FINITE GENERATION OF RECURSIVELY ENUMERABLE SETS

JULIA ROBINSON

Suppose we wish to build up the class of recursively enumerable sets by starting with the set $\mathfrak{N}$ of natural numbers and constructing new sets from those already obtained using as little auxiliary machinery as possible. One way would be to start with a finite number of functions $F_1, \cdots, F_k$ (of one variable, from and to $\mathfrak{N}$) such that every recursively enumerable set can be obtained from $\mathfrak{N}$ by constructing new sets $F_j[\mathfrak{J}]$ where $\mathfrak{J}$ is a previously obtained set. We can think of $F_1, \cdots, F_k$ as unary operations on sets of natural numbers. Any set $\mathcal{S}$ obtained in this way is the range of a function F obtained by composition from $F_1, \cdots, F_k$. If we consider the values of $F_1, \cdots, F_k$ as given, then the number of steps needed to compute Fn does not depend on n. Hence for all $x \in \mathcal{S}$, there exists a proof that $x \in \mathcal{S}$ of bounded length in terms of $F_1, \cdots, F_k$ (just as there is a one-step proof that a composite number is composite in terms of multiplication).

We say a set of natural numbers is *generated* by $F_1, \cdots, F_k$ if it is the range of a function obtained by composition from $F_1, \cdots, F_k$. Also a class $\mathcal{C}$ of sets is *generated* by $F_1, \cdots, F_k$ if every nonempty set of $\mathcal{C}$ is generated by $F_1, \cdots, F_k$ and every set generated by $F_1, \cdots, F_k$ is in $\mathcal{C}$.

EXAMPLE. Let $G_0, G_1, \cdots$ be the primitive recursive functions listed systematically so that the function G given by

$$G(2^n(2x+1)-1) = G_n x$$

is recursive. Then

$$G_n = G(SD)^n D$$

where $Sx = x+1$ and $Dx = 2x$. Every nonempty recursively enumerable set is the range of some primitive recursive function and hence the range of $G(SD)^n D$ for some n. Since S, D, and G are recursive functions, any function obtained from them by composition will be recursive and will have a recursively enumerable range. Thus G, S, and D generate the class of recursively enumerable sets. However, they do not form an interesting set of generators since all the work of listing recursively enumerable sets is done in computing G.

Received by the editors August 22, 1967.

In this paper, we give sets of easy-to-compute generators for the classes of recursively enumerable sets and diophantine sets, i.e. sets which are existentially definable in terms of $+$ and $\cdot$. General theorems describing classes of sets which can be finitely generated are proved in [**3**].

Let $Ix=x$, $Ox=0$, $Tx=2^x$, $Zx=0^x$, and $E(x, y)$ be the characteristic function of equality, i.e. $E(x, y)=0^{|x-y|}$. We shall also use pairing functions. If $J(x, y)$ maps the set of ordered pairs of natural numbers onto $\mathfrak{N}$ in a one-to-one way, then J and its inverse functions K and L, given by $KJ(x, y)=x$ and $LJ(x, y)=y$, are called pairing functions. We shall always take J to be the Cantor pairing function given by

$$J(x, y) = \tfrac{1}{2}((x + y)^2 + 3x + y)$$

unless otherwise stated. Its inverse functions are probably the simplest functions which assume every natural number infinitely often. They can easily be computed recursively by the equations

$$\begin{array}{lll} & K0 = L0 = 0, & \\ (1) & K(x + 1) = 0, \quad L(x + 1) = Kx + 1 & \text{if } Lx = 0, \\ & K(x + 1) = Kx + 1, \quad L(x + 1) = Lx - 1 & \text{if } Lx \neq 0. \end{array}$$

EXAMPLE. Let J, K, and L be any pairing functions and let G be given by $G(J(x, n))=G_nx$. Then $G_nx=GJ(K, SL)^nJ(I, O)x$. In the example above, we took $J(x, n)=2^n(2x+1)-1$, so $J(x, n+1) =2J(x, n)+1$ and $J(x, 0)=2x$.

