Characterizing Streaming Decidability of CSPs via Non-Redundancy

Amatya Sharma
University of Michigan

Santhoshini Velusamy
University of Waterloo

Thanks to NSERC Discovery grant, NSF award and ECF supplement, and University of Michigan Rackham Conference Travel Grant

Constraint Satisfaction Problems

A fixed local predicate/relation: $R \subseteq D^k$

Variables: $x_1, \ldots, x_n$

Constraints: $R(x_{i_1}, \ldots, x_{i_k})$

An assignment is satisfying if it satisfies every constraint.

Examples

$\neq$ → graph coloring / bipartiteness

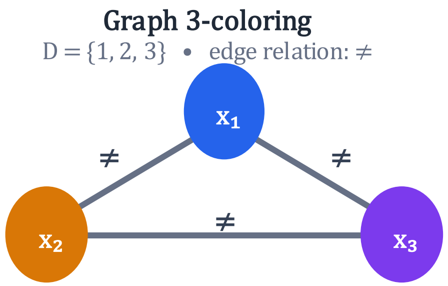
NAE_k → hypergraph 2-coloring

OR_k → k-SAT

$\oplus_k$ → linear equations

Constraint languages: more generally, allow a finite family $\Gamma = \{R_1, \ldots, R_t\}$.
Example: Unique Games: each permutation π gives the relation $R_\pi = \{(a, \pi(a)) : a \in [q]\}$.

Our results hold for finite Γ; for this talk, we fix one relation R.



The streaming satisfiability problem

INPUT: A CSP INSTANCE

A fixed local predicate/relation: $R \subseteq D^k$

Variables: $x_1, \dots, x_n$

Constraints arrive one at a time

C_1	C_2	C_3	$\cdots$	C_m

time

Each constraint applies R to some k variables.

one pass • minimize space • time unrestricted

ALGORITHM'S WORKING MEMORY

small memory

M_t

a summary of everything seen so far

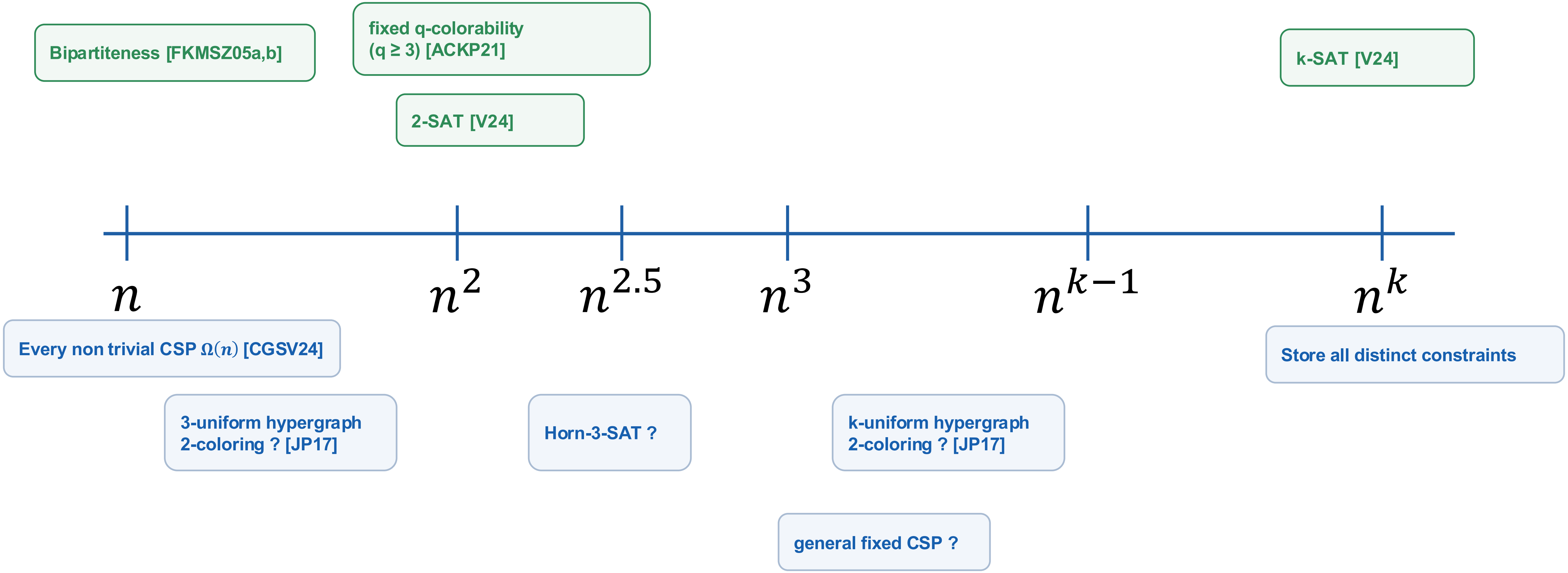
As each constraint arrives:

