The halting problem is decidable on a set of asymptotic probability one

Joel David Hamkins
The City University of New York
http://jdh.hamkins.org

Alexei Miasnikov
The City University of New York
http://www.cs.gc.cuny.edu/∼amyasnikov

February 21, 2005

Abstract. The halting problem for Turing machines is decidable on a set of asymptotic probability one. The proof is sensitive to the particular computational model.

The halting problem for Turing machines is perhaps the canonical undecidable set. Nevertheless, we prove that there is an algorithm deciding almost all instances of it. The halting problem is therefore among the growing collection of those exhibiting the "black hole" phenomenon of complexity theory, by which the difficulty of an unfeasible or undecidable problem is confined to a very small region, a black hole, outside of which the problem is easy.

We use the most natural method for measuring the size of a set of Turing machine programs, namely, that of *asymptotic density.* The asymptotic density or probability of a set B of Turing machine programs is the limit of

MSC: 03D10; 68Q17. Keywords: Turing machines, halting problem, decidability. The first author is affiliated with the College of Staten Island of CUNY and The CUNY Graduate Center, and his research has been supported by grants from the Research Foundation of CUNY. The second author is affiliated with The City College of New York and The CUNY Graduate Center.

the proportion of all n-state programs in B as n increases. That is, if P_n is the set of all n-state programs, then the asymptotic probability of B is

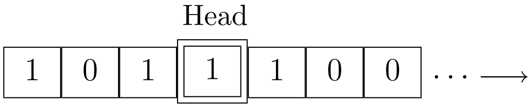
$$\mu(B) = \lim_{n\to\infty} \frac{|B \cap P_n|}{|P_n|},$$

if this limit exists. If B has asymptotic probability one, for example, then for sufficiently large n, more than 99% of all n-state programs are in B, and so on as close to 100% as desired.

Main Theorem 1 *There is a set B of Turing machine programs such that*

1. *B has asymptotic probability one.*
2. *B is polynomial time decidable.*
3. *The halting problem $H \cap B$ is polynomial time decidable.*

The proof is sensitive to the particular computational model. We use the Turing machine model with a finite program directing the operation of a head reading and writing 0s and 1s while moving on a one-way infinite tape.



The Turing machine has n states $Q = \{\, q_1, \ldots, q_n \,\}$, with q_1 designated as the *start* state, plus a separate designated *halt* state, which is not counted as one of the n states. A Turing machine *program* is a function

$$p : Q \times \{\, 0, 1 \,\} \to (Q \cup \{\, \mathit{halt} \,\}) \times \{\, 0, 1 \,\} \times \{\, L, R \,\}.$$

The transition $p(q, i) = \langle r, j, R \rangle$, for example, directs that when the head is in state q reading symbol i, it should change to state r, write symbol j, and move one cell to the right. The computation of a program proceeds by iteratively performing the instructions of such transition rules, halting when the *halt* state is reached. If the machine attempts to move left from the leftmost cell, then the head falls off the tape and all computation ceases. Since the domain of the program has size $2n$ and the target space has size $4(n+1)$, we can easily count the number of programs:

Lemma 1.1 *The number of n-state Turing machine programs is* $(4(n{+}1))^{2n}$.

As a warm-up exercise, let us calculate the asymptotic probability of the set of programs having no transition reaching the *halt* state. Such a criterion is clearly linear time decidable, and no computation by such a program can ever reach the *halt* state.

Lemma 1.2 *The collection of programs having no transition reaching the halt state has asymptotic probability* $1/e^2$*, which is about* 13.5%*.*

Proof: If p has no transition reaching the *halt* state, then $p : Q \times \{\,0,1\,\} \to Q \times \{\,0,1\,\} \times \{\,L,R\,\}$. Since this target set has size $4n$, the total number of such functions is $(4n)^{2n}$. The asymptotic proportion of all n-state programs with this property is therefore

$$\lim_{n\to\infty} \frac{(4n)^{2n}}{(4(n+1))^{2n}} = \lim_{n\to\infty} \left(\frac{n}{n+1}\right)^{2n} = \lim_{n\to\infty} \left[(1-\frac{1}{n+1})^n\right]^2 = 1/e^2.$$

Therefore, the asymptotic probability that a Turing machine program does not engage the *halt* state is $1/e^2$. □

Definition 2 The *halting problem* is the set H of programs p that halt when computing on input 0, on a tape initially filled with 0s.

