Formalizing a Diophantine Representation of the Set of Prime Numbers

Karol Pąk
University of Białystok, Poland

Cezary Kaliszyk
Universität Innsbruck, Austria

Abstract

The DPRM (Davis-Putnam-Robinson-Matiyasevich) theorem is the main step in the negative resolution of Hilbert's 10th problem. Almost three decades of work on the problem have resulted in several equally surprising results. These include the existence of diophantine equations with a reduced number of variables, as well as the explicit construction of polynomials that represent specific sets, in particular the set of primes. In this work, we formalize these constructions in the Mizar system. We focus on the set of prime numbers and its explicit representation using 10 variables. It is the smallest representation known today. For this, we show that the exponential function is diophantine, together with the same properties for the binomial coefficient and factorial. This formalization is the next step in the research on formal approaches to diophantine sets following the DPRM theorem.

2012 ACM Subject Classification Theory of computation → Interactive proof systems

Keywords and phrases DPRM theorem, Polynomial reduction, prime numbers

Digital Object Identifier 10.4230/LIPIcs.ITP.2022.26

Category Short Paper

Supplementary Material Formalization can be found at:
Software (Formalization): `http://cl-informatik.uibk.ac.at/cek/itp2022/`

Funding ERC starting grant no. 714034 *SMART* and Cost action CA20111 *EuroProofNet*.

Acknowledgements We would like to thank Yuri Matiyasevich for his comments on the previous version of this paper.

1 Introduction

Hilbert's 10th problem (H10) asks whether there exists an algorithm[1] that can determine if a diophantine equation has a solution over the integers. A major step towards the negative resolution of the problem was achieved by the *Davis conjecture*, stating that the notions of diophantine sets and recursively enumerable sets coincide. This is the case since recursively enumerable sets without algorithms for recognizing their elements have already been known. Indeed, by the Davis Normal Form Theorem [3], for every recursively enumerable set $R \subseteq \mathbb{N}^m$ there exist a number n together with a polynomial P over $m+n+2$ variables (n of the variables are parameters and $m+2$ are unknowns) with integer coefficients, such that

$$\forall_{a_1,\ldots,a_n} R(a_1,\ldots,a_n) \iff \exists_x \forall_{y\leq x} \exists_{x_1\leq x,\ldots,x_m\leq x} P(a_1,\ldots,a_n,x,y,x_1,\ldots,x_m) = 0 \quad (1)$$

[1] Today interpreted as an adequate RAM program or equivalently a Turing machine searching for solutions.

Eliminating the single universal quantifies from Equation 1, it becomes the condition defining a diophantine relation. However, the research undertaken by Robinson, Davis, and Putnam to eliminate this quantifier required almost 30 years. It is therefore not surprising that they created several theorems that give a negative solution to the problem but under certain assumptions, which enable such quantifier elimination. One of such assumptions is that the exponential function can be defined in a diophantine way. This has been eliminated by Matiyasevich, who definitively completed the proof of the DPRM-theorem [4, 9], all every recursively enumerable set of natural numbers is diophantine.

Our work formalizes multiple consequences and results that originated from the proof of the DPRM, all concerning diophantine equations. We formalize the fact that the exponential function is diophantine [14], together with the same property for the binomial coefficient and factorial. These allow the proof of DPRM-theorem [12] in a post-Matiyasevich approach proposed by Smorynski [16], where the concept of recursively enumerable is defined using Equation 1.

We also formalize Robinson's [15] conditions for the DPRM-theorem, namely that the set of primes is representable by a diophantine polynomial. This polynomial is sufficient to prove the DPRM-theorem in the Robinson approach. On the other hand, the existence of this polynomial is guaranteed, but not its explicit statement. In fact, stating it explicitly has been a long-standing challenge and the problem of finding the polynomial defining the prime numbers with a minimum number of variables is an open problem in number theory.

In 1971, Yuri Matiyasevich proposed the construction of a diophantine polynomial of degree 37 with 24 variables (one of the variables is a parameter and 23 are unknowns) that defined the set of prime numbers. This result has been improved together with Robinson [11] to a polynomial with 14 variables including 13 unknowns. To do this, they showed that every diophantine polynomial, can be reduced to 13 unknowns. Wada et al. [6] have later reduced the polynomial to 12 variables, however, the rank of the polynomial is 13,697.

