The High-School Math-Club Proof for DavisPutnamRobinson *

PETER FREYD
pjf@upenn.edu

January 21, 2020

An **exponential polynomial** is a function of the form

$$P\langle x_1, x_2, \ldots, x_n\rangle \;=\; R\langle x_1, x_2, \ldots, x_n, 2^{x_1}, 2^{x_2}, \ldots, 2^{x_n}\rangle$$

where $R\langle x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n\rangle$ is an ordinary polynomial with integer coefficients. An **exponential diophantine set** ("**E-D set**") is a set of natural numbers that appears as the set of non-negative values of an exponential polynomial with the variables ranging over the natural numbers.

The purpose of this note is to give an elementary proof of this theorem first proved by Davis, Putnam and Robinson:[1]

Theorem:

A set of natural numbers is recursively enumerable iff it is an exponential diophantine set. [2]

One direction, of course, is immediate: any E-D set is obviously recursively enumerable. For the other direction we need a few more definitions.

All variables will be understood to refer to natural numbers. An n-ary predicate on the variables $x_1, x_2, \ldots, x_n$ is an **exponential diophantine predicate** ("**E-D predicate**") if it is of the form

$$\exists_{y_1,y_2,\ldots,y_m}\,\{P\langle x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_m\rangle = 0\}$$

where P is an exponential polynomial.

An n-ary (partial) function, f, is an **exponential diophantine (partial) function** ("**E-D (partial) function**") if the $(n{+}1)$–ary predicate $f(x_1, x_2, \ldots, x_n) = x_{n+1}$ is an E-D predicate.

---

* It's a fond position of mine that proofs actually exist only in their performance. This proof—in its arm-waving form—dates from the 70s. It was my pleasure to describe it to Hilary Putnam on the occasion of his receiving an honorary degree (in 1985) from Penn—it happened during the cross-campus academic procession. he said, to my surprise, that he always knew it had to be trivialized some day. (Some years earlier I had started describing it to John H. Conway. As with Hilary I jumped immediately to the three ★s on page 3 below—for John there wasn't much I had to tell him; he enjoyed finishing the proof on his own). The written version below is best viewed as something of a score, indeed something of a "figured base" score for a performance.

[1] Martin Davis, Hilary Putnam, Julia Robinson: "The decision problem for exponential diophantine equations," *Ann. of Math.* (2) 74 1961 425–436.

[2] "Recursively enumerable" is the traditional way of saying that a set's elements can be listed—albeit not in order—by a digital computer.

Clearly every E-D set is the extent of a unary E-D predicate. Conversely, given a unary E-D predicate on one free variable, x,

$$\exists_{y_1,y_2,\ldots,y_m}\{P\langle x, y_1, y_2, \ldots, y_m\rangle = 0\}$$

we obtain its extent as the set of non-negative values of

$$x - (1 + x)P^2\langle x, y_1, y_2, \ldots, y_m\rangle.$$

E-D predicates are easily seen to be closed under existential quantification and (finite) conjunction using addition of squares. As an example:

$$\exists_y\{P\langle x, y\rangle = 0\} \wedge \exists_y\{Q\langle x, y\rangle = 0\} \quad \Leftrightarrow \quad \exists_{y,z}\{P^2\langle x, y\rangle + Q^2\langle x, z\rangle = 0\}.$$ [3]

An ordinary polynomial applied to an exponential polynomial yields an exponential polynomial, but such is not the case for the composition of exponential polynomials. Nonetheless, when the value of an E-D function is substituted for a variable in an E-D predicate the result is still an E-D predicate because if $f\langle x_1, \ldots, x_n\rangle = x_{n+1}$ and $\Phi\langle y_1, \ldots, y_k, \ldots, y_m\rangle$ are E-D predicates then $\Phi\langle y_1, \ldots, f(x_1, \ldots, x_n), \ldots, y_m\rangle$ is equivalent to $\exists_z\{\,[f\langle x_1, \ldots, x_n\rangle = z] \wedge [\Phi\langle y_1, \ldots, z, \ldots, y_m\rangle]\,\}$.

A corollary is that E-D functions are closed under composition.

