$, since the PTM-$\omega$-consistency of $S$ over $T$ follows similarly in each remark.

**Remark 1** (Restriction of the related formulas of PTM-$\omega$-consistency) PTM-$\omega$-consistency is defined only for $\Sigma_i^P$, $\Pi_i^P$ or $\Delta_i^P$-formulas. This restriction is introduced from the fact that if $\varphi(\mathbf{e}^*,\mathbf{x})$ has a bounded quantifier $Qw < a$ with $|a| = 2^{|x|^c}$ for a constant $c$, then no PTM can even read $\#\mathbf{a}$ numeralwise. Since the notion of PTM-$\omega$-consistency is introduced to characterize a property of the provability of a PTM in theory $T$, such a restriction seems reasonable.

Actually, the proof of P $\neq$ EXP may imply that PA or a PT-extension of PA is PTM-$\omega$-inconsistent for formula $\varphi(\mathbf{e}^*,\mathbf{x})$ corresponding to the *formulation* of P $\neq$ EXP, which has a bounded quantifier with $\exists w < a$ with $|a| = 2^{|x|^c}$ for constant $c$. (In other words, the asymptotic polynomial-time unprovability of P $\neq$ EXP does not imply the formal unprovability of P $\neq$ EXP.)

**Remark 2** (Inequivalence of PTM-$\omega$-consistency and $\omega$-consistency) PTM-$\omega$-consistency and $\omega$-consistency do not imply each other.

First, we show that PTM-$\omega$-consistency does not imply $\omega$-consistency. If we assume that PTM-$\omega$-consistency of $T$ for $\varphi(\mathbf{e}^*,\mathbf{x})$ implies $\omega$-consistency of $T$ for $\varphi(\mathbf{e}^*,\mathbf{x})$, PTM-$\omega$-consistency of $T$ for $\varphi(\mathbf{e}^*,\mathbf{x})$ implies consistency of $T$, since if $T$ is inconsistent, $T$ is $\omega$-inconsistent for $\varphi(\mathbf{e}^*,\mathbf{x})$. That is, the inconsistency of $T$ implies PTM-$\omega$-*inconsistency* of $T$ for $\varphi(\mathbf{e}^*,\mathbf{x})$. However, the inconsistency of $T$ implies PTM-$\omega$-*consistency* of $T$ for any formula, since if $T$ is inconsistent, $T$ can prove any sentence and Eq.(57) does not hold, which implies that $T$ cannot be PTM-$\omega$-inconsistent. This is contradiction. Therefore, PTM-$\omega$-consistency does not imply $\omega$-consistency.

Next, we show that $\omega$-consistency does not imply PTM-$\omega$-consistency. Here, we assume that

$$\mathfrak{N} \models \overline{\mathrm{P}{\neq}\mathrm{NP}}.$$

It follows that theory $T = \mathrm{PA} + \overline{\mathrm{P{\neq}NP}}$ is $\omega$-consistent since PA is $\omega$-consistent, and clearly

$$T \vdash \overline{\mathrm{P{\neq}NP}}.$$

We now assume that $T$ is PTM-$\omega$-consistent for $\Delta_2^P$. Then,

$$T \nvdash \overline{\mathrm{P{\neq}NP}},$$

by Theorem 67. This is a contradiction. Therefore, $T$ is PTM-$\omega$-inconsistent for $\Delta_2^P$, while $T$ is $\omega$-consistent, if $\mathfrak{N} \models \overline{\mathrm{P{\neq}NP}}$. That is, $\omega$-consistency does not imply PTM-$\omega$-consistency, assuming that $\mathfrak{N} \models \overline{\mathrm{P{\neq}NP}}$.

**Remark 3** (Relationship between PTM-$\omega$-consistency and $\omega$-consistency) Although PTM-$\omega$-consistency and $\omega$-consistency do not imply each other, as described above, the computational resource unbounded version of PTM-$\omega$-consistency for $\Delta_1$-formulas is equivalent to $\omega$-consistency for $\Delta_1$-formulas.

Now we define a computational resource unbounded version of PTM-$\omega$-consistency, TM-$\omega$-consistency, as follows: Theory $T$ is TM-$\omega$-inconsistent for $\varphi(\mathbf{e}^*, \mathbf{x})$, if the following two conditions hold simultaneously.

$$\forall e \in \mathbb{N}\ \exists e^* \in \mathbb{N}\ \ \exists \ell \in \mathbb{N}\ \forall n \geq \ell\ \forall f \in \mathcal{R}\ \ \mathrm{TM}_e(n)\ \nvdash_T\ \exists \mathbf{x}\ (\mathbf{n} \leq \mathbf{x} < \mathbf{n} + f(|\mathbf{n}|))\ \ \varphi(\mathbf{e}^*, \mathbf{x}), \tag{60}$$
$$\exists e \in \mathbb{N}\ \ \forall e^* \in \mathbb{N}\ \ \forall \ell \in \mathbb{N}\ \ \exists n \geq \ell\ \ \ \mathrm{TM}_e(n)\ \vdash_T\ \exists \mathbf{x} \geq \mathbf{n}\ \ \varphi(\mathbf{e}^*, \mathbf{x}), \tag{61}$$

where $\mathcal{R}$ is a set of primitive recursive functions. Here $T$ is a consistent primitive recursive extension of PA.

See Section 2.7 for the definition of $\mathrm{TM}_e(n) \vdash_T\ \ldots$, and see Definition 63 for a generalized version of PTM-$\omega$-consistency.

Eq. (60) is equivalent to

$$\exists e^* \in \mathbb{N}\ \ \exists \ell \in \mathbb{N}\ \forall n \geq \ell\ \ \forall f \in \mathcal{R}\ \ \ T\ \nvdash\ \exists \mathbf{x}\ (\mathbf{n} \leq \mathbf{x} < \mathbf{n} + f(|\mathbf{n}|))\ \ \varphi(\mathbf{e}^*, \mathbf{x}), \tag{62}$$

since

$$\begin{aligned} & \exists e \in \mathbb{N}\ \ \forall e^* \in \mathbb{N}\ \ \ \forall \ell \in \mathbb{N}\ \exists n \geq \ell\ \exists f \in \mathcal{R}\ \ \ \mathrm{TM}_e(n)\ \vdash_T\ \exists \mathbf{x}\ (\mathbf{n} \leq \mathbf{x} < \mathbf{n} + f(|\mathbf{n}|))\ \ \varphi(\mathbf{e}^*, \mathbf{x}) \\ \Leftrightarrow\ \ & \forall e^* \in \mathbb{N}\ \ \forall \ell \in \mathbb{N}\ \exists n \geq \ell\ \exists f \in \mathcal{R}\ \ \ T\ \vdash\ \exists \mathbf{x}\ (\mathbf{n} \leq \mathbf{x} < \mathbf{n} + f(|\mathbf{n}|))\ \ \varphi(\mathbf{e}^*, \mathbf{x}). \end{aligned} \tag{63}$$

($\Rightarrow$ is trivial, and $\Leftarrow$ can be shown by constructing a TM that searches all proof trees, $\pi$, of $\exists \mathbf{x}\ (\mathbf{n} \leq \mathbf{x} < \mathbf{n} + f(|\mathbf{n}|))\ \ \varphi(\mathbf{e}^*, \mathbf{x})$ for all $(e^*, \ell, n, f) \in \mathbb{N}^3 \times \mathcal{R}$ in the order of the value of $e^* + \ell + n + |\#f| + |\#\pi|$ from 0 to greater. )

Eq. (61) is equivalent to

$$\forall e^* \in \mathbb{N}\ \ \forall \ell \in \mathbb{N}\ \ \exists n \geq \ell\ \ \ T\ \vdash\ \ \exists \mathbf{x} \geq \mathbf{n}\ \varphi(\mathbf{e}^*, \mathbf{x}). \tag{64}$$

Since $\varphi(\mathbf{e}^*, \mathbf{x})$ is a $\Delta_1$-formula and $T$ is a consistent extension of PA, Eq. (62) implies

$$\exists e^* \in \mathbb{N}\ \ \exists \ell \in \mathbb{N}\ \forall x \geq \ell\ \ \ T\ \vdash\ \ \neg\varphi(\mathbf{e}^*, \mathbf{x}). \tag{65}$$

Hence, if $T$ is TM-$\omega$-inconsistent for $\Delta_1$-formula $\varphi(\mathbf{e}^*, \mathbf{x})$, $T$ is $\omega$-inconsistent for $\varphi(\mathbf{e}^*, \mathbf{x})$, since there exists $(e^*, n) \in \mathbb{N}^2$ such that

$$\begin{aligned} & \forall x \geq n\ \ \ T\ \vdash\ \ \neg\varphi(\mathbf{e}^*, \mathbf{x})\ \ \ \ \wedge \\ & T\ \vdash\ \ \exists \mathbf{x} \geq \mathbf{n}\ \varphi(\mathbf{e}^*, \mathbf{x}) \end{aligned}$$

from Eqs. (65) and (64).

On the other hand, if $T$ is $\omega$-inconsistent for $\Delta_1$-formula $\psi(\mathbf{x})$, $T$ is TM-$\omega$-inconsistent for $\varphi(\mathbf{e}^*, \mathbf{x})$ ($\equiv \psi(\mathbf{x})$ for all $e^* \in \mathbb{N}$), since

$$\begin{aligned} &\forall x \in \mathbb{N} \ \ T \vdash \ \neg\psi(\mathbf{x}) \quad \wedge \quad T \vdash \ \exists \mathbf{x} \ \psi(\mathbf{x}) \\ \Rightarrow \quad & \exists e^* \in \mathbb{N} \ \ \forall n \geq 0 \ \ \forall f \in \mathcal{R} \quad T \not\vdash \ \exists \mathbf{x} \ (\mathbf{n} \leq \mathbf{x} < \mathbf{n} + f(|\mathbf{n}|)) \ \ \varphi(\mathbf{e}^*, \mathbf{x}) \\ & \wedge \quad \forall e^* \in \mathbb{N} \ \ \forall n \in \mathbb{N} \ \ (\exists n \geq n) \quad T \vdash \ \ \exists \mathbf{x} \geq \mathbf{n} \ \ \varphi(\mathbf{e}^*, \mathbf{x}) \end{aligned}$$

Thus, TM-$\omega$-consistency for $\Delta_1$-formulas is equivalent to $\omega$-consistency for $\Delta_1$-formulas.

**Remark 4** (Provability of PTM-$\omega$-consistency) Is PA (or another reasonable theory $T$) PTM-$\omega$-consistent for the related formula? Unfortunately, we have not proven the PTM-$\omega$-consistency of PA for $\Delta_2^P$. Moreover, as shown in Theorem 73, no PTM-$\omega$-consistent theory $T$, which is a consistent PT-extension of PA, can prove the PTM-$\omega$-consistency of PA, although PTM-$\omega$-consistency of PA for $\Delta_2^P$ seems to be as natural as the $\omega$-consistency of PA.

**Remark 5** (Characterization of PTM-$\omega$-consistency through axioms and deduction) Assume that PA is PTM-$\omega$-consistent, and that $T$ is a theory constructed by adding an axiom $X$ to PA and is PTM-$\omega$-*inconsistent*. Then,

$$\exists e \in \mathbb{N} \ \forall e^* \in \mathbb{N} \ \forall \ell \in \mathbb{N} \ \exists n \geq \ell \quad \mathrm{PTM}_e(n) \ \vdash_{\mathrm{PA}} \ X \ \rightarrow \ \exists \mathbf{x} \geq \mathbf{n} \ \ \varphi(\mathbf{e}^*, \mathbf{x}), \tag{66}$$

$$\forall e \in \mathbb{N} \ \exists e^* \in \mathbb{N} \ \exists \ell \in \mathbb{N} \ \forall n \geq \ell \ \forall c \in \mathbb{N} \quad \mathrm{PTM}_e(n) \ \not\vdash_{\mathrm{PA}} \ \exists \mathbf{x} \ (\mathbf{n} \leq \mathbf{x} < \mathbf{n} + |\mathbf{n}|^{\mathbf{c}}) \ \ \varphi(\mathbf{e}^*, \mathbf{x}). \tag{67}$$

We then assume that the deduction of Eq.(66) is asymptotically polynomial-time, i.e.,

$$\begin{aligned} &\exists e' \in \mathbb{N} \ \ \forall e^* \in \mathbb{N} \ \ \forall \ell \in \mathbb{N} \ \exists n \geq \ell \ \ \neg Q_1 x_1 \in \mathbb{N} \cdots \neg Q_k x_k \in \mathbb{N} \ \ \exists x \geq n \\ &\mathrm{PTM}_{e'}(x_1, \cdots, x_k, x) \ \vdash_{\mathrm{PA}} \ Y(\mathbf{x}_1, \cdots, \mathbf{x}_k) \ \rightarrow \ \ \varphi(\mathbf{e}^*, \mathbf{x}), \end{aligned} \tag{68}$$

where $X \equiv Q_1\mathbf{x}_1 \cdots Q_k\mathbf{x}_k \ Y(\mathbf{x}_1, \cdots, \mathbf{x}_k)$, $Q_i$ $(i = 1, \ldots, k)$ are quantifiers and $Y(x_1, \cdots, x_k)$ is a $\Delta_1$ formula. Here note that a polynomial (in the size of input) number of application of logical axioms, Modus Ponens and Generalization rules is an asymptotically polynomial-time deduction.

