Hard CSPs have hard gaps
at location 1

Andrei Krokhin
Durham University, UK

Joint work with
Peter Jonsson and Fredrik Kuivinen
Linköping University, Sweden
Constraints

- $D$ – a finite set with $|D| > 1$;
- $R_D^{(m)} = \text{subsets of } D^m$, $R_D = \bigcup_{m=1}^{\infty} R_D^{(m)}$.

Definition 1 A constraint over a set of variables $V = \{x_1, x_2, \ldots, x_n\}$ is a pair of the form $C = (x, \varphi)$ where

- $x = (x_{i_1}, \ldots, x_{i_m})$ is the constraint scope,
- $\varphi \in R_D^{(m)}$ is the constraint relation.

The constraint $C$ is said to be satisfied by an assignment $f : V \rightarrow D$ if $(f(x_{i_1}), \ldots, f(x_{i_m})) \in \varphi$. 
The Constraint Satisfaction Problem

**CSP**

**Instance:** A collection $C_1, \ldots, C_q$ of constraints over $V$.

**Question:** Is there an assignment $f : V \rightarrow D$ satisfying all these constraints?

**Max CSP**

**Instance:** A collection $C_1, \ldots, C_q$ of constraints over $V$.

**Goal:** Find an assignment $f : V \rightarrow D$ satisfying maximum number of the constraints?
Parameterisation of CSP and Max CSP

Definition 2 A constraint language is finite subset of $R_D$.

For a constraint language $\Gamma$, CSP($\Gamma$) and Max CSP($\Gamma$) consist of all CSP and MAX CSP, respectively, instances in which all constraint relations belong to $\Gamma$.

Research Programme

Classify the complexity and approximability of the problems CSP($\Gamma$) and Max CSP($\Gamma$).

Disclaimer: We assume $P \neq NP$ throughout.
The bounded occurrence property

Definition 3 Let CSP(Γ)-k (Max CSP(Γ)-k) denote the problem CSP(Γ) (Max CSP(Γ), respectively) restricted to instances where the number of occurrences of each variable (counted with multiplicity of constraints) is bounded by k.

NB. This is very similar to restricting graph problems to classes of graphs of bounded degree.

Definition 4 We say that CSP(Γ)-B (Max CSP(Γ)-B) is hard (in some sense) if there exists a number k such that CSP(Γ)-k (Max CSP(Γ)-k, resp.) is hard in that sense.
**Example: 2-Col and Max Cut**

Let $D = \{0, 1\}$ and $\Gamma = \{\text{neq}\}$ where $(x, y) \in \text{neq}$ iff $x \neq y$.

Then CSP(\(\Gamma\)) is the **2-COLOURABILITY** problem and Max CSP(\(\Gamma\)) is precisely the **Max Cut** problem.

For an instance $\mathcal{I}$ of CSP(\(\Gamma\)) over $V = \{x_1, \ldots, x_n\}$, consider a (multi)graph $G_{\mathcal{I}} = (V, E)$ with $E$ consisting of constraint scopes in $\mathcal{I}$.

Clearly, $\mathcal{I}$ is satisfiable iff $G_{\mathcal{I}}$ is 2-colourable.

Moreover, computing maximum cut in $G_{\mathcal{I}}$ is the same as maximising the number of satisfied constraints in $\mathcal{I}$.

**Complexity:** 2-Col is in \(\mathbf{P}\), Max Cut-3 is \(\mathbf{NP}\)-hard.
Example: 3-Sat and Max 3-Sat

Let $D = \{0, 1\}$ and let $\Gamma_{3\text{sat}} = \{\varrho_0, \varrho_1, \varrho_2, \varrho_3\}$ where

- $\varrho_0 = \{0, 1\}^3 \setminus \{(0, 0, 0)\}$
  $$x \lor y \lor z$$
- $\varrho_1 = \{0, 1\}^3 \setminus \{(0, 0, 1)\}$
  $$x \lor y \lor \overline{z}$$
- $\varrho_2 = \{0, 1\}^3 \setminus \{(0, 1, 1)\}$
  $$x \lor \overline{y} \lor \overline{z}$$
- $\varrho_3 = \{0, 1\}^3 \setminus \{(1, 1, 1)\}$
  $$\overline{x} \lor \overline{y} \lor \overline{z}$$

It is easy to see that CSP($\Gamma_{3\text{sat}}$) is precisely 3-Sat and
Max CSP($\Gamma_{3\text{sat}}$) is precisely Max 3-Sat.