The main result of the current work is the formalization of the polynomial over 10 variables that defines the prime numbers together with a large number of formalized results in numeral analysis (263 proved top-level MML theorems totaling 922KB) necessary for this result. This polynomial, proposed by Matiyasevich [10], is today the smallest known[2] polynomial for the open problem. This work improves on the work by the first author, presented at the FMM 2021 workshop [13], where the 26-variable polynomial was defined in Mizar without any further properties.

2 DPRM Formalizations and Their Relation to Number Theory

Larchey-Wendling and Forster [8] formalized the DPRM theorem in Coq and Bayer et al. [1] in Isabelle/HOL. Both formalizations develop register machines, Minsky machines, advanced properties of the Pell equation and prove that exponentiation is diophantine. The first work proved the bounded universal quantification theorem in order to reach the final DPRM theorem, and then additionally showed the undecidability of `H10` and discusses other undecidable problems as future work. The second work proves the DPRM theorem following the approach proposed by Matiyasevich in [7] instead of proving the bounded universal quantification theorem. There, the discussed future work is to extend it to register machines to prove the undecidability of the Halting problem. Carneiro's formalization in Lean [2] uses Pell equations to prove the key lemma of Matiyasevich, stating that exponentiation is diophantine.

[2] Private email exchange with Yuri Matiyasevich, January 2022.

Our work focuses on the applications of `H10` listed by Sun [17] and instead of register machines we focus on the theory of polynomials. By `H10`, $\exists_{x_1,\ldots,x_\nu\in\mathbb{Z}} P(a,x_1,\ldots,x_\nu)=0$ is undecidable for some diophantine polynomial P, $\nu\in\mathbb{N}$. In 1970 Matiyasevich justified $\nu<200$. Further ingenious number-theoretic ideas allowed him to prove together with Robinson [11], $\nu\geq 13$ by developing general diophantine polynomial reduction methods. Observe, that when $\nu\geq 13$, the unknowns range over positive (or non-negative) integers. In 1975, Matiyasevich further announced that $\nu\geq 9$ and Jones gave a complete proof [5], but in both cases (Matiyasevich's announcement and Jones's proof) the range of all unknowns is again limited to the positive (or non-negative) integers. Sun [17] improved this result by modifying the range of the unknowns. He showed the case $\nu\leq 9$ with only one limitation, namely $x_1,\ldots,x_8\in\mathbb{Z}$ and $x_9\in\mathbb{N}$, and thus finally obtained $\nu\leq 11$ where the unknowns range over all integers. For this reason, we focus our work on Sun's result, even if the condition $\nu\leq 9$ seems better than $\nu\leq 11$. This is also our main justification for the work to construct $J_{1+q,\mathbb{C}}$ with an arbitrary q. We use it for $q=3$, $q=7$ as required in [11], and $q=17$ in Sun's number theoretic results [17].

3 Preliminaries

We shortly remind the definition of *diophantine sets*, that will be used in the formalization. A diophantine polynomial in k variables $v_1,v_2,\ldots,v_k$ is a linear combination of monomials with non-zero coefficients of the shape $c\cdot v_1^{p_1}v_2^{p_2}v_3^{p_3}\ldots v_j^{p_j}$, where the coefficients c are integers, the exponents p_i are natural numbers, and v_i are variables. The variables will be separated into parameters and unknowns as follows. A diophantine equation, is an equation of the form $P(x_1,\ldots,x_j,y_1,\ldots,y_k)=0$, where P is a diophantine polynomial and $x_1,\ldots,x_j$, $y_1,\ldots,y_k$ indicate the 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 unknowns $y_1,\ldots,y_k\in\mathbb{N}$ such that $P(x_1,\ldots,x_j,y_1,\ldots,y_k)=0$. Similarly, an n-ary predicate $\mathcal{P}$ is diophantine, iff the set of n-tuples for which the predicate $\mathcal{P}$ is satisfied is diophantine. In particular, the divisibility relation $\mathcal{P}(a,b)\equiv a\mid b$ and congruence $\mathcal{P}(a,b,c)\equiv a\mid b$ are diophantine, as $\mathcal{P}(a,b)\iff\exists_x\, ax-b=0$ and $\mathcal{P}(a,b,c)\iff\exists_x\, a-b-cx=0$.

