On the Approximation Resistance of a Random Predicate

Johan Håstad

KTH Numerical Analysis
and Computer Science

November 12, 2007

History of talk

Originally given at the APPROX-conference in Princeton, August 2007.

Survey of the general area, stating results but giving very few details. Ask for details.

Boolean CSPs

Constraint Satisfaction Problems where each input is a bit.

Same predicate appears P in all constraints.

A k-ary predicate P that accepts t of the 2^k inputs.

Most famous example of CSPs, k-sat

Disjunctions of k literals, m constraints, n variables.

2-Sat: $(x_1 \vee \bar{x}_2) \wedge (x_2 \vee x_7) \wedge \ldots \wedge (\bar{x}_1 \vee \bar{x}_{11})$.

$k = 2, t = 3$

3-Sat: $(x_1 \vee \bar{x}_2 \vee x_3) \wedge (x_2 \vee x_7 \vee \bar{x}_8) \wedge \ldots \wedge (\bar{x}_1 \vee x_8 \vee \bar{x}_{11})$.

$k = 3, t = 7$.

3-Lin

System of linear equations modulo 2 with at most three variables in each equation.

$$\left\{\begin{array}{ccccccccc} x_1 & + & x_2 & + & x_3 & & & = & 1 \\ x_1 & + & x_2 & & & & & = & 1 \\ x_1 & + & x_2 & + & & & x_4 & = & 1 \\ & & x_2 & + & & & x_4 & = & 0 \\ x_1 & + & & & x_3 & + & x_4 & = & 0 \\ & & x_2 & + & x_3 & + & x_4 & = & 1 \\ x_1 & + & & & x_3 & & & = & 0 \end{array}\right. \qquad \bmod 2$$

m equations n variables.

Many predicates, majority, not-all-equal etc

We have 2^{2^k} predicates on k inputs. It is specified by an answer on each k bit string.

Equivalent predicates

Please note that negations are allowed for free, and so are permutations of the inputs.

We get families of equivalent predicates, each containing up to $2^k \cdot k!$ different predicates.

The number of families is $2^{2^k(1-o(1))}$.

Number of different predicates

Can be calculated by computer.

k	2	3	4
Predicates	14	254	65534
Non-EQ	3	16	400

Only counting non-constant predicates.

Max-CSPs

Finding optimal solution is almost always, [S78], NP-complete and we are interested in approximation problem.

Max-CSP: Find the assignment that satisfies that maximum number of constraints.

Max-CSPs

Finding optimal solution is almost always, [S78], NP-complete and we are interested in approximation problem.

Max-CSP: Find the assignment that satisfies that maximum number of constraints.

Given a list of m k-tuples of literals find an assignment that makes as many as possible of the resulting k-tuples of bits satisfy P.

Approximation ratio

An algorithm has approximation ratio α if for any instance

$$\frac{\text{Value of found solution}}{\text{Value of optimal solution}} \geq \alpha$$

For randomized algorithms, expectation over internal coinflips, always worst case inputs.

Easy result for Max-3Sat

$$\varphi = (x_1 \vee \bar{x}_2 \vee x_3) \wedge (x_2 \vee x_7 \vee \bar{x}_8) \wedge \ldots \wedge (\bar{x}_1 \vee x_8 \vee \bar{x}_{11})$$

A random assignment satisfies each clause with probability $7/8$.

Easy result for Max-3Sat

$$\varphi = (x_1 \vee \bar{x}_2 \vee x_3) \wedge (x_2 \vee x_7 \vee \bar{x}_8) \wedge \ldots \wedge (\bar{x}_1 \vee x_8 \vee \bar{x}_{11})$$

A random assignment satisfies each clause with probability $7/8$.

On average we satisfy $7m/8$ clauses which gives a $7/8$-approximation.

Easy result for Max-3Sat

$$\varphi = (x_1 \vee \bar{x}_2 \vee x_3) \wedge (x_2 \vee x_7 \vee \bar{x}_8) \wedge \ldots \wedge (\bar{x}_1 \vee x_8 \vee \bar{x}_{11})$$

A random assignment satisfies each clause with probability $7/8$.

On average we satisfy $7m/8$ clauses which gives a $7/8$-approximation.

How do we do this deterministically?