Complexity: Both problems are $\mathbf{NP}$-hard.
The complexity classification problem

Conjecture 1 (Feder, Vardi ’98) \textit{Dichotomy conjecture:} Each problem \( \text{CSP}(\Gamma) \) is either in \( \mathbf{P} \) or else \( \mathbf{NP} \)-complete.

Theorem 1 (Bulatov, Jeavons, K. ’05) \textit{If} \( \Gamma \) \textit{has property} \( (G\text{-set}) \) \textit{then} \( \text{CSP}(\Gamma) \) \textit{is} \( \mathbf{NP} \)-complete.

Conjecture 2 (BJK’05) \textit{Algebraic dichotomy conjecture:} \textit{If} \( \Gamma \) \textit{does not have property} \( (G\text{-set}) \) \textit{then} \( \text{CSP}(\Gamma) \) \textit{is in} \( \mathbf{P} \).

Theorem 2 (Bulatov ’03-06) \textit{Conjecture 2 holds when} \( |D| \leq 3 \) \textit{or when} \( \Gamma \) \textit{contains all unary relations.}
A property equivalent to \((G\text{-set})\)

Assume wlog that \(\Gamma\) is a core, and let \(C_D = \{\{d\} | d \in D\}\).
Recall the relations \(\varrho_0, \varrho_1, \varrho_2, \varrho_3\) from \(\Gamma_{3sat}\).

Then \(\Gamma\) has property \((G\text{-set})\) iff there exist

1. a subset \(U\) of \(D\) and a function \(h : U \rightarrow \{0, 1\}\), and
2. four pp-formulas (=conjunctive queries) over \(\Gamma \cup C_D\) expressing precisely the relations

\[
h^{-1}(\varrho_j) = \{(a, b, c) \in U^3 | (h(a), h(b), h(c)) \in \varrho_j\}.
\]
A property opposite to $(G\text{-SET})$

A weak near-unanimity (WNU) operation on $D$ is an $n$-ary ($n \geq 2$) operation which satisfies the identities $f(x, \ldots, x) = x$ and $f(y, x, \ldots, x) = \ldots = f(x, \ldots, x, y)$.

Examples: $\min(x_1, \ldots, x_n)$, $x_1 + \ldots + x_n + x_{n+1} \pmod{n}$.

Recall that a polymorphism of $\Gamma$ is an operation that preserves every relation in $\Gamma$.

Theorem 3 (Maróti, McKenzie ’07)
A core $\Gamma$ does not have property $(G\text{-SET})$ iff it has a WNU polymorphism of some arity.
The approximability classification problem

Fact 1  For each problem Max CSP(Γ), there exist a constant $c_Γ \leq 1$ and a poly-time $c_Γ$-approximation algorithm (i.e., producing a solution of value at least $c_Γ \cdot \text{OPT}(I)$ for every instance $I$ of Max CSP(Γ)).

Problem 1  Characterise sets $Γ$ such that

- Max CSP(Γ) is in $\text{PO}$ (i.e., $c_Γ = 1$)
- Max CSP(Γ) is $\text{NP}$-hard and
  - Max CSP(Γ) has a PTAS – polynomial time approximation scheme (i.e., $c_Γ$ can be chosen arbitrarily close to 1)
  - $c_Γ \leq \delta < 1$ – “hard to approximate”
Hard gap at location 1

Definition 5 A problem $\text{Max CSP}(\Gamma)$ is said to have a hard gap at location 1 if, for some fixed $\alpha < 1$, it is $\text{NP}$-hard to distinguish between

- instances where all constraints can be satisfied, and
- those where at most $\alpha$-fraction can be satisfied.