$$M_t = Update(M_{t-1}, C_t)$$

Goal: decide whether some assignment satisfies every constraint.

For a fixed relation R, how much memory is necessary and sufficient?

Motivation: different predicates give different exponents



What property of R determines the exponent?

Main result: the answer is non-redundancy

For every fixed relation R, the optimal one-pass streaming space is

$$\widetilde{\Theta}(\mathrm{NRD}_n(R))$$

This gives tight bounds for the cases left open on the previous slide:

3-uniform hypergraph 2-coloring	$\tilde{\Theta}(n^2)$
k-uniform hypergraph 2-coloring	$\tilde{\Theta}(n^{k-1})$
Horn-3-SAT	$\tilde{\Theta}(n^3)$

So what is non-redundancy?

Non-redundancy

Definition

An instance is non-redundant if removing any constraint changes the set of satisfying assignments.

$\boxed{\mathrm{NRD}_n(R) :=}$ largest size of a non-redundant instance on n variables.

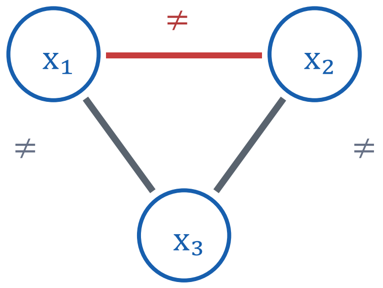
[BCK20, C22]

Witness Assignment

For every constraint in a non-redundant instance, there is an assignment that satisfies all the other constraints, but violates this one.

This is exactly the perspective we will use later in the lower bound.

Bipartiteness/XOR example



$(x_1 = 0,\ \ x_2 = 0,\ \ x_3 = 1)$

= Witness assignment for edge (x_1, x_2)

$$\mathrm{NRD}_n(\mathrm{XOR}) = \Omega(n)$$

Non-redundancy

Definition

An instance is non-redundant if removing any constraint changes the set of satisfying assignments.

$\mathrm{NRD_n(R)}$:= largest size of a non-redundant instance on n variables.

[BCK20, C22]

Witness Assignment

For every constraint in a non-redundant instance, there is an assignment that satisfies all the other constraints, but violates this one.

This is exactly the perspective we will use later in the lower bound.

2-SAT example

Put one clause on every unordered pair {i,j}:

$$C_{ij} = x_i \vee x_j$$

$$\binom{n}{2} = \Theta(n^2)$$

Witness for C_{ij}:

$$x_i = x_j = 0, \qquad x_\ell = 1 \; (\ell \neq i,j)$$

Every other clause contains a variable set to 1.
Only C_{ij} is violated.

$$\mathrm{NRD}_n(2\text{-SAT}) = \Theta(n^2)$$

For every fixed R, the optimal one-pass streaming space is $\widetilde{\Theta}(\mathrm{NRD}_n(R))$

Upper bound: keep only a non-redundant subinstance

Maintain a subinstance with exactly the same satisfying assignments as all constraints seen so far.

current subinstance I

C_1 $\cancel{C_2}$ C_3 C_4 $+$ C

new constraint

$\downarrow$

Delete constraints one by one as long as it doesn't change the satisfying assignments.
An old constraint may become redundant.
Redundancy checks may be expensive; time is unrestricted in this model.

