On the Power of PP

Nikolai K.Vereshchagin
The Institute of New Technologies
Kirovogradskaja 11, Moscow 113587, Russia
e-mail: amuchnik@globlab.msk.su

Abstract

It is proved that $\mathbf{MA} \subset \mathbf{PP}$ *(relativizable) and that* $\mathbf{AM}^A \cap \mathbf{co\text{-}AM}^A \not\subset \mathbf{PP}^A$ *for some oracle* A.

1 Introduction

Recently two interesting results about the class **PP** were obtained. We mean the result of S.Toda [1] that polynomial hierarchy **PH** is polynomially Turing reducible to **PP** and the result that **PP** is closed under polynomial truth table reductions (see [2] and [3]). These results make more interesting to study **PP**. Another reason to study **PP** is that this class has the following interpretation. Random input r of the probabilistic machine M that recognizes a language L can be regarded as a voter and the output $M(x, r)$ of M on the input word x and random input r can be regarded as the opinion of voter r about whether x is in L. From this point of view **PP** is the class of all languages L such that membership of x in L can be determined via election with $2^{poly(|x|)}$ voters, every voter being polynomial time bounded.

In this paper we prove one (simple) positive theorem about **PP** and one negative theorem:

Theorem 1. $\mathbf{MA} \subset \mathbf{PP}$.

Theorem 2. $\mathbf{AM}^A \cap \mathbf{co\text{-}AM}^A \not\subset \mathbf{PP}^A$ *for some oracle* A.

Theorem 1 is relativizable. Theorem 2 shows that theorem 1 cannot be strengthened to relativezable inclusion $\mathbf{AM} \subset \mathbf{PP}$ (remember that $\mathbf{MA} \subset \mathbf{AM}$ [4]). Another meaning of theorem 2 is that Toda's result that PH is Turing reducible to **PP** cannot be strengthened to relativizable inclusion

PH $\subset$ **PP** because **AM** $\subset \Pi_2$ [4] (note that Toda's proof is relativizable). From theorems 1 and 2 we can deduce that $\mathbf{AM}^A \cap \textbf{co-AM}^A \not\subset \mathbf{MA}^A$ for some oracle A, but the latter result is easier than theorem 2.

2 Definitions

We'll consider languages over the binary alphabet $\mathbf{B} = \{0, 1\}$. The set of all binary words of length n is denoted by $\mathbf{B}$. Functions with binary values are called predicates. Instead of $P(x) = 1$ where P is a predicate we'll write simply $P(x)$. All Turing machines output 0, 1.

Definition 1. *A language L belongs to* **PP** *iff there is a polynomial time probabilistic Turing machine M such that $x \in L \Leftrightarrow \mathrm{Prob}[M(x, r) = 1] > 1/2$ where the probability is taken over the uniform distribution in the set of random inputs r of M.*

Remark. We can easily prove that in definition 1 the threshold 1/2 may be replaced with any other constant or with any rational number of the form $a(x)/2^{s(|x|)}$ where s is a polynomial and $a\colon \mathbf{B}^* \to \mathbf{N}$ is polynomially computable function (integers are written in binary notation).

Definition 2. *$L \in$ **MA** iff there are a polynomial p and polynomially computable predicate $Q(x, r, s)$ such that*

$$x \in L \Rightarrow \exists s \in \mathbf{B}^{p(|x|)} \; Prob_r[Q(x, r, s)] > 2/3$$

$$x \notin L \Rightarrow \forall s \in \mathbf{B}^{p(|x|)} \; Prob_r[Q(x, r, s)] < 1/3,$$

where probability is taken over uniform distribution in $\mathbf{B}^{p(|x|)}$.

Definition 3. *$L \in$ **AM** iff there are a polynomial p and polynomial computable predicate $Q(x, r, s)$ such that*

$$x \in L \Rightarrow Prob[\exists s \in \; \mathbf{B}^{p(|x|)} Q(x, r, s)] > 2/3$$

$$x \notin L \Rightarrow Prob[\exists s \in \; \mathbf{B}^{p(|x|)} Q(x, r, s)] < 1/3$$

where probability is taken over uniform distribution in $r \in \mathbf{B}^{p(|x|)}$.

Theorem [4]. **MA** $\subset$ **AM**.

3 Results

Theorem 1. **MA** $\subset$ **PP**.