We need to build a repertoire of E-D predicates and functions. $x = y$ is an E-D predicate (because, of course, it's the same as $x - y = 0$); $x \leq y$ (because $x \leq y$ iff $\exists_z\, x + z = y$);[4] $x < y$ (because $x < y$ iff $x + 1 \leq y$); and using Pascal [5] notation:

$$x \,\texttt{DIV}\, y = z \quad \Leftrightarrow \quad \exists_u\{(u < y) \wedge (yz + u = x)\};$$
$$x \,\texttt{MOD}\, y = u \quad \Leftrightarrow \quad \exists_z\{(u < y) \wedge (yz + u = x)\}.$$

If E-D functions did not include x^y (for positive x) we would change the definition of exponential diophantine. But, in fact, x^y is equal for positive x to $2^{(x+1)(y+1)y}$ MOD $(2^{(x+1)(y+1)} - x)$. This mysterious-looking fact is entirely elementary: first, $2^z \equiv x \bmod (2^z - x)$ for any large z (obviously); next raise each side of the congruence to the y^{th} power: $2^{zy} \equiv x^y \bmod (2^z - x)$; third, and finally, replace z with a term on x and y large enough so that $x^y < (2^z - x)$.

We'll use a power of a number other than 2 just once:

[3] We won't need it, but for for the record use multiplication for disjunction:

$$\exists_y\{P\langle x, y\rangle = 0\} \vee \exists_y\{Q\langle x, y\rangle = 0\} \quad \Leftrightarrow \quad \exists_y\{\,P\langle x, y\rangle Q\langle x, y\rangle = 0\,\}$$

[4] The existential is, note, unique. Indeed, all existentials henceforth will be unique, thus we could have defined E-D predicates to require unique existentials.

[5] Back in the 70s—when this proof was devised—the computer language Pascal was an AP subject.

We obtain binomial coefficients as an E-D function with:

$$\binom{n}{r} = \left((2^{n+2}+1)^n \texttt{ DIV } 2^{(n+2)r}\right) \texttt{ MOD } 2^{n+2}$$

This is, perhaps, most easily seen by using the binomial theorem (of all things) on $\left((K+1)^n \texttt{ DIV } K^r\right) \texttt{ MOD } K$ for very large K and then noticing that $K = 2^{n+2}$ is large enough.

$$\begin{aligned}
(K+1)^n &= K^n \quad + \cdots + \binom{n}{r+1} K^{r+1} + \binom{n}{r} K^r \ + \cdots + 1 \\
(K+1)^n \texttt{ DIV } K^r &= K^{n-r} + \cdots + \binom{n}{r+1} K \quad + \binom{n}{r} \\
\left((K+1)^n \texttt{ DIV } K^r\right) \texttt{ MOD } K &= \binom{n}{r}
\end{aligned}$$

(We won't need it, but an E-D formula for the factorial function is now easily obtainable.[6] And from that—and Wilson's theorem—we could obtain the set of primes as an E-D set, to wit, the set of non-negative values of $(x+2)(1-2(x!-1-y(x+2))^2)$. [7])

⋆ ⋆ ⋆

Our point of departure is the following. Every natural number names a unique finite set of natural numbers via binary encoding: given a set, S, its name is $\sum_{i\in S} 2^i$; given a number, n, the set it names is $\{\, i \mid (n \texttt{ DIV } 2^i) \texttt{ MOD } 2 = 1 \,\}$. hence any predicate or function on finite sets of natural numbers gives rise to a predicate or function on numbers. We will reserve upper-case italics for natural numbers that are being used as names of finite sets of natural numbers. We will write $A \subseteq B$ to mean that the set named by A is contained in the set named by B. We will write $A \perp\!\!\!\perp B$ to mean that the sets named by A and B are disjoint.

A wonderful fact known from ancient times is:

Lemma:

