Proc. Natl. Acad. Sci. USA
Vol. 95, pp. 95–97, January 1998
Computer Sciences

Limit, logic, and computation

MICHAEL H. FREEDMAN

Microsoft Research 9N, 1 Microsoft Way, Redmond, WA 48052

Contributed by Michael H. Freedman, October 23, 1997

ABSTRACT **We introduce "ultrafilter limits" into the classical Turing model of computation and develop a paradigm for interpreting the problem of distinguishing the class *P* from *NP* as a logical problem of decidability. We use *P*(*NP*) to denote decision problems which can be solved on a (nondeterministic) Turing machine in polynomial time. The concept is that in an appropriate limit it may be possible to prove that problems in *P* are still decidable, so a problem whose limit is undecidable would be established as lying outside of *P*.**

The Turing machine T represents an abstraction of the principles of mechanical computation. The machine consists of a head and a tape. The head is capable of being in one of a finite number of "internal states" $\{q_i\}$ and can read and overwrite a symbol $\in \{S_j\}$ from a finite set of symbols and then shift one block left or right along the tape. It contains a finite internal program which directs its operations. At any time the *complete state* of T is the record on the tape together with the internal state. Consider a problem Q, with a yes/no answer, for which infinitely many instances exist, for example, the satisfiability of Boolean formulae. The decision problem Q is said to lie in class P if there is an internal program which will correctly answer all instances I of Q "yes" ("no") by halting on symbol (0) after a number of operations which is some fixed polynomial function in the number of bits of the input. One says Q lies in NP (nondeterministic polynomial time) if there is an "existential" program operating on I plus a number of "guess bits" which correctly answers all instances I in polynomial time. The existential program is deemed to answer "yes" if for some setting of the guess bits the machine halts on 1.

Since the *P*/*NP* problem has at its core the distinction between polynomial and exponential growth, it is natural to look for perspective to other models within mathematics where this dichotomy is manifest. In complex analysis an exponential function, e.g., 2^z, has an essential singularity at infinity, in contrast to the continuous branched structure at infinity exhibited by a polynomial. This dichotomy is mirrored in cardinal arithmetic (1) where the function 2^x is discontinuous at every limit cardinal α, for which no smaller cardinal β has a power set $P(\beta)$ equinumerous with $P(\alpha)$, that is,

$$\lim_{x<\alpha}(2^x) \not\cong 2^\alpha = 2^{\lim_{x<\alpha}(x)}.$$

This last fact, for $\alpha = \aleph_0$, influenced Sipser (2) in his thinking that the distinction between analytic (projections of Borel) sets and coanalytic sets (the complement of an analytic set) in analysis might provide a tool for distinguishing *NP* sets (sets accepted by *NP*-time Turing machines) from co-*NP* sets (the complement of *NP* sets).

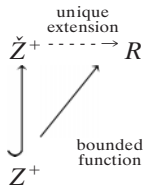
The theory of group presentations may be taken as an analog of computation. Milnor (3) and Schwarzc (4) introduced the notion of the growth of a finitely generated group G. The group G has *polynomial growth* (exponential growth), if it has a presentation in which the number of distinct elements of G which can be written as words of length $= \ell$ in the generators and their inverses is $\leq P(\ell)$ for some polynomial P ($\geq b^\ell$ for some base $b > 1$). It is easy to show that both these properties are in fact independent of the presentation and depend only on the group G.

Within this theory we will explain how taking the appropriate limit transforms a distinction in growth rates into a dimensional dichotomy. A celebrated theorem of Gromov's (5) states that G has polynomial growth iff it contains a nilpotent subgroup of finite index. The proof considers a sequence of base-pointed metric spaces $\{(G_\varepsilon, id)\}$, $\varepsilon \to 0$, where G_ε has metric $\text{dist}(g_1, g_2) = \varepsilon$ (minimum word length $(g_1 g_2^{-1})$). This sequence has a convergent subsequence, in the Gromov-Hausdorff sense, if and only if G has polynomial growth, in which case the limiting metric space (Y, y_0) is *finite*-dimensional.* The proof proceeds by representing G into isometries (Y), which is a Lie group by the Montgomery-Zippen theorem. Ultimately, the limit Y is seen to be a nilpotent Lie group endowed with a Carno-metric. For example, if $G =$ integers Z, then Y is the real line R. If $G = Z^n$, then $Y = R^n$. If G is the discrete Heisenberg group

$$\left\{ \begin{vmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{vmatrix}, \quad x, y, z \text{ integers} \right\},$$

then Y is the continuous Heisenberg group with x, y, and z real.