We now assume that $X$ can be asymptotically proven by a PTM over PA. Then,

$$\exists e'' \in \mathbb{N} \ Q_1 x_1 \in \mathbb{N} \cdots Q_k x_k \in \mathbb{N} \quad \mathrm{PTM}_{e''}(x_1, \cdots, x_k) \ \vdash_{\mathrm{PA}} \ Y(\mathbf{x}_1, \cdots, \mathbf{x}_k). \tag{69}$$

From Eqs. (68) and (69), we obtain

$$\exists e \in \mathbb{N} \ \ \forall e^* \in \mathbb{N} \ \ \forall \ell \in \mathbb{N} \ \exists n \geq \ell \ \ \exists x \geq n \ \ \mathrm{PTM}_e(x) \ \vdash_{\mathrm{PA}} \ \varphi(\mathbf{e}^*, \mathbf{x}),$$

Hence,

$$\exists e \in \mathbb{N} \ \ \forall e^* \in \mathbb{N} \ \ \forall \ell \in \mathbb{N} \ \exists n \geq \ell \ \exists c \in \mathbb{N} \ \ \mathrm{PTM}_e(n) \ \vdash_{\mathrm{PA}} \ \ \exists \mathbf{x} \ (\mathbf{n} \leq \mathbf{x} < \mathbf{n} + |\mathbf{n}|^{\mathbf{c}}) \ \ \varphi(\mathbf{e}^*, \mathbf{x}),$$

where $\exists \mathbf{x} \ (\mathbf{n} \leq \mathbf{x} < \mathbf{n} + |\mathbf{n}|^{\mathbf{c}}) \ \varphi(\mathbf{e}^*, \mathbf{x}) \ = \ \varphi(\mathbf{e}^*, \mathbf{n})$, when $c = 0$ (i.e., $\exists c \in \mathbb{N}$). This contradicts Eq.(67). Therefore, if a theory $T$, which is PA $+X$, is PTM-$\omega$-*inconsistent*, $X$ cannot be asymptotically proven by any PTM, assuming that PA is PTM-$\omega$-consistent and the deduction of Eq.(66) can be done asymptotically by a PTM.

Here it is worth noting that any (true) axiom $X$ *can* be asymptotically proven by a *resource unbounded* TM over PA. The point in this remark is that $X$ *cannot* be asymptotically proven by any *polynomial-time bounded* TM (i.e., PTM) over PA.

**Remark 6** (Generalization of PTM-$\omega$-consistency: $\mathcal{C}$-$\omega$-consistency)

We now generalize the concept of PTM-$\omega$-consistency to $\mathcal{C}$-$\omega$-consistency, where $\mathcal{C}$ is a (uniform) computational class.

Here, we introduce some concepts regarding $\mathcal{C}$. Let $\mathrm{U}_{\mathcal{C}}$ be a universal Turing machine specified to $\mathcal{C}$ in a manner similar to $\mathrm{U}_{\mathrm{PTM}}$. Here we omit the precise definition of $\mathrm{U}_{\mathcal{C}}$, by which $\mathcal{C}$ is specified. Each Turing machine in $\mathcal{C}$ is specified by $e \in \mathbb{N}$ as $\mathrm{U}_{\mathcal{C}}(e, \cdot)$. We now introduce the following notation:

$$\begin{aligned} &\mathcal{C}_e(a) \vdash_T \varphi(\mathbf{a}) \\ &\quad \Leftrightarrow \quad \mathrm{U}_{\mathcal{C}}(e, (p, \#\Phi, a)) = \#\pi \ \wedge \ \mathrm{U}_{\mathrm{PTM}}(v_T, (\#\varphi(\mathbf{a}), \#\pi)) \text{ accepts.} \end{aligned}$$

If the truth of $\mathbf{Axiom}_T(\mathbf{n})$ (see Section 2.3) can be correctly decided by an algorithm of class $\mathcal{C}$ in $|n|$, on input $n$, we say that $T$ is $\mathcal{C}$-axiomizable. If $T$ is an extension of $T_0$ and $\mathcal{C}$-axiomizable, then we say that $T$ is a $\mathcal{C}$-extension of $T_0$.

**Definition 63.** *($\mathcal{C}$-$\omega$-consistency)*

*Let theory $S$ be a $\mathcal{C}$-extension of theory $T$. $S$ is $\mathcal{C}$-$\omega$-inconsistent for $\Delta_1$-formula $\varphi(\mathbf{e}^*, \mathbf{x})$ over $T$, if the following two conditions hold simultaneously.*

$$\forall e \in \mathbb{N}\ \exists e^* \in \mathbb{N}\ \exists \ell \in \mathbb{N}\ \forall n \geq \ell\ \forall f \in F_{\mathcal{C}} \quad \mathcal{C}_e(n) \not\vdash_T \exists \mathbf{x}\ (\mathbf{n} \leq \mathbf{x} < \mathbf{n} + f(|\mathbf{n}|)) \ \ \varphi(\mathbf{e}^*, \mathbf{x}), \tag{70}$$

$$\exists e \in \mathbb{N}\ \forall e^* \in \mathbb{N}\ \forall \ell \in \mathbb{N}\ \exists n \geq \ell \quad \mathcal{C}_e(n) \ \vdash_S \exists \mathbf{x} \geq \mathbf{n} \ \ \varphi(\mathbf{e}^*, \mathbf{x}), \tag{71}$$

*where $F_{\mathcal{C}}$ is a set of primitive recursive functions, $f$, such that $\mathrm{U}_{\mathcal{C}}(e, x)$ can do an existential search with $f(|x|)$ steps (e.g., decide $\exists y (x \leq y < x + f(|x|))\ g(y) = 0$ by search of $y$ for $x, x+1, \ldots, x + f(|x|)$).*

*Theory $S$ is $\mathcal{C}$-$\omega$-consistent for $\varphi(\mathbf{e}^*, \mathbf{x})$ over $T$, if theory $S$ is not $\mathcal{C}$-$\omega$-inconsistent for $\varphi(\mathbf{e}^*, \mathbf{x})$ over $T$.*

*Theory $S$ is $\mathcal{C}$-$\omega$-consistent for $\Sigma_i^P$ ($\Pi_i^P$, $\Delta_i^P$, resp.) over $T$, if $S$ is $\mathcal{C}$-$\omega$-consistent for any $\Sigma_i^P$ ($\Pi_i^P$, $\Delta_i^P$, resp.) formula $\varphi(\mathbf{e}^*, \mathbf{x})$ over $T$.*

*Theory $T$ is $\mathcal{C}$-$\omega$-consistent for $\varphi(\mathbf{e}^*, \mathbf{x})$ ($\Sigma_i^P$, $\Pi_i^P$, $\Delta_i^P$, resp.), if $T$ is PTM-$\omega$-consistent for $\varphi(\mathbf{e}^*, \mathbf{x})$ ($\Sigma_i^P$, $\Pi_i^P$, $\Delta_i^P$, resp.) over $T$.*

The following definition is equivalent to the above: Theory $S$ is $\mathcal{C}$-$\omega$-consistent for $\varphi(\mathbf{e}^*, \mathbf{x})$ over $T$, if the following condition holds.

$$\begin{aligned} &\forall e \in \mathbb{N}\ \ \exists e^* \in \mathbb{N}\ \ \exists \ell \in \mathbb{N}\ \ \forall n \geq \ell\ \ \forall f \in F_{\mathcal{C}} \quad \mathcal{C}_e(n) \ \not\vdash_T\ \exists \mathbf{x}\ (\mathbf{n} \leq \mathbf{x} < \mathbf{n} + f(|\mathbf{n}|)) \ \ \varphi(\mathbf{e}^*, \mathbf{x}) \\ \Rightarrow\ &\forall e \in \mathbb{N}\ \ \exists e^* \in \mathbb{N}\ \ \exists \ell \in \mathbb{N}\ \ \forall n \geq \ell \quad \mathcal{C}_e(n) \ \not\vdash_S\ \exists \mathbf{x} \geq \mathbf{n} \ \ \varphi(\mathbf{e}^*, \mathbf{x}). \end{aligned} \tag{72}$$

## 7.2 Unprovability of $\overline{\mathbf{P{\neq}NP}}$ under PTM-$\omega$-Consistency

**Lemma 64.** *Let theory $T$ be a consistent PT-extension of PA. Then,*

$$\forall e \in \mathbb{N}\ \ \exists e^* \in \mathbb{N}\ \ \exists n \in \mathbb{N}\ \ \forall x \geq n \quad \mathrm{PTM}_e(x) \ \not\vdash_T\ \neg\mathrm{DecSAT}(\mathbf{e}^*, \mathbf{x}).$$

*Proof.* Since

$$\mathrm{DecSAT}(\mathbf{e}, \mathbf{x}) \ \equiv \ \mathrm{CD}[\mathrm{SAT}(\mathbf{x})](\mathbf{e}, \lceil \mathcal{SAT} \rceil, \mathbf{x})$$

(see Section 6) we obtain this theorem immediately from Theorem 45.

$\dashv$

**Lemma 65.** *Let theory $T$ be a consistent PT-extension of PA.*

*Let* $\Sigma[e^*, c] \equiv \{\exists \mathbf{x}\ (\mathbf{a} \le \mathbf{x} < \mathbf{a} + |\mathbf{a}|^{\mathbf{c}})\ \ \neg\mathrm{DecSAT}(\mathbf{e^*}, \mathbf{x}) \mid a \in \mathbb{N}\}$, *and* $\mathrm{Size}_{\Sigma[e^*,c]}(a) = |a|^{c+1}$.

$$\forall e \in \mathbb{N}\ \ \exists e^* \in \mathbb{N}\ \ \forall n \in \mathbb{N}\ \ \forall c \in \mathbb{N}\quad \mathrm{PTM}_e(n)\ \not\vdash_T\ \exists \mathbf{x}\ (\mathbf{n} \le \mathbf{x} < \mathbf{n} + |\mathbf{n}|^{\mathbf{c}})\ \ \neg\mathrm{DecSAT}(\mathbf{e^*}, \mathbf{x}).$$

*Proof.* Let $\mathcal{E}$ be a subset of $\mathbb{N}$ such that $e \in \mathcal{E}$ if and only if $\mathrm{U_{PTM}}(e, \cdot)$ is a PTM as follows:

- Let $\varphi(\mathbf{x}) \equiv \mathrm{SAT}(\mathbf{x})$. Let $\Phi \equiv \{\varphi(\mathbf{a}) \mid a \in \mathbb{N}\}$, and $\Phi' \equiv \{\neg\varphi(\mathbf{a}) \mid a \in \mathbb{N}\}$. Let $\Psi[c] \equiv \{\psi(\mathbf{c}, \mathbf{a}, \mathbf{s}) \mid a \in \mathbb{N}\ \wedge\ s < 2^{m^c - 1}\ \}$, where

$$\psi(\mathbf{c}, \mathbf{a}, \mathbf{s})\ \equiv\ \psi(\mathbf{a}, \mathrm{Bit}(\mathbf{s}, \mathbf{0})) \wedge \psi(\mathbf{a} + \mathbf{1}, \mathrm{Bit}(\mathbf{s}, \mathbf{1})) \wedge \cdots \wedge \psi(\mathbf{a} + |\mathbf{a}|^{\mathbf{c}} \dot{-} \mathbf{1}, \mathrm{Bit}(\mathbf{s}, |\mathbf{a}|^{\mathbf{c}} \dot{-} \mathbf{1})),$$
$$\psi(\mathbf{x}, \mathbf{y}) \equiv (\varphi(\mathbf{x}) \wedge \mathbf{y} = \mathbf{0})\ \vee\ (\neg\varphi(\mathbf{x}) \wedge \neg(\mathbf{y} = \mathbf{0})),$$
$$\mathrm{Size}_{\Phi}(a) = \mathrm{Size}_{\Phi'}(a) = |a|, \quad \text{and} \quad \mathrm{Size}_{\Psi[c]}(a) = |a|^{c+1}.$$

- Syntactically check whether the input has the form of $(d, \#\Psi[c], (a, s))$, then follow the specification below. Otherwise, there is no particular specification on the input.
- For all $i = 0, 1, \ldots, |a|^c - 1$, simulate either one of $\mathrm{U_{PTM}}(e, (d, \#\Phi, a+i))$ and $\mathrm{U_{PTM}}(e, (d, \#\Phi', a + i))$ by some rule (e.g., $\mathrm{U_{PTM}}(e, (d, \#\Phi, a + i))$ is simulated if and only if $a + i$ is even.)
- $\mathrm{U_{PTM}}(e, (d, \#\Phi, a+i))$ accepts (and $\mathrm{U_{PTM}}(e, (d, \#\Phi, a+i))$ rejects) if and only if $\mathrm{U_{PTM}}(e, (d, \#\Phi', a + i))$ rejects (and $\mathrm{U_{PTM}}(e, (d, \#\Phi, a + i))$ accepts).
- Accept $(d, \#\Psi[c], (a, s))$ if and only if

$$\mathrm{Bit}(s, i) = 0 \quad \Leftrightarrow \quad \mathrm{U_{PTM}}(e, (d, \#\Phi, a + i)) \ \ \text{accepts},$$

  for all $i = 0, 1, \ldots, |a|^c - 1$.