$A \perp B$ *iff* $\binom{A+B}{A}$ *is odd.*

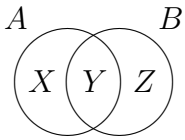
The proof uses the formula for the power of a prime p that divides $n!$, to wit $\sum_{i=1}^{\infty} n \texttt{ DIV } p^i$. Note that for any A, B and $C \neq 0$ it is the case that $A \texttt{ DIV } C + B \texttt{ DIV } C \leq (A+B) \texttt{ DIV } C$ (because $\lfloor x \rfloor + \lfloor y \rfloor \leq \lfloor x+y \rfloor$ [8]). hence, $(A+B)!/A!B!$ is odd iff $A \texttt{ DIV } 2^i + B \texttt{ DIV } 2^i = (A+B) \texttt{ DIV } 2^i$ each i. In some implementations of Pascal, $A \texttt{ DIV } 2^i$ is denoted $A \texttt{ SHR } i$, meaning that the number A, written in binary notation, is shifted to the right by removing the right-most i binary digits. It isn't hard to see that $A \texttt{ SHR } i + B \texttt{ SHR } i = (A+B) \texttt{ SHR } i$ iff no "carry" takes place from the $(i-1)^{\text{th}}$ column to the i^{th} column when adding A and B in binary arithmetic. hence the oddness of $\binom{A+B}{A}$ is equivalent to no carries occurring in any column and that's just what $A \perp B$ is saying.

[6] Since $r! = \lim_{n\to\infty} n^r \binom{n}{r}^{-1}$ and the convergence is from above, we have that $n^r \texttt{ DIV } (((2^n+1)^n \texttt{ DIV } 2^{rn}) \texttt{ MOD } 2^n) = r!$ for large enough n. Indeed, $n = 2^{r^2}$ is large enough (hence $r! = 2^{r^3} \texttt{ DIV } (((2^{2^{r^2}}+1)^{2^{r^2}} \texttt{ DIV } 2^{r^2 2^{r^2}}) \texttt{ MOD } 2^{2^{r^2}})$.)

[7] I had the pleasure of learning a version of this in an early 60s conversation with Hilary Putnam; I had the even greater pleasure of relaying it to a lot of other mathematicians.

[8] The triangle inequality backwards!

Using the Venn diagram



we obtain:

$$A \cap B = Y \;\Leftrightarrow\; \exists_{X,Z}\Big\{(X \perp Y) \wedge (X \perp Z) \wedge (Y \perp Z) \wedge (X + Y = A) \wedge (Z + Y = B)\Big\}$$

$$A \setminus B = X \;\Leftrightarrow\; \exists_{Y,Z}\Big\{(X \perp Y) \wedge (X \perp Z) \wedge (Y \perp Z) \wedge (X + Y = A) \wedge (Z + Y = B)\Big\}$$

$$A \subseteq B \;\Leftrightarrow\; A \cap B = A \;\; [9]$$

We can now proceed to encode the behavior of a Turing machine into E-D predicates. We consider a Turing machine whose tape-symbols are $\{0, 1\}$ and whose states are $\{0, 1, \ldots, \mathsf{F}\}$, where 0 will be a strict initial state and F a unique dead final state. We wish to know the machine's "successful numbers," defined here to be those numbers n such that the machine reaches state F when started in state 0 on a tape that is all 0 except for a string of 1's of length n, the initial position being the right-most 1 (understanding that if $n = 0$ the initial position doesn't matter). For technical reasons (to become clear) we will insist that the machine be modified, if necessary, so that for every successful n it requires at least $n+2$ steps to reach F and it makes at least one left and one right move.

We take it as known that every recursively enumerable set arises as a set of successful numbers for some such Turing machine.

The behavior of the machine is determined by three functions, $\boldsymbol{\sigma}, \boldsymbol{\mu}, \boldsymbol{\nu}$, each defined on the cartesian product $\{0, 1\} \times \{0, 1, \ldots, \mathsf{F}–1\}$.

Upon reading symbol i in state s:
$\sigma\langle i, s\rangle \in \{0, 1\}$ is the symbol to be written;
$\mu\langle i, s\rangle \in \{\ell, r\}$ is then the prescribed motion of the machine (we will think of the tape as stationary and the machine as moving left or right);
$\nu\langle i, s\rangle \in \{1, 2, \ldots, \mathsf{F}\}$ is the next state.

We will suppose that $\boldsymbol{n}$ is a successful number. We shall list a sequence of variables and equations that describe this success. The first is $\boldsymbol{t}$, the number of steps it takes for the machine to reach state F. We will use t also as the radius of the tape, that is, we view the tape as being of length $2t$, with the initial position as the left member of the two center positions. The condition that $t > n + 1$, insures, among other things, that there's room for all of the initial 1s.

All further variables will be upper-case and will be viewed as names of finite sets of natural numbers. The first of these is $\boldsymbol{T}$ which will record the entire history of the tape. We will consider T as consisting of t blocks, each block consisting of $2t$ binary digits. The right-most block will be the initial tape. (See the example at the end.)