Easy result for Max-3Lin

$$\left\{\begin{array}{ccccccccccc} x_1 & + & x_2 & + & x_3 & & & = & 1 & & \\ x_1 & + & x_2 & & & & & = & 1 & & \\ x_1 & + & x_2 & + & & & x_4 & = & 1 & & \\ & & x_2 & + & & & x_4 & = & 0 & & \bmod 2 \\ x_1 & + & & & x_3 & + & x_4 & = & 0 & & \\ & & x_2 & + & x_3 & + & x_4 & = & 1 & & \\ x_1 & + & & & x_3 & & & = & 0 & & \end{array}\right.$$

A random assignment satisfies each clause with probability $1/2$.

We get an $1/2$-approximation algorithm.

In general

It is easy to approximate Max-P within $t2^{-k}$.

A random assignment satisfies on the average $t2^{-k}m$ constraints and this can be found deterministically.

In general

It is easy to approximate Max-P within $t2^{-k}$.

A random assignment satisfies on the average $t2^{-k}m$ constraints and this can be found deterministically.

The trivial approximation ratio.

Approximation resistance

A predicate P is approximation resistant if $\forall \epsilon > 0$ it is hard to approximate max-CSP(P) within $\epsilon + t2^{-k}$.

Approximation resistance

A predicate P is approximation resistant if $\forall \epsilon > 0$ it is hard to approximate max-CSP(P) within $\epsilon + t2^{-k}$.

A predicate P is approximation resistant on satisfiable instances if $\forall \epsilon > 0$ it is hard distinguish instances where we can satisfy all constraints from those where we can only satisfy a fraction $\epsilon + t2^{-k}$ of the constraints.

Hereditary properties

A predicates P is hereditary approximation resistant if whenever $P(x) \Rightarrow Q(x)$ then Q is also approximation resistant.

My view

Approximation resistance on satisfiable instances is possibly the ultimate hardness for a CSP.

Efficient computation cannot tell whether we can satisfy all constraints or only the fraction obtained by a random assignment.

My view

Approximation resistance on satisfiable instances is possibly the ultimate hardness for a CSP.

Efficient computation cannot tell whether we can satisfy all constraints or only the fraction obtained by a random assignment.

Are there such predicates?

Binary constraints

Constraints on two variables, $k = 2$.

Semidefinite programming [GW95,LLZ] gives an .8740 approximation algorithm in general.

There are hence no approximation resistant predicates on two binary variables.

More general fact

Extends to give non-approximation resistance of binary predicates over all domains sizes [H05] and binary constraints.

The case $k = 3$

Max-3-Lin is hereditary approximation resistant [H01], and this gives all approximation resistant predicates [Z98] on three inputs.

The case $k = 3$

Max-3-Lin is hereditary approximation resistant [H01], and this gives all approximation resistant predicates [Z98] on three inputs.

What does this give for Max-3-Sat, Max-3-Maj, Max-3-NAE?

Satisfiable instances

Max-3-Sat is approximation resistant on satisfiable instances.

Unknown what happens for the "not two ones predicate" on satisfiable instances.

The case $k = 4$

Partial classification by Hast [H05].

400 essentially different predicates.

- 79 approximation resistant.
- 275 not approximation resistant.
- 46 not classified.

# Acc	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Non-res	1	4	6	19	27	50	50	52	27	26	9	3	1	0	0
Res	0	0	0	0	0	0	0	16	6	22	11	15	4	4	1
Unkn	0	0	0	0	0	0	6	6	23	2	7	1	1	0	0

Satisfiability ignored.

Questions

How common is approximation resistance?

Can we find big classes of approximation resistant predicates?

Questions

How common is approximation resistance?

Can we find big classes of approximation resistant predicates?

What about a random predicate?

Random predicates

A random predicate from space $R_{p,k}$ accepts each input with probability p (and has $t \approx p2^k$).

Is a random predicate for $p = 1/2$ likely to be approximation resistant?

Predicate P_{ST}^{1}

A predicate given by a subspace of dimension $l_1 + l_2$ with $k = l_1 + l_2 + l_1 l_2$.

Showed to be approximation resistant by Samorodnitsky and Trevisan [ST00] and hereditary so by Hast [H05].

Predicate P^1_{ST}

A predicate given by a subspace of dimension $l_1 + l_2$ with $k = l_1 + l_2 + l_1 l_2$.

Showed to be approximation resistant by Samorodnitsky and Trevisan [ST00] and hereditary so by Hast [H05].