Note, that the number of variables used in a polynomial defining a diophantine property includes the explicitly stated parameters a,b,c but also the implicitly appearing unknown x. Informally, a function is referred to as diophantine if the relation between its arguments and its results is. In particular, the key lemma of Matiyasevich, stating that the exponential function is diophantine means that the relation $\mathcal{P}(a,b,c)\equiv a=b^c$ is.

4 Pell Equation

Even if Matiyasevich originally showed the key lemma using properties of the Fibonacci sequence, further results and publications in the domain use the Pell equation instead. The Pell equation states $x^2-Dy^2=1$ with a non-square parameter D. If $D=a^2-1$, we can explicitly give all the solutions of this equation via Lucas sequences: $x=\chi_a(n), y=\psi_a(n)$:

$$\begin{aligned}
&\chi_a(0)=1, && \chi_a(1)=a, && \chi_a(n+2)=2a\chi_a(n+1)-\psi_a(n),\\
&\psi_a(0)=0, && \psi_a(1)=1, && \psi_a(n+2)=2a\psi_a(n+1)-\psi_a(n).
\end{aligned}$$

which is the approach used in the HOL-Light, the Lean [2] and Mizar [14] formalizations. Alternatively, one can consider the equation $(ax-y)^2-(a^2-1)x^2=1$. This, transformed as $x^2-bxy+y^2=1$ with $b=2a$ can be used to build the sequence of solutions $\alpha_b(n)$. That

sequence has the interesting property: if $x^2 - bxy + y^2 = 1$ then either $x = \alpha_b(m), y = \alpha_b(m+1)$ or $x = \alpha_b(m+1), y = \alpha_b(m)$. The latter approach is used in the Coq [8] and Isabelle [1] formalizations. The similarity can be analysed by noticing the relation $\alpha_b(n) = \psi_{2a}(n)$. Of course ψ is more general, while α has more properties. However, the condition $\alpha_b(k) \mid \alpha_n(m) \Leftrightarrow k \mid m$ (see Equation (3.23) in [7]) is explicitly stated in the publication describing the Coq formalization [8] and the previous version of the Isabelle formalization [1] while $\psi_a(k) \mid \psi_n(m) \Leftrightarrow k \mid m$ is proved in HOL Light, Lean [2] and in Mizar [14][3].

Irrespective of the considered sequence, all formalizations prove that $a = \alpha_b(c)$ or $a = \psi_b(c)$ can be represented using an (implicit) diophantine relation, stated as a combination of less complicated diophantine relations. Additionally these sub-relation use additional explicit unknowns and may also have implicit ones in these relations. For example Bayer et al. [1] express $3 < b \wedge a = \alpha_b(c)$ using 6 explicit unknowns and 15 relations including, e.g., 4 uses of equivalence $\equiv$. Carneiro [2] uses 5 additional unknowns explicitly and several congruences. Similarly, our previous work used 6 explicitly given additional unknowns [14], reduced to 5 explicit unknowns and a single implicit ones [13] in order to achieve the 26-variable polynomial (3). Here, in order to formalize the best known 10-variable polynomial proposed in [10], we use the representation proposed by Matiyasevich and Robinson [11]: Two explicit unknowns i, j and 3 implicit ones corresponding to the relations $=\square$ ($=\square$ is a one-argument relation, which is true when the argument is a square of a natural number, i.e., $x = \square \Leftrightarrow \exists_{n \in \mathbb{N}} x = n^2$), $\mid$ and $\leq$.

Note that the Theorem 1 also depends on the parameter $e \in \mathbb{N}$, which can be simply eliminated by $e = 0$. However, we use the original formulation to simplify the comparison of conditions in Theorem 1 and Theorem 2, where we substitute $e = L-1$.