If G has faster than polynomial growth, $\{G_\varepsilon, id\}$ will not approach a limit in the Gromov-Hausdorff sense, but an ultrafilter limit can be forcibly extracted. Let $(X_i, *_i)$ be any sequence of pointed metric spaces $i = 1, 2, 3, \ldots$. Consider *admissible* sequences $\{x_i \in X|$, there exists a constant c so that $\text{dist}(*_i, x_i) < ic$. Let ω be any non-principle ultrafilter in $\check{Z}^+$, i.e., ω belongs to the growth $\check{Z}^+ \backslash Z^+$ in the Stone-Cĕch compactification $\check{Z}^+$ of Z^+. Using the universal property



,

Gromov (6) defines a unique real number

$$d(\{x_i\}, \{x_i'\}) = \frac{\text{dist}(x_i, x_i')}{i}\,[\omega]$$

which induces a pseudo-metric on the admissible sequences. Dividing by the equivalence relation—points with distance =

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

*Convergent can be taken to mean that there is a pointed metric space (Y, y_0), so that for every radius r there exists an $\varepsilon > 0$, so that the ball of radius r about id in G_ε, $(B_r(id \subset G_\varepsilon), id)$ imbeds into (Y, y_0) $(1 + 1/r)$-quasi-isometrically (i.e., with metric distortion between factors of $(1 + 1/r)$ and $(1 - 1/r)$), and that every point of Y is in the closure of the images of these balls.

0 are equivalent—yields a metric space X, which we call the ω-limit of (X_i). It has been conjectured that the homeomorphism type of the ω-limit is independent of the choice of non-principle ultrafilter ω.

If the sequence $((X_i, \frac{1}{i}\text{dist}_i), *_i)$ is convergent in the Gromov-Hausdorff sense, then this limit is also the ω-limit; however, the ω-limit exists in complete generality. For example, when applied to the constant sequence $\{G, id\}$, G a word-hyperbolic group (the generic case for groups of nonpolynomial growth; ref. 6), then the ω-limit is an $\mathbb{R}$-tree (a space in which there is a unique imbedded interval joining every two points). Although of covering dimension one, this $\mathbb{R}$-tree is enormously large, in the sense that there is no countable basis for its topology and its Hausdorff dimension is *infinite*.

The paradigm: "polynomial growth implies a well-behaved limit," if applied to the *P/NP* problem, would take the schematic following form:

> A polynomial time algorithm T solving a finite-decision problem Q should "converge" to some "continuous procedure" for solving an infinitary version of Q, whereas an exponential-time algorithm should not be expected to have any sensible limit.

Applications of Paradigm

There is a toy model of computation, the search of a database, in which this paradigm applies. Consider the databases consisting of the positive orthant of Z^n and W, where Z^n is the integer lattice in Euclidean n-space and W is the universal unrooted 3-valent tree with edges of length = 1. (W could also be taken to be a co-compact lattice in any hyperbolic space H^n, $n \geq 2$, and all the following assertions would remain true but be slightly more technical to check.) Writing each integer $r \in Z$ in base 2, the kth component $f_k(r)$ of a map $f: Z^+ \cup 0 \to Z^n$ is defined by reading only the digits congruent to $k \bmod n$. Now fix any non-principle ultrafilter $\omega \in \check{Z}^+$. Regarding $Z^+ \cup 0$ as a sequence of spaces where the jth copy has its standard metric multiplied by $(j)^{-(n-1)}$, and regarding Z^n as the constant sequence of spaces, a Hölder-$1/n$ continuous ω-limit $\bar{f}: R^+ \cup 0 \to R^n$ is obtained from applying the limit construction to domain and range simultaneously. The map f may be interpreted as a particularly efficient† search of the positive orthant of Z^n; the rescaling of Z amounts to a speed-up of the search, so that the ball of radius j in the jth Z^n is searched in time proportional to j. Finally, $\bar{f}$ is the limiting solution to the infinitary version of the search problem in which all points in the positive orthant of R^n must be visited. The map $\bar{f}$ is the Peano-Hilbert curve.