The next variable, $\boldsymbol{P}$, will record the history of the machine's position. Equations to come will insure that it consists of t blocks each block consisting of $2t$ digits all 0 except for exactly one 1.

[9] It might be noticed that necessarily $X = 0$, $Y = A$, $Z = B - A$ and hence $A \subseteq B$ *iff* $\binom{B}{A}$ *is odd.* A direct argument is available based on the fact that $A \subseteq B$ iff when A is subtracted from B in binary arithmetic no "borrowing" occurs. (We could, of course, also define $A \cup B$ but all unions we'll need will be disjoint unions and ordinary addition will suffice.)

$\boldsymbol{E}$ is for our convenience. It marks the end of each block and is defined by

$$E \;=\; (2^{2t})^{t-1} + \cdots + (2^{2t})^2 + (2^{2t}) + 1 \;=\; (2^{2t^2} - 1)\ \mathtt{DIV}\ (2^{2t} - 1)$$

It has a unique 1 in each block, to wit, at the far right end of that block. All further variables denote subsets of E. (We'll be using this space for input, output and the internal states of the machine.) The first of these variables is $\boldsymbol{R}$, the "symbol being read." It is defined by the two E-D predicates:

$$R \subseteq E \quad \text{and} \quad (2^{2t} - 1)R \cap P = T \cap P$$

To see what this last equation is doing, imagine T written in binary arithmetic with P written immediately below it. In each block the unique 1 of P may be viewed as pointing to the binary digit of T above it. We want R to have a copy—in each block—of that digit in the right end of the block. Note first— keeping $R \subseteq E$ in mind—that $(2^{2t} - 1)R$ fills each block either with all 0s or 1s depending on the digit of R occurring at the right end of that block. $(2^{2t} - 1)R \cap P$ consists of at most one 1 in each block and if 1 then its position is the same as the machine's. It's a 1 iff the digit of R appearing at the right end of that block is a 1.

$\boldsymbol{W} \subseteq E$ is for the "symbol to be written."

Perhaps the most substantive of the equations (it determines T given P, W, t and n):[10]

$$T = \left[\Big((T \setminus P) + \big((2^{2t} - 1)W \cap P\big) \Big)\, 2^{2t} + (2^n - 1)2^t \right] \mathtt{MOD}\ 2^{2t^2}$$

Reading left to right: $T \setminus P$ describes the tape with the symbol being read in each block erased; $(2^{2t} - 1)W$ fills each block with all 0s or 1s depending on the digit of W that occurs at the right end of that block; $(2^{2t} - 1)W \cap P$ is all 0 except for one 1 just in those blocks where the symbol to be written is 1 and the position of each such 1 is the same as the machine's: $(T \setminus P) + \big((2^{2t} - 1)W \cap P\big)$ thus describes the result of erasing in each block the symbol being read and replacing it with the symbol to be written; $\Big((T \setminus P) + \big((2^{2t} - 1)W \cap P\big)\Big)\, 2^{2t}$ is the result of shifting the last result $2t$ digits to the left: $\Big((T \setminus P) + \big((2^{2t} - 1)W \cap P\big)\Big)\, 2^{2t} + (2^n - 1)2^t$ inserts the initial tape; the final expression (the one involving $\mathtt{MOD}\ 2^{2t^2}$) describes the result of then throwing away the very last tape (which, note, will be unneeded since the final state will have been reached).

$M \subseteq E$ encodes the motion instruction:

$$P = \left[\Big(P \cap \big((2^{2t} - 1)M\big) \Big)\, 2^{2t+1} + \Big(P \setminus \big((2^{2t} - 1)M\big) \Big)\, 2^{2t-1} + 2^t \right] \mathtt{MOD}\ 2^{2t^2}$$

Reading left to right: $P \cap \big((2^{2t} - 1)M\big)$ is the result of erasing all digits in P except those that occur in a block in which M has a 1 at the right end of that block; 2^{2t+1} will shift the modified position variable one block plus one position to the left; $P \setminus \big((2^{2t} - 1)M\big)$ is the result of erasing all digits in P except those that occur in a block in which M has a 0 at the right end of that block; 2^{2t-1} will shift the modified position variable one block minus one position to the left; $+2^t$ inserts the initial position; thus $\Big(P \cap \big((2^{2t} - 1)M\big)\Big)\, 2^{2t+1} + \Big(P \setminus \big((2^{2t} - 1)M\big)\Big)\, 2^{2t-1} + 2^t$ establishes the initial position and then moves it either one position left or right for each succeeding block depending on the digits in M; the final expression is the result of then throwing away the very last position marker.