Note that $\mathcal{E}$ can be primitive recursive by adopting a syntactically checkable canonical coding of the above-mentioned specification on $\mathrm{U_{PTM}}$. In other words, only $e$, for which PTM $\mathrm{U_{PTM}}(e, \cdot)$ is specified in the canonical coding, is in $\mathcal{E}$. (Even if $\mathrm{U_{PTM}}(e', \cdot)$ has the same functionality as $\mathrm{U_{PTM}}(e, \cdot)$ with $e \in \mathcal{E}$, unless $e'$ adopts the canonical coding, $e' \notin \mathcal{E}$.

*Claim.* For any $c \in \mathbb{N}$ and for any $e \in \mathcal{E}$,

$$\begin{aligned}&\mathrm{PA} \vdash\ \ \forall \mathbf{n}\ \ \forall \mathbf{s} < \mathbf{2}^{|\mathbf{n}|^{\mathbf{c}} \dot{-} \mathbf{1}} \\ &\quad (\ \mathrm{CA}[\psi(\mathbf{c}, \mathbf{n}, \mathbf{s})](\mathbf{e}, \ulcorner\Psi[c]\urcorner, \mathbf{n}, \mathbf{s}) \quad \rightarrow \quad \forall \mathbf{x}\ (\mathbf{n} \le \mathbf{x} < \mathbf{n} + |\mathbf{n}|^{\mathbf{c}})\ \ \mathrm{CD}[\varphi(\mathbf{x})](\mathbf{e}, \ulcorner\Phi\urcorner, \mathbf{x})\ ).\end{aligned} \tag{73}$$

*Proof.* From the construction of $\mathrm{U_{PTM}}(e, \cdot)$ with $e \in \mathcal{E}$, for any constant $c \in \mathbb{N}$,

$$\begin{aligned}&\mathrm{PA} \vdash\ \ \forall \mathbf{n}\ \ \forall \mathbf{s} < \mathbf{2}^{|\mathbf{n}|^{\mathbf{c}} \dot{-} \mathbf{1}} \\ &\quad (\ \text{PTM-Acc}[\psi(\mathbf{c}, \mathbf{n}, \mathbf{s})](\mathbf{e}, \ulcorner\Psi[c]\urcorner, \mathbf{n}, \mathbf{s}) \\ &\qquad \leftrightarrow \quad \forall \mathbf{i} < |\mathbf{n}|^{\mathbf{c}}\ \ (\ \mathrm{Bit}(\mathbf{s}, \mathbf{i}) = \mathbf{0} \quad \leftrightarrow \quad \text{PTM-Acc}[\varphi(\mathbf{n} + \mathbf{i})](\mathbf{e}, \ulcorner\Phi\urcorner, \mathbf{n} + \mathbf{i})\ )\ ).\end{aligned} \tag{74}$$

In addition, from the construction of $\mathrm{U_{PTM}}(e, \cdot)$ with $e \in \mathcal{E}$,

$$\mathrm{PA} \vdash\ \ \forall \mathbf{x}\ (\ \text{PTM-Acc}[\varphi(\mathbf{x})](\mathbf{e}, \ulcorner\Phi\urcorner, \mathbf{x})\ ) \quad \leftrightarrow \quad \neg\text{PTM-Acc}[\varphi(\mathbf{x})](\mathbf{e}, \ulcorner\Phi'\urcorner, \mathbf{x})\ )\ ). \tag{75}$$

On the other hand,

$$\begin{aligned}&\mathrm{PA} \vdash\ \ \forall \mathbf{n}\ \ \forall \mathbf{s} < \mathbf{2}^{|\mathbf{n}|^{\mathbf{c}} \dot{-} \mathbf{1}} \\ &\quad (\ \psi(\mathbf{c}, \mathbf{n}, \mathbf{s}) \quad \leftrightarrow \quad \forall \mathbf{i} < |\mathbf{n}|^{\mathbf{c}}\ \ (\ \mathrm{Bit}(\mathbf{s}, \mathbf{i}) = \mathbf{0} \quad \leftrightarrow \quad \varphi(\mathbf{n} + \mathbf{i})\ )\ ).\end{aligned} \tag{76}$$

By Eqs. (74), (75) and (76), we obtain

$$
\begin{aligned}
\mathrm{PA} \vdash\ & \forall \mathbf{n}\ \ \forall \mathbf{s} < \mathbf{2}^{|\mathbf{n}|^{c} \dot{-} \mathbf{1}} \\
& (\ \mathrm{CA}[\psi(\mathbf{c},\mathbf{n},\mathbf{s})](\mathbf{e},\ulcorner\Psi[c]\urcorner,\mathbf{n},\mathbf{s}) \\
& \quad \leftrightarrow\ \ (\ \mathrm{PTM\text{-}Acc}[\psi(\mathbf{c},\mathbf{n},\mathbf{s})](\mathbf{e},\ulcorner\Psi[c]\urcorner,\mathbf{n},\mathbf{s})\ \ \wedge\ \ \psi(\mathbf{c},\mathbf{n},\mathbf{s})\ ) \\
& \quad \rightarrow\ \ \forall \mathbf{i} < |\mathbf{n}|^{\mathbf{c}}\ \ (\ (\ \mathrm{Bit}(\mathbf{s},\mathbf{i}) = \mathbf{0}\ \ \leftrightarrow\ \ \mathrm{PTM\text{-}Acc}[\varphi(\mathbf{n}+\mathbf{i})](\mathbf{e},\ulcorner\Phi\urcorner,\mathbf{n}+\mathbf{i})\ ) \\
& \qquad\qquad\qquad \wedge\ \ (\ \mathrm{Bit}(\mathbf{s},\mathbf{i}) = \mathbf{0}\ \ \leftrightarrow\ \ \varphi(\mathbf{n}+\mathbf{i})\ )\ ) \\
& \quad \rightarrow\ \ \forall \mathbf{i} < |\mathbf{n}|^{\mathbf{c}}\ \ (\ \varphi(\mathbf{n}+\mathbf{i})\ \ \leftrightarrow\ \ \mathrm{PTM\text{-}Acc}[\varphi(\mathbf{n}+\mathbf{i})](\mathbf{e},\ulcorner\Phi\urcorner,\mathbf{n}+\mathbf{i})\ ) \\
& \quad \leftrightarrow\ \ \forall \mathbf{x}\ (\mathbf{n} \le \mathbf{x} < \mathbf{n}+|\mathbf{n}|^{\mathbf{c}})\ \ \mathrm{CD}[\varphi(\mathbf{x})](\mathbf{e},\ulcorner\Phi\urcorner,\mathbf{x})\ )\ ).
\end{aligned}
$$

$\dashv$

Therefore, for any $c \in \mathbb{N}$ and for any $e \in \mathcal{E}$,

$$
\begin{aligned}
\mathrm{PA} \vdash\ & \forall \mathbf{n}\ \ \forall \mathbf{s} < \mathbf{2}^{|\mathbf{n}|^{c} \dot{-} \mathbf{1}} \\
& (\ (\ \ \exists \mathbf{x}\ (\mathbf{n} \le \mathbf{x} < \mathbf{n}+|\mathbf{n}|^{\mathbf{c}})\ \neg\mathrm{CD}[\varphi(\mathbf{x})](\mathbf{e},\ulcorner\Phi\urcorner,\mathbf{x})\ \ \rightarrow\ \ \neg\mathrm{CA}[\psi(\mathbf{c},\mathbf{n},\mathbf{s}](\mathbf{e},\ulcorner\Psi[c]\urcorner,\mathbf{n},\mathbf{s})\ ).
\end{aligned}
\tag{77}
$$

We now assume that there exists $e \in \mathbb{N}$ such that

$$
\forall e^* \in \mathbb{N}\ \exists n \in \mathbb{N}\ \exists c \in \mathbb{N} \quad \mathrm{PTM}_e(n)\ \vdash_T\ \exists \mathbf{x}\ (\mathbf{n} \le \mathbf{x} < \mathbf{n}+|\mathbf{n}|^{\mathbf{c}})\ \ \neg\mathrm{DecSAT}(\mathbf{e}^*,\mathbf{x}). \tag{78}
$$

In other words, we assume there exists $e \in \mathbb{N}$ such that

$$
\forall e^* \in \mathbb{N}\ \exists n \in \mathbb{N}\ \exists c \in \mathbb{N} \quad \mathrm{PTM}_e(n)\ \vdash_T\ \exists \mathbf{x}\ (\mathbf{n} \le \mathbf{x} < \mathbf{n}+|\mathbf{n}|^{\mathbf{c}})\ \ \neg\mathrm{CD}[\varphi(\mathbf{x})](\mathbf{e}^*,\ulcorner\Phi\urcorner,\mathbf{x}) \tag{79}
$$

Then we can construct $\mathrm{U_{PTM}}(e',\cdot)$ with $e' \in \mathcal{E}$ by using $\mathrm{U_{PTM}}(e,\cdot)$ as follows:

- (Input :) $(p, \#\Gamma[e^*,c], (n,s))$, where $\Gamma[e^*,c] \equiv \{\neg\mathrm{CA}[\psi(\mathbf{c},\mathbf{a},\mathbf{t})](\mathbf{e},\ulcorner\Psi[c]\urcorner,\mathbf{a},\mathbf{t}) \mid (a,t) \in \mathbb{N}^2\ \wedge\ t < 2^{|a|^c - 1}\}$.
- (Output: ) Gödel number of a proof tree of $\neg\mathrm{CA}[\psi(\mathbf{c},\mathbf{n},\mathbf{s})](\mathbf{e},\ulcorner\Psi[c]\urcorner,\mathbf{n},\mathbf{s})$ or 0.
- Run the following computation

  $$
  \mathrm{U_{PTM}}(e,(p,\#\Sigma[e^*,c],n)) = z, \quad \mathrm{U_{PTM}}(v_T,(\#\eta,z)),
  $$

  where $\eta\ \ \equiv\ \ \exists \mathbf{x}\ (\mathbf{n} \le \mathbf{x} < \mathbf{n}+|\mathbf{n}|^{\mathbf{c}})\ \ \neg\mathrm{CD}[\varphi(\mathbf{x})](\mathbf{e}^*,\ulcorner\Phi\urcorner,\mathbf{x})$.
- Compute the proof (say $\pi_2$) of

  $$
  \forall \mathbf{y}\ \ \forall \mathbf{t} < \mathbf{2}^{|\mathbf{n}|^{c} \dot{-} \mathbf{1}}\ (\ \exists \mathbf{x}\ (\mathbf{y} \le \mathbf{x} < \mathbf{y}+|\mathbf{n}|^{\mathbf{c}})\ \neg\mathrm{CD}[\varphi(\mathbf{x})](\mathbf{e}^*,\ulcorner\Phi\urcorner,\mathbf{x})\ \ \rightarrow\ \ \neg\mathrm{CA}[\psi(\mathbf{c},\mathbf{y},\mathbf{t})](\mathbf{e}^*,\ulcorner\Psi[c]\urcorner,\mathbf{y},\mathbf{t})\ ),
  $$

  since there exists a proof of this formula by Eq. (77) if $e^* \in \mathcal{E}$. (The computation time is finite.)
- Check whether $\mathrm{U_{PTM}}(v_T,(\#\eta,z))$ accepts or rejects. If it rejects, output 0 and halt. If it accepts, then combine $\pi_1$ ($z = \#\pi_1$) and $\pi_2$ and make a new proof tree, $\pi_3$, of $\neg\mathrm{CA}[\psi(\mathbf{c},\mathbf{n},\mathbf{s})](\mathbf{e},\ulcorner\Psi[c]\urcorner,\mathbf{n},\mathbf{s})$, as follows:

  $$
  \begin{aligned}
  \pi_3\ \ \equiv\ & < \neg\mathrm{CA}[\psi(\mathbf{c},\mathbf{y},\mathbf{t})](\mathbf{e}^*,\ulcorner\Psi[c]\urcorner,\mathbf{y},\mathbf{t}),\ \text{Modus Ponens} > \\
  & \qquad [\ \pi_1,\ < \text{Formula A},\ \text{Modus Ponens} >\ [\ \pi_2,\ \text{Axiom X}\ ]\ ],
  \end{aligned}
  $$

  where Formula A is

  $$
  \exists \mathbf{x}\ (\mathbf{n} \le \mathbf{x} < \mathbf{n}+|\mathbf{n}|^{\mathbf{c}})\ \ \neg\mathrm{CD}[\varphi(\mathbf{x})](\mathbf{e},\ulcorner\Phi\urcorner,\mathbf{x})\quad \rightarrow\quad \neg\mathrm{CA}[\psi(\mathbf{c},\mathbf{n},\mathbf{s})](\mathbf{e},\ulcorner\Psi[c]\urcorner,\mathbf{n},\mathbf{s})
  $$

and Axiom X is a logical axiom,

$$\forall \mathbf{y}\ \forall \mathbf{t} < \mathbf{2^{|n|^c \dot{-} 1}}(\ \exists \mathbf{x}\,(\mathbf{y} \le \mathbf{x} < \mathbf{y}{+}|\mathbf{n}|^{\mathbf{c}})\ \neg\mathrm{CD}[\varphi(\mathbf{x})](\mathbf{e}^*, \ulcorner\Phi\urcorner, \mathbf{x}) \ \rightarrow\ \neg\mathrm{CA}[\psi(\mathbf{c},\mathbf{n},\mathbf{s})](\mathbf{e}^*, \ulcorner\Psi[c]\urcorner, \mathbf{n}, \mathbf{s})\ )$$

$$\rightarrow\quad (\ \exists \mathbf{x}\ (\mathbf{n} \le \mathbf{x} < \mathbf{n} + |\mathbf{n}|^{\mathbf{c}})\ \neg\mathrm{CD}[\varphi(\mathbf{x})](\mathbf{e}^*, \ulcorner\Phi\urcorner, \mathbf{x}) \quad\rightarrow\quad \neg\mathrm{CA}[\psi(\mathbf{c},\mathbf{n},\mathbf{s})](\mathbf{e}^*, \ulcorner\Psi[c]\urcorner, \mathbf{n}, \mathbf{s})\ ).$$

– Output $\pi_3$ for the proof tree of formula $\neg\mathrm{CA}[\psi(\mathbf{c},\mathbf{n},\mathbf{s})](\mathbf{e}^*, \ulcorner\Psi[c]\urcorner, \mathbf{n}, \mathbf{s})$.

