POLYNOMIAL HIERARCHY, BETTI NUMBERS AND A REAL ANALOGUE OF TODA'S THEOREM

SAUGATA BASU AND THIERRY ZELL

1. Brief Description

Toda proved in 1989 that the (discrete) ***polynomial time hierarchy,* PH**, is contained in the class $\mathbf{P}^{\#\mathbf{P}}$, namely the class of languages that can be decided by a Turing machine in polynomial time given access to an oracle with the power to compute a function in the ***counting complexity class* #P**. This result which relates the counting complexity class $\#\mathbf{P}$ with the polynomial hierarchy, thereby illustrating the power of counting, is considered to be a seminal result in computational complexity theory.

In the late eighties Blum, Shub and Smale introduced the notion of Turing machines over more general fields, thereby generalizing the classical problems of computational complexity theory such as $\mathbf{P}$ vs $\mathbf{NP}$ to corresponding problems over arbitrary fields (such as the reals complex, p-adic numbers etc.) If one considers languages accepted by a Blum-Shub-Smale machine over a finite field one recovers the classical notions of discrete complexity theory. Over the last two decades there has been a lot of research activity towards proving real as well as complex analogues of well known theorems in discrete complexity theory. The first steps in this direction were taken by the authors Blum, Shub, and Smale (henceforth B-S-S) themselves, when they proved the $\mathbf{NP}_{\mathrm{R}}$-completeness of the problem of deciding whether a real polynomial equation in many variables of degree at most four has a real solution (this is the real analogue of Cook-Levin's theorem that the satisfiability problem is $\mathbf{NP}$-complete in the discrete case), and subsequently through the work of several researchers (Koiran, Bürgisser, Cucker, Meer to name a few) a well-established complexity theory over the reals as well as complex numbers have been built up, which mirrors closely the discrete case. The goal of this paper is to prove a real analogue of Toda's well known theorem.

In order to formulate a real analogue of Toda's theorem it is first necessary to define precisely real counter-parts of the discrete polynomial time hierarchy $\mathbf{PH}$ and the discrete complexity class $\#\mathbf{P}$. The real analogue of the discrete polynomial time hierarchy is easy to define (completely analogous to the discrete case) once we have identified the class $\mathbf{P}_{\mathrm{R}}$ (i.e. the class of languages accepted by a real machine (in the sense of B-S-S) in polynomial time). Note that a "language" over the reals is given by a sequence of semi-algebraic sets $(S_n \subset \mathrm{R}^{k(n)})_{n>0}$, with $k(n)$ some fixed polynomial in n (the semi-algebraic sets S_n play the role of "strings of length $k(n)$" in the discrete theory). (A subset $S \subset \mathrm{R}^n$ is semi-algebraic if it can be described by a Boolean formula whose atoms are a finite number of polynomial equalities and inequalities.) Once the class $\mathbf{P}_{\mathrm{R}}$ has been so identified, the real polynomial hierarchy, $\mathbf{PH}_{\mathrm{R}}$, is defined syntactically in precisely the same way as the discrete polynomial hierarchy.

For technical reasons (explained in the paper) we need to define a *compact version* of the real polynomial hierarchy which we denote by $\mathbf{PH}^{c}_{\mathrm{R}}$. One can think of this class (without losing too much) as the languages in $\mathbf{PH}_{\mathrm{R}}$ whose individual parts $S_n \subset \mathrm{R}^{k(n)}$ are all compact (i.e. closed and bounded) semi-algebraic sets (actually, all the quantified variables in the quantified formula defining the S_n are also required to vary over compact sets). The inclusion $\mathbf{PH}^{c}_{\mathrm{R}} \subset \mathbf{PH}_{\mathrm{R}}$ is a straight-forward consequence of their respective definitions.

The definition of the counting class $\#\mathbf{P}_{\mathrm{R}}$ is more involved (and also more interesting). Before defining the real analogue of the class $\#\mathbf{P}$, let us recall its definition in the discrete case.

Definition 1.1. Let $k = \mathbb{Z}/2\mathbb{Z}$. We say that a sequence of functions $(f_n : k^n \to \mathbb{N})_{n>0}$ is in the class $\#\mathbf{P}$ if there exists a language $L = (L_n \subset k^n)_{n>0}$, with $L \in \mathbf{P}$, as well as a polynomial $m(n)$, such that $f_n(\mathbf{x}) = \mathrm{card}(L_{m+n,\mathbf{x}})$ for each $\mathbf{x} \in k^n$, where $L_{m+n,\mathbf{x}} = L_{m+n} \cap \pi^{-1}(\mathbf{x})$ and $\pi : k^{m+n} \to k^n$ is the projection along the first m co-ordinates.

In other words, f_n *counts the number of points in the fibers,* $L_{m+n,\mathbf{x}}$*, of a language* $L \in \mathbf{P}$. (The geometric language might look unnecessary but it is very helpful towards obtaining the right analogue in the real case).

In order to define real analogues of counting complexity classes of discrete complexity theory, it is thus necessary to identify the proper notion of "counting" in the context of semi-algebraic geometry. Counting complexity classes over the reals have been defined previously by Meer, and studied extensively by other authors namely Burgisser and Cucker. These authors used a straightforward generalization to semi-algebraic sets of counting in the case of finite sets – namely the counting function took the value of the cardinality of a semi-algebraic set if it happened to be finite, and ∞ otherwise. This is in our view not a fully satisfactory generalization since the count gives no information when the semi-algebraic set is infinite, and most interesting semi-algebraic sets have infinite cardinality.