LEMMA 1. *The functions Ox, Zx, $x+y$, $x\cdot y$, and $E(x, y)$ are all obtainable from J, K, S, and D by substitution.*

PROOF. R. M. Robinson [**4**, p. 665], derived the remarkable identity for all x and y with $x\geqq y$:

$$(2) \qquad x - y = KDDJ(SSSKDDJ(SSSKDDJ(y, DDx), x), Dy).$$

He then used it to define $x+y$ by substitution from J, K, S, and D. We shall let $x-y$ denote the function on the right of (2) for *all* x and y. Then $Ox=x-x$ and $x\cdot y=(J(0, x+y)-J(0, x))-J(0, y)$. To define Z, we make use of (1) to see that

$$SJ(0, 0) = J(0, 1), \qquad SJ(0, x + 1) = J(1, x).$$

Hence $KSJ(0, x)=\operatorname{sgn} x$. Since $K2=1$ and $K3=0$, we have $Zx =KSSKSJ(0, x)$. Finally, $E(x, y)=Z((x-y)+(y-x))$.

THEOREM 1. *Every nonempty recursively enumerable set is the range of a function of one variable obtained from S, D, T, K, and L by com-*

position and pairing. Conversely, the range of such a function is recursively enumerable. If T is omitted from the set of initial functions then just the nonempty diophantine sets are obtained.

PROOF. Every recursively enumerable set $\mathcal{S}$ is exponential diophantine and conversely. (See Davis, Putnam, and Robinson [1].) Hence

$$x \in \mathcal{S} \leftrightarrow \bigvee_{y_1, \cdots, y_n} F(x, y_1, \cdots, y_n) = G(x, y_1, \cdots, y_n),$$

where F and G are suitably chosen terms built up from $x, y_1, \cdots, y_n$, and particular natural numbers by means of $+$, $\cdot$, and T. (See [1, Corollary 5].) Let a be any element of $\mathcal{S}$. Then $\mathcal{S} = \mathcal{R}(H)$ where

$$H(x, y_1, \cdots, y_n) = x \cdot E(F(x, y_1, \cdots, y_n), G(x, y_1, \cdots, y_n)) \\ + a \cdot ZE(F(x, y_1, \cdots, y_n), G(x, y_1, \cdots, y_n)).$$

Let $M = H(K, KL, \cdots, KL^{n-1}, L^n)$. Then M is a function of one variable obtained from O, S, K, L, T, and Z by forming new functions F from previously obtained functions A and B by taking $F = AB$, $F = A + B$, $F = A \cdot B$, and $F = E(A, B)$. Also $\mathcal{R}(M) = \mathcal{R}(H) = \mathcal{S}$. By Lemma 1, we see that every such function M can be obtained from K, L, S, D, T, and $I = J(K, L)$ by composition and pairing. On the other hand, M is primitive recursive so $\mathcal{R}(M)$ is recursively enumerable.

The proof that every nonempty diophantine set is obtained if we omit T from the set of initial functions, is obtained from the above proof by omitting T throughout. A function F is said to be diophantine if there is a polynomial P with integer coefficients such that

$$y = Fx \leftrightarrow \bigvee_{u_1, \cdots, u_n} P(x, y, u_1, \cdots, u_n) = 0.$$

Hence if F is diophantine then $\mathcal{R}(F)$ is a diophantine set. Indeed,

$$y \in \mathcal{R}(F) \leftrightarrow \bigvee_{x, u_1, \cdots, u_n} P(x, y, u_1, \cdots, u_n) = 0.$$

Also, if F and G are diophantine then FG is diophantine. Clearly, J, K, L, S, and D are diophantine functions so the range of any function obtained from K, L, S, and D by composition and pairing is diophantine.

REMARK. It is not known whether all recursively enumerable sets are diophantine. Hence we do not know whether T is necessary to obtain all recursively enumerable sets.

Let $C = J(KL, J(LL, K))$. Then $C^3 = I$, since $CJ(x, J(y, z)) = J(y, J(z, x))$. For any function A, we define $A^* = J(K, AL)$. Then for all A and B

$$(3) \qquad (AB)^* = A^*B^*,$$

$$(4) \qquad J(A, B) = B^*J(L, K)A^*J(I, I),$$

$$(5) \qquad A^{**} = CJ(L, K)A^*J(L, K)C^2.$$

These formulas can be easily checked by carrying out the compositions indicated on the right using the fact that $J(FK, GL)J(M, N) = J(FM, GN)$ and $J(FL, GK)J(M, N) = J(FN, GM)$ etc.