$\downarrow$

I' C_1 C_3 C_4 C non-redundant

Since I′ is non-redundant,

$$|I'| \leq \mathrm{NRD}_n(R)$$

Each constraint takes $O(\log n)$ bits to store.

$$\boxed{O(\mathrm{NRD}_n(R)\ \log n)\text{ space}}$$

Streaming lower bounds from INDEX

One-way communication INDEX

Alice	one message →	Bob
$x \in \{0,1\}^N$		x_i?

Streaming sat CSP(R)

Alice: $C_1, \ldots, C_N$ → (s)-bit state → Bob: $C'_1, \ldots, C'_m$

Any randomized one-way protocol requires $\Omega(N)$ bits. [KN97]

$$s\text{−space streaming CSP}(R) \;\Rightarrow\; s\text{−bit one−way communication for INDEX}$$

So: encode NRD-bit INDEX into streaming satisfiability

$\Rightarrow \Omega(\mathrm{NRD})$ space.

2-SAT encodes INDEX on all variable pairs

$N = \binom{n}{2}$

INDEX bits are indexed by unordered pairs {a,b}.

Alice

For every pair {a,b}:

$$z_{a,b} = 1 \Rightarrow \text{stream } C_{a,b} = (x_a \lor x_b)$$

→ s-bit state →

Bob receives (i,j)

Append clauses:

$$(\neg x_i \lor \neg x_i), \qquad (\neg x_j \lor \neg x_j)$$

forces $x_i = x_j = 0$

$$z_{i,j} = 1 \Rightarrow C_{i,j} \text{ present} \Rightarrow \text{UNSAT}$$

$$z_{i,j} = 0 \Rightarrow C_{i,j} \text{ absent} \Rightarrow \text{SAT}$$

SAT witness: $x_i = x_j = 0, \quad x_\ell = 1 \ (\ell \neq i, j)$ Exactly the private witness.

$$\Omega(N) = \Omega(n^2) \text{ space}$$

2-SAT: encoding INDEX on variable pairs

$N = \binom{n}{2}$ INDEX bits are indexed by unordered pairs {a,b}.

Alice

For every pair {a,b}:

$z_{a,b} = 1 \Rightarrow$ stream $C_{a,b} = (x_a \lor x_b)$

→ s-bit state →

Bob receives (i,j)

Append clauses:

$(\neg x_i \lor \neg x_i)\ (\neg x_j \lor \neg x_j)$

forces $x_i = x_j = 0$

$z_{i,j} = 1 \Rightarrow C_{i,j}$ present $\Rightarrow$ UNSAT

$z_{i,j} = 0 \Rightarrow C_{i,j}$ absent $\Rightarrow$ SAT

SAT witness: $x_i = x_j = 0,\ x_\ell = 1\ (\ell \neq i,j)$

$\Omega(N) = \Omega(n^2)$ space

Generalizing the 2-SAT reduction

Take a largest non-redundant instance $I = \{C_1, \dots, C_N\}$, $N = \mathrm{NRD}_n(R)$.

For every C_i, choose a witness assignment a_i: a_i satisfies every other constraint, but violates C_i.

Alice

$z \in \{0,1\}^N$

For every j:

$z_j = 1 \Rightarrow$ stream C_j

Bob receives (i,j)

Append clauses:

force the witness assignment a_i

$z_i = 1 \Rightarrow C_i$ present $\Rightarrow$ UNSAT

because the forced witness a_i violates C_i.

$z_i = 0 \Rightarrow C_i$ absent $\Rightarrow$ SAT

because a_i satisfies every other C_j.

If Bob can force a_i, INDEX gives $\Omega(N) = \Omega(\mathrm{NRD}_n(R))$. Exactly the private witness.

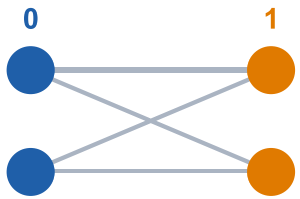
But can Bob always force a_i using constraints from R?

