Formalization of Prime Representing Polynomial in Mizar

Karol Pąk

Abstract

The aim of our work is to show, using the Mizar system that our techniques invented to formalize the unsolvability of Hilbert's tenth problem in a Matiyasevich way, can be reused to prove that an assumption used by Julia Robinson demonstrates the same result independently.

We present our formalization that the set of prime numbers is representable by a polynomial formula.

Keywords

Prime number, Diophantine, Representing Polynomial

1. Introduction

Martin Davis, YuriMatiyasevich, Hilary Putnam and Julia Robinson [DPR61, Mat70] have proven that every recursively enumerable set is diophantine, and hence prove the Hilbert's Tenth Problem in the negative: *there is no algorithmic way of determining whether some arbitrary diophantine equation has a solution.* This is known as the MRDP-theorem (due to Matiyasevich, Robinson, Davis, and Putnam). This problem took seventy years to resolve, during which many attempts have been made to solve the problem. It is therefore not surprising that Julia Robinson and Martin Davis, with a contribution from Hilary Putnam, created several theorems that give a negative solution to the problem but under some assumptions. One of these assumptions that *the exponential function can be defined in a diophantine way* has been eliminated by Yuri Matiyasevich using a trick with clever use of Fibonacci numbers, who definitively completed the proof of the MRDP-theorem.

In our work, we focus on another theorem under some, currently eliminated assumption, proposed by Julia Robinson [Rob69]. She proved that *if the set of prime numbers was diophantine, then every recursively enumerable set would be diophantine.* We do this for two main reasons. First, the set of prime numbers can be representable by a complicated polynomial formula (proposed in [JSWW76]) and consequently, the set is diophantine. We can investigate the possibilities of the Mizar system [? GKN15] to prove that explicitly given polynomial with 26 variables determine the set of prime numbers. We also use a trick with Mizar schemes (see [GKN10]) that go beyond first-order logic to show a sophisticated proof of the existence of such a polynomial without formulating it explicitly. Second, the proof of the assumption requires

Fifth Workshop on Formal Mathematics for Mathematicians, July 30–31, 2021, Timisoara, Romania

karol@mizar.org (K. Pąk)
http://alioth.uwb.edu.pl/~pakkarol/ (K. Pąk)
0000-0002-7099-1669 (K. Pąk)

nearly all the techniques invented to prove the MRDP-theorem that we formalized in the Mizar system [Pąk19b] and seems a natural continuation of the formal approach to diophantine sets.

2. Diophantine sets

Obviously, we need to begin by quickly explaining what we mean by *diophantine*. A diophantine polynomial in the k variables $x_1, x_2, \ldots, x_k$ is defined in informal mathematical practice as finite sum of expressions of the type $c_i v_1 v_2 v_3 \ldots v_j$ where the coefficients c_i are integers (positive or negative) and v_i are variables. A diophantine equation, in traditional form is an equation of the form $P(x_1, \ldots, x_j, y_1, \ldots, y_k) = 0$, where P is a diophantine polynomial, $x_1, \ldots, x_j, y_1, ..., y_k$ indicate parameters and unknowns, respectively. A set $D \subseteq \mathbb{N}^n$ of n-tuples is called diophantine if there exists a $n + k$–variable diophantine polynomial P such that $\langle x_1, \ldots, x_n \rangle \in D$ if and only if there exist variables $y_1, \ldots, y_k \in \mathbb{N}$ such that $P(x_1, \ldots, x_j, y_1, \ldots, y_k) = 0$.

In the context of the MRDP-theorem, we repeatedly refer to concepts of diophantine polynomials, equations, and sets, but in a slightly inverted way. Instead of being given an equation and seeking its solutions, we will give a set of solutions and seek a corresponding diophantine equation. In particular, the set of numbers which are either even or multiples of three is diophantine, since $(x - 2y)(x - 3y)$ takes a zero for each x, which is either even or a multiple of three. Similarly, the set of numbers which are even and multiples of three is diophantine, since $(x - 2y)^2 + (x - 3z)^2$ or simply $x - 6y$ takes a zero for such x. It is easy to see that in the general case a diophantine polynomial is not determined uniquely by a given diophantine set. So we might ask what is the smallest possible degree and/or what is the smallest possible number of parameters in a diophantine polynomial to determine a given diophantine set. In our simple example the question is straightforward, but it is not so for the set of prime numbers. In 1971, Yuri Matiyasevich give the construction of a diophantine polynomial with 24 variables and degree 37 that determines the set of prime numbers. Using the Skolem substitution method [Dav73] we can reduce the degree to 5. However, this procedure increases the number of variables. Currently, the smallest known number of variables to represent primes is 12 and is proposed by Yuri Matiyasevich and Julia Robinson in [MR75], but the degree of the polynomial is more than a few thousand (more than 6,000 from our estimate).