For the purposes of defining the halting problem H, one should specify whether it officially counts as halting or not, if the head should happen to fall off the left edge of the tape. For definiteness, let us regard such computations as having not officially halted, as the *halt* state was not reached. Thus, we regard H as the set of programs that eventually reach the *halt* state from an initially empty tape. To be even more specific, if the head falls off the tape while executing the transition $p(q,i) = \langle r,j,L\rangle$, then we do not regard the state r as having been achieved, since this step was not completed.

Proof of Main Theorem: We now prove the Main Theorem. Let B be the set of programs that on input 0 either halt before repeating a state or fall off the tape before repeating a state. Clearly, B is polynomial time decidable, since we need only run a program p for at most n steps, where n is the number of states in p, to determine whether or not it is in B. It is equally clear that the halting problem is polynomial time decidable for programs p

in B, since again we need only simulate p for n steps to know whether it halted or fell off. What remains is to prove that this behavior occurs with asymptotic probability one.

Lemma 1.3 *For any fixed input and fixed $k \geq 0$, the set of programs not repeating states within the first k steps has asymptotic probability one.*

Proof: Just to be clear, we count a computation that halts or falls off the tape as satisfying the property, if it does so before repeating a state. We calculate for large n the proportion of all n-state programs having this property, by induction on k. When $k = 0$, then all programs have the property. Suppose that the set B_k of programs having the desired property for k has asymptotic probability one, and consider B_{k+1}. Fix any ϵ, and choose n large enough so that B_k has proportion more than $1 - \epsilon/2$ of all n-state programs. Among all n-state programs p in B_k, consider the probability that p is in B_{k+1}. If p leads to a computation where the head has already fallen off the tape, then of course $p \in B_{k+1}$. Otherwise, the first k steps of computation by p have led to the successive states $q_{i_0}, q_{i_1}, \ldots, q_{i_k}$, which have not yet repeated. The $(k+1)^{\text{th}}$ step of computation involves a transition rule $p(q_{i_k}, j_k) = (q_{i_{k+1}}, j_{k+1}, m_k)$, giving respectively the new state, the new bit to write on the tape and the direction to move the head. In order for p to be in B_{k+1}, it suffices that $q_{i_{k+1}}$ must not be one of the previously used states $\{\, q_{i_0}, q_{i_1}, \ldots, q_{i_k} \,\}$. Since there are $(n + 1) - (k + 1) = n - k$ many other equally likely states to choose from, the proportion of all n-state programs agreeing with p on the first k steps and satisfying the additional requirement is $\frac{n-k}{n}$. The proportion of all n-state programs in B_{k+1}, consequently, is at least $(1 - \epsilon/2)(\frac{n-k}{n})$. Since $\frac{n-k}{n}$ goes to 1 as n becomes large, we may choose n large enough so that this proportion is at least $1 - \epsilon$. Thus, B_{k+1} has asymptotic probability one, as desired. □

Proceeding with the main argument, let B_k be the set of programs that do not repeat a state within their first k steps of computation. The key idea is that for the first k steps of computation, the programs in B_k behave statistically like a random walk with uniform probability of going left or right. The reason is that if a program lands in a totally new state q, reading some symbol i, then among the programs landing in that situation and agreeing with the computation so far, exactly half of them will opt to move left and half will move right, precisely because nothing about state q has yet been determined. Because of this, we may make use of Polya's classical result on random walks, which we mention without proof.

Lemma 1.4 (Polya [**?**], see also e.g. [**?**]) *In the random walk with equal likelihood of moving left or right on a one-way infinite tape, beginning on the left-most cell, the probability of eventually falling off the left edge is* 1.

This is the famous *recurrence* phenomenon, because it asserts that such a random walk has probability one of eventually returning to its starting point. It follows that with probability one the random walk reaches any given fixed position of the tape. Interestingly, the recurrence property holds for random walks in dimensions one and two, but not in dimensions three or higher.

Putting everything together, let us show that B has asymptotic probability one. Fix any $\epsilon > 0$, and by Lemma 1.4 find some large k such that with probability exceeding $\sqrt{1-\epsilon}$, the k-step random walk falls off the left edge of the tape. By Lemma 1.3, let n be large enough so that B_k contains more than the proportion $\sqrt{1-\epsilon}$ of all n-state programs. Combining these facts with the observation that programs in B_k operate statistically like random walks for their first k steps of computation (or until they halt, if this is sooner), as far as the head position is concerned, we conclude that proportion at least $(\sqrt{1-\epsilon})^2 = 1-\epsilon$ of all n-state programs exhibit the desired property. So the set B of all such programs has asymptotic probability one, and the theorem is proved. □

Let us now clarify matters by untangling the two possibilities for programs in B, namely, (1) the programs that halt before repeating a state and (2) the programs that fall off the tape before repeating a state. The fact is that behavior (1) is very rare and behavior (2) occurs with asymptotic probability one.

