EFFICIENT COMPUTATIONAL PROOFS AND INAPPROXIMABILITY

JOHAN HÅSTAD

Consider the classical NP-complete problem of Boolean satisfiability. We are given a Boolean formula and we are interested whether there is an assignment to the n occurring Boolean variables in the formula that satisfies the formula. The fact that this problem is NP-complete means that unless NP=P there is no efficient (i.e. polynomial time) algorithm that answers this question correctly for all formulas.

The fact that satisfiability is said to belong to NP is due to the fact that if the formula is indeed satisfiable then there is a short proof of this fact that can be verified efficiently. This proof is simply the description of an assignment that does satisfy the formula.

This proof has the property that a verifier needs to read n bits and one would suspect that this is essentially the best that can be achieved for a deterministic verifier. This fact changes drastically if we allow the verifier to be probabilistic. This is a called a Probabilistically Checkable Proof (PCP) and in such a proof we require the verifier always to accept a correct proof of a correct statement but to reject any proof for an incorrect statement with probability at least 1/2. As this probability is only over the random choices of the verifier the error probability can be decreased by running the verifier repeatedly.

The PCP-theorem tells us that any NP-statement, such as satisfiability, admits a PCP where the verifier reads a constant number of bits, independently of the size of the formula being proved. This is a mathematically exact form of saying that random spot checks can be made efficiently.

In the PCP-theorem the verifier uses very limited randomness, in that only $O(\log n)$ random bits are needed. This fact makes it possible to derive very strong consequences on the approximability of NP-hard optimization problems.

Among many results let us mention a two. One can prove that if NP$\neq$P then given a linear system of equations modulo 2, one cannot efficiently tell whether we can simultaneously satisfy a fraction $(1-\epsilon)$ of the equations or that no assignment satisfies more than a fraction $(\frac{1}{2}+\epsilon)$. Given a graph with n nodes it is computationally hard to determine whether its largest independent set has $n^{1-\epsilon}$ nodes or only n^{ϵ}. In both cases ϵ is an arbitrarily small positive number.

We will informally describe the PCP-theorem, some in-approximability results and the connection between the two areas.