In our Mizar formalization, we chose a diophantine polynomial with 26 variables to represent primes that is given in [JSWW76]. We show that for any positive integer k so that $k + 1$ is prime it is necessary and sufficient that there exist other natural variables a-z for which the polynomial

$$\begin{gathered}
[wz + h + j - q]^2 + [(gk + g + k)(h + j) + h - z]^2 + [(2k)^3(2k + 2)(n + 1)^2 + 1 - f^2]^2 + \\
[p + q + z + 2n - e]^2 + [e^3(e + 2)(a + 1)^2 + 1 - o^2]^2 + [x^2 - (a^2 - 1)y^2 - 1]^2 + \\
[16(a^2 - 1)r^2y^2y^2 + 1 - u^2)^2 + [((a + u^2(u^2 - a))^2 - 1)(n + 4dy)^2 + 1 - (x + cu)^2]^2 + \\
[m^2 - (a^2 - 1)l^2 - 1]^2 + [k + i(a - 1) - l]^2 + [n + l + v - y]^2 + \\
[p + l(a - n - 1) + b(2a(n + 1) - (n + 1)^2 - 1) - m]^2 + \\
[q + y(a - p - 1) + s(2a(p + 1) - (p + 1)^2 - 1) - x]^2 + [z + pl(a - p) + t(2ap - p^2 - 1) - pm]^2
\end{gathered} \tag{1}$$

equals zero.

3. Prime representing polynomial

The proof that (1) determine the set of prime numbers is based on two concepts: the special case of Pell's Equation that has the form $x^2 - (a^2 - 1)y^2 = 1$, where $a > 1$ and Wilson's theorem which characterizes the primes in terms of the factorial function, i.e., for any positive integers k holds $k + 1$ is prime if and only if $k + 1|k! + 1$. Note that we had to prove that the equality $y = x!$ is diophantine since it is one of the key steps to proving the MRDP-theorem and the theorem is already formulated in [Pąk19a] as follows:

```
theorem :: HILB10_4:31
  for i1,i2 be Element of n st n<>0 holds
      {p where p is n-element XFinSequence of NAT: p.i1 = p.i2!}
    is diophantine Subset of n-xtuples_of NAT;
```

Note that *i* **be** Element **of** *n* means in particular that *i* is in the domain of a zero-based indices, *n*-tuples *p*, since *n*={0,1,2,. . .,*n*-1} in the standard construction of the natural numbers developed in the Mizar library. Rather than showing full proof that the set of prime numbers is diophantine, we just show a trick with Mizar schemes (second-order theorems) that provide the ability to combine diophantine relation using conjunctions and alternatives as well as a special case to substitution (see [AP18]):

```
scheme :: HILB10_3:sch 4
 Substitution{P[Nat,Nat,natural object,Nat,Nat,Nat],
  F(Nat,Nat,Nat)->natural object}:
    for i1,i2,i3,i4,i5 holds {p: P[p.i1,p.i2,F(p.i3,p.i4,p.i5),p.i3,p.i4,p.i5]}
      is diophantine Subset of n-xtuples_of NAT
provided
  for i1,i2,i3,i4,i5,i6 holds {p: P[p.i1,p.i2,p.i3,p.i4,p.i5,p.i6]}
    is diophantine Subset of n-xtuples_of NAT
and
  for i1,i2,i3,i4 holds {p: F(p.i1,p.i2,p.i3) = p.i4}
     is diophantine Subset of n-xtuples_of NAT;
```