Turn now to ω-limit (W, w_0) where some vertex of W has been chosen as basepoint. There are $2^{\aleph_0}$ edge paths leaving w_0 and heading toward infinity. These define an uncountable set of sequences $\{w_{i,j}\}$, $i \in Z^+$, $j \in 2^{\aleph_0}$, whose mutual ω-distances $d(\{w_{i,j}\}, j, \{w_{i,j'}\}) = 2$ are all two. This implies that the ω-limit $\bar{W} = \bar{W}$ has no countable basis for its topology. Consequently, $\bar{W}$ is not equal to the image under any continuous map of any second countable space, e.g., $R^+ \cup 0$. Thus no discrete search of W can be constructed so that a rescaled limit leads to a continuous search (i.e., epimorphism) of $\bar{W}$. In these models we see "polynomial time" converging to continuous and "exponential time," failing to define an appropriate limit, echoing the observations in complex analysis and cardinal arithmetic.

Let N_k be the set of Boolean formulae which are conjuctions of k-fold disjuncts of a finite alphabet of literals; k-sat denotes the satisfaction problem for fomuli in N_k. It is well known (7) that 2-sat lies in P, whereas 3-sat is NP-complete. In "k-sat on Groups and Undecidability" (unpublished work), an infinitary version of k-sat is introduced which depends on a fixed infinite group. Truth assignments for group elements are sought subject to a family of disjunctive clauses closed under right multiplication by a finite index subgroup $H \subset G$. It is shown that for this extension of the satisfaction problem for $G \cong Z \oplus Z$, 2-sat remains decidable while 3-sat becomes undecidable in ZFC. While supporting the paradigm, the proof does not argue on the basis of the inclusion 2-sat $\subset P$ and therefore does not immediately generalize. In view of the first two examples, w-limits seem to be a promising general approach to constructing decidable‡ limits of problems in P.

What follows are two speculations on how the introduction of ω-limits could have a role in distinguishing P from NP. The sketched arguments should be read as design criteria for limits that we do not know how to construct. The first is based on the preservation of connectivity under a continuous map. The second searches for a bridge to Gödel's incompleteness theorem. Fix a non-principal ultrafilter $\omega \in \check{Z}^+$ and let $\mathbb{F}$ be the set of finite Boolean formulae on an infinite alphabet organized into a pointed metric space, or perhaps some weaker structure, in some manner which we do not yet know how to specify. Let $\bar{\mathbb{F}}$ be an ω-limit or some similar limit of $\mathbb{F}$. For the argument-plan we propose, the definition of the structure and the details of the limit taken must be such that $\bar{\mathbb{F}}$ is connected. However, it is necessary that if we restrict to $\mathbb{F}' \subset \mathbb{F}$, a class of formulae which can be checked for satisfiability in poly-time, the limit yields a disconnected space $\bar{\mathbb{F}}'$. We suppose that the ω-satisfiability (1) or unsatisfiability (0) of $\bar{f} = \{f_i\} \in \bar{\mathbb{F}}$ can be defined by applying ω to the satisfiability of each f_i in a sequence defining $\bar{f}$. One can imagine that a polynomial time algorithm T could be rewritten "efficiently"—as in our choice of scanning function f into the positive orthant of Z^n—so that it would converge to a "continuous decision procedure" $\bar{T}: \bar{\mathbb{F}} \to \{0, 1\}$, contradicting the connectivity of $\mathbb{F}$. The picture is that T would evolve $\bar{f} \in \bar{\mathbb{F}}$ through a succession (variable t) of complete state sequences (variable i) $\{S_{i,t}\}$ defining a path with parameter $t \in [0, 1]$ in S, the ω-limit of the discrete space of sequences of complete states $\{S_i\}$. Thus T would define a homotopy $\bar{T}: \bar{\mathbb{F}} \times I \to S$ whose end $\bar{T}(F \times 1)$ must be disconnected according to yes/no on ω-satisfiability. This would contradict the topological connectivity of $\bar{\mathbb{F}}$.