LEMMA 2. *If F can be obtained from $A_1, \cdots, A_n$, K, and L by composition and pairing, then F can be obtained from $A_1^*, \cdots, A_n^*$, K, $J(L, K)$, $J(I, I)$, and $J(KL, J(LL, K))$ by composition alone. Here J, K, and L can be arbitrary pairing functions.*

PROOF. Let $\mathfrak{A}$ be the least class of functions closed under composition which contains $A_1^*, \cdots, A_n^*$, K, $J(L, K)$, $J(I, I)$, and C. We wish to show that if F is obtained from $A_1, \cdots, A_n$, K, and L by composition and pairing, then F belongs to $\mathfrak{A}$. Since $F = LF^*J(I, I)$ and $L = KJ(L, K)$, it is sufficient to show that F^* belongs to $\mathfrak{A}$, and this will be done by induction.

I. Suppose $F = A_j$ then F^* belongs to $\mathfrak{A}$ by definition. Also $K^* = LC^2$ and $L^* = J(L, K)LC$. Hence both K^* and L^* belong to $\mathfrak{A}$.

II. Suppose $F = AB$ where A^* and B^* belong to $\mathfrak{A}$. Then $F^* = A^*B^*$ so F^* belongs to $\mathfrak{A}$.

III. Suppose $F = J(A, B)$ where A^* and B^* belong to $\mathfrak{A}$. Then $F = B^*J(L, K)A^*J(I, I)$.

Hence

$$F^* = B^{**}J(L, K)^*A^{**}J(I, I)^*.$$

By (5), B^{**} and A^{**} can be obtained by composition from C, $J(L, K)$, B^*, and A^*. Hence we need only show that $J(L, K)^*$ and $J(I, I)^*$ belong to $\mathfrak{A}$. Now

$$J(L, K)^* = J(K, J(LL, KL)) = CJ(KL, J(K, LL)) = CJ(KL, L^*).$$

Hence by (4),

$$J(L, K)^* = CL^{**}J(L, K)K^*L^*J(I, I).$$

Also

$$J(I, I)^* = J(K, J(L, L)) = L^{**}CJ(I, I).$$

Finally, K^{**} and L^{**} belong to $\mathfrak{A}$ by (5) and I, hence F^* belongs to $\mathfrak{A}$.

THEOREM 2. *The class of recursively enumerable sets is generated by K, $J(K, SL)$, $J(K, DL)$, $J(K, TL)$, $J(L, K)$, $J(I, I)$, $J(KL, J(LL, K))$.*

This set of generators with $J(K, TL)$ deleted generates the class of diophantine sets.

THEOREM 2 is an immediate consequence of Lemma 2 and Theorem 1.

REMARK. By the lemma on page 714 of [2], the functions K, $J(L, K)$, and $J(I, I)$ can be replaced by the two functions $J(LK, KL)$ and $J(L, I)$. (Recall that $K=J(KK, LK)$.) Indeed, the total number of generators can be reduced to two,

$$K \quad \text{and} \quad J(L, J(F_1, J(F_2, \cdots, J(F_{n-1}, F_n))) \cdots),$$

where $F_1, \cdots, F_n$ are the generators other than K.

LEMMA 3. *If F can be obtained by composition and pairing from $A_1, \cdots, A_n$, K, and L then $F=KB$ for some function B obtained by composition from $A_1^*, \cdots, A_n^*$, $J(L, K)$, $J(I, I)$, $J(KL, J(LL, K))$, $J(L, K)^*$, $J(I, I)^*$, and $J(KL, J(LL, K))^*$. Here J, K, and L can be arbitrary pairing functions.*

PROOF. By Lemma 2,

$$F = B_0KB_1K \cdots B_{t-1}KB_t \tag{6}$$

for some $B_0, \cdots, B_t$ obtained from $A_1^*, \cdots, A_n^*$, $J(L, K)$, $J(I, I)$, and $J(KL, J(LL, K))$ by composition. We can take $t>0$ since $B_0=B_0KJ(I, I)$. (If B_i is the identity function, then $B_i=J(L, K)^2$.) Now

$$J(L, K)A^*J(L, K) = J(AK, L).$$

Hence $AK=KJ(L, K)A^*J(L, K)$. Thus each K in (6) can be brought to the front in turn. For example,

$$\begin{aligned} F &= B_0KB_1K \cdots \\ &= KJ(L, K)B_0^*J(L, K)B_1K \cdots \\ &= K^2J(L, K)J(L, K)^*B_0^{**}J(L, K)^*B_1^*J(L, K)B_2K \cdots \\ &\;\;\vdots \\ &= K^tH \end{aligned}$$

where H is obtained from the functions listed in the lemma by composition. Finally,

$$KK = KJ(KL, J(LL, K))J(L, K) = KCJ(L, K)$$

so

$$F = K^t H = K(CJ(L, K))^{t-1}H.$$

THEOREM 3. *Every nonempty recursively enumerable set is the range of a function KB where B is obtained by composition from*