Therefore, if we assume Eq. (79), there exists $e' \in \mathcal{E}$ such that

$$\exists n \in \mathbb{N}\ \exists c \in \mathbb{N}\ \forall s < 2^{|n|^c - 1}\ \forall e^* \in \mathcal{E}\ \ \mathrm{PTM}_{e'}(n,s)\ \vdash_T\ \neg\mathrm{CA}[\psi(\mathbf{c},\mathbf{n},\mathbf{s})](\mathbf{e}^*), \ulcorner\Psi[c]\urcorner, \mathbf{n}, \mathbf{s}).$$

Then, since $e' \in \mathcal{E}$ implies $g(e') \in \mathcal{E}$, there exists $e' \in \mathcal{E}$ such that

$$\exists n \in \mathbb{N}\ \exists c \in \mathbb{N}\ \forall s < 2^{|n|^c - 1}\ \ \mathrm{PTM}_{e'}(n,s)\ \vdash_T\ \neg\mathrm{CA}[\psi(\mathbf{c},\mathbf{n},\mathbf{s})](g(\mathbf{e}'), \ulcorner\Psi[c]\urcorner, \mathbf{n}, \mathbf{s}).$$

Here, for any $n \in \mathbb{N}$, there exists $s < 2^{|n|^c - 1}$ such that

$$\mathfrak{N} \models\ \psi(\mathbf{c},\mathbf{n},\mathbf{s}).$$

Hence, there exist $e' \in \mathbb{N},\ n \in \mathbb{N},\ c \in \mathbb{N},\ s \in \mathbb{N}$ such that

$$\begin{aligned}&\mathfrak{N} \models\ \psi(\mathbf{c},\mathbf{n},\mathbf{s}) \quad \wedge\\ &\mathrm{PTM}_{e'}(n,s)\ \vdash_T\ \neg\mathrm{CA}[\psi(\mathbf{c},\mathbf{n},\mathbf{s})](g(\mathbf{e}'), \ulcorner\Psi[c]\urcorner, \mathbf{n}, \mathbf{s}).\end{aligned}$$

This contradicts Corollary 46, so, Eq. (78) does not hold.

Since the contradiction occurs when $e^* = g(e')$ with $e' \in \mathcal{E}$, we now define $g^*$ as follows:

$$g^*(e) \equiv \begin{cases} g(e) & \text{if } e \in \mathcal{E},\\ g(e') & \text{if } e \notin \mathcal{E}.\end{cases}$$

Since deciding whether $e \in \mathcal{E}$ or not is primitive recursive and the transformation of $e$ to $e'$ is also primitive recursive, function $g^*$ is primitive recursive. Then,

$$\forall e \in \mathbb{N}\ \ \forall n \in \mathbb{N}\ \ \forall c \in \mathbb{N}\quad \mathrm{PTM}_e(n)\ \not\vdash_T\ \exists \mathbf{x}\ (\mathbf{n} \le \mathbf{x} < \mathbf{n} + |\mathbf{n}|^{\mathbf{c}})\ \neg\mathrm{DecSAT}(g^*(\mathbf{e}), \mathbf{x}).$$

⊣

**Lemma 66.** *Let theory $T$ be a consistent PT-extension of PA and* PTM-$\omega$*-consistent for* $\Delta_2^P$*. Then*

$$\forall e \in \mathbb{N}\ \ \exists e^* \in \mathbb{N}\ \ \exists \ell \in \mathbb{N}\ \ \forall n \ge \ell\ \ \mathrm{PTM}_e(n)\ \not\vdash_T\ \ \exists \mathbf{x} \ge \mathbf{n}\ \ \neg\mathrm{DecSAT}(\mathbf{e}^*, \mathbf{x}).$$

*Proof.* From Lemma 65

$$\forall e \in \mathbb{N}\ \ \exists e^* \in \mathbb{N}\ \ \forall n \ge 0\ \ \forall c \in \mathbb{N}\quad \mathrm{PTM}_e(n)\ \not\vdash_T\ \exists \mathbf{x}\ (\mathbf{n} \le \mathbf{x} < \mathbf{n} + |\mathbf{n}|^{\mathbf{c}})\ \neg\mathrm{DecSAT}(\mathbf{e}^*, \mathbf{x}).$$

Since $\mathrm{SAT}(\mathbf{x}) \in \Sigma_1^P$ and $\neg\mathrm{SAT}(\mathbf{x}) \in \Pi_1^P$, then $\neg\mathrm{DecSAT}(\mathbf{e}^*, \mathbf{x}) \in \Sigma_2^P$ and $\neg\mathrm{DecSAT}(\mathbf{e}^*, \mathbf{x}) \in \Pi_2^P$. That is, $\neg\mathrm{DecSAT}(\mathbf{e}^*, \mathbf{x}) \in \Delta_2^P$.

Therefore, if $T$ is PTM-$\omega$-consistent for $\Delta_2^P$, $T$ is PTM-$\omega$-consistent for $\neg\mathrm{DecSAT}(\mathbf{e}^*, \mathbf{x}) \in \Delta_2^P$.

We then obtain, from the definition of PTM-$\omega$-consistency,

$$\forall e \in \mathbb{N}\ \ \exists e^* \in \mathbb{N}\ \ \exists \ell \in \mathbb{N}\ \ \forall n \ge \ell\ \ \mathrm{PTM}_e(n)\ \not\vdash_T\ \ \exists \mathbf{x} \ge \mathbf{n}\ \ \neg\mathrm{DecSAT}(\mathbf{e}^*, \mathbf{x}).$$

$\dashv$

**Theorem 67.** *Let theory $T$ be a consistent PT-extension of PA and* PTM-$\omega$*-consistent for $\Delta_2^P$.*

$$T \nvdash \ \overline{\mathrm{P} \neq \mathrm{NP}}.$$

*Namely, there exists no proof of* $\overline{\mathrm{P} \neq \mathrm{NP}}$ *in $T$.*

*Proof.* Assume that

$$T \vdash \ \overline{\mathrm{P} \neq \mathrm{NP}},$$

i.e.,

$$T \vdash \ \forall \mathbf{e}^* \ \ \forall \mathbf{n} \ \ \exists \mathbf{x} \geq \mathbf{n} \ \ \neg \mathrm{DecSAT}(\mathbf{e}^*, \mathbf{x}). \tag{80}$$

We can then construct PTM $\mathrm{U_{PTM}}(e, \cdot)$ as follows:

- (Input: ) $(p, \#\Sigma[e^*], n) \in \mathbb{N}^3$, where $\Sigma[e^*] \equiv \{\exists \mathbf{x} \geq \mathbf{a} \ \ \neg \mathrm{DecSAT}(\mathbf{e}^*, \mathbf{x}) \mid a \in \mathbb{N}\}$.
- (Output: ) Gödel number of a proof tree of $\exists \mathbf{x} \geq \mathbf{n} \ \ \neg \mathrm{DecSAT}(\mathbf{e}^*, \mathbf{x})$.
- Find a proof, $\pi$, of formula $\forall \mathbf{y} \ \ \forall \mathbf{z} \ \ \exists \mathbf{x} \geq \mathbf{z} \ \ \neg \mathrm{DecSAT}(\mathbf{y}, \mathbf{x})$, where $\pi$ exists according to the assumption, Eq. (80). Here, the size of $\pi$ is constant in $|n|$.
- Construct the following proof tree of $\exists \mathbf{x} \geq \mathbf{n} \ \ \neg \mathrm{DecSAT}(\mathbf{e}^*, \mathbf{x})$:

$$< \exists \mathbf{x} \geq \mathbf{n} \ \ \neg \mathrm{DecSAT}(\mathbf{e}^*, \mathbf{x}), \text{Modus Ponens} > \ [\ \pi, \ \text{Axiom X} \ ],$$

where Axiom X is a logical axiom,

$$\forall \mathbf{y} \ \ \forall \mathbf{z} \ \ (\ \exists \mathbf{x} \geq \mathbf{z} \ \ \neg \mathrm{DecSAT}(\mathbf{y}, \mathbf{x}) \ ) \quad \rightarrow \quad (\ \exists \mathbf{x} \geq \mathbf{n} \ \ \neg \mathrm{DecSAT}(\mathbf{e}^*, \mathbf{x}) \ )$$

- Output the Gödel number of the proof tree.

Clearly, PTM $\mathrm{U_{PTM}}(e, \cdot)$ outputs a correct value for all $(e^*, n) \in \mathbb{N}^*$. Therefore, we obtain

$$\exists e \in \mathbb{N} \ \ \forall e^* \in \mathbb{N} \ \ \forall n \in \mathbb{N} \quad \mathrm{PTM}_e(n) \ \vdash_T \quad \exists \mathbf{x} \geq \mathbf{n} \ \ \neg \mathrm{DecSAT}(\mathbf{e}^*, \mathbf{x}).$$

This contradicts Lemma 66. Thus,

$$T \nvdash \ \overline{\mathrm{P} \neq \mathrm{NP}}.$$

$\dashv$

**Remark:** Theorem 67 and its generalization imply the results by Baker, Gill and Solovay [1] and by Hartmanis and Hopcroft [13, 14].

First, let assume the following proposition, which is a generalization of Theorem 67 and will be formally given in Part 2 of this paper.

**Proposition 68.** *Let $\mathcal{C}$ be a (uniform) computational class (see Remark 6 of Definition 62), and theory $T$ be a consistent $\mathcal{C}$-extension of PA and $\mathcal{C}$-$\omega$-consistent for QBF.*

*Then, $T$ cannot prove any super-$\mathcal{C}$-computational-lower bound.*

We now assume that a relativizable proof of $\overline{\mathrm{P} \neq \mathrm{NP}}$ exists for any oracle $A$ and that it is formalized in PA (or more generally, a $\omega$-consistent theory $T$).

Then, for any oracle $A$

$$\mathrm{PA} \vdash \overline{\mathrm{P}^A \neq \mathrm{NP}^A}.$$

From Proposition 68, PA should be PTM$^A$-*$\omega$-inconsistent* for any oracle $A$. Hence, PA should be TM-$\omega$-inconsistent, which is equivalent to $\omega$-inconsistent (see Remark 3 of Definition 62). That is, PA is not $\omega$-consistent. This is a contradiction. Thus, there exists no relativizable proof of $\overline{\mathrm{P} \neq \mathrm{NP}}$ in PA, which corresponds to the result by Baker, Gill and Solovay [1].