▶ **Theorem 1** (`HILB10_8:19`). *Let $A, B, C \in \mathbb{N}$ with $A > 1$, $B > 0$ and $e \in \mathbb{N}$. Then $C = \psi_A(B)$ if and only if there exists $i, j \in \mathbb{N}$ and auxiliary unknowns $D, E, F, G, H, I \in \mathbb{Z}$ such that*

$$DFI =\square \wedge F \mid (H - C) \wedge B \leq C \tag{2}$$

and $D = (A^2-1)C^2+1$, $E = 2(i+1)D(e+1)C^2$, $F = (A^2-1)E^2 + 1$, $G = A+F(F-A)$, $H = B+2jC$, $I = (G^2-1)H^2+1$, where the auxiliary unknowns can be replaced by polynomials over A, B, C, i, j and e.

Therefore, $C = \psi_A(B)$ can be represented as $0 = (DFI-\alpha^2)^2 + (F\beta-H+C)^2(F\beta+H-C)^2 + (B+\gamma-C)^2$, where $\alpha, \beta, \gamma \in \mathbb{N}$ are hidden unknowns.

5 Prime Numbers

The main idea behind the construction of a polynomial representing the prime numbers uses Wilson's theorem, i.e., for any positive integer k, $k+1$ is prime if and only if $k+1 \mid k! + 1$. Note that in Mizar we had to formalize the fact that $y = x!$ is a diophantine relation, since it

[3] See the complete formalization statement of the theorem `Y_DIVIDES` in HOL-Light `https://github.com/jrh13/hol-light/blob/master/Examples/pell.ml`, the theorem `y_dvd_iff` in Lean `https://github.com/leanprover-community/mathlib/blob/master/src/number_theory/pell.lean`, theorems `HILB10_1:34` and `HILB10_1:36` in Mizar `http://mizar.uwb.edu.pl/version/current/html/hilb10_1.html`.

is one of the key steps to proving the DPRM-theorem in our approach. A proof that only focuses on the existence of this polynomial can be expressed in a surprisingly concise way (less than 100 Mizar lines of proof) using higher-order schemes. Compare this with more than 2000 lines required to prove that for any $k \in \mathbb{N}^+$ holds $k+1$ is prime if and only if, there exist $a - z \in \mathbb{N}$ unknowns for which

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

equals zero.

To get closer to the 10 variables, we define the notion of prime numbers following [10]:

▶ **Theorem 2** (`HILB10_8:23`). *Let $k \in \mathbb{N}$. Then k is prime if and only if there exists $f, i, j, m, u \in \mathbb{N}^+$, $r, s, t \in \mathbb{N}$ unknowns and auxiliary unknowns $A - I, L, M, S - W, Q \in \mathbb{Z}$ such that*

$$\begin{aligned}&DFI = \square \wedge (M^2-1)S^2+1 = \square \wedge ((MU)^2-1)T^2+1 = \square \wedge\\&(4f^2-1)(r-mSTU)^2 + 4u^2S^2T^2 < 8fuST(r-mSTU) \wedge\\&FL \mid (H-C)Z + F(f+1)Q + F(k+1)((W^2-1)Su - W^2u^2+1)\end{aligned} \tag{4}$$

and $A = M(U+1)$, $B = W+1$, $C = r+W+1$, $D = (A^2-1)C^2+1$, $E = 2iC^2LD$, $F = (A^2-1)E^2+1$, $G = A+F(F-A)$, $H = B+2(j-1)C$, $I = (G^2-1)H^2+1$, $W = 100fk(k+1)$, $U = 100u^3W^3+1$, $M = 100mUW+1$, $S = (M-1)s+k+1$, $T = (MU-1)t+W-k+1$, $Q = 2MW-W^2-1$, $L = (k+1)Q$.

One can verify, that the simplest polynomial specified by Equation 4, uses 8 unknowns explicitly along with one implicit one for each relation (three occurrences of $=\square$, inequality, and divisibility). Together with k this gives a total of 14 variables. The fact, that this does not require 5 implicit unknowns, was a striking solution proposed in [10]. In the next section, we will reduce the 5 implicit unknowns used for Equation 4 to a single one.