Theorem 3 *The asymptotic probability one behavior of Turing machine programs, on fixed input, is that the head falls off the tape before repeating a state.*

Proof: First, we generalize Lemma 1.3 to exclude the possibility of halting.

Lemma 3.1 *For any fixed input and fixed $k \geq 0$, the set of programs not repeating states and not halting within the first k steps has asymptotic probability one.*

Proof: Let C_k be the desired set of programs, which includes the programs that fall off the tape within the first k steps of computation, provided that they do so before repeating a state. As in Lemma 1.3, we show inductively that C_k has asymptotic density one. When $k = 0$, this is trivial. Let us now calculate the probability that a program p is in C_{k+1}, given that it is in C_k. If the head fell off within k steps, then p will also be in C_{k+1}. Otherwise, as in Lemma 1.3, the first k steps of computation exhibit states $q_{i_0}, q_{i_1}, \ldots, q_{i_k}$, which have not yet repeated. The $(k+1)^{\text{th}}$ step of computation involves a transition $p(q_{i_k}, j_k) = (q_{i_{k+1}}, j_{k+1}, m_k)$, which will place p into C_{k+1} if $q_{i_{k+1}}$ is a new state and not the *halt* state. Since there are $n - (k+1)$ remaining states to choose from, the probability that p will be in C_{k+1} is at least $\frac{n-(k+1)}{n}$. Since this probability goes to 1 as n goes to infinity, we conclude that C_{k+1} has asymptotic probability one. □

Now fix any $\epsilon > 0$. Select k large enough so that the random walk in k steps has probability exceeding $\sqrt{1-\epsilon}$ of falling off the left edge. By Lemma 3.1, take n sufficiently large so that the proportion of all n-state programs that do not halt in k steps and do not repeat a state in k-steps is at least $\sqrt{1-\epsilon}$. Thus, as in the Main Theorem, these computations behave statistically like random walks, as far as the head position is concerned, and so the proportion of all n-state machines that fall off the tape in k steps before repeating a state or halting is at least $\sqrt{1-\epsilon}\sqrt{1-\epsilon} = 1 - \epsilon$, as desired. □

Corollary 4 *The halting problem H has asymptotic probability zero. And the complement of H contains a decidable set of asymptotic probability one.*

Proof: If the head falls off the tape, then the computation cannot reach the *halt* state, and so the program is not in H. So H has density zero. By Theorem 3, the set of programs that fall off the tape before repeating a state is clearly polynomial time decidable and has asymptotic probability one. □

The previous Corollary is sensitive the formalism that computations for which the head falls off the tape are not counted as halting. If one wishes instead to count them as halting, then the conclusion would be that the corresponding version of H would have asymptotic probability one and contain a decidable set of asymptotic probability one.

Because the computational behavior identified in Theorem 3, with the head falling off the tape before a state is repeated, is both typical and trivial, many other well-known undecidability problems for Turing machines can also be decided almost everywhere. We give two examples.

Definition 5 Let FIN be the set of programs computing functions with finite domain and COF be the set of programs with cofinite domain. These sets are well known to be undecidable (see e.g. [**?**]).

Corollary 6 *There is a set of programs B such that:*

1. *B has asymptotic probability one.*
2. *B is polynomial time decidable.*
3. *$FIN \cap B$ is polynomial time decidable.*
4. *$COF \cap B$ is polynomial time decidable.*

Proof: For the purposes of computing functions, we assume that input on a Turing machine tape is given by a string of 1s, that is, in unary form. Let B be the set of programs that fall off the tape before halting or repeating a state, on a tape initially filled entirely with 1s. This set is clearly polynomial time decidable, and by Theorem 3, it has asymptotic probability one. But any program that falls off the tape will have had a chance to inspect only finitely many of the 1s on the tape before doing so, and so the program will have this same behavior provided that there is a sufficiently long string of 1s on the tape as input. So every program in B is in FIN and none are in COF. On B, therefore, these questions are decidable. □

The proof of Corollary 6 shows that almost every program computes a finite function. In other words, FIN has asymptotic probability one and COF has asymptotic probability zero. If one takes the domains of the computable functions as the natural enumeration of the computably enumerable (c.e.) sets, then this means that almost every c.e. set is finite.