11

The obstacle: Bob may not be able to force one assignment

Example: Bipartitness / XOR $R(x,y) = [x \neq y]$

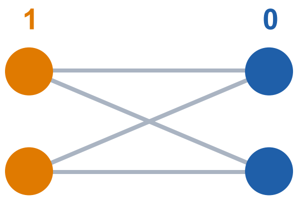
$a = (0,1,1,0, \ldots)$



flip every bit

$\leftrightarrow$

$\bar{a} = (1,0,0,1, \ldots)$



$$a \text{ satisfies } I \iff \bar{a} \text{ satisfies } I$$

for every XOR instance I.

$\downarrow$

~~force exactly a_i~~

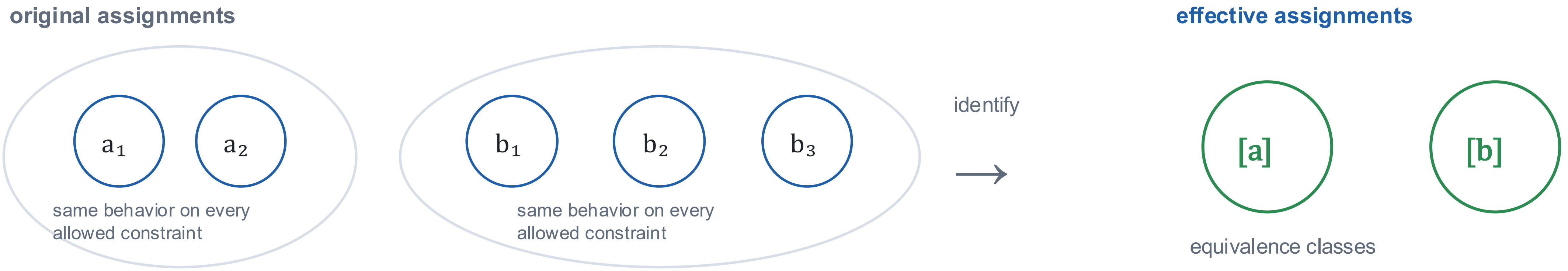
There may be no valid CSP instance whose unique satisfying assignment is a_i.

If R cannot distinguish two assignments, Bob should not need to distinguish them either.

What can R actually observe about an assignment?

The repair: identify indistinguishable assignments

If the CSP cannot distinguish two assignments, then they are effectively the same assignment.



Key fact

Bob can append a CSP instance whose satisfying assignments are exactly $[a_i]$.

Bob gets i $\longrightarrow$ $\mathrm{Sat}(I\_Bob) = [a_i]$ $\longrightarrow$ same INDEX reduction works $\longrightarrow$ $\Omega(\mathrm{NRD}_n(R))$ Lower bound

The One Ring: Non-Redundancy

Grazie!

On the 2027 postdoc market — happy to chat!

Streaming satisfiability

$\tilde{\Theta}(\mathrm{NRD})$

Approximate sparsification

preserve every assignment's value
using few weighted constraints

$\tilde{\Theta}(\mathrm{NRD})$

[BG25]

NRD

Streaming Approximate MinCSP

minimize number of
unsatisfied constraints

$\tilde{\Theta}(\mathrm{NRD})$

[BP**S**V26]

The same structural parameter appears across different CSP objectives. What other objectives?

Open problems

Understanding NRD

Known: symmetric Boolean CSPs classified through arity 4; almost all at arity 5.
[BGP26, **S**V26b, BGJLW26]

Simplest open cases

$$R = \{0,2,3\} \quad \text{or} \quad R = \{1,2,4\}$$

$$n^2 \lesssim \mathrm{NRD}_n(R) \lesssim n^3$$

What is the correct exponent?

Multiple passes

Partial p-pass bounds and tradeoffs are known.

$\Omega(\mathrm{NRD}(R)/p)$-type lower bounds?

and nontrivial space–pass tradeoffs for special cases such as 2-SAT

What is the right p-pass characterization for CSP satisfiability?