Proof. Let $L \in \mathbf{MA}$ and let p and Q are correspondingly polynomial and predicate from definition 2. Using standard amplification we can construct a new polynomial p_1 and a new polynomially computable predicate Q_1 such that

$$x \in L \Rightarrow \exists s \in \mathbf{B}^{p(|x|)} \operatorname{Prob}[Q_1(x,r,s)] > 1 - 4^{-p(|x|)}$$

$$x \notin L \Rightarrow \forall s \in \mathbf{B}^{p(|x|)} \operatorname{Prob}[Q_1(x,r,s)] > 4^{-p(|x|)}$$

where probability is taken over the uniform distribution in $r \in \mathbf{B}^{p_1(|x|)}$. Consider now the uniform distribution on pairs $\langle r, s\rangle \in \mathbf{B}^{p(|x|)+p_1(x)}$. We have $x \in L \Rightarrow$

$$\operatorname{Prob}[Q_1(x,r,s)] > 2^{-p(|x|)}(1 - 4^{-p(|x|)}) > 4^{-p(|x|)}$$

and

$$x \notin L \Rightarrow \operatorname{Prob}[Q_1(x,r,s)] < 4^{-p(|x|)}.$$

Using the Remark we get $L \in \mathbf{PP}$.

Theorem 2. *There is an oracle A such that*

$$\mathbf{AM}^A \cap \textbf{co-AM}^A \not\subset \mathbf{PP}^A.$$

Proof. For simplicity of notation we'll construct an oracle A such that $\mathbf{AM}^A \not\subset \mathbf{PP}^A$. The proof can be easily transformed into the proof of the theorem.

Let A be a language and let $n \in \mathbf{N}$. We will consider the value of A on the words of length $2n$ as the matrix of order $2^n \times 2^n$ with coefficients 0, 1. Denote this matrix by A_n. Call A_n 1-goof iff $> 2/3$ rows of A_n contain at least one 1 and call A_n 0-good iff $< 1/3$ rows of A_n contain at least one 1. Call A_n good if it is 1-good or 0-good. Associate with any oracle A the language $L(A) = \{1^n \mid A_n \text{ is 1-good}\}$. We'll construct an oracle A such that A_n is good for all $n \in \mathbf{N}$ and $L(A) \notin \mathbf{PP}^A$. From the former condition we can easy deduce that $L(A) \in \mathbf{AM}^A$.

To ensure $L(A) \notin \mathbf{PP}^A$ let us enumerate all polynomial probabilistic machines and denote i-th machine by PP_i. Define for beginning A in such a way that A_n is good for all $n \in \mathbf{N}$. We will perform steps with numbers 0, 1, 2,.... On the i-th step we'll ensure that $L(A)$ differs from the language recognized by PP_i^A. To this end we will change the value of A on finite number of words in such a way that for some $n \in \mathbf{N}$ holds

$$1^n \in L(A) \not\Leftrightarrow \text{Prob}[PP_i^A(1^n, r) = 1] > 1/2 \tag{1}$$

After changing we will fix the value of A on all words which the truth value of (1) depends on. This means that on later steps we will not change the value of A on these words.

Let us describe i-th step. Choose n such that no value of A (the oracle constructed on $(i-1)$th step) on words with length $2n$ is fixed and sufficiently large (how large must be n we'll see in the end of the proof). Denote by M_n the set of all 0–1-matrices of order $2^n \times 2^n$. If $B \in M_n$ then denote by $A[B]$ the oracle obtained from A by replacing A_n with B. Let us prove that there is a good $B \in M_n$ such that for $A[B]$ holds (1). Suppose the contrary: for all good $B \in M_n$

$$B \text{ is 1-good } \Leftrightarrow \text{Prob}[PP_i^{A[B]}(1^n, r) = 1] > 1/2 \tag{2}$$

We'll deduce a contradiction. Denote for brevity $PP_i^{A[B]}(1^n, r)$ by $P(B, r)$.

We'll construct two probability distributions μ, ν on M_n such that the matrix B taken at random with respect to μ with high probability is 1-good and the matrix B taken at random with respect to ν with high probability is 0-good. More precisely μ and ν will satisfy

$$\text{Prob}_\mu[B \text{ is 1-good}] > 1 - 2^{-h(n)} \tag{3}$$

$$\text{Prob}_\nu[B \text{ is 0-good}] > 1 - 2^{-h(n)} \tag{4}$$

where $h(n)$ grows superpolynomially.