6 The Polynomial Reduced to a Single Unknown Variable

We first notice $a = \square \equiv \exists_{x \in \mathbb{N}}\, 0 = x^2 - a = \prod(x \pm \sqrt{a})$, where the product considers all sign combinations. Of course, the product is a polynomial, but its factors are not. Additionally, a can be negative, so we need to consider $\mathbb{C}$ as the domain and take into account the non-uniqueness of the square root (however, since $a \in \mathbb{Z}$, there is only one root in the first quadrant of the complex plane).

▶ **Theorem 3.** *Suppose $A_1, \ldots, A_q \in \mathbb{Z}$. Then $A_1 = \square, \ldots, A_q = \square$ if and only if*

$$0 = \prod(X \pm \sqrt{A_1} \pm \sqrt{A_2}W \pm \ldots \pm \sqrt{A_q}W^{q-1})$$

for some $X \in \mathbb{Z}$ where $W = 1 + A_1^2 + \ldots + A_q^2$.

Theorem 3 is formulated in [11] and used for $q = 7$, however, the formalization additionally requires a justification that the product over 2^q possible combinations of signs eliminates similar elements giving a linear combination of $\binom{2^{q-1}+q-1}{q-1}$ monomials with non-zero coefficients. For this, we will define a helper polynomial $J_{n,\mathcal{R}}$ in Theorem 4. There, all factors that included square roots will appear in even powers, which will eliminate these roots. This allows using $J_{q+1,\mathbb{C}}(r_0, \sqrt{r_1 W^2}, \ldots, \sqrt{r_{n-1} W^{2q-2}})$ as an appropriate $\mathbb{Z}$-valued polynomial over $q+1$, used in Theorem 3 following Matiyasevich's elegant adaptation in order to ensure the satisfiability of all 5 predicates from *Equation* 4. Our current formalization does not include Theorem 3 in its full generality (ongoing work with most of the needed lemmas complete), as we only use it with $q = 3$ and substitution the constant $W = 2$ instead of the polynomial $W = 1 + r_1^2 + \ldots + r_q^2$.

▶ **Theorem 4** (`POLYNOM9:def 10`). *Let $\mathcal{R}$ be a commutative ring, $n \in \mathbb{N}$ with $n > 1$. There exists an $\mathcal{R}$-valued polynomial over n variables $J_{n,\mathcal{R}}$ obeying the following conditions:*

- *$J_{n,\mathcal{R}}(r_1, r_2, \ldots, r_n) = \prod(r_1 \pm r_2 \pm \ldots \pm r_n)$ for all $r_1, r_2, \ldots, r_n \in \mathcal{R}$,*
- *let $p_\alpha R^\alpha$ be any monomial with nonzero coefficients of $J_{n,\mathcal{R}}$, where $R^\alpha = r_1^{\alpha_1} \cdot r_2^{\alpha_2} \cdot \ldots \cdot r_n^{\alpha_n}$. Then every power of α_i is even, the sum of the factors $\sum_{i=1}^n \alpha_i$ is equal to 2^{n-1}, coefficient p_α is an integer multiple of $1_\mathcal{R}$, i.e., is equal to $1_\mathcal{R}+1_\mathcal{R}+\ldots+1_\mathcal{R}$ or $-1_\mathcal{R}-1_\mathcal{R}\ldots-1_\mathcal{R}$ and the coefficient of $r_1^{2^{n-1}}$ equals $1_\mathcal{R}$.*

The existence of a polynomial that has these properties is quite an involved proof by induction. We only show here the outline of the most difficult part. By the induction hypothesis: $\prod(r_1 \pm \ldots \pm r_n \pm r_{n+1}) = \prod(r_1 \pm \ldots \pm (r_n + r_{n+1})) \cdot \prod(r_1 \pm \ldots \pm (r_n - r_{n+1})) = \sum_\alpha c_\alpha R^\alpha (r_n + r_{n+1})^{2i_\alpha} \cdot \sum_\beta c_\beta R^\beta (r_n - r_{n+1})^{2i_\beta}$ where R^α represent products of $r_1, \ldots, r_{n-1}$ to even powers. We multiply the sums as follows. If $i_\alpha = i_\beta$, then $c_\alpha R^\alpha (r_n + r_{n+1})^{2i_\alpha} c_\beta R^\beta (r_n - r_{n+1})^{2i_\beta} = c_\alpha c_\beta R^\alpha R^\beta (r_n^2 - r_{n+1}^2)^{i_\alpha}$. If $i_\alpha < i_\beta$ (the case $i_\alpha > i_\beta$ is similar) we add each two summands

