

## Hard on Average Constraint Satisfaction Problems

Many randomly generated constraint satisfaction problems (random K-SAT, coloring of a random graph, maximum independent set of a random graph) exhibit a gap between algorithmic and existential results.

What is the source of the apparant hardness? Is there an underlying phase transition?

It turns out that the clue is given by the intricate geometry of the solution space. Overlap Gap Property originating from the theory of Spin Glasses is also related.

Consider  $\mathbb{G}(n, p)$ . It is easy to show that the size of the largest clique is  $\log_{\frac{1}{p}} n$ . A trivial greedy algorithm can also find a clique of size  $\log_{\frac{1}{p}} n$ . Since 1976 no better algorithm have been found.

Independent sets in sparse random graphs also show the same behaviour.

Gamarnik and Sudan have shown in 2013 that there exists a model with the Overlap Gap Property (OGP), which was used to disprove Hatami, Lovasz and Szegedy conjecture regarding the power of local algorithms, based on i.i.d factors.

Using a similar idea of multi-OGP, local algorithms fail for independent sets with size  $\frac{(1+\epsilon) \log d}{d} s$ .

The same idea is also useful in the study of the Random NAE-K-SAT problem.

Given a boolean formula in the form of a conjunction of disjunction clauses, in the proble of NAE-K-SAT we say that a formula is satisfiable if there exists a truth assignment such that in every clause there exists one satisfying and one violating variable.

Coja-Oghlan and Panagiotou have found a necessary and and sufficient condition to find a satisfying solution. However, an algorithm is only known for simpler conditions.

Multi-OGP for random K-SAT and Sequential Local Algorithms can also be used to prove theoretical bounds on the efficiency of some algorithms, as shown by Gamarnik and Sudan in 2014. In this sense, for example, there is a condition which, if satisfied, prevents sequential local algorithms from finding a satisfying assignment. Similarly, there is a condition by which the WALKSAT algorithm fails to find a satisfying assignment.

Another application of the idea is to the problem of finding a maximum submatrix of a Gaussian matrix. Bhamidi, Def and Nobel have found that there is a global optimal solution to the problem. Gamarnik and Li analysed the best of the three algorithms, and found that, again, there is a gap between the theoretical and existential solution.

Spase high-dimensional linear regression provides another set of problems with the OGP property.

Suppose a model is given  $Y = X\beta^* + W$ , where  $X$  and  $W$  are i.i.d Gaussian. How can we recover  $\beta^*$  from observed  $X$  and  $Y$ , provided the sample is *sparse*?

A natural approach is to solve a hard quadratic minimization problem. Is the optimal solution  $\beta_0$  a good approximation of the ground truth  $\beta^*$ ? How can we solve this problem efficiently?

Recently, there has been a spectacular success achived in tecniques based on convex optimisation (see Donoho, Tanner, Wainwright, Hastie, Tibshirani, Candes and Tao). However, there is a condition stemming from information theory so that there is no way to recover the solution.