Gives many approximation resistant predicates but is it enough for a random predicate?

A rough calculation

If we accept t inputs the probability that it implies a random predicate is 2^{-t}.

We have at most $k!2^k$ equivalent predications.

Need $t \leq k \log k$ while we have $t \approx 2^{\sqrt{k}}$ for P^1_{ST}.

Predicate P_{ST}^2

A predicate given by a subspace of dimension d with $2^{d-1} < k \leq 2^d - 1$.

Assuming the Unique Games Conjecture (UGC) showed to be approximation resistant by Samorodnitsky and Trevisan [ST05].

Unique Games Conjecture

Made by Khot [K02] , a binary CSP over a large alphabet L.

Constraints are permutations $\pi_{ij}(x_i) = x_j$ for some pairs i, j.

Problem: Distinguish instances where we can satisfy fraction $1 - \epsilon$ from those where we can only satisfy fraction δ.

Unique Games Conjecture

Made by Khot [K02] , a binary CSP over a large alphabet L.

Constraints are permutations $\pi_{ij}(x_i) = x_j$ for some pairs i, j.

Problem: Distinguish instances where we can satisfy fraction $1 - \epsilon$ from those where we can only satisfy fraction δ.

Conjecture: $\forall \epsilon, \delta > 0$ there $\exists$ alphabet size L for which the problem is NP-complete.

Unique Games Conjecture

Made by Khot [K02] , a binary CSP over a large alphabet L.

Constraints are permutations $\pi_{ij}(x_i) = x_j$ for some pairs i, j.

Problem: Distinguish instances where we can satisfy fraction $1 - \epsilon$ from those where we can only satisfy fraction δ.

Conjecture: $\forall \epsilon, \delta > 0$ there $\exists$ alphabet size L for which the problem is NP-complete.

A very open conjecture.

Main result

Theorem: Assuming the unique games conjecture a random predicate from $R_{1/2,k}$ is with high probability, for sufficiently large k, approximation resistant.

Main result

Theorem: Assuming the unique games conjecture a random predicate from $R_{1/2,k}$ is with high probability, for sufficiently large k, approximation resistant.

Extends to $p = k^{-c}$ for $1/2 \leq c \leq 1$, $c \approx k2^{-d}$.

Proof 1

Assuming UGC P^2_{ST} is hereditary approximation resistant.

Extending the proof of Samorodnitsky and Trevisan.

Key Lemma

Lemma: For $S \subseteq [d]$ functions f_S such that

- One function (almost) unbiased, $|E[f_S(x)]| \leq \delta$.
- No two functions have high common influence, $\max(inf_i(f_{S_1}), inf_i(f_{S_2})) \leq \epsilon$.

$$\left| E_{x_1 \ldots x_d} \left[\prod_{S \subseteq [d]} f_S(\prod_{i \in S} x_i) \right] \right| \leq \delta + (2^d - 2)\sqrt{\epsilon},$$

i.e. Gowers uniformity norm is small.

Key Lemma

Lemma: For $S \subseteq [d]$ functions f_S such that

- One function (almost) unbiased, $|E[f_S(x)]| \leq \delta$.
- No two functions have high common influence, $\max(inf_i(f_{S_1}), inf_i(f_{S_2})) \leq \epsilon$.

$$\left| E_{x_1 \ldots x_d} \left[\prod_{S \subseteq [d]} f_S(\prod_{i \in S} x_i) \right] \right| \leq \delta + (2^d - 2)\sqrt{\epsilon},$$

i.e. Gowers uniformity norm is small.

New simpler more direct proof compared to [ST05].

Proof 2

Prove that if Q is random from $R_{1/2,k}$ then it is likely that there is a P^2_{ST}-equivalent predicate P' such that $P' \Rightarrow Q$.

Proof 2

Prove that if Q is random from $R_{1/2,k}$ then it is likely that there is a P^2_{ST}-equivalent predicate P' such that $P' \Rightarrow Q$.

Second moment method using only P^2_{ST}-equivalent predicates that are very different.

Summary

- Approximation resistance is a very strong notion of hardness.
- If the Unique Games Conjecture is true then a vast majority of predicates are approximation resistant.

Open problems

1. Prove result without the unique games conjecture.
2. Prove approximation resistance on satisfiable instances.
3. Classify more predicates with respect to approximation resistance.