$$c_\alpha R^\alpha (r_n + r_{n+1})^{2i_\alpha} c_\beta R^\beta (r_n - r_{n+1})^{2i_\beta} + c_\beta R^\beta (r_n + r_{n+1})^{2i_\beta} c_\alpha R^\alpha (r_n + r_{n+1})^{2i_\alpha} = c_\alpha c_\beta R^\alpha R^\beta \cdot (r_n^2 - r_{n+1}^2)^{i_\alpha} \cdot \sum_{i=0}^{i_\beta - i_\alpha} 2 \cdot \binom{2(i_\beta - i_\alpha)}{2i} r_n^{2i} r_{n+1}^{2(i_\beta - i_\alpha - i)}. \quad (5)$$

In both cases, we obtain a polynomial, where all variables are raised to even powers, which completes the most involved part of the proof. ⌟

The complete formalized proof of this theorem includes all the required sign combinations and required 143 helper lemmas and 13K lines of proofs.

The next step in the simplification of the polynomial given in Theorem 3 (following [10]), proceeds by defining $K_1(y, x_1, x_2, x_3)$ to be $J_{4,\mathbb{C}}(-y, \sqrt{x_1}, \sqrt{4x_2}, \sqrt{16x_3})$ and proving $\exists_{y \in \mathbb{Z}} K_1(y, x_1, x_2, x_3) = 0 \Leftrightarrow x_1 = \square \wedge x_2 = \square \wedge x_3 = \square$ under the assumptions $x_1, x_2, x_3 \in \mathbb{N}$ and $2 \nmid x_1, 2 \nmid x_2$. Note, that the substitution (compare Equation 4)

$$x_1 = (M^2 - 1)S^2 + 1,\ x_2 = ((MU)^2 - 1)T^2 + 1,\ x_3 = DFI$$

satisfies these assumptions.

The next step in the informal proof performs a rational substitution in the integer polynomial and justifies that this is again an integer polynomial. This requires some work with the type system in the formalization. Indeed, we perform the substitution $y := y - \frac{r}{p}$ in K_1, where p, r are the new variables. In order to construct the polynomial $K_2(y, x_1, x_2, x_3, p, r)$ to be $p^8 \cdot K_1(y - \frac{r}{p}, x_1, x_2, x_3)$ for $y, x_1, x_2, x_3, p, r \in \mathbb{R}$ where $p \neq 0$, the formalization is split into two stages. First, we define $K_2'(y, x_1, x_2, x_3, z) = K_1(y - z, x_1, x_2, x_3)$, where the power

α of z in each monomial in K_2' is ≤ 8 and replace z^α by $p^\alpha r^{8-\alpha}$ obtaining K_2. This way, we obtain a polynomial, where $K_2(y, x_1, x_2, x_3, p, r) = 0$ ensures $p|r$ for every $p, r \in \mathbb{N}, p \neq 0$, but both factors in Equation 4 are non-negative and $FL > 0$.

The final theorem confirms that K is a polynomial:

▶ **Theorem 5** (POLYNOM9:77). *Let $x_1, x_2, x_3, p, r, n \in \mathbb{N}$, $v \in \mathbb{Z}$ where $2 \nmid x_1$, $2 \nmid x_2$, $p > 0$, $n > \sqrt{x_1} + 2\sqrt{x_2} + 4\sqrt{x_3} + r$. Then $\exists_{y \in \mathbb{N}} K(y, x_1, x_2, x_3, p, r, n, v) = 0$ iff $x_1 = \square \wedge x_2 = \square \wedge x_3 = \square \wedge p \mid r \wedge 0 \leq v$, where we have $K(y, x_1, x_2, x_3, p, r, n, v) = K_2(y - nv, x_1, x_2, x_3, p, r)$.*

With K, we can represent Equation 4 using a single implicit unknown and we can represent primes using the 10-variable polynomial with the following substitutions in K:

$$\begin{aligned} v &= 8fuST(r - mSTU) - ((4f^2 - 1)(r - mSTU)^2 + 4u^2S^2T^2) - 1, \\ n &= MS + 2MUT + 4A^2CEGH + 2(HL + FfQ + Fk(W^2Su + W^2u^2)). \end{aligned}$$

