On Approximability, Convergence, and Limits of CSP Problems

Roland Markó
Hausdorff Center for Mathematics
University of Bonn

Ph.D. Thesis Defense
July 18, 2016

Outline

1. Introduction
   - Constraint satisfaction problems
   - Approximation and testing
   - Convergence of graph sequences and limits
2. Main results
   - Limits of colored hypergraphs
   - Testing and approximation of graph parameters
   - Nondeterministic testing
3. Further research

Introduction

Constraint Satisfaction Problem (CSP)

Example for an rCSP instance F

Variable set $V(F)$: $x_1 \quad x_2 \quad x_3 \quad x_4 \quad x_5$

Constraint set $C(F)$:

$x_1 \vee x_2 \vee \neg x_4$	$x_2 \vee \neg x_3 \vee \neg x_5$	$x_3 \oplus x_4 \oplus x_5$

Objective: make as many constraints true as possible

The constraint $x_1 \vee x_2 \vee \neg x_4$ is a Boolean map $\omega = (f; x_e)$ on $\{0,1\}^{V(F)}$ specified by

- the constrained variables vector $x_e = (x_1, x_2, x_4) \in V(F)^3$
- the constraint type function $f(a,b,c) = a \vee b \vee \neg c$

General setting: $\mathcal{C}(K, r)$ is the set of all constraint-types $f : K^r \to \{0,1\}$

Constraint Satisfaction Problem (CSP)

Example for an rCSP instance F

Variable set $V(F)$:

x_1	x_2	x_3	x_4	x_5
1	1	0	1	0

Constraint set $C(F)$:

$x_1 \vee x_2 \vee \neg x_4$ $\quad$ $x_2 \vee \neg x_3 \vee \neg x_5$ $\quad$ $x_3 \oplus x_4 \oplus x_5$

Objective: make as many constraints true as possible

The constraint $x_1 \vee x_2 \vee \neg x_4$ is a Boolean map $\omega = (f; x_e)$ on $\{0,1\}^{V(F)}$ specified by

- the constrained variables vector $x_e = (x_1, x_2, x_4) \in V(F)^3$
- the constraint type function $f(a,b,c) = a \vee b \vee \neg c$

General setting: $\mathcal{C}(K, r)$ is the set of all constraint-types $f : K^r \to \{0,1\}$

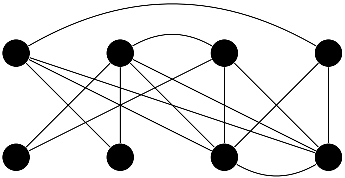
Constraint Satisfaction Problem (CSP)

- CSP decision problem: Can one turn all constraints of an instance F simultaneously to be true?
- MAX-CSP optimization problem:

$$\mathrm{MAX}{-}r\mathrm{CSP}(F) = \max_{l \in K^{V(F)}} \sum_{\omega=(f;x_e)\in C(F)} \omega(l),$$

Both are NP-hard problems in general

Example: MAX-CUT graph parameter: Maximize the number of crossing edges between two parts of a vertex bipartition



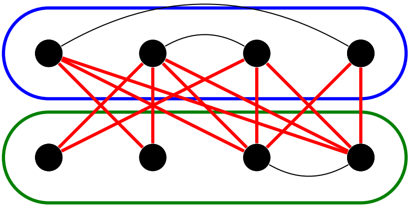
Constraint Satisfaction Problem (CSP)

- CSP decision problem: Can one turn all constraints of an instance F simultaneously to be true?
- MAX-CSP optimization problem:

$$\mathrm{MAX}{-}r\mathrm{CSP}(F) = \max_{l \in K^{V(F)}} \sum_{\omega=(f;x_e)\in C(F)} \omega(l),$$

Both are NP-hard problems in general

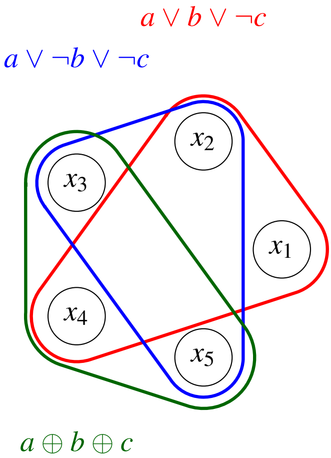
Example: MAX-CUT graph parameter:
Maximize the number of crossing edges between two parts of a vertex bipartition



Constraint Satisfaction Problem (CSP)

Colored graphs and CSP instances

- an rCSP instances on some domain K can be represented by a colored directed r-uniform hypergraph: edges colors are (subsets of) the constraint types $\mathcal{C}(K, r)$
- the decision/optimization problem translates to the problem of determining graph properties/parameters

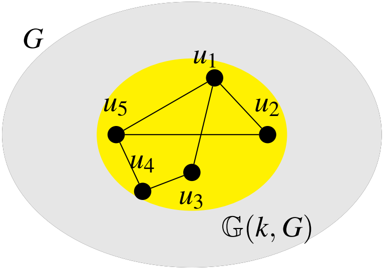


Testing properties

Motivation: Random vertex sampling in approximation algorithms (Arora-Karger-Karpinski ('95))