Similarly, we can also obtain a result corresponding to that by Hartmanis and Hopcroft [13, 14] as follows:

First we assume that $T$ is a $\omega$-consistent theory. Then, we can construct TM $M$ such that

$$\begin{aligned} & \exists e \in \mathbb{N}\, \forall e^* \in \mathbb{N}\, \forall \ell \in \mathbb{N}\, \exists n \geq \ell\, \exists f \in \mathcal{R} \quad \mathrm{PTM}_e^{L(M)}(n)\ \vdash_T\ \exists \mathbf{x}\ (\mathbf{n} \leq \mathbf{x} < \mathbf{n} + f(|\mathbf{n}|))\ \ \varphi(\mathbf{e}^*, \mathbf{x}) \\ \Leftrightarrow\ \ & \forall e^* \in \mathbb{N}\ \ \forall \ell \in \mathbb{N}\ \ \exists n \geq \ell\ \ \exists f \in \mathcal{R}\quad T\ \vdash\ \exists \mathbf{x}\ (\mathbf{n} \leq \mathbf{x} < \mathbf{n} + f(|\mathbf{n}|))\ \ \varphi(\mathbf{e}^*, \mathbf{x}), \end{aligned} \tag{81}$$

and

$$\begin{aligned} & \exists e \in \mathbb{N}\ \ \forall e^* \in \mathbb{N}\ \ \forall \ell \in \mathbb{N}\ \ \exists n \geq \ell \quad \mathrm{PTM}_e^{L(M)}(n)\ \vdash_T\ \exists \mathbf{x} \geq \mathbf{n}\ \ \varphi(\mathbf{e}^*, \mathbf{x}) \\ \Leftrightarrow\ \ & \forall e^* \in \mathbb{N}\ \ \forall \ell \in \mathbb{N}\ \ \exists n \geq \ell \quad T\ \vdash\ \ \exists \mathbf{x} \geq \mathbf{n}\ \varphi(\mathbf{e}^*, \mathbf{x}). \end{aligned} \tag{82}$$

(For a method of constructing $M$, see the description just after Eq. (63) in Remark 3 of Definition 62.)

We now assume that

$$T \vdash\ \overline{\mathrm{P}^{L(M)} \neq \mathrm{NP}^{L(M)}}.$$

From Proposition 68, $T$ should be PTM$^{L(M)}$-*$\omega$-inconsistent*. However, from the construction of TM $M$, PTM$^{L(M)}$-*$\omega$-inconsistent* is equivalent to *$\omega$-inconsistent*, since Eqs. (81) and (82) hold. That is, $T$ should be *$\omega$-inconsistent*. This is a contradiction. Thus, for any $\omega$-consistent theory $T$, there exists a TM $M$ such that

$$T \nvdash\ \overline{\mathrm{P}^{L(M)} \neq \mathrm{NP}^{L(M)}},$$

which corresponds to the result by Hartmanis and Hopcroft [13, 14].

### 7.3 Unprovability of Super-Polynomial-Time Lower Bounds in PSPACE under PTM-$\omega$-Consistency

We can obtain the following theorem in a manner similar to that used in Section 7.2.

**Theorem 69.** *Let language $L$ be in PSPACE. Let theory $T$ be a consistent PT-extension of PA and* PTM-*$\omega$-consistent for QBF.*

$$T \nvdash\ \forall \mathbf{e}\ \ \forall \mathbf{n}\ \ \exists \mathbf{x} \geq \mathbf{n} \quad \neg \mathrm{CD}[\varphi_L(\mathbf{x})](\mathrm{e}, \ulcorner \Phi_L \urcorner, \mathbf{x}),$$

*Namely, there exists no proof of any super-polynomial-time computational lower bound of $L$ in $T$.*

## 8 Unprovability of PTM-$\omega$-Consistency

This section shows that the independence of P vs NP from $T$ by proving PTM-$\omega$-consistency of $T$ for a $\Delta_2^P$-formula (i.e., through Theorem 67) cannot be proven in theory $S$, where $S$ is a consistent PT-extension of $T$ and is PTM-$\omega$-consistent for $\Delta_2^P$. This result is based on the second incompleteness theorem of polynomial-time proofs, Theorem 21.

Let $T$ be a consistent PT-extension of PA, and assume that P$\neq$NP is true. Then, if $T$ is proven to be PTM-$\omega$-consistent for $\neg\mathrm{DecSAT}(\mathbf{e}^*, \mathbf{x})$, Theorem 67 will imply that $T$ is independent from P$\neq$NP. To prove the PTM-$\omega$-consistency of $T$ for $\neg\mathrm{DecSAT}(\mathbf{e}^*, \mathbf{x})$, it is sufficient to prove that

$$\forall e \in \mathbb{N} \ \ \exists e^* \in \mathbb{N} \ \ \exists \ell \in \mathbb{N} \ \ \forall n \geq \ell \quad \mathrm{PTM}_e(n) \ \not\vdash_T \ \exists \mathbf{x} \geq \mathbf{n} \ \ \neg\mathrm{DecSAT}(\mathbf{e}^*, \mathbf{x}), \tag{83}$$

since it has been already proven that

$$\forall e \in \mathbb{N} \ \ \exists e^* \in \mathbb{N} \ \ \exists \ell \in \mathbb{N} \ \ \forall n \geq \ell \ \ \forall c \in \mathbb{N} \quad \mathrm{PTM}_e(n) \ \not\vdash_T \ \exists \mathbf{x} \ (\mathbf{n} \leq \mathbf{x} < \mathbf{n} + |\mathbf{n}|^{\mathbf{c}}) \ \ \neg\mathrm{DecSAT}(\mathbf{e}^*, \mathbf{x})$$

by Lemma 65.

This section shows that theory $S$ cannot prove Eq.(83) formally, if $S$ is a consistent PT-extension of $T$ and PTM-$\omega$-consistent for $\Delta_2^P$ over $T$. That is, the PTM-$\omega$-consistency of $T$ for $\neg\mathrm{DecSAT}(\mathbf{e}^*, \mathbf{x})$ cannot be proven in $S$. In other words, the independence of P vs NP from $T$ by proving the PTM-$\omega$-consistency of $T$ cannot be proven in $S$. Here, the formal sentence of Eq.(83) in PA is

$$\forall \mathbf{e} \ \exists \mathbf{l} \ \forall \mathbf{n} \geq \mathbf{l} \quad \neg\mathrm{Pr}_T[\exists \mathbf{x} \geq \mathbf{n} \ \ \neg\mathrm{DecSAT}(h(\mathbf{e}), \mathbf{x})](\mathbf{e}, \lceil \Phi(e) \rceil, \mathbf{n}), \tag{84}$$

where $h$ is a primitive recursive function[4], and $\Phi(e) \equiv \ \{\exists \mathbf{x} \geq \mathbf{a} \ \neg\mathrm{DecSAT}(h(\mathbf{e}), \mathbf{x}) \mid a \in \mathbb{N}\}$.

This result is based on the incompleteness theorem of polynomial-time proofs, Theorem 21. To obtain this result, however, a slight modification is required for Theorem 21 as follows:

**Lemma 70.** *Let theory $T$ be a consistent PT-extension of PA, and $\Psi(e') \equiv \{\psi(\mathbf{e}', \mathbf{a}) \mid a \in \mathbb{N}\}$.*

$$\forall e \in \mathbb{N} \ \ \exists e' \in \mathbb{N} \ \ \forall x \in \mathbb{N} \quad \mathrm{PTM}_e(x) \not\vdash_T \ \ \neg\mathrm{Pr}_T[\psi(\mathbf{e}', \mathbf{x})](\mathbf{e}', \lceil \Psi(e') \rceil, \mathbf{x}).$$

*Proof.* First, Eq. (12) is obtained in the same manner as that of Theorem 21.

We then obtain

$$\mathrm{PA} \vdash \ \forall \mathbf{x} \ \forall \mathbf{y} \ \ ( \ \rho_{e,T}(\mathbf{x}) \wedge \neg\rho_{e,T}(\mathbf{x}) \ \ \rightarrow \ \ \psi(\mathbf{y}, \mathbf{x}) \ ), \tag{85}$$

in place of Eq. (13).

We then obtain the following claim (in a manner similar to Corollary 14):

*Claim.* Let $\Phi \equiv \{\varphi(\mathbf{a}) \mid a \in \mathbb{N}\}$ and $\Psi(e) \equiv \{\psi(\mathbf{e}, \mathbf{a}) \mid a \in \mathbb{N}\}$, where $e \in \mathbb{N}$. Suppose that $T$ is a consistent PT-extension of PA. We assume

$$T \vdash \ \forall \mathbf{x} \ \forall \mathbf{y} \ \ (\varphi(\mathbf{x}) \rightarrow \psi(\mathbf{y}, \mathbf{x})).$$

Then, for all $e_1 \in \mathbb{N}$ there exists $e_2 \in \mathbb{N}$ such that

$$\forall e \in \mathbb{N} \quad \mathrm{PA} \vdash \quad \forall \mathbf{x} \ ( \ \mathrm{Pr}_T[\varphi(\mathbf{x})](\mathbf{e}_1, \lceil \Phi \rceil, \mathbf{x}) \ \ \rightarrow \ \ \mathrm{Pr}_T[\psi(\mathbf{e}, \mathbf{x})](\mathbf{e}_2, \lceil \Psi(e) \rceil, \mathbf{x}) \ ). \tag{86}$$

[4] In Eq.(83), there exists a primitive recursive function $h$ such that $e^* = h(e)$ for all $e \in \mathbb{N}$: i.e.,

$$\forall e \in \mathbb{N} \ \ \forall \ell \in \mathbb{N} \ \ \exists n \geq \ell \quad \mathrm{PTM}_e(n) \ \not\vdash_T \ \exists \mathbf{x} \geq \mathbf{n} \ \ \neg\mathrm{DecSAT}(h(\mathbf{e}), \mathbf{x}).$$

*Proof.* From the first derivability condition (D.1) of a traditional proof theory [2] and the assumption of this lemma, we obtain

$$\mathrm{PA} \vdash \ \mathrm{Pr}_T(\ulcorner \forall \mathbf{x}\ \forall \mathbf{y}\ \ (\varphi(\mathbf{x}) \rightarrow \psi(\mathbf{y},\mathbf{x}))\urcorner).$$

Then, PTM $\mathrm{U}_{\mathrm{PTM}}(e_2, \cdot)$ is constructed by using PTM $\mathrm{U}_{\mathrm{PTM}}(e_1, (p, \#\Phi, \cdot))$ as follows:

1. (Input :) $(p, \#\Psi(e), x)$
2. (Output: ) Gödel number of a proof tree of $\psi(\mathbf{e}, \mathbf{x})$ or 0.
3. Run the following computation

$$\mathrm{U}_{\mathrm{PTM}}(e_1, (p, \#\Phi, x)) = z, \quad \mathrm{U}_{\mathrm{PTM}}(v_T, (\#\varphi(\mathbf{x}), z)).$$

4. Compute the proof (say $\pi_2$) of $\forall \mathbf{w}\ \forall \mathbf{y}\ \ (\varphi(\mathbf{w}) \rightarrow \psi(\mathbf{y}, \mathbf{w}))$, since there exists a proof for the predicate from the assumption.
5. Check whether $\mathrm{U}_{\mathrm{PTM}}(v_T, (\#\varphi(\mathbf{x}), z))$ accepts or rejects. If it rejects, output 0 and halt. If it accepts, then combine $\pi_1$ $(z = \#\pi_1)$ and $\pi_2$ and make a new proof tree, $\pi_3$, for $\psi(\mathbf{e}, \mathbf{x})$, as follows:

$$\pi_3 \ \equiv \ < \psi(\mathbf{e}, \mathbf{x}), \text{Modus Ponens} > [\pi_1, < \varphi(\mathbf{x}) \rightarrow \psi(\mathbf{e}, \mathbf{x}), \text{Modus Ponens} > [\pi_2, \text{Axiom X}]],$$

   where Axiom X is a logical axiom, " $\forall \mathbf{w}\ \forall \mathbf{y}\ \ (\varphi(\mathbf{w}) \rightarrow \psi(\mathbf{y}, \mathbf{w})) \ \rightarrow \ (\varphi(\mathbf{x}) \rightarrow \psi(\mathbf{e}, \mathbf{x}))$".
6. Output $\pi_3$ for the proof tree of formula $\psi(\mathbf{e}, \mathbf{x})$.

The other part of the proof can be completed in an analogous manner to that in Lemma 13 except for the constructions of functions $h$ and $g$ to meet the above-mentioned construction of $\mathrm{U}_{\mathrm{PTM}}(e_2, \cdot))$ in this proof.

$\dashv$

Therefore, by setting $e \leftarrow e_2$ in Eq. (88), for all $e_1 \in \mathbb{N}$ there exists $e_2 \in \mathbb{N}$ such that

$$\mathrm{PA} \vdash \quad \forall \mathbf{x}\ (\ \mathrm{Pr}_T[\varphi(\mathbf{x})](\mathrm{e}_1, \ulcorner \Phi \urcorner, \mathbf{x}) \ \rightarrow \ \mathrm{Pr}_T[\psi(\mathrm{e}_2, \mathbf{x})](\mathrm{e}_2, \ulcorner \Psi(e_2) \urcorner, \mathbf{x})\ ). \tag{87}$$

Then, applying Eq. (85) to Eq. (87), we obtain that for any $e^{+++} \in \mathbb{N}$, there exists $e' \in \mathbb{N}$ such that

$$\mathrm{PA} \vdash \ \forall \mathbf{x}\ (\ \mathrm{Pr}_T[\rho_{e,T}(\mathbf{x}) \wedge \neg\rho_{e,T}(\mathbf{x})](\mathrm{e}^{+++}, \ulcorner \mathcal{G}^{+++} \urcorner, \mathbf{x}) \ \rightarrow \ \mathrm{Pr}_T[\psi(\mathrm{e}', \mathbf{x})](\mathrm{e}', \ulcorner \Psi(e') \urcorner, \mathbf{x})\ ), \tag{88}$$

in place of Eq. (14).