[10] And is the only appearance of n in any of the conditions.

The remaining variables, $\boldsymbol{S}_0,\boldsymbol{S}_1,\ldots,S_{\mathsf{F}-1}$ encode the states. In each block all but one of these variables will be all 0 and in the one exceptional case there will be just one 1 and it will be at the right end of the block. hence we require that $S_s \perp S_{s'}$ for all $s < s' < \mathsf{F}$ and $S_0 + S_1 + \cdots + S_{\mathsf{F}-1} = E$ (implying, in particular, that $S_s \subseteq E$ for all s). S_0 is for convenience: we set it as $S_0 = 1$.

The remaining equations embody the defining functions of the machine. W and M are determined by the values of σ and μ, which, in turn, are determined by $S_0, S_1, \ldots, S_{\mathsf{F}-1}$ and R:

$$W = \sum_{\{\, s \mid \sigma\langle 1,s\rangle = 1 \,\}} S_s \cap R \quad + \sum_{\{\, s \mid \sigma\langle 0,s\rangle = 1 \,\}} S_s \setminus R$$

$$M = \sum_{\{\, s \mid \mu\langle 1,s\rangle = \ell \,\}} S_s \cap R \quad + \sum_{\{\, s \mid \mu\langle 0,s\rangle = \ell \,\}} S_s \setminus R$$

The W-equation says that there's a 1 in the right end of a block in W iff one of two conditions holds: the symbol being read is 1 and the state is such that σ calls for 1 to be written when 1 is being read; the symbol being read is 0 and the state is such that σ calls for 1 to be written with 0 is being read. Essentially the same analysis holds for the M-equation.

$S_1, S_2, \ldots, S_{\mathsf{F}-1}$ are determined by the values of ν, which, in urn, are determined by the "earlier" values of $S_0, S_1, \ldots, S_{\mathsf{F}-1}$ and R. For each $s' = 1, 2, \ldots, \mathsf{F}–1$:

$$S_{s'} = \left[\sum_{\{\, s \mid \nu\langle 1,s\rangle = s' \,\}} S_s \cap R \quad + \sum_{\{\, s \mid \nu\langle 0,s\rangle = s' \,\}} S_s \setminus R \right] 2^{2t}$$

The analysis of the S-equations is as for the W- and M-equations with the added feature of shifting everything $2t$ digits to the left. Note that $S_0 + S_1 + \cdots + S_{\mathsf{F}-1} = E$ requires that state F be called for at the t^{th} step (and not before).

(The disjointness of the Ss did not require an explicit assumption—it is a consequence of the final equations. The fact that the Ss are all less than 2^{2t} is what implies that the dead state is reached; the fact that the sum of the Ss equals E is needed only for the promise that all existentials are unique.)

Our assumption has been that n is a successful number. It should be clear that in that case there is a value for each of the variables $t, T, P, R, W, M, S_1, S_2, \ldots, S_{\mathsf{F}-1}$.

For the converse add the conditions $t \geqq 2$, $T \perp E$ and $P \perp E$ (which in the forward direction are consequences of the special conditions that we imposed on the Turing machine "for technical reasons"). Given a set of values for the variables that satisfy the E-D equations we obtain a description of the behavior of a machine that succeeds in reaching state F starting with the initial tape for n.

Example →

Example

$n = 2 \qquad t = 5$

T*ape*	=	0010100000	0011100000	0001100000	0001100000	0001100000.
P*osition*	=	0000100000	0001000000	0010000000	0001000000	0000100000.
E*nd*	=	0000000001	0000000001	0000000001	0000000001	0000000001.
R*ead*	=	0000000001	0000000001	0000000000	0000000001	0000000001.
W*rite*	=	0000000000	0000000000	0000000001	0000000001	0000000001.
M*ove*	=	0000000000	0000000000	0000000000	0000000001	0000000001.

$$f$$

Available at

`http://www.math.upenn.edu/~pjf/diophantine.pdf`