Sampling from dense graphs

$\mathbb{G}(k, G)$: induced subgraph of G on a set $S \subset V$ of cardinality k chosen uniformly at random



$\mathcal{P}$ graph property: family of graphs invariant under relabeling vertices

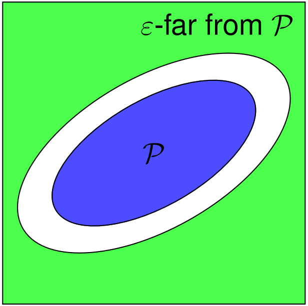
G is ε-far from $\mathcal{P}$: we have to add or remove at least $\varepsilon|V(G)|^2$ edges to obtain a member of $\mathcal{P}$

Testing properties

Graph property testing

$\mathcal{P}$ is testable if there is a $\mathcal{P}'$ and $q(\varepsilon)$ s.t.

- $G \in \mathcal{P} \implies \mathbb{G}(k, G) \in \mathcal{P}'$ with probability at least $2/3$ for all k
- G is ε-far from $\mathcal{P} \implies \mathbb{G}(k, G) \notin \mathcal{P}'$ with probability at least $2/3$ for $k \geq q(\varepsilon)$



Examples for testable properties

- triangle-freeness (Ruzsa-Szemerédi ('78))
- partition problem properties (Goldreich-Goldwasser-Ron ('98)), for example:
  - MAX-CUT is at least δn^2
  - k-colorability
  - the largest clique has size at least δn

Testing parameters

Definition

The graph parameter f is testable if for any $\varepsilon > 0$ there exists a $q(\varepsilon) \geq 1$ such that for any graph G and $k \geq q(\varepsilon)$

$$\mathbb{P}(|f(G) - f(\mathbb{G}(k,G))| > \varepsilon) < \varepsilon.$$

The smallest q satisfying the definition is the sample complexity of f and is denoted by q_f.

Motivation from approximation algorithms:

Theorem (Alon,F. de la Vega,Kannan and Karpinski ('02))

MAX-rCSP is testable with sample size $O(\varepsilon^{-4}\log(1/\varepsilon))$.

Convergence of graph sequences and limits

Definition (Lovász and Szegedy)

The sequence $(G_n)_{n\geq 1}$ of simple graphs converges if for every $k \geq 1$ the sequences $(\mathbb{G}(k, G_n))_{n\geq 1}$ of random graphs converge in distribution.

For a non-trivial meaning the graphs need to be dense

The limit object

- $W \colon [0,1]^2 \to [0,1]$ measurable and symmetric ($W(x,y) = W(y,x)$) is called a graphon
- Sampling from W: $\mathbb{G}(k, W)$
  - $V(\mathbb{G}(k, W)) = [k]$
  - generate independent $X_1, \ldots, X_k$ uniform $[0,1]$ random variables
  - make conditionally independent coin flips with success prob. $W(X_i, X_j)$ to decide upon inclusion of ij in $E(\mathbb{G}(k, W))$

Convergence of graph sequences and limits

Theorem (Lovász and Szegedy ('06))

For every convergent sequence $(G_n)_{n\geq 1}$ of simple graphs there is a graphon W s.t. for every $k \geq 1$ the sequences

$$\mathbb{G}(k, G_n) \xrightarrow{d} \mathbb{G}(k, W).$$

Every graphon is a limit of some sequence.

Example: Erdős-Rényi random graphs, $G(n,p) \xrightarrow{n\to\infty} W$ almost surely, where $W \equiv p$

Application of the graph limit theory:

Theorem (Borgs, Chayes, Lovász, Sós and Vesztergombi ('12))

The parameter f is testable iff for every convergent $(G_n)_{n\geq 1}$ the sequence $(f(G_n))_{n\geq 1}$ also converges.

Main problems

What are the limit objects of convergent CSP instance sequences?

What hypergraph parameters are efficiently testable?

What hypergraph properties are testable, what is their sample complexity?

Limits of colored hypergraphs

Limits of sequences of combinatorial objects

Previous work:

- Lovász-Szegedy('06): graphs
- Elek-Szegedy('12): r-uniform hypergraphs (r-graphs)
- Diaconis-Janson('08): directed graphs
- Lovász-Szegedy('12): compact edge colored graphs

Limits of compact colored hypergraphs

Theorem (Karpinski and Markó)

For a compact Polish space $\mathcal{K}$, the $\mathcal{K}$-colored r-graph limits are functions of the form

$$W\colon [0,1]^{2^{[r]}\setminus\{\emptyset\}} \to \mathcal{K}.$$

Two proof methods, both use the Riesz representation theorem similar to Lovász-Szegedy('12) :