A more mathematically natural notion of "counting" a semi-algebraic set $S \subset \mathrm{R}^k$ is computing a well-studied discrete topological invariant of S called the Euler-Poincaré characteristic (denoted $\chi(S)$). One reason this is natural for mathematicians (a reason we prefer not to emphasize here) is that the Euler-Poincaré characteristic maps a semi-algebraic set to its image in the ***Grothendieck ring of semi-algebraic sets*** which is isomorphic to $\mathbb{Z}$ – and is thus the "right" way of counting semi-algebraic sets from the *motivic* point of view.

Unfortunately, the Euler-Poincaré characteristic fails to distinguish between empty and non-empty semi-algebraic sets, since a non-empty semi-algebraic set (e.g. an odd dimensional sphere) can have have vanishing Euler-Poincaré characteristic. Thus even though it is elegant, the above fact seems to rule out using the Euler-Poincaré characteristic as a substitute for the counting function in real complexity theory.

Fortunately, we can make up for this deficiency rather easily by replacing the Euler-Poincaré characteristic by a more general topological invariant – namely, the whole sequence of ***Betti numbers***, $b_i(S)$, of S, where $b_i(S)$ is the rank of the singular homology group $\mathrm{H}_i(S) = \mathrm{H}_i(S, \mathbb{Z}))$ of S. We refrain here from defining the homology groups of S, other than noting that $b_0(S)$ is the number of connected components of S and for compact S, the Euler-Poincaré characteristic is given by the alternating sum $\chi(S) = \sum_{i \geq 0} (-1)^i b_i(S)$. In order to be more succinct we let $P_S \in \mathbb{Z}[T]$ denote the ***Poincaré polynomial*** of S, namely

$$P_S(T) \overset{\text{def}}{=} \sum_{i \geq 0} b_i(S)\, T^i. \tag{1.1}$$

Notice that for $S \subset \mathrm{R}^k$, $\deg(P_S) \leq k-1$. Also, it is easy to see that the Poincaré polynomial, $P_S(T)$, carries more complete information about S than its Euler-Poincaré characteristic. Indeed, the number of semi-algebraically connected components, $b_0(S)$, of S is obtained by setting T to 0, and in case S is compact we also recover $\chi(S)$ by setting T to -1 in $P_S(T)$. Since $b_0(S) > 0$ if and only if S is non-empty, P_S, unlike $\chi(S)$, can distinguish between empty and non-empty semi-algebraic sets. In particular, in case S is a finite set of points, P_S also contains the information regarding the cardinality of S which in this case equals $b_0(S) = P_S(0)$.

Definition 1.2 (The class $\#\mathbf{P}^{\dagger}_{\mathrm{R}}$)**.** We say a sequence of functions $(f_n : \mathrm{R}^n \to \mathbb{Z}[T])_{n>0}$ is in the class $\#\mathbf{P}^{\dagger}_{\mathrm{R}}$, if there exists a language $S = (S_n \subset \mathrm{R}^n)_{n>0}$, with $S \in \mathbf{P}_{\mathrm{R}}$, as well as a polynomial $m(n)$, such that $f_n(\mathbf{x}) = P_{S_{m+n,\mathbf{x}}}$ for each $\mathbf{x} \in \mathrm{R}^n$, where $S_{m+n,\mathbf{x}} = S_{m+n} \cap \pi^{-1}(\mathbf{x})$ and $\pi : \mathrm{R}^{m+n} \to \mathrm{R}^n$ is the projection along the first m co-ordinates. (We denote our class $\#\mathbf{P}^{\dagger}_{\mathrm{R}}$ to avoid any possible confusion with previous authors' work.)

Notice the formal similarity between Definitions 1.2 and 1.1, namely that in both cases the functions f_n *counts the fibers above* $\mathbf{x}$, but the notion of counting is different in each case.

We can now state the main result of this paper.

Theorem 1.3 (Real analogue of Toda's theorem)**.**

$$\mathbf{PH}^c_{\mathrm{R}} \subset \mathbf{P}_{\mathrm{R}}^{\#\mathbf{P}^{\dagger}_{\mathrm{R}}}.$$

We remark that while it is quite easy (indeed trivial) to prove the inclusions of the (compact versions) of the classes $\mathbf{NP}_{\mathrm{R}}$ and $\mathbf{coNP}_{\mathrm{R}}$ in $\mathbf{P}_{\mathrm{R}}^{\#\mathbf{P}_{\mathrm{R}}^{\dagger}}$, the inclusions of the higher levels are not obvious at all and require more sophisticated tools from algebraic topology. In fact we find the fact that such an inclusion can be proved at all quite unexpected. The situation in this regard is very similar to the proof of Toda's original result, where the proof for the higher levels of the polynomial hierarchy is also rather intricate. However, Toda's proof is combinatorial in nature, and there is no direct way of extending such a proof to real complexity classes. The proof of Theorem 1.3 above proceeds along completely different lines and is fundamentally topological in nature. We refer the reader to Section 1.4 of the main paper for a rough outline of the argument, and to Section 4 for the full proof.

Department of Mathematics, Purdue University, West Lafayette, IN 47906, U.S.A.
E-mail address: `sbasu@math.purdue.edu`

School of Mathematics and Computing Sciences, Lenoir-Rhyne University, Hickory, NC 28603
E-mail address: `thierry.zell@lr.edu`