Let us denote by $\mathbf{E}_\mu$, $\mathbf{E}_\nu$, and $\mathbf{E}$ correspondingly the average with respect to distributions μ, ν and uniform distribution in the set of r's. Let us prove that (3) and (4) yield

$$\mathbf{E}_\mu \mathbf{E} P(B, r) > \mathbf{E}_\nu \mathbf{E} P(B, r) \tag{5}$$

Indeed if μ will be concentrated only on 1-good matrices, we would have $\mathbf{E}_\mu \mathbf{E} P(B, r) > \mathbf{E}_\mu \frac{1}{2} = \frac{1}{2}$ and more precisely $\mathbf{E}_\mu \mathbf{E} P(B, r) \geq \frac{1}{2} + 2^{-poly(n)}$. If ν will be concentrated only on 0-good matrices we would have $\mathbf{E}_\nu \mathbf{E} P(B, r) < \frac{1}{2}$. As the gap $2^{-poly(n)}$ is less than $2^{-h(n)}$ we see that for sufficiently large n, (5) follows from (3) and (4). On the other hand, μ and ν will be such that

$$\mathbf{E}_\mu P(B, r) = \mathbf{E}_\nu P(B, r) \text{ for all } r. \tag{6}$$

Evidently this contradicts (5).

Distribution μ and ν will be constructed as follows. Let σ be some probability distribution in the segment $[0,1]$. Let us associate with σ a probability distribution in M_n denoted by $\alpha(\sigma)$. A random matrix B with respect to $\alpha(\sigma)$ is generated as follows. Pick independently random $p_1,\dots,p_{2^n}$ in $[0,1]$ with respect to σ. Then for each $j \le 2^n$ take j-th row of B as the sequence of 2^n Bernoulli tests with probability of 1 equal to p_j . More formally, for any matrix $(c_{jl}) \in M_n$

$$\mathrm{Prob}_{\alpha(\sigma)}[B=(c_{jl})] = \prod_{j=1}^{2^n} (\int_0^1 (\prod_{l=1}^{2^n} x_j^{c_{jl}})\, d\sigma(x_j))$$

where $x^0 = 1-x$, $x^1 = x$.

Let us denote by k the maximal number of queries which machine PP_i makes to oracle on inputs of the form $1^n, r$. Note that $k \le \mathit{poly}(n)$. Consider the first k moments of σ:

$$m_1(\sigma) = \int_0^1 x\, d\sigma(x),\ m_2(\sigma) = \int_0^1 x^2\, d\sigma(x), \dots,$$

$$m_k(\sigma) = \int_0^1 x^k\, d\sigma(x).$$

We claim that for all r, $\mathbf{E}_{\alpha(\sigma)} P(B,r)$ is a polynomial in $m_1(\sigma)$, $m_2(\sigma)$, $\dots, m_k(\sigma)$ (the coefficients of this polynomial depend only on PP_i, n and r).

Let us prove this claim. Remember that $P(B,r) = PP_i^{A[B]}(1^n,r)$. Let us fix some $B \in M_n$ and simulate the work of PP_i on $1^n, r$ with oracle $A[B]$. Let us write down the questions to B (i.e. the questions of length $2n$ to the oracle) made during this work and also the answers. Denote by $u_1,\dots,u_k$ the questions and by $b_1,\dots,b_k$ the answers (thus $u_1,\dots,u_k \in \mathbf{B}^{2n}$, $b_1,\dots,b_k \in \mathbf{B}$). Let us call the sequence $(u_1,\dots,u_k,b_1,\dots,b_k)$ the *protocol* on B and denote it by $Prot(B)$. Denote also

$$Prot = \{Prot(B) \mid B \in M_n,\ P(B,r)=1\}.$$

Obviously

$$Prot(B_1) = Prot(B_2) \Rightarrow P(B_1,r) = P(B_2,r).$$

Therefore we have

$$\mathbf{E}_{\alpha(\sigma)} P(B,r) = \sum_{v \in Prot} \mathrm{Prob}_{\alpha(\sigma)}[Prot(B) = v].$$