Hence, we obtain: for any $e \in \mathbb{N}$, there exists $e' \in \mathbb{N}$ such that

$$\mathrm{PA} \vdash \ \forall \mathbf{x}\ \ (\ \neg\mathrm{Pr}_T[\psi(\mathrm{e}', \mathbf{x})](\mathrm{e}', \ulcorner \Psi(e') \urcorner, \mathbf{x}) \ \rightarrow \ \rho_{e,T}(\mathbf{x})\ ), \tag{89}$$

in place of Eq. (15).

The remaining part of the proof of this lemma is the same as that of Theorem 21

$\dashv$

**Lemma 71.** *Let theory $T$ be a consistent PT-extension of PA. Let $\Phi(e') \equiv \{\exists \mathbf{x} \geq \mathbf{a}\ \neg\mathrm{DecSAT}(h(\mathrm{e}'), \mathbf{x}) \mid a \in \mathbb{N}\}$ and $h$ be a primitive recursive function.*

$$\begin{aligned} &\forall e \in \mathbb{N} \ \ \exists e' \in \mathbb{N} \ \ \exists m \in \mathbb{N} \ \ \forall \ell \geq m \ \ \forall c \in \mathbb{N} \\ &\mathrm{PTM}_e(\ell) \ \nvdash_T \ \exists \mathbf{n}\ (\mathbf{l} \leq \mathbf{n} < \mathbf{l} + |\mathbf{l}|^{\mathrm{c}}) \ \ \neg\mathrm{Pr}_T[\exists \mathbf{x} \geq \mathbf{n} \ \ \neg\mathrm{DecSAT}(h(\mathrm{e}'), \mathbf{x})](\mathrm{e}', \ulcorner \Phi(e') \urcorner, \mathbf{n}). \end{aligned} \tag{90}$$

*Proof.* First, we show the following claim:

*Claim.* Let $\Psi(d,c) \equiv \{\forall \mathbf{n}\ (\mathbf{a} \le \mathbf{n} < \mathbf{a} + |\mathbf{a}|^{\mathbf{c}})\ \ \psi(\mathbf{d},\mathbf{n}) \mid a \in \mathbb{N}\}$ and $\Psi(d) \equiv \{\psi(\mathbf{d},\mathbf{a}) \mid a \in \mathbb{N}\}$, where $\psi(\mathbf{d},\mathbf{a})$ is any formula, $c \in \mathbb{N}$ and $d \in \mathbb{N}$. Then,

$$
\begin{aligned}
&\forall e \in \mathbb{N}\ \exists \widetilde{e} \in \mathbb{N}\ \forall d \in \mathbb{N} \\
&\quad \mathrm{PA} \vdash\ \forall \mathbf{m}\ \forall \mathbf{c}\ \ (\ \ \mathrm{Pr}_T[\forall \mathbf{n}(\mathbf{m} \le \mathbf{n} < \mathbf{m} + |\mathbf{m}|^{\mathbf{c}})\ \ \psi(\mathbf{d},\mathbf{n})](\mathbf{e}, \ulcorner \Psi(d,c) \urcorner, \mathbf{m}) \\
&\qquad\qquad \to\ \ \forall \mathbf{n}\ (\mathbf{m} \le \mathbf{n} < \mathbf{m} + |\mathbf{m}|^{\mathbf{c}})\ \ \mathrm{Pr}_T[\psi(\mathbf{d},\mathbf{n})](\widetilde{\mathbf{e}}, \ulcorner \Psi(d) \urcorner, \mathbf{n})\ ) \qquad (91)
\end{aligned}
$$

*Proof.* First, we construct TM $\mathrm{U}(\widetilde{e},\cdot)$ using PTM $\mathrm{U}_{\mathrm{PTM}}(e,\cdot)$ as follows:

1. (Input :) $(p, \#\Psi(d), n)$
2. (Output: ) Gödel number of a proof tree of $\psi(\mathbf{d},\mathbf{n})$ or nothing (does not halt).
3. Let $M_0(c) = \min\{m \in \mathbb{N} \mid m \le n < m + |m|^c\}$ and $M_0(c) = \max\{m \in \mathbb{N} \mid m \le n < m + |m|^c\}$
4. Set $c \leftarrow 0$ and $m \leftarrow M_0(c)$.
5. Run the following computation
$$\mathrm{U}_{\mathrm{PTM}}(e, (p, \#\Psi(d,c), m) = z,\quad \mathrm{U}_{\mathrm{PTM}}(v_T, (\#\rho, z)),$$
where $\rho \equiv \forall \mathbf{x}\ (\mathbf{m} \le \mathbf{x} < \mathbf{m} + |\mathbf{m}|^{\mathbf{c}})\ \ \psi(\mathbf{d},\mathbf{x})$
6. Check whether $\mathrm{U}_{\mathrm{PTM}}(v_T, (\#\rho, z))$ accepts or rejects. If it accepts, go to 7. If it rejects, set $m \leftarrow m+1$ and check whether $m > M_1$. If $m \le M_1$, then go to 5. Otherwise, set $c \leftarrow c+1$, compute $M_0(c)$ and $M_0(1)$, $m \leftarrow M_0(c)$, and go to 5.
7. . Make a new proof tree, $\pi_2$, for $\psi(\mathbf{d},\mathbf{n})$, from proof tree $\pi_1$ ($z = \#\pi_1$) for $\rho$, as follows:
$$\pi_2 \ \ \equiv\ \ <\psi(\mathbf{d},\mathbf{n}), \text{Modus Ponens}> [\pi_1, \text{Axiom Y}],$$
where Axiom Y is a logical axiom, " $\forall \mathbf{x}\ (\mathbf{m} \le \mathbf{x} < \mathbf{m} + |\mathbf{m}|^{\mathbf{c}})\ \ \psi(\mathbf{d},\mathbf{x})\ \ \to\ \ \psi(\mathbf{d},\mathbf{n}))$".
Output $\pi_2$ for the proof tree of formula $\psi(\mathbf{d},\mathbf{n})$.

Here, if $c$ is a constant in $|n|$, then TM $\mathrm{U}(\widetilde{e},\cdot)$ should be PTM in $|n|$.
Therefore, from the construction of $\mathrm{U}(\widetilde{e},\cdot)$, we obtain

$$
\begin{aligned}
&\forall e \in \mathbb{N}\ \exists \widetilde{e} \in \mathbb{N}\ \forall d \in \mathbb{N} \\
&\quad \mathrm{PA} \vdash\ \forall \mathbf{m}\ \forall \mathbf{c}\ \ (\ \ \mathrm{Pr}_T[\forall \mathbf{n}(\mathbf{m} \le \mathbf{n} < \mathbf{m} + |\mathbf{m}|^{\mathbf{c}})\ \ \psi(\mathbf{d},\mathbf{n})](\mathbf{e}, \ulcorner \Psi(d,c) \urcorner, \mathbf{m}) \\
&\qquad\qquad \to\ \ \forall \mathbf{n}\ (\mathbf{m} \le \mathbf{n} < \mathbf{m} + |\mathbf{m}|^{\mathbf{c}})\ \ \mathrm{Pr}_T[\psi(\mathbf{d},\mathbf{n})](\widetilde{\mathbf{e}}, \ulcorner \Psi(d) \urcorner, \mathbf{n})\ )
\end{aligned}
$$

$\dashv$

We now assume that

$$
\begin{aligned}
&\exists e \in \mathbb{N}\ \ \forall e' \in \mathbb{N}\ \ \forall m \in \mathbb{N}\ \ \exists \ell \ge m\ \ \exists c \in \mathbb{N} \\
&\mathrm{PTM}_e(\ell) \vdash_T\ \exists \mathbf{n}\ (\mathbf{l} \le \mathbf{n} < \mathbf{l} + |\mathbf{l}|^{\mathbf{c}})\ \ \neg\mathrm{Pr}_T[\exists \mathbf{x} \ge \mathbf{n}\ \ \neg\mathrm{DecSAT}(h(\mathbf{e}'),\mathbf{x})](\mathbf{e}', \ulcorner \Phi(e') \urcorner, \mathbf{n}).
\end{aligned}
$$

Then, from Eq.(91),

$$
\begin{aligned}
&\exists e \in \mathbb{N}\ \ \forall e' \in \mathbb{N}\ \ \forall m \in \mathbb{N}\ \ \exists \ell \ge m\ \ \exists c \in \mathbb{N} \\
&\mathrm{PTM}_e(\ell)\ \vdash_T\ \ \neg\mathrm{Pr}_T[\forall \mathbf{n}(\mathbf{l} \le \mathbf{n} < \mathbf{l} + |\mathbf{l}|^{\mathbf{c}})\exists \mathbf{x} \ge \mathbf{n}\ \ \neg\mathrm{DecSAT}(h(\mathbf{e}'),\mathbf{x})](\mathbf{e}', \ulcorner \Phi(e',c) \urcorner, \mathbf{l}).
\end{aligned}
$$

This contradicts Lemma 70.

$\dashv$

We obtain the following lemma immediately from Lemma 71 and the PTM-$\omega$-consistency of $S$.

**Lemma 72.** *Let theory $T$ be a consistent PT-extension of PA, $S$ be a consistent PT-extension of $T$, and $S$ be PTM-$\omega$-consistent for $\Delta_2^P$ over $T$. Let $\Phi(e) \equiv \{\exists \mathbf{x} \geq \mathbf{a}\ \neg\mathrm{DecSAT}(h(\mathbf{e}), \mathbf{x}) \mid a \in \mathbb{N}\}$ and $h$ be a primitive recursive function.*

$$\forall e \in \mathbb{N}\ \exists e' \in \mathbb{N}\ \exists m \in \mathbb{N}\ \forall \ell \geq m \quad \mathrm{PTM}_e(\ell) \not\vdash_S \exists \mathbf{n} \geq \mathbf{l}\ \neg\mathrm{Pr}_T[\exists \mathbf{x} \geq \mathbf{n}\ \neg\mathrm{DecSAT}(h(\mathbf{e}'), \mathbf{x})](\mathbf{e}', \ulcorner\Phi(e')\urcorner, \mathbf{n}).$$

**Theorem 73.** *Let theory $T$ be a consistent PT-extension of PA, and $S$ be a consistent PT-extension of $T$ and PTM-$\omega$-consistent for $\Delta_2^P$ over $T$.*

$$S \not\vdash \quad \forall \mathbf{e}\ \ \exists \mathbf{l}\ \ \forall \mathbf{n} \geq \mathbf{l} \quad \neg\mathrm{Pr}_T[\exists \mathbf{x} \geq \mathbf{n}\ \ \neg\mathrm{DecSAT}(h(\mathbf{e}), \mathbf{x})](\mathbf{e}, \ulcorner\Phi(e)\urcorner, \mathbf{n}).$$

*Namely, the PTM-$\omega$-consistency of $T$ for $\neg\mathrm{DecSAT}(\mathbf{e}^*, \mathbf{x})$, which is sufficient to prove $T \not\vdash \overline{\mathrm{P} \neq \mathrm{NP}}$, cannot be proven in $S$ (see Eqs. (83) and (84)).*

*Proof.* Let assume that

$$S \vdash \quad \forall \mathbf{e}\ \ \exists \mathbf{l}\ \ \forall \mathbf{n} \geq \mathbf{l} \quad \neg\mathrm{Pr}_T[\exists \mathbf{x} \geq \mathbf{n}\ \ \neg\mathrm{DecSAT}(h(\mathbf{e}), \mathbf{x})](\mathbf{e}, \ulcorner\Phi(e)\urcorner, \mathbf{n}).$$

Then

$$\forall e \in \mathbb{N}\quad S \vdash \quad \exists \mathbf{l}\ \ \forall \mathbf{n} \geq \mathbf{l} \quad \neg\mathrm{Pr}_T[\exists \mathbf{x} \geq \mathbf{n}\ \ \neg\mathrm{DecSAT}(h(\mathbf{e}), \mathbf{x})](\mathbf{e}, \ulcorner\Phi(e)\urcorner, \mathbf{n}).$$

This implies

$$\forall e \in \mathbb{N}\quad S \vdash \quad \forall \mathbf{l}\ \ \exists \mathbf{n} \geq \mathbf{l} \quad \neg\mathrm{Pr}_T[\exists \mathbf{x} \geq \mathbf{n}\ \ \neg\mathrm{DecSAT}(h(\mathbf{e}), \mathbf{x})](\mathbf{e}, \ulcorner\Phi(e)\urcorner, \mathbf{n}).$$

This means that there exists $e^* \in \mathbb{N}$ such that

$$\forall e \in \mathbb{N}\ \ \forall \ell \in \mathbb{N} \quad \mathrm{PTM}_{e^*}(\ell) \vdash_S \quad \exists \mathbf{n} \geq \mathbf{l} \quad \neg\mathrm{Pr}_T[\exists \mathbf{x} \geq \mathbf{n}\ \ \neg\mathrm{DecSAT}(h(\mathbf{e}), \mathbf{x})](\mathbf{e}, \ulcorner\Phi(e)\urcorner, \mathbf{n}).$$

This contradicts Lemma 72.