Since the truncated subtraction (difference or zero) represented in Mizar as -′ is diophantine

```
theorem :: HILB10_3:20
  for a,b,c be Integer, i1,i2,i3 be Element of n holds
      {p where p is n-element XFinSequence of NAT: a*p.i1 = b*p.i2-'c}
    is diophantine Subset of n -xtuples_of NAT;
```

we conclude in particular that {*p*: *p.i1* = *p.i2*-′1} is also diophantine. Combining these with HILB10_4:31 and using Substitution we obtain that {*p*: *p.i1* = (*p.i2*-′1)!} is diophantine. This is proved by writing $F = \lambda i_1\, i_2\, i_3.\ i_2 \mathbin{-'} 1$, $P = \lambda i_1\, i_2\, i_3\, i_4\, i_5\, i_6.\ i_4 = i_3!$. Note that most of the arguments of P, F are unused. We have decided on such a solution in order to avoid repeating the *substitution* schemes for individual cases of arity, since such arity of P, F was sufficient to apply all substitutions done in the MRDP-theorem. In the same manner we can see that $\{p\colon\ p.i1 = (p.i2 -' 1)! + 1\}$ is diophantine writing $F = \lambda i_1\, i_2\, i_3.\ (i_2 -' 1)!$, $P = \lambda i_1\, i_2\, i_3\, i_4\, i_5\, i_6.\ i_4 = 1{*}i_3{+}1$ and using the following theorem:

```
theorem :: HILB10_3:15
  for a,b be Integer, i1,i2 be Element of n holds
      {p where p is n-element XFinSequence of NAT: p.i1 = a*p.i2+b}
    is diophantine Subset of n-xtuples_of NAT;
```

Next, using again the Substitution with the fact that the congruence is diophantine and writing $F = \lambda i_1\, i_2\, i_3.\ (i_2 -' 1)!{+}1$, $P = \lambda i_1\, i_2\, i_3\, i_4\, i_5\, i_6.\ 1{*}i_3, 0{*}i_4$ are_congruent_mod $1{*}i_4$, we obtain that $\{p\colon\ (p.i -' 1)! + 1 \bmod p.i = 0\}$ is diophantine.

```
theorem :: HILB10_3:3
  for a,b,c be Integer, i1,i2,i3 be Element of n holds
      {p where p is n-element XFinSequence of NAT:
         a*p.i1,b*p.i2 are_congruent_mod c*p.i3}
    is diophantine Subset of n-xtuples_of NAT;
```

We continue in this fashion with HILB10_3:7 and obtain that $\{p\colon\ p.i > 0\}$ is diophantine.

```
theorem :: HILB10_3:7
  for a,b,c be Integer, i1,i2 be Element of n holds
      {p where p is n-element XFinSequence of NAT: a*p.i1 > b*p.i2+c}
    is diophantine Subset of n-xtuples_of NAT;
```

```
scheme :: HILB10_3:sch 3
 IntersectionDiophantine{n()->Nat,P,Q[XFinSequence]}:
 {p where p is n()-element XFinSequence of NAT: P[p] & Q[p]}
   is diophantine Subset of n()-xtuples_of NAT
provided
  {p where p is  n()-element XFinSequence of NAT: P[p]}
     is diophantine Subset of n()-xtuples_of NAT
and
  {p where p is n()-element XFinSequence of NAT: Q[p]}
     is diophantine Subset of n()-xtuples_of NAT;
```

Finally, using the IntersectionDiophantine scheme we can conclude that the intersection of these sets, that is equal to $\{p\colon\ (p.i -' 1)! + 1 \bmod p.i = 0\ \&\ p.i > 1\}$ is diophantine. Then the proof that the set of prime numbers is diophantine is easy to complete by applying the Wilson's theorem [Pąk21].

```
theorem :: HILB10_6:4
  for i being Element of n holds
      {p where p is n-element XFinSequence of NAT: p.i is prime}
    is diophantine Subset of n-xtuples_of NAT
```

Using such techniques invented to prove the MRDP-theorem in [Pąk19b] we needed less than 100 lines of code to complete the, but the prime representing polynomial is deeply hidden in the proof, e.g., in the constructions used in the schemes. Moreover, we need a more sophisticated list of arithmetical properties than the one used in [Pąk19a] to to reduce the number of variables to 26 which occur in (1).