Thus it suffices to prove that for every $v \in Prot$, $\mathrm{Prob}_{\alpha(\sigma)}[Prot(B) = v]$ is a polynomial in $m_1(\sigma)$, $m_2(\sigma)$, $\dots, m_k(\sigma)$. Let us fix some $v \in Prot$, $v = (u_1, \dots, u_k, b_1, \dots, b_k)$. Evidently

$$\mathrm{Prob}_{\alpha(\sigma)}[Prot(B) = v] =$$

$$= \mathrm{Prob}_{\alpha(\sigma)}[B(u_i) = b_i, i = 1, \dots, k].$$

Remember that each u_i is considered to be a pair of numbers of a row and a column in B, denote the number of the row by l_i. Denote for each $s \le 2^n$ by c_s the number of different u_i, $i = 1, \dots, k$ such that $l_i = s_i$ and $b_i = 1$ and by d_s the number of different u_i, $i = 1, \dots, k$ such that $l_i = s_i$ and $b_i = 0$. Then

$$\mathrm{Prob}_{\alpha(\sigma)}[Prot(B) = v] = \prod_{s=1}^{2^n} \int_0^1 x^{c_s}(1-x)^{d_s}\, d\sigma(x).$$

Evidently, $\int_0^1 x^{c_s}(1-x)^{d_s}\, d\sigma(x)$ is a linear combination of $m_r(\sigma) = \int_0^1 x^r\, d\sigma(x)$, $r = 0, 1, \dots, k$ (because $c_s + d_s \le k$). Thus the claim is proved.

Therefore if we take two probability distributions σ and τ in $[0,1]$ such that

$$m_i(\sigma) = m_i(\tau) \text{ for } i = 1, 2, \dots, k, \tag{7}$$

and take $\mu = \alpha(\sigma)$ and $\nu = \alpha(\tau)$ we will get (6). In order to satisfy (3) we'll take σ such that

$$\mathrm{Prob}_\sigma[p \ge 2^{-n+4}] \ge \frac{4}{5} \tag{8}$$

Let us prove that (8) implies (3). Suppose that σ satisfies (8). Let B be a random matrix with respect to $\alpha(\sigma)$. Denote by q the probability that a fixed row of B has only zeros. Obviously, $q \le \frac{1}{5} + \frac{4}{5}(1 - 2^{-n+4})^{2^n} \approx \frac{1}{5} + \frac{4}{5}e^{-16} < \frac{1}{4}$ (for large n). From the Law of large numbers it follows that with probability $1 - 2^{-const \cdot 2^n}$ the frequence of nonzero rows in B is greater than $2/3$. In order to satisfy (4) we will take $\nu = \alpha(\tau)$ such that

$$\mathrm{Prob}_\tau[p = 0] \ge \frac{3}{4}. \tag{9}$$

Let us prove that (9) implies (4). Let τ satisfy (9) and let B be a random matrix with respect to $\alpha(\tau)$. Then the probability q that a fixed row of B has only zeros is greater than $3/4$. Therefore we can reason as in above case. Thus it remains to prove the following lemma.

Lemma. *Let $p(n)$ be a polynomial. Then for all sufficiently large n there are probability distributions σ and τ in $[0,1]$ satisfying the conditions (8), (9) and (7) for $k = p(n)$.*

Proof. We'll define σ explicitly and τ implicitly by using a criterion on the existence of a measure in $[0,1]$ with given moments (a measure differs from a probability distribution with that a measure of entire segment $[0,1]$ can be different from 1; thus the probability distribution can be defined as any measure μ such that $\mu([0,1]) = \int_0^1 1 \cdot d\mu(x) = 1$).

Let $m = (m_0, m_1, \dots, m_k)$ be a sequence of real numbers. Let $a(x) = \sum_{i=0}^k a_i x^i$ be a polynomial of degree $\le k$. Define $(m, a(x))$ to be equal to $m_0 a_0 + m_1 a_1 + \cdots + m_k a_k$. The following theorem is due to M.Riesz. In paper [5] this theorem is proved for the infinite sequences of moments and measures in the set of reals. Riesz' proof is good also in our case. See also [6] (and [7] in Russian).