Let us turn now to the question of whether the conclusions of the main theorem hold for other models of computability.

Corollary 7 *The conclusion of the main theorem also holds for the following models of computability:*

1. *Single tape Turing machines with an arbitrary finite alphabet, operating on a one-way infinite tape.*

2. *Multi-tape Turing machines with an arbitrary finite alphabet, operating on one-way infinite tapes.*

3. *Turing machines with a head moving on a half-plane or quarter-plane grid of cells with an arbitrary finite alphabet.*

Proof: For the multi-tape model of (2), we assume that there is a single head moving back and forth, reading and writing on all columns at once. The corollary is proved merely by observing that the calculations of Lemma 1.3 do not fundamentally rely on the size of the alphabet, and so in the case of a general alphabet, it is still true that for any k the set of programs that adopt new states for their first k moves has asymptotic probability one. Because these programs therefore act like a random walk for the first k steps, the probability that they fall off the left end of the tape can be made as close to 1 as possible. So (1) holds. The multi-tape model of (2) is functionally equivalent to having a larger alphabet, if one regards a entire column of cell values as a single element of a larger alphabet. So (2) holds. For (3), we observe that first, the analogue of Lemma 1.3 remains true, and second, the desired conclusion now follows from the two-dimensional generalization of Lemma 1.4, by which random walks on a the half-plane (or any smaller portion of the plane), eventually fall off the edge with probability one. □

The reader will have already observed, of course, that our argument does *not* work with Turing machines using doubly-infinite tapes, where there is no possibility of the head falling off. And neither does it work with the one-way infinite tape models that allow computation somehow to continue after attempting to move left from the left-most cell. We admit that this situation is unsatisfactory, because one doesn't like results in computability theory to be sensitive to the choice of computational model.

Question 8 *Does the conclusion of the Main Theorem hold for all models of Turing machine computation?*

For this question the focus, of course, is on the models with two-way infinite tapes. One can weaken the desired conclusion by asking only that the halting problem be decided on a set of large probability, rather than probability one.

If one weakens this too much, by asking only to decide the problem on a set of nonzero probability, then it becomes a trivial consequence of Lemma 1.2.

Theorem 9 *For any model of Turing machine, including those with two-way infinite tapes, there is a set B of Turing machine programs such that*

1. *B has nonzero asymptotic probability.*
2. *B is polynomial time decidable.*
3. *The halting problem $H \cap B$ is polynomial time decidable.*

Proof: The set of programs arising in Lemma 1.2, which have no transition leading to the *halt* state, has asymptotic probability $1/e^2$; but clearly no such program is in H. □

Question 10 *In Theorem 9, how large can the probability of B be? Can one always decide the halting problem in an asymptotic majority of cases?*

tape Turing machine model, where if nonzero number less than one. We can not rely on whether the tapes are asymptotic density at least 13.5% on

We close the paper with a few elementary cautionary observations. First, the relation of Turing equivalence does not respect asymptotic density.

Theorem 11 *The relation of Turing equivalence does not respect the property of having asymptotic probability one in the natural numbers. Indeed, for any set A of natural numbers there is a set B that is Turing equivalent to A and has any prescribed asymptotic density (or nonconvergent density).*

Proof: If a set A has asymptotic density one in the natural numbers, then the complement of A, which is Turing equivalent to A, has asymptotic density zero. But also, any set A is Turing equivalent to a set with asymptotic density zero, by simply multiplying its second member by 2, it's third member by 3, and so on, so as to stretch it out to density zero. The complement of this set, which is also Turing equivalent, has asymptotic density one. Intermediate densities can be achieved by adding regular blocks of numbers in a computable pattern, so as to achieve a given intermediate density, while the true information is coded on a thin set of density zero. By alternating blocks of numbers with large empty stretches, one can arrange that the asymptotic density of the set does not converge, and even that the upper density is 1 while the lower density is 0. Meanwhile, the true information of the set is coded on a thin set, of density zero, which does not upset those calculations. □

Second, the notion of what happens "almost everywhere" can be highly sensitive to what are otherwise unimportant differences in formalism. For example, if one takes as the basic model of computability a suitable generalization of C++ programs, then most would agree that for the usual purposes it is an irrelevant formalism whether one excludes programs at the outset that have syntax errors preventing them from compiling, or instead takes them to compute the empty function. But if they were officially counted, then because clearly there are far more programs with errors than without, it would mean that almost every program would be trivial in this way. For such a model, all interesting phenomena would occur on a set of asymptotic density zero.