For this purpose, we formalize additional properties of the special case of Pell's Equation by following the idea presented in [JSWW76] as follows:

```
theorem :: HILB10_6:24
  for a be non trivial Nat for y,n be Nat st 1 <= n holds y = Py(a,n) iff
    ex c,d,r,u,x be Nat st
      [x,y] is Pell s_solution of a^2-'1 & u^2 = 16*(a^2-1)*r^2*y^2*y^2+1 &
      (x+c*u)^2 = ((a+u^2*(u^2-a))^2-1)*(n+4*d*y)^2+1 & n <= y;
```

```
theorem :: HILB10_6:31
  for f,k be positive Nat holds f = k! iff
    ex j,h,n,p,q,w,z be positive Nat st
      q = w*z+h+j & z = f*(h+j)+h & (2*k)|^3*(2*k+2)*(n+1)|^2+1 is square &
      p = (n+1)|^k & q = (p+1)|^n & z = p|^(k+1);
```

where the truncated subtraction, the second power, the nth power are represented as -', ^2, |^n, respectively. Now we are ready to express and prove in the Mizar system that the set of prime numbers is representable by the polynomial formula (1).

```
theorem :: HILB10_6:33
  for k be positive Nat holds
     k+1 is prime iff ex a,b,c,d,e,f,g,h,i,j,l,m,n,o,p,q,r,s,t,u,w,v,x,y,z be Nat st
          0 = (w*z+h+j-q)^2 + ((g*k+g+k)*(h+j)+h-z)^2 +
              ((2*k)|^3*(2*k+2)*(n+1)|^2+1-f^2)^2 + (p+q+z+2*n-e)^2 +
              (e|^3*(e+2)*(a+1)|^2+1-o^2)^2 + (x^2-(a^2-'1)*y^2-1)^2 +
              (16*(a^2-1)*r^2*y^2*y^2+1-u^2)^2 +
              (((a+u^2*(u^2-a))^2-1)*(n+4*d*y)^2+1-(x+c*u)^2)^2 +
              (m^2-(a^2-'1)*l^2-1)^2 + (k+i*(a-1)-l)^2 + (n+l+v-y)^2 +
              (p+l*(a-n-1)+b*(2*a*(n+1)-(n+1)^2-1)-m)^2 +
              (q+y*(a-p-1)+s*(2*a*(p+1)-(p+1)^2-1)-x)^2 +
              (z+p*l*(a-p)+t*(2*a*p-p^2-1)-p*m)^2;
```

4. Conclusions

Our formalization has so far focused on the polynomial proposed in [JSWW76]. We showed formally in the Mizar system that the polynomial determines the set of prime numbers, hence the set is diophantine. Now we are working on reducing the number of variables in the considered polynomial to 12 as has been done by Yuri Matiyasevich and Julia Robinson in [MR75].

References

[AP18] Marcin Acewicz and Karol Pąk. Basic diophantine relations. *Formalized Mathematics*, 26(**2**):175–181, 2018.

[Dav73] Martin Davis. Hilbert's tenth problem is unsolvable. *The American Mathematical Monthly*, 80(3):233–269, 1973.

[DPR61] Martin Davis, Hilary Putnam, and Julia Robinson. The decision problem for exponential diophantine equations. *Annals of Mathematics*, 74:425–436, 1961.

[GKN10] Adam Grabowski, Artur Korniłowicz, and Adam Naumowicz. Mizar in a nutshell. *J. Formalized Reasoning*, 3(2):153–245, 2010.

[GKN15] Adam Grabowski, Artur Korniłowicz, and Adam Naumowicz. Four decades of Mizar. *Journal of Automated Reasoning*, 55(3):191–198, 2015.

[JSWW76] James P. Jones, Daihachiro Sato, Hideo Wada, and Douglas Wiens. Diophantine representation of the set of prime numbers. *The American Mathematical Monthly*, 83(6):449–464, 1976.

[Mat70] Yuri Matiyasevich. Enumerable sets are diophantine. *Doklady Akademii Nauk SSSR (in Russian)*, 191:279–282, 1970.

[MR75] Yuri Matiyasevich and Julia Robinson. Reduction of an arbitrary diophantine equation to one in 13 unknowns. *Acta Arithmetica*, 27:521–553, 1975.

[Pąk19a] Karol Pąk. Diophantine sets. Part II. *Formalized Mathematics*, 27(**2**):197–208, 2019.

[Pąk19b] Karol Pąk. Formalization of the MRDP theorem in the Mizar system. *Formalized Mathematics*, 27(**2**):209–221, 2019.

[Pąk21] Karol Pąk. Prime Representing Polynomial. *Formalized Mathematics*, 29(**4**):221–228, 2021.

[Rob69] Julia Robinson. Diophantine decision problems. *Studies in number theory*, 6:76—116, 1969.