Riesz' theorem. *Two following conditions are equivalent:*

(i) *There is a measure μ in $[0,1]$ such that for all $i \in \{0, 1, \dots, k\}$, $\int_0^1 x^i \, d\mu(x) = m_i$.*

(ii) *For all polynomials $a(x)$ of degree $\le k$, nonnegative on $[0,1]$, it holds $(m, a(x)) \ge 0$.*

If k is even then (ii) *is equivalent to the condition*

(iii) *For all polynomials $b(x)$, $c(x)$, $\deg b(x) \le k/2$, $\deg c(x) \le k/2 - 1$, it holds $(m, b(x)^2) \ge 0$, $(m, x(1-x)c(x)^2) \ge 0$.*

The implication (ii)⇒(iii) is obvious. The implication (i)⇒(ii) is simple and we'll use its proof in the sequel. Let us prove it. Assume that (i) is true and let μ be a measure satisfying (i). Let $a(x) = \sum_{i=0}^k a_i x^i$ be a polynomial nonnegative on $[0,1]$. Then

$$(m, a(x)) = \sum_{i=0}^{k} a_i m_i = \sum_{i=0}^{k} a_i \int_0^1 x^i \, d\mu(x)$$

$$= \int_0^1 a(x) \, d\mu(x) \ge 0.$$

For the seek of completeness we'll also prove that (iii)⇒(ii) and (ii)⇒(i) in the Appendix.

Our plan is as follows. We'll define a probability distribution σ on $[0,1]$ such that

$$\int_0^1 b(x)^2\, d\sigma(x) \geq \frac{3}{4} b(0)^2 \text{ for all polynomials} \tag{10}$$
$$b(x) \text{ with degree } \leq k/2.$$

Then we'll define the sequence $m = (m_0, \ldots, m_k)$ by equalities $m_0 = 1/4$, $m_1 = m_1(\sigma)$, …, $m_k = m_k(\sigma)$.

This sequence m satisfies (iii) because if $b(x)$ has degree $\leq \frac{k}{2}$ then $(m, b(x)^2) = \int_0^1 b(x)^2\, d\sigma(x) - \frac{3}{4} b(0)^2 \geq 0$ (because $m_0 = \frac{1}{4} = m_0(\sigma) - \frac{3}{4}$) and if $c(x)$ has degree $\leq \frac{k}{2} - 1$ then

$$(m, x(1-x)c(x)^2) = \int_0^1 x(1-x)c(x)^2\, d\sigma(x) \geq 0$$

(because the polynomial $x(1-x)c(x)^2$ has no constant term).

By Riesz' theorem there is a measure μ in $[0,1]$ such that $m_0(\mu) = \int_0^1 1\, d\mu(x) = \frac{1}{4}$ and for all $i \in \{1, 2, \ldots, k\}$, $m_i(\mu) = \int_0^1 x^i\, d\mu(x) = \int_0^1 x^i\, d\sigma(x) = m_i(\sigma)$. Let $g_\mu(x)$ be distribution function of μ, i.e. $g_\mu(x) = \mu([0,x])$. Consider the function $f(x) = \frac{3}{4} + g_\mu(x)$. Then $f(x)$ is the distribution function of some probability distribution τ in $[0,1]$.

Evidently τ satisfies the required conditions.

Thus it remains to construct σ such that (8) and (10) hold. Let us denote for brevity 2^{-n+4} by θ. Let us define σ by equalities $\mathrm{Prob}_\sigma[p = \theta] = 4/5$ and if $A \subset [0,1] \backslash \{\theta\}$ then $\mathrm{Prob}_\sigma[p \in A] = \frac{1}{5} \int_{x \in A} \rho(x)\, dx$ where $\rho(x) = \frac{c_1}{\sqrt{1-(1-2x)^2}}$ and c_1 is chosen in such a way that $\int_0^1 \rho(x)\, dx = 1$. In other words, σ is the probability distribution such that for all $A \subset [0,1]$ $\mathrm{Prob}_\sigma[p \in A] = \frac{4}{5} \chi_A(\theta) + \frac{1}{5} \int_0^1 \chi_A(x) \rho(x)\, dx$, where χ_A stands for the characteristic function of A.

Evidently (8) is true.