(7) $S^*,\ D^*,\ T^*,\ J(L, K),\ J(I, I),\ C,\ J(L, K)^*,\ J(I, I)^*,\ C^*.$

The range of B is a primitive recursive set. If T^* is deleted from the set of functions (7) *then just the nonempty diophantine sets are obtained. In this case, both the range of B and its complement are diophantine.*

PROOF. In light of Lemma 3 and Theorem 1, we need only show that $\mathcal{R}(B)$ is a primitive recursive set in the first case and $\mathcal{C}\mathcal{R}(B)$ is diophantine in the second case. (Here $\mathcal{C}\mathcal{R}(B)$ is the complement of the range of B.) Notice that $Gx \geqq x$ for G equal to S^*, D^*, T^*, $J(I, I)^*$, or $J(I, I)$. This is clear for the *-functions since $J(x, y) \leqq J(x, z)$ whenever $y \leqq z$. Thus $S^*x = J(K, SL)x \geqq J(K, L)x = x$, etc. Also $J(I, I)x = J(x, x) \geqq x$. Hence if G is one of these functions, then

$$x \in \mathcal{R}(GH) \leftrightarrow \bigvee_{y \leqq x} (Gy = x \wedge y \in \mathcal{R}(H)).$$

Furthermore, if G is one of the remaining functions of (7), then G is a primitive recursive permutation such that G^{-1} is also primitive recursive. Indeed, $C^{-1} = C^2$, $(C^*)^{-1} = C^*C^*$, $J(L, K)^{-1} = J(L, K)$, and $(J(L, K)^*)^{-1} = J(L, K)^*$. Hence

$$x \in \mathcal{R}(GH) \leftrightarrow G^{-1}x \in \mathcal{R}(H).$$

Therefore by induction starting with $H = I$, we see that $\mathcal{R}(B)$ is primitive recursive.

In the diophantine case, all the functions of (7) with T^* excluded are diophantine. Hence if B is obtained from them by composition, $\mathcal{R}(B)$ is diophantine. If M is strictly monotone, then

$$x \in \mathcal{C}\mathcal{R}(M) \leftrightarrow \bigvee_{y} (My < x < M(y + 1)), \tag{8}$$

$$x \in \mathcal{R}(M^*) \leftrightarrow \bigvee_{y} (My = Lx), \tag{9}$$

$$x \in \mathcal{C}\mathcal{R}(M^*) \leftrightarrow \bigvee_{y} (My < Lx < M(y + 1)). \tag{10}$$

If G is a univalent function, then

$$\mathcal{C}\mathcal{R}(GH) = \mathcal{R}(G \restriction \mathcal{C}\mathcal{R}(H)) \cup \mathcal{C}\mathcal{R}(G). \tag{11}$$

Suppose $\mathcal{R}(H)$ and $\mathcal{C}\mathcal{R}(H)$ are diophantine, and G is one of the functions S^*, D^*, $J(I, I)$, and $J(I, I)^*$. Then $\mathcal{C}\mathcal{R}(G)$ is diophantine by (8) or (10). Hence by (11),

$$x \in \mathcal{CR}(GH) \leftrightarrow \bigvee_{y} ((y \in \mathcal{CR}(H) \wedge Gy = x) \vee x \in \mathcal{CR}(G)),$$

so $\mathcal{CR}(GH)$ is diophantine. If G is one of the permutations of (7), then

$$x \in \mathcal{CR}(GH) \leftrightarrow \bigvee_{y} (Gy = x \wedge y \in \mathcal{CR}(H)).$$

Hence by induction $\mathcal{CR}(B)$ is diophantine.

References

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

2. Julia Robinson, *General recursive functions*, Proc. Amer. Math. Soc. **1** (1950), 703–718.

3. ———, *Finitely generated classes of sets*, Proc. Amer. Math. Soc. (to appear).

4. Raphael M. Robinson, *Primitive recusive functions*. II, Proc. Amer. Math. Soc. **6** (1955), 663–666.

University of California, Berkeley