Thus,

$$S \not\vdash \quad \forall \mathbf{e}\ \ \exists \mathbf{l}\ \ \forall \mathbf{n} \geq \mathbf{l} \quad \neg\mathrm{Pr}_T[\exists \mathbf{x} \geq \mathbf{n}\ \ \neg\mathrm{DecSAT}(h(\mathbf{e}), \mathbf{x})](\mathbf{e}, \ulcorner\Phi(e)\urcorner, \mathbf{n}).$$

$\dashv$

## 9 Unprovability of the Security of Computational Cryptography

This section will show that the security of any computational cryptographic scheme is unprovable in the standard notion of the modern cryptography, where an adversary is modeled to be a polynomial-time Turing machine.

First we will introduce a very fundamental cryptographic problem, the intractability of totally inverting a one-way function by a deterministic PTM (polynomial-time Turing machine). Modern computational cryptography is based on the assumption of the existence of one-way functions [11].[5]

[5] Although a one-way function is usually defined against probabilistic PTMs, one-wayness against deterministic PTMs is more fundamental than that against probabilistic PTMs. For example, if a function is one-way against probabilistic PTMs, then the function will be also one-way against deterministic PTMs. That is, proving the one-wayness of a function against probabilistic PTMs always implies proving that against deterministic PTMs. However, the reverse is not always true.

In other words, to prove any level of security of such a computational cryptosystem implies proving the one-wayness (the intractability of total inversion by any deterministic PTM) of an underlying function. Therefore, if it is impossible to prove the one-wayness of any function, it will be also impossible to prove any level of security of any computational cryptographic scheme.

This section will show that the intractability of totally inverting a function by a deterministic PTM is impossible to prove formally in the standard modern cryptographic setting.

**Definition 74.** *Let $n \in \mathbb{N}$, and $f_n : \mathbb{Z}/n\mathbb{Z} \to \mathbb{Z}/n^c\mathbb{Z}$ be a function with parameter $n$ and $\mathcal{F} \equiv \{f_n \mid n \in \mathbb{N}\}$ be a set of functions, where $c$ is a constant.*

*$\mathcal{F}$ is called one-way if there is no (deterministic) PTM $\mathrm{U_{PTM}}(e,\cdot)$ such that for all $x = (n,y,z) \in \mathbb{N} \times \mathbb{Z}/n^c\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$ $\mathrm{U_{PTM}}(e,x)$ outputs $w \in \mathbb{Z}/n\mathbb{Z}$ if there exists $w$ such that $y = f_n(w)$, and outputs nothing otherwise. Here $\mathrm{Size}(x) = \mathrm{Size}(n,y,z) = |n|$.*

**Definition 75.** *Let $\mathcal{F} \equiv \{f_n \mid n \in \mathbb{N}\}$ be a set of functions (see Definition 74). Let $\mathrm{Inv}$ be an inversion oracle (a deterministic algorithm or a table) such that $\mathrm{Inv}(n,y)$ outputs one of $\{w \mid y = f_n(w)\}$ if there exists $w$ such that $y = f_n(w)$, and outputs nothing otherwise.*

*$\mathcal{F}$ is called decisionally one-way if, for any inversion oracle Inv, there is no (deterministic) PTM $\mathrm{U_{PTM}}(e,\cdot)$ such that, for all $x = (n,y,z) \in \mathbb{N} \times \mathbb{Z}/n^c\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$, $\mathrm{U_{PTM}}(e,x)$ accepts if and only if $\mathrm{Inv}(n,y) > z$. (Note that $w$ is uniquely determined for each Inv and $(n,y)$.)*

**Lemma 76.** *$\mathcal{F}$ is one-way if and only if $\mathcal{F}$ is decisionally one-way.*

*Proof.* It is trivial that if a PTM can invert $f_n$, it can also solve the corresponding decisional problem.

On the other hand, we will show that if a PTM can solve the decisional problem of $f_n$, then there exists a PTM that can invert $f_n$. In other words, $f_n$ can be completely inverted by using the solution of the decisional problem as a black-box $|n|$ times. Here, we use binary search. Given a problem $(n,y)$ to invert $f_n$, queries to $\mathrm{U_{PTM}}(e,\cdot)$ are $(n,y,\lfloor n/2 \rfloor)$, $(n,y,\lfloor (3/4)n \rfloor)$ (if the answer to the previous query is accept), .... Repeating this binary search $|n|$ times yields an integer $v \in \mathbb{Z}/n\mathbb{Z}$. Then, check whether $y = f_n(v)$ holds or not. If it holds, set $w = v$. Otherwise, set $w$ to a null string (or decide that there exists no value of $w \in Zn$ such that $y = f_n(w)$).

Therefore, $\mathcal{F}$ is *one-way* if and only if no PTM can solve the corresponding decisional problem.

⊣

As shown in Lemma 76, the one-wayness of function family $\mathcal{F}$ can be characterized by the intractability of the decisional problem, which can be also characterized by a formula in theory $T$ as shown below.

**Definition 77.** *Let $R_{\mathcal{F}}^{Inv} \subset \mathbb{N}^4$ be a relation with respect to inversion oracle Inv such that $(e,n,y,z) \in R_{\mathcal{F}}^{Inv}$ if and only if $(n,y,z) \in \mathbb{N} \times \mathbb{Z}/n^c\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$, and $\mathrm{U_{PTM}}(e,(n,y,z))$ accepts if and only if $\mathrm{Inv}(n,y) > z$.*

**Lemma 78.** *If $\mathcal{F}$ is one-way, P$\neq$NP.*

*Proof.* $\mathcal{F}$ is one-way, if and only if, for any inversion oracle *Inv*, there is no (deterministic) PTM $\mathrm{U_{PTM}}(e,\cdot)$ such that, for all $(n,y,z) \in \mathbb{N} \times \mathbb{Z}/n^c\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$, $\mathrm{U_{PTM}}(e,(n,y,z))$ accepts if and only if $(e,n,y,z) \in R_{\mathcal{F}}^{Inv}$.

Language $\{(e,n,y,z) \in R_{\mathcal{F}}^{Inv}\}$ is clearly NP. Therefore, if P = NP, there is no one-way function family.

$\dashv$

We then obtain the following lemma immediately from Lemma 78 and Theorem 67.

**Lemma 79.** *For any theory $T$ which is a consistent PT-extension of PA and* PTM-$\omega$*-consistent for $\Delta_2^P$, there exists no proof of the one-wayness of any function family in $T$.*

To study the (im)possibility of proving the one-wayness of a function family, $\mathcal{F} \equiv \{f_n \mid n \in \mathbb{N}\}$, we need to make a model of provers (and adversaries). We now present a reasonable model of provers.

**Definition 80.** *(Model of a prover in computational cryptography)*

*A prover is a PTM, which, given the (finite size of) description of a cryptographic problem, outputs a proof of the problem in a theory $T$ that is a constant PT-extension of PA and PTM-$\omega$-consistent for $\Delta_2^P$.*

This model should be justified by the fact that an adversary is modeled to be a PTM in the definition of the one-wayness of a function family in the above definition, which is the standard setting in modern cryptography. In other words, the models of a prover and adversary should be equivalent, since both prover and adversary are theoretical models of our human being who analyzes the security of a one-way function family to prove the security or to break it. The key part of this model is that theory $T$ available for a prover to prove the security should be PTM-$\omega$-consistent, since a prover is assumed to be a PTM. This is because PTM-$\omega$-*inconsistent* theory may include an unreasonably strong axiom (e.g., $\overline{\mathrm{P}{\neq}\mathrm{NP}}$ itself) that no PTM can prove asymptotically in PA.

We now obtain the following theorem from Lemma 79.

**Theorem 81.** *Under the prover model of Definition 80, there exists no proof of the one-wayness of any function family.*

Note that PTM is just a one possible model of the feasible computation for our human being. Even if in the future we have to change the feasible computation model of our human being, the impossibility result of Theorem 81 remains unchanged, because the computational models of prover and adversary should be equivalent in any feasible computation model. We will show similar results in various computational classes in Part 2 of this paper.

In addition, combining the result [15] with Theorem 67 yields the following consequence:

**Theorem 82.** *Under the prover model of Definition 80, there exists no proof of the existence of a (black-box) reduction from a one-way permutation to a secret key agreement.*

## 10 Proof Complexity

In order to characterize the computational complexity to recognize the feasibility (triviality) of a theory to prove a statement, this section introduces a proof complexity.

**Definition 83.** *We say that the proof complexity of $\varphi$ is $O(\mathcal{C})$ if there exists a proof of $\varphi$ in a theory $T$ that is a consistent PT-extension of PA and $\mathcal{C}'$-$\omega$-consistent for any class $\mathcal{C}'$ that includes $\mathcal{C}$. We say that the proof complexity of $\varphi$ is $\Omega(\mathcal{C})$ if there exists no proof of $\varphi$ in any theory $T$ that is a consistent PT-extension of PA and $\mathcal{C}$-$\omega$-consistent.*

We now assume that PA is $\mathcal{C}$-$\omega$-consistent for any computational class $\mathcal{C}$, which includes a computational class with constant-time complexity, $O(1)$.

Then, we obtain the following result:

- Let $\overline{\mathrm{P}{\neq}\mathrm{EXP}}$ be a sentence that formalizes the statement of P≠EXP in PA in a manner similar to that in Section 6.
  The proof complexity of $\overline{\mathrm{P}{\neq}\mathrm{EXP}}$ is $O(1)$ (i.e., a *constant*-time), since $\overline{\mathrm{P}{\neq}\mathrm{EXP}}$ can be proven in PA.
- Let Con(PA) be a sentence that formalizes the consistency of PA in PA. In other words,
  $$\mathrm{Con(PA)} \equiv \forall \mathbf{x}\ \neg \text{PTM-Acpt}(\mathbf{v}_{\mathrm{PA}}, \ulcorner \bot \urcorner, \mathbf{x}).$$
  (For the notation and related result, see Lemmas 7 and 9.)
  The proof complexity of Con(PA) is $O(\mathrm{P})$ (i.e., a *polynomial*-time), since Con(PA) cannot be proven in PA by the second Gödel incompleteness theorem, but it has a polynomial-time proof over PA (Lemma 9), and there exists a PTM-$\omega$-consistent theory $T$ (e.g., $T \equiv \mathrm{PA} + \mathrm{Con(PA)}$) that proves Con(PA).
- The proof complexity of $\overline{\mathrm{P} \neq \mathrm{NP}}$ is $\Omega(\mathrm{P})$ (i.e., *super-polynomial*-time), since $\overline{\mathrm{P} \neq \mathrm{NP}}$ cannot be proven in a PTM-$\omega$-consistent theory (Theorem 67).
- The proof complexity of $\sigma$ ($\equiv\ \forall \mathbf{e}\ \exists \mathbf{l}\ \forall \mathbf{n} \geq \mathbf{l}\ \ \neg \mathrm{Pr}_T[\exists \mathbf{x} \geq \mathbf{n}\ \ \varphi(\mathbf{x})](\mathbf{e}, \ulcorner \Phi \urcorner, \mathbf{n})$) is $\Omega(\mathrm{P})$ (i.e., *super-polynomial*-time), from Theorem 73. Therefore, the proof complexity of the independence of P vs NP from $T$ by proving the PTM-$\omega$-consistency is also $\Omega(\mathrm{P})$.

## 11 Informal Observations

If we assume Hypotheses 1 and 2 in Section 1.3, our main theorem implies that P vs NP is independent from PA. As the next step, it is natural to try to prove Hypothesis 2 (PTM-$\omega$-consistency of PA). Since PA cannot prove Hypothesis 2, a theory $T$ to prove Hypothesis 2 should include an axiom, $X$, outside PA. What axiom is appropriate for this purpose?

Usually it is not so easy for mathematicians/logicians to select/determine an appropriate axiom that would be widely recognized as feasible. An extreme strategy is to adopt Hypothesis 2 itself as the new axiom, but such an axiom would not be accepted as feasible. Then, what is the criterion of a feasible axiom? Unfortunately we now have no candidate. Here we note that consistency and $\omega$-consistency are too weak as such a criterion since PA + Hypothesis 2 is $\omega$-consistent (i.e., consistent) if Hypothesis 2 is true. Currently, the feasibility of an axiom is decided only by whether it is widely accepted by many mathematicians/logicians to be feasible.

Our result may suggest a criterion for the feasibility of an axiom/theory.