With the abbreviations expanded, this gives a diophantine polynomial *Poly* of degree > 6000 over parameter k and unknowns $f, i, j, m, u, r, s, t, y$, so we present only the non-expanded version with the following property. Let $k \in \mathbb{N}^+$. Then $k+1$ is prime if and only if there exists a 10-element vector of natural numbers v such that the first element is equal to k ($v.1 = k$) and v is a root of *Poly* ($\mathsf{eval}(Poly, v) = 0.\mathbf{F}_\mathbb{R}$). In the formal proof, rather than specify the existence of a parameter k and 9 unknowns, we simplify this by using a 10-element vector with one element equal to k. The final statement in Mizar is:

```
theorem :: POLYNOM9:85
  ex Poly be INT-valued Polynomial of 10,F_R st
    for k be positive Nat holds
      k+1 is prime iff ex v being natural-valued Function of 10,F_R st
                              v.1 = k & eval(Poly,v) = 0.F_R;
```

This is already the minimal polynomial and completes our goal.

References

1. Jonas Bayer, Marco David, Abhik Pal, Benedikt Stock, and Dierk Schleicher. The DPRM theorem in Isabelle (short paper). In John Harrison, John O'Leary, and Andrew Tolmach, editors, *10th International Conference on Interactive Theorem Proving, ITP 2019*, volume 141 of *LIPIcs*, pages 33:1–33:7. Dagstuhl, 2019. doi:10.4230/LIPIcs.ITP.2019.33.
2. Mario Carneiro. A Lean formalization of Matiyasevič's theorem, 2018. arXiv:1802.01795.
3. Martin Davis. Arithmetical problems and recursively enumerable predicates. *J. Symb. Log.*, 18(1):33–41, 1953. doi:10.2307/2266325.
4. Martin Davis, Hilary Putnam, and Julia Robinson. The decision problem for exponential diophantine equations. *Annals of Mathematics*, 74:425–436, 1961. doi:10.2307/1970289.
5. James P. Jones. Universal diophantine equation. *The Journal of Symbolic Logic*, 45(3):549–571, 1982.
6. 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.
7. Michael Lamoureux, editor. *On Hilbert's Tenth Problem*, volume 1. Pacific Institute for the Mathematical Sciences, PIMS Distinguished Chair Lectures, 2000.
8. Dominique Larchey-Wendling and Yannick Forster. Hilbert's Tenth Problem in Coq. In Herman Geuvers, editor, *4th International Conference on Formal Structures for Computation and Deduction (FSCD 2019)*, volume 131 of *LIPIcs*, pages 27:1–27:20, Dagstuhl, Germany, 2019. Dagstuhl. doi:10.4230/LIPIcs.FSCD.2019.27.

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

10 Yuri Matiyasevich. Primes are nonnegative values of a polynomial in 10 variables. *Journal of Soviet Mathematics*, 15:33–44, 1981. doi:10.1007/BF01404106.

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

12 Karol Pąk. Formalization of the MRDP theorem in the Mizar system. *Formalized Mathematics*, 27(**2**):209–221, 2019. doi:10.2478/forma-2019-0020.

13 Karol Pąk. Formalization of prime representing polynomial in Mizar. In *FMM 2021 workshop*, 2021. URL: http://alioth.uwb.edu.pl/~pakkarol/articles/KP-FMM2021.pdf.

14 Karol Pąk. The Matiyasevich Theorem. Preliminaries. *Formalized Mathematics*, 25(**4**):315–325, 2017. doi:10.1515/forma-2017-0029.

15 Julia Robinson. Diophantine decision problems. *Studies in number theory*, 6:76–116, 1969.

16 Craig Alan Smorynski. *Logical Number Theory I, An Introduction*. Universitext. Springer-Verlag Berlin Heidelberg, 1991.

17 Zhi-Wei Sun. Further results on Hilbert's Tenth Problem. *Science China Mathematics*, 64:281–306, 2021. doi:10.1007/s11425-020-1813-5.