The first-order theory A of arithmetic is known to contain weak fragments A^- for which there exists a decision procedure (carried out within first-order arithmetic, say by a Turing machine) for the provability in A of statements in A^-. The best known example (and a high tide mark of Hilbert's program to axiomatize mathematics) is Presberger Arithmetic PrA (8), which is essentially Peano arithmetic§ absent multiplication. Without multiplication, indexing of formulae cannot be achieved, so Presberger Arithmetic escapes Gödel's incompleteness theorem. It seems plausible that a suitable ultrafilter limit could resurrect multiplication, since multiplication by any *fixed* integer is explicitly expressible as a repeated addition. Thus one might have a schematic formula on the level of ω-limits:

$$\omega - PrA \equiv \omega - A.$$

In A, the Gödel sentence "there exists an integer x_0 which codes for a proof of $0 = 1$" requires only a single unbounded existential quantifier. Suppose that we can construct a fragment of arithmetic A^- in which (i) the problem Q, of deciding

†The function $\bar{f}$ is efficient in the sense that from any initial point $f(i_0)$, the sequence $f(i_0), \ldots, f(i_0 + c_n r^n)$ will completely explore the ball of radius r in Z^n about $f(i_0)$ for some constant c_n depending only on the dimension n.

‡Perhaps "decidable" should be replaced here with some notion "continuously decidable" as discrete proofs limit to continuous objects.

§This denotes the standard axiomatization of arithmetic which features the rules for addition, multiplication, and mathematical induction.

(in A) the validity of sentences of A^- with only one unbound existential quantifier, lies in NP and (*ii*) $\omega - A^- \equiv \omega - A$. With regard to (*i*) we note that ref. 9 proves that PA is a bit too strong a fragment; a nondeterministic Turing machine must run at least $\geq 2^{\ell^c}$, for some $c > 0$, to decide such sentences of length $= \ell$. One may have to look to systems as weak as A^- = quantified Boolean formulae to achieve this condition. Note that a Boolean formula can be written to specify the ith bit of multiplication so some aspect of arithmetic is retained even at this level. The paradigm that things polynomial have well behaved limits would then suggest that a polynomial-time algorithm for Q would yield a decision procedure (suitably interpreted) for "ω-sentences" in $\omega - A^-$ with a single unbounded quantifier. By (*ii*) such ω-sentences would include Gödel-like ω-sentences and hence be undecidable. Such a contradiction would show P to be strictly smaller than NP. The philosophy is that within an appropriate limit, quick should become decidable, whereas slow may become undecidable.

I would like to express my thanks to Sam Buss and Hugh Woodin for stimulating conversations on the material presented. This work was partially supported by National Science Foundation Grant DMS 9501105, by Microsoft Research, and by the University of California, San Diego.

1. Kunen, K. (1980) *Set Theory: An Introduction to Independence Proofs* (North-Holland, Amsterdam), pp. 27–35.
2. Sipser, M. (1985) in *Theory of Algorithms*, eds. Lovasz, L. & Szemeredi, E., Colloq. Math. Soc. János Bolyai (North-Holland, Amsterdam), Vol. 44, pp. 387–391.
3. Milnor, J. (1968) *J. Diff. Geom.* **2,** 1–7.
4. Schwarzc, A. (1955) *Dokl. Ak. Nauk, USSR* **105,** 32–37.
5. Gromov, M. (1981) *Publications Mathématiques, IHES* **53,** 53–73.
6. Gromov, M. (1994) *Asymptotic Invariants of Infinite Groups*, L. M. S. Lecture Notes Series (London Math Society, London).
7. Hopcraft, J. & Ullman, J. (1974) *Introduction to Automata Theory, Languages, and Computation* (Addison-Wesley, New York).
8. Enderton, H. (1972) *A Mathematical Introduction to Logic* (Academic, New York), p. 188.
9. Furer, M. (1982) *Theor. Comput. Sci.* **18,** 105–111.