Let us prove (10). Let $b(x)$ be a polynomial of degree $\leq k/2$. Then $\int_0^1 b(x)^2\, d\sigma(x) = \frac{4}{5} b(\theta)^2 + \frac{1}{5} \int_0^1 b(x)^2 \rho(x)\, dx$. We claim that either $\frac{4}{5} b(\theta)^2$ or $\frac{1}{5} \int_0^1 b(x)^2 \rho(x)\, dx$ is greater than $\frac{3}{4} b(0)^2$ (for sufficiently large n). Indeed, assume that $\frac{1}{5} \int_0^1 b(x)^2 \rho(x)\, dx \leq \frac{3}{4} b(0)^2$ that is

$$\int_0^1 \left(\frac{b(x)}{b(0)} \right)^2 \rho(x)\, dx \leq \frac{15}{4}$$

Let us prove that $\frac{4}{5} b(\theta)^2 \geq \frac{3}{4} b(0)^2$. In fact, we'll prove that $\frac{b(\theta)}{b(0)}$ is exponentially close to 1. Let us substitute $1 - 2x = y$ for convenience. Then we have

$$\int_{-1}^{1} d(y)^2\gamma(y)\,dy \le \frac{15}{2}, \tag{11}$$

where $d(y) = \frac{b(\frac{1-y}{2})}{b(0)}$, $\gamma(y) = \frac{c_1}{\sqrt{1-y^2}}$. Thus we have to prove that $d(1-2\theta)$ is close to $d(1) = 1$. Let $m = k/2$ and $T_0(x)$, $T_1(x)$, $\dots, T_m(x)$ be $(m+1)$ first Chebyshev's polynomials, i.e. $T_i(\cos t) = \cos it$ for all $t \in [0, \pi]$. The density $\rho(x)$ is chosen in such a way that $T_i(y)$ are orthogonal with density $\gamma(y)$, moreover they are almost orthonormal: $\int_{-1}^{1} T_i(y)T_j(y)\gamma(y)\,dy$ is equal to 0 if $i \ne j$, is equal to c_2 if $i = j \ne 0$ and is equal to c_3 if $i = j = 0$ where c_2, c_3 are some positive constants. It is well known that the polynomials T_0, T_1, $\dots, T_m$ form a basis in the space of all polynomials of degree $\le m$. Let $d_0, \dots, d_m$ be the coefficients of the polynomial $d(y)$ in this basis. Then (11) yields that $\int_{-1}^{1}(\sum_{i=0}^{m} d_iT_i(y))^2\gamma(y)\,dy = \sum_{i=0}^{m} d_i^2 \int_{-1}^{1} T_i(y)^2\gamma(y)\,dy \le \frac{15}{2}$. Hence for some constant c_4 we have $|d_0|, \dots, |d_m| \le c_4$. Let us deduce from this that $|d(1-2\theta) - d(1)|$ is small. Let us denote 2θ by ν. We have $|d(1-\nu) - d(1)| \le \sum_{i=0}^{m} |d_i||T_i(1-\nu) - T_i(1)|$.

We claim that $|T_i(1-\nu)-T_i(1)| = i^2\nu(1+o(1))$ as $i^2\nu \to 0$. Suppose that this is already proved. Then for sufficiently large n we have $|d(1-\nu)-d(1)| \le (m+1) \cdot c_4 \cdot m^2 \cdot \nu \cdot 2$. Therefore

$$|d(1-\nu) - d(1)| \le poly(n)2^{-n+4}.$$

Hence

$$\frac{b(\theta)}{b(0)} = d(1-\nu) \ge 1 - poly(n)2^{-n+4} \ge \frac{99}{100}$$

for sufficiently large n.

Thus it remains to prove that $T_i(1-\nu)-T_i(1) = i^2\nu(1+o(1))$ as $i^2\nu \to 0$.

Let $\alpha \in [0, \pi]$ be defined by equality $\cos\alpha = 1-\nu$. Then $\nu = \frac{\alpha^2}{2}(1+o(1))$ as $\nu \to 0$. Hence

$$T_i(1-\nu) = T_i(\cos\alpha) = \cos i\alpha = 1 - \frac{\alpha^2 i^2}{2}(1+o(1)) =$$

$$1 - i^2\nu(1+o(1)) = T_i(1) - i^2\nu(1+o(1)).$$

This completes the proof of the lemma.