Although axiom $X$ is outside PA (i.e., PA cannot prove $X$), there exists an *asymptotic* proof of $X$ over PA, if $X$ is true. In other words, a Turing machine can produce an asymptotic proof of $X$ over PA. We then consider the computational complexity of a Turing machine that can produce an asymptotic proof of $X$ over PA. According to Theorem 73, theory $T = \mathrm{PA} + X$ to prove Hypothesis 2 should be PTM-$\omega$-*inconsistent*, and Remark 5 of Definition 62 shows that $X$ *cannot* be asymptotically proven by any *polynomial-time bounded* TM (i.e., PTM) over PA, under some assumption.

If the computational capability of human beings (along with our available/feasible computing facilities) is modeled as a polynomial-time Turing machine, which is widely accepted as a feasible computation model, our result implies that no human being can prove axiom $X$ asymptotically over PA. This may imply that axiom $X$ cannot be perceived as a feasible (or trivially true) statement by human beings, since it is beyond our capability to prove (or recognize the truth of) it even asymptotically over PA. If so, a theory $T$ in which Hypothesis 2 can be proven should include an axiom that cannot be perceived as feasible by human beings. That is, Hypothesis 2 cannot be proven in any feasible theory $T$, which is widely recognized to be feasible by mathematicians/logicians (i.e., human beings). In other words, even if Hypotheses 1 and 2 are true and P vs NP is independent from PA, such an independency cannot be proven (through proving Hypothesis 2) in any feasible theory $T$ for us. Similarly, even if Hypothesis 1 is true, $\overline{\mathrm{P}{\neq}\mathrm{NP}}$ may not be proven in any feasible theory for human beings.

Con(PA), which is a formal sentence representing the consistency of PA in PA, is also unprovable in PA. That is, a theory $T$ to prove Con(PA) should include an axiom, $Y$, outside PA. In contrast with the above-mentioned case of proving Hypothesis 2, Con(PA) can be asymptotically proven by a polynomial-time (more precisely, linear-time) Turing machine over PA (Lemma 9), and can be proven in a PTM-$\omega$-consistent theory, PA + Con(PA), if PA is PTM-$\omega$-consistent. Although Con(PA) would not be accepted as a feasible axiom, the fact that Con(PA) can be proven in a PTM-$\omega$-consistent theory may imply the existence of a feasible axiom, $Y$, for us such that $T$ = PA + $Y$ can prove Con(PA) and $T$ is PTM-$\omega$-consistent. Actually, Gentzen [10] proved Con(PA) in a feasible theory for us, which is in ZF (formal theory of set theory) and whose additional axiom, $Y$, to PA is regarding transfinite induction (corresponding to the axiom of foundation in ZF).

The relationship between Gödel's incompleteness theorem and our result is similar to that between recursion theory and computational complexity theory. Recursion theory studies (un)computability on Turing machines, which are widely accepted as the most general computation model (the Church-Turing thesis), while computational complexity theory studies (un)computability on a much more restricted computation model, a *feasible computation model for us (human beings)*, i.e., polynomial-time Turing machines (PTMs). The major difference in the computation model of recursion theory and the computational complexity theory is that the former is resource *unbounded*, while the latter is resource *bounded* (polynomial-time bounded).

Gödel's incompleteness theorem is a result on unprovabilty in the most general formal theories, consistent theories (or slightly restricted theories, $\omega$-consistent theories), that include PA, while our main theorem is a result on unprovabilty in much more restricted formal theories, *feasible formal theories for us (human beings)*, i.e., PTM-$\omega$-consistent theories, that include PA. The major difference in the formal theory of Gödel's incompleteness theorem and our main theorem is that the former considers only the feasibility of the theory for a resource *unbounded* machine (i.e., consistency or $\omega$-consistency), while the latter considers the feasibility of the theory for a resource *bounded* (polynomial-time bounded) machine (i.e., PTM-$\omega$-consistency). In fact, as shown in Remark 3 of Definition 62, the resource *unbounded* version of PTM-$\omega$-consistency is $\omega$-consistency.

Here, it is worth noting that it should be controversial to decide the feasibility of a theory by PTM-$\omega$-consistency, where all axioms and deductions in a theory should be asymptotically proven by a PTM, but that it might be similar to the situation in computational complexity theory where it should have been controversial to characterize a feasible computation by class P, since class P clearly includes many infeasible computations for us such as $n^{10000}$ computational complexity in input size $n$.

Therefore, it may be reasonable to consider that class P is introduced to characterize an *infeasible* computation, rather than to characterize a *feasible* computation. That is, we consider that a computation outside P is *infeasible*, or an infeasible computation is characterized as a super-polynomial-time computation class (super-P), since almost all computations in super-P are actually infeasible except a very small fraction of super-P such as a computation with $O(n^{\log\log\log n})$

complexity (In contrast, almost all computations in P are infeasible such that a computation with $n^c$ complexity is infeasible for $c > 20$, and only a small fraction of P is feasible).

Similarly, it may be reasonable to consider that PTM-$\omega$-*inconsistency*, rather than PTM-$\omega$-*consistency*, is introduced to characterize *infeasible* theories. In fact, as we mentioned above, it is considered to be difficult for us (or PTMs) to perceive the feasibility (triviality) of an axiom of a PTM-$\omega$-*inconsistent* theory, since an axiom of a PTM-$\omega$-*inconsistent* theory *cannot* be proven even asymptotically by any PTM over PA, under some assumption (Remark 5 of Definition 62). Our main theorems imply that P$\neq$NP (or any super-polynomial-time lower bound in PSPACE) is provable only in such an *infeasible* theory.

Note that our results do not deny the possibility of proving P=NP in a feasible theory for us, if P=NP is true.

Gödel's second incompleteness theorem has a positive significance in that it helps us to separate two distinct theories, $T$ and $S$, because $T \vdash \mathrm{Con}(S)$ implies that $T \neq S$ (and $T \supset S$) since $S \not\vdash \mathrm{Con}(S)$ by Gödel's second incompleteness theorem. Using this idea, the results of this paper may provide some hint of the computational capability of human beings.

Let $M$ be a machine whose computational capability is unknown. If $C$ is a computational class, our result helps us to characterize the computational power of $M$ relative to $C$, because $M \vdash_T \mathrm{SuperLowerBound}(C)$ where theory $T$ is feasible for $M$ implies that the computational power of $M$ should be beyond $C$. Here $\mathrm{SuperLowerBound}(C)$ denotes a formula to represent the super-$C$ computational lower bound in PA. If we assume $M$ to be a computational model of human beings, then our obtained computational lower bound result of $M \vdash_T \mathrm{SuperLowerBound}(C)$ in a feasible theory $T$ for us implies the upper bound of our computational power. For example, we have already obtained a proof of a super-$\mathrm{AC}^0$ lower bound [9, 24]. This fact means that the computational power of human beings may exceed $\mathrm{AC}^0$.

This result may also give us some hint as to why all known results of computational lower bounds inside PSPACE are limited to very weak or restricted computational classes. If the computational capability of human beings is considered to far exceed the target computational class for lower bound proof (e.g., the target class is $\mathrm{AC}^0$), then it is likely that we may produce a proof of the lower bound statement in a feasible theory for us. However, if our computational capability is comparable to (or is not much beyond) the target computational class for lower bound proof, then it may be very unlikely that we can provide its proof in a feasible theory for us. In other words, the best result of computational lower bounds may suggest the computational capability of human beings.

## 12 Concluding Remarks

This paper introduced a new direction for studying computational complexity lower bounds; resource bounded unprovability (Sections 2 and 3) and resource bounded undecidability (Sections 4 and 5). This approach can be generalized to various systems by generalizing verification machines, $\mathrm{U}_{\mathrm{PTM}}(v_T, \cdot)$ in proof systems (Section 2) and $\mathrm{U}(v, \cdot)$ in decision systems (Section 4).

As mentioned in Section 11, the relationship between Gödel's incompleteness theorem and our result is similar to that between recursion theory and computational complexity theory. Recursion theory studies (un)computability on the most general computation model, Turing machines (TMs), while computational complexity theory studies (un)computability on a much more restricted computation model, a *feasible computation model for us*, i.e., polynomial-time Turing machines (PTMs), where PTMs are a resource (polynomial-time) bounded version of TMs. Gödel's incompleteness theorem is a result on unprovabilty in the most general formal theories, consistent theories (or slightly restricted theories, $\omega$-consistent theories), that includes PA, while our main theorem is a result on unprovabilty in much more restricted formal theories, *feasible theories for us*, i.e., PTM-$\omega$-consistent theories, that includes PA, where PTM-$\omega$-consistent theories are a resource (polynomial-time) bounded version of $\omega$-consistent theories.

In Part 2, we will extend these results to other computational classes and show that: for all $i \geq 1$, a super-$\Pi_i^P$ lower bound and a super-$\Sigma_i^P$ lower bound cannot be proven in a $\Sigma_i^P$-$\omega$-consistent theory and a $\Pi_i^P$-$\omega$-consistent theory, respectively. For all $i \geq 1$, a super-AC$^{i-1}$ lower bound and a super-NC$^i$ lower bound cannot be proven in an AC$^{i-1}$-$\omega$-consistent theory and an NC$^i$-$\omega$-consistent theory, respectively. In addition, Part 2 will present similar results on probabilistic and quantum computational classes, since a probabilistic TM and quantum TM can be simulated by a classical deterministic TM; they can be formulated in PA in a manner similar to that in Part 1. Thus, for example, we will show that a super-BPP lower bound cannot be proven in a BPP-$\omega$-consistent theory and that a super-BQP lower bound cannot be proven in a BQP-$\omega$-consistent theory.

## Acknowledgments

The authors would like to thank Noriko Arai, Toshiyasu Arai, Amit Sahai, Mike Sipser, Jun Tarui and Osamu Watanabe for their invaluable comments and suggestions. We would also like to thank anonymous reviewers of ECCC and FOCS'04 for valuable comments on previous versions of our manuscript.

## References

1. T.P. Baker, J. Gill and R. Solovay, Relativizations of the P=?NP Questions, SIAM J.Comput., Vol.4, No.4, pp.431-442, 1975.
2. J. Barwise, Mathematical Logic, (especially, Section D.1 "The Incompleteness Theorems," by C. Smorynski), North Holland, 1977.
3. P. Beame and T. Pitassi, Propositional Proof Complexity: Past, Present and Future, Tech. Rep. TR98-067, ECCC, 1998.
4. S. Ben-David and S. Halevi, On the Independence of P versus NP, Technion, TR 714, 1992.
5. S. Buss, Bounded Arithmetic, Bibliopolis, Napoli, 1986.
6. N.C.A. da Costa and F.A. Doria, Consequence of an Exotic Definition, Applied Mathematics and Computation, 145, pp.655-665, 2003.
7. H.B. Enderton, A Mathematical Introduction to Logic, Academic Press, 2001.
8. R. Fagin, Generalized First Order Spectra and Polynomial-time Recognizable Sets, Complexity of Computation, ed. R. Karp, SIAM-AMS Proc. 7, pp.27–41, 1974.
9. M. Furst, J.B. Saxe and M. Sipser, Parity, Circuits and the Polynomial-time Hierarchy. Math. Syst. Theory, 17, pp.13-27, 1984.
10. G. Gentzen, Die gegenwärtige Lage in der mathematischen Grundlagenforshung. Neue Fassung des Widerspruchsfreitsbeweises für die reine Zahlentheorie, Leipzig, 1938.
11. O. Goldreich, Foundations of Cryptography, Vol.1, Cambridge University Press, 2001.
12. O. Goldreich, Modern Cryptography, Probabilistic Proofs and Pseudorandomness, Springer-Verlag, 1999.
13. J. Hartmanis and J. Hopcroft, Independence Results in Computer Science, SIGACT News, 8, 4, pp.13-24, 1976.
14. J. Hartmanis, Feasible Computations and Provable Complexity Problems, SIAM, 1978.
15. R. Impagliazzo and S. Rudich, Limits on the Provable Consequences of One-Way Permutations, Proc. of STOC'89, 1989.

16. K. Iwama and H. Morizumi, An Explicit Lower Bound of $5n - o(n)$ for Boolean Circuits, Proc. of MFCF, pp.353-364, 2002.
17. S. Kurz, M.J. O'Donnell and S. Royer, How to Prove Representation-Independent Independence Results, Information Processing Letters, 24, pp.5-10, 1987.
18. J. Krajicek, Bounded Arithmetic, Propositional Logic, and Complexity Theory, Cambridge University Press, 1995.
19. P. Pudlák, The Lengths of Proofs, Chapter VIII, Handbook of Proof Theory, S. Buss Ed., pp.547-637, Elsevier, 1998.
20. A.A. Razborov, Resolution Lower Bounds for Perfect Matching Principles, Proc. of Computational Complexity, IEEE, pp. 29-38, 2002.
21. A.A. Razborov and S. Rudich, Natural Proofs, JCSS, Vol.55, No.1, pp.24–35, 1997.
22. J.R. Shoenfield, Mathematical Logic, Association for Symbolic Logic, 1967.
23. M. Sipser, Introduction to the Theory of Computation, PWS Publishing Company, 1997.
24. R. Smolensky, Algebraic Methods in the Theory of Lower Bounds of Boolean Circuit Complexity, Proc. of STOC'87, pp.77-82, 1987.