Fact 2 If $\text{Max CSP}(\Gamma)$ has a hard gap at location 1 then

- $c_\Gamma \leq \alpha < 1$ — hard to approximate (even when restricted to satisfiable instances);
- $\text{Max CSP}(\Gamma)$ cannot have a PTAS;
- $\text{CSP}(\Gamma)$ cannot be in $\text{P}$.
Relating to the PCP theorem

Theorem 4 (Arora et al’ 98, Arora,Safra ’98, Dinur’07)
The following equivalent statements hold:

1. \( \text{NP} \subseteq \text{PCP}[\log n, 1] \),

2. for some constraint language \( \Gamma \) over some \( D \),
   \( \text{Max CSP}(\Gamma) \) has a hard gap at location 1,

3. \( \text{Max 3-Sat} \) has a hard gap at location 1.

The proof of equivalence of the statements is quite easy (half a page), while the proof of validity is very hard.

Recent combinatorial proof of (2) by Dinur deals entirely with CSPs.
Main result

Theorem 5  If $\Gamma$ has property \((G\text{-set})\) then the problem Max CSP($\Gamma$)-$B$ has a hard gap at location 1.

Note that if the algebraic dichotomy conjecture holds then Max CSP($\Gamma$) has a hard gap at location 1 for all $\Gamma$ with hard CSP($\Gamma$).

Corollary 1  If $\Gamma$ has property \((G\text{-set})\) then the problem Max CSP($\Gamma$)-$B$ is hard to approximate even when it is restricted to satisfiable instances. In particular, Max CSP($\Gamma$)-$B$ has no PTAS.
Key elements in proof

- Recall that property (G-set) for a core $\Gamma$ implies that $\Gamma \cup C_D$ pp-expresses pre-images of relations from $\Gamma_{3sat}$.
- Hard gap for MAX 3-SAT—$B$ is the base case.
- Moving to pre-images for free
- Show that the presence of a hard gap is preserved when adding $C_D$ and pp-expressed relations
Adding pp-expressed relations

Lemma 1 (Jeavons’98) If a constraint language $\Gamma$ pp-expresses a relation $\varrho$ then $\text{CSP}(\Gamma \cup \{\varrho\})$ poly-time reduces to $\text{CSP}(\Gamma)$.

The above also holds in the bounded occurrence setting.

Lemma 2 If a constraint language $\Gamma$ pp-expresses $\varrho$ and $\text{Max CSP}(\Gamma \cup \{\varrho\}) - k$ has hard gap at location 1 then, for some $k'$, $\text{Max CSP}(\Gamma) - k'$ has hard gap at location 1.

The gap parameter $\alpha$ becomes $\alpha' = \alpha + (1 - \alpha)(1 - 1/N)$ where $N$ is the number of relations in pp-expression for $\varrho$. 
Adding $\mathcal{C}_D$

**Lemma 3 (Bulatov, Jeavons, AK ’05)** *If $\Gamma$ is a core then $\text{CSP}(\Gamma \cup \mathcal{C}_D)$ poly-time reduces to $\text{CSP}(\Gamma)$.*

The transformation in this lemma does not preserve the bounded occurrence property.

The proof (of the main theorem) gets around this.

All in all, the new gap parameter $\alpha'$ can be computed from

- the **Max 3-sat** gap parameter,
- the size of the domain $|D|$,
- a certain constant from expander graph construction,
- the number of relations in 5 pp-expressions from $\Gamma$. 
One application

**Theorem 6** Let $\varrho \in R_D$ be non-empty and let $\Gamma = \{\varrho\}$. If $(d, \ldots, d) \in \varrho$ for some $d \in D$ then $\text{Max CSP}(\Gamma)$ is trivial. Otherwise, $\text{Max CSP}(\Gamma) - B$ is hard to approximate.

$\text{Max Cut}$ (= Max CSP($\{\text{neq}\}$)) is hard to approximate. Theorem 6 can be seen as a generalisation of this.

The proof is based on the main theorem, and uses the bounded occurrence property (in the main theorem) in an essential way.