4 Appendix

1. Proof of implication (ii)⇒(i) in Riesz' theorem.

Assume that (ii) is true. Let us enumerate $q_1, q_2, \ldots$ all rational numbers from $[0,1]$ and define $r_i(x)$ to be a function on $[0,1]$ such that $r_i(x)$ is equal to 0 if $0 \le x \le q_i$ and equal to 1 if $q_i < x \le 1$. Consider the linear space L over $\mathbf{R}$ consisting of all functions $f(x)$ on $[0,1]$ of the form

$$f(x) = \sum_{i \in I} s_i r_i(x) + a(x) \tag{12}$$

where I is a finite set of natural numbers, $s_i \in \mathbf{R}$ and $a(x)$ is a polynomial with degree $\le k$. Let K be the set of all $f \in L$ such that f is nonnegative on $[0,1]$.

Claim. *There is a linear functional l defined on L such that l is nonnegative on K and $l(a(x)) = (m, a(x))$ for all polynomials $a(x)$ of degree $\le k$.*

Proof of the claim. Let us define L_i to be the set of all functions $f(x)$ of the form (12) with $I = \{1, 2, \ldots, i\}$ and define l_0 to be the functional defined on L_0 (the set of all polynomials of degree $\le k$) as $l_0(a(x)) = (m, a(x))$. Then (ii) means that l_0 is nonnegative on $K \cap L_0$. Using the induction we'll prove that there is a sequence $l_0, l_1, l_2 \ldots$ of linear functionals such that l_i is defined on L_i, is nonnegative on $L_i \cap K$ and extends l_{i-1}. Then as l we can take the union of all l_i, $i \in \mathbf{N}$.

Let the functional l_i be already defined and nonnegative on $L_i \cap K$. Obviously we have to define the value of l_{i+1} only on $r_{i+1}(x)$. Suppose that this value is equal to v. One can easily verify that in this case l_{i+1} is nonnegative on $L_{i+1} \cap K$ iff v satisfies two conditions

(a) $v \le l_i(f(x))$ for all $f(x) \in L_i$ such that

$$r_{i+1}(x) \le f(x) \text{ for all } x \in [0,1],$$

(b) $l_i(g(x)) \le v$ for all $g(x) \in L_i$ such that

$$r_{i+1}(x) \ge g(x) \text{ for all } x \in [0,1].$$

Let us prove that there is $v \in \mathbf{R}$ satisfying (a) and (b). Let us denote

$$A = \{l_i(f(x)) \mid f(x) \in L_i,\ \forall x \in [0,1]\ r_{i+1}(x) \le f(x)\}$$

$$B = \{l_i(g(x)) \mid g(x) \in L_i,\ \forall x \in [0,1]\ g(x) \le r_{i+1}(x)\}$$

Evidently it is sufficient to prove that $A \neq \emptyset$, $B \neq \emptyset$ and $\forall v_1 \in A\ \forall v_2 \in B$ $v_1 \geq v_2$.

As $r_{i+1}(x)$ is bounded and L_0 contains all constant functions, we have $A \neq \emptyset$, $B \neq \emptyset$. If $v_1 \in A$, $v_1 = l_i(f(x))$ and $v_2 \in B$, $v_2 = l_i(g(x))$, then $(f(x) - g(x)) \in K$ therefore $v_1 = l_i(f(x)) \geq l_i(g(x)) = v_2$.

The claim is proved.

Now let us consider the function g defined on $[0,1]$ by $g(x) = \sup\{l(r_i) \mid q_i \leq x, i \in \mathbf{N}\}$. We can easily prove that $g(x)$ is monotone and continuous from the right ($\lim_{y\to x+0} g(y) = g(x)$). Hence g is the distribution function for some measure μ in $[0,1]$, i.e. there is a measure μ in $[0,1]$ such that $\mu([0,x]) = g(x)$ for all $x \in [0,1]$. Obviously for all i, $\int_0^1 r_i(x)\, d\mu(x) = l(r_i(x))$. From this and the nonnegativeness of l on K we can easily deduce that $\int_0^1 x^i\, d\mu(x) = l(x^i) = m_i$ for all $i \in \{0, 1, \ldots, k\}$.

2. Proof of implication (iii)$\Rightarrow$(ii) in Riesz' theorem. This implication easily follows from the fact that for even k every polynomial of degree $\leq k$ which is nonnegative on $[0,1]$ has the form $a(x)^2 + x(1-x)b(x)^2$ where $\deg a(x) \leq k/2$, $\deg b(x) \leq k/2 - 1$. The latter fact in turn follows from the fact that each polynomial nonnegative on the set $\{y \in \mathbf{R} \mid y \geq 0\}$ has the form $p(y)^2 + yq(y)^2$. Indeed, suppose that the latter assertion is true and $c(x)$ is a polynomial with degree $\leq k$ nonnegative on $[0,1]$. Then the polynomial $c(\frac{y}{1+y})(1+y)^k$ is nonnegative on $[0,+\infty[$ therefore for some polynomials $p(y)$ and $q(y)$ it holds $c(\frac{y}{1+y})(1+y)^k = p(y)^2 + yq(y)^2$. Evidently $\deg p \leq k/2$, $\deg q \leq k/2 - 1$. Substituting $y = \frac{x}{1-x}$ we get $c(x) = p(\frac{x}{1-x})^2(1-x)^k + x(1-x)^{k-1}q(\frac{x}{1-x})^2$. Evidently

$$a(x) = p(\frac{x}{1-x})(1-x)^{k/2} \text{ and}$$
$$b(x) = q(\frac{x}{1-x})(1-x)^{k/2-1}$$

are polynomials of degrees correspondingly $\leq k/2$ and $\leq k/2 - 1$. Thus it remains to prove that every polynomial $r(y)$ which is nonnegative on $[0,+\infty[$ has the form $r(y) = p(y)^2 + yq(y)^2$. Let us define P to be the set of all polynomials having such form.

Let $r(y)$ is nonnegative on $[0,+\infty[$. Obviously, it is sufficient to prove two assertions: (a) $r(y)$ can be represented as the product of polynomials from P and (b) if $r_1(y) \in P$ and $r_2(y) \in P$ then $r_1(y) \cdot r_2(y) \in P$. Let us decompose $r(y)$ into the product of polynomials irreducible over $\mathbf{R}$

$$r(y) = A \cdot (y + a_1)^{i_1} \cdots (y + a_n)^{i_n} \times$$

$$\times (y^2 + b_1 y + c_1)^{j_1} \cdots (y^2 + b_m y + c_m)^{j_m}$$

Evidently $A > 0$. Let us take arbitrary $k \leq n$. Then $a_k \geq 0$ or i_k is even. If $a_k \geq 0$ then $y + a_k \in P$ as $y + a_k = (\sqrt{a_k})^2 + y \cdot 1^2$. If i_k is even then $(y + a_k)^{i_k} \in P$. Let us take arbitrary $k \leq m$. Obviously $c_k > 0$ because $s(y) = y^2 + b_k y + c_k$ is irreducible. We have $s(y) = (y - \sqrt{c_k})^2 + y(2\sqrt{c_k} + b_k)$. Since $s(y)$ is irreducible, we have $s(\sqrt{c_k}) = \sqrt{c_k}(2\sqrt{c_k} + b_k) \geq 0$ therefore $s(y) \in P$.

The assertion (b) follows from the equality

$$(p(y)^2 + yq(y)^2)(s(y)^2 + yt(y)^2) =$$
$$= (p(y)s(y) - yq(y)t(y))^2 + y(p(y)t(y) + q(y)s(y))^2.$$

Acknowledgments

The author is sincerely grateful to Vladimir Borisenko, Frederic Green, Lane Hemachandra, Andrey Muchnik, Alexander Razborov, Alexander Shen and Yuri Nikolaevich Tyurin.

References

[1] S.Toda, "On the computational power of **PP** and ♯**P**". Proc. of 30th Symp. on Found. of Comp. Sci. (1989), pp.514–519.

[2] R.Beigel, N.Reingold and D.Spielman, "**PP** is closed under intersection". Proc. of 23rd ACM Symp. on Th. of Comp. (1991), pp.1-9.

[3] L.Fortnow, N.Reingold. "**PP** is closed under truth table reductions." 6th IEEE Conf. on Structure in Complexity Theory, 1991, pp.13-15.

[4] L.Babai, "Trading group theory for randomness". Proc. 17th ACM Symp. on Theory of Comp. (1985), pp.421-429.

[5] M.Riesz. Sur le probléme des moments. Troisieme Note. Arkiv för mat., astr. och fys., 1923, v.17.

[6] R.Riesz et B.Sz.-Nagy. Lecons d'analyse functionelle. Akademiai Kiado, Budapest 1972, 6th ed. (There are English and Russian translations.)

[7] A.I.Akhiezer. The classical problem of moments and some related topics in calculus. Moscow, Fizmatgiz, 1961. (Russian.)