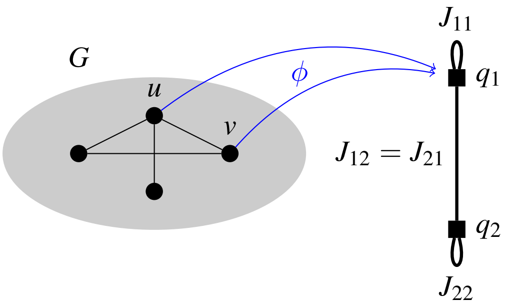
- Generalization of the ultralimit construction of Elek-Szegedy: direct structural consequences, proof of a weighted hypergraph version of the Regularity Lemma
- Generalization of the exchangeability correspondence of Diaconis-Janson using the representation theorem for exchangeable random arrays

Testing and approximation of graph parameters

Ground state energy

Example for an energy model (Spin model)

- $G = (V, E)$ is a graph
- q_1 and q_2 are the possible states
- J is a 2×2 symmetric interaction matrix



A map $\phi\colon V \to \{q_1, q_2\}$ assigns states to the vertices

Energy of G w.r.t ϕ and J: $\mathcal{E}_\phi(G, J) = \frac{1}{|V(G)|^2} \sum_{uv \in E(G)} J_{\phi(u)\phi(v)}$

Ground state energy (GSE) of G w.r.t J: $\hat{\mathcal{E}}(G, J) = \max_\phi \mathcal{E}_\phi(G, J)$

Testing the ground state energy

Related to polynomial time approximation schemes (PTAS) for MAX-rCSP

Previous work:

- Borgs-Chayes-Lovász-Sós-Vesztergombi('12): GSEs of graphs are testable

Theorem (Karpinski and Markó)

Given:

- $\mathcal{K}$: *compact Polish color set*
- E: *finite layer set*
- $G = (G^e)_{e \in E}$: E*-tuple of* $\mathcal{K}$*-colored* r*-graphs*
- $J = (J^e)_{e \in E}$: r*-arrays with* $J^e \in C(\mathcal{K})^{q \times \cdots \times q}$

For $k \geq \Theta^4 \log(\Theta) q^r$ *with* $\Theta = \frac{2^{r+7} q^r r}{\varepsilon}$ *we have*

$$\mathbb{P}(|\hat{\mathcal{E}}(G, J) - \hat{\mathcal{E}}(\mathbb{G}(k, G), J)| > \varepsilon |E| \, \|J\|_\infty) < \varepsilon.$$

Testing the ground state energy

Further contribution: positive testability results related to the ground state energy notion

- microcanonical GSE
- GSE with bias towards certain states
- GSE for weighted graphs with unbounded weights
- free energies = log-partition function per vertex

Nondeterministic testing

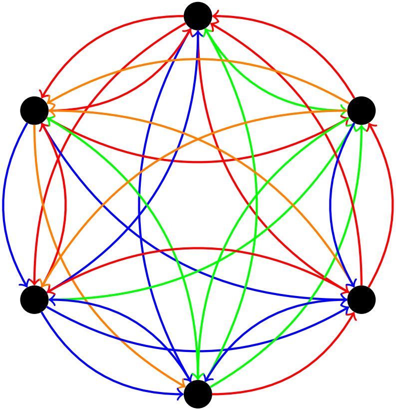
Nondeterministic testing

The complete edge-k-colored directed graph $\mathbf{G}$ is a (k, m)-coloring of G, if starting from $\mathbf{G}$ after

- erasing all edges of $\mathbf{G}$ colored with an element of $[m+1, \ldots, k]$
- discarding the coloring, orientation, and multiplicity of the remaining edges

we end up with G.

Notation: $\mathbf{G}' = G$



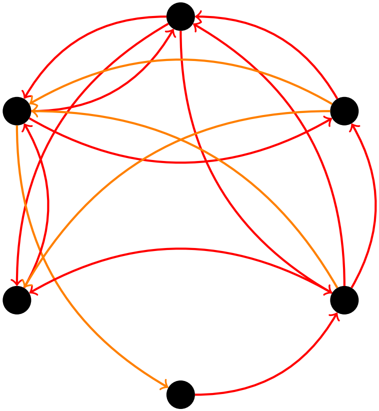
Nondeterministic testing

The complete edge-k-colored directed graph $\mathbf{G}$ is a (k, m)-coloring of G, if starting from $\mathbf{G}$ after

- erasing all edges of $\mathbf{G}$ colored with an element of $[m + 1, \ldots, k]$
- discarding the coloring, orientation, and multiplicity of the remaining edges

we end up with G.

Notation: $\mathbf{G}' = G$



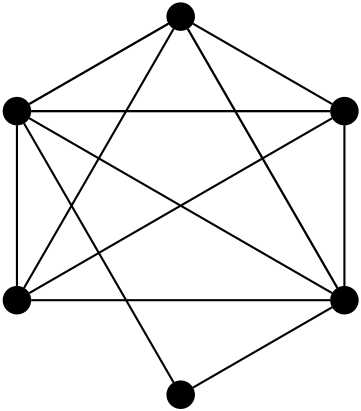
Nondeterministic testing

The complete edge-k-colored directed graph $\mathbf{G}$ is a (k, m)-coloring of G, if starting from $\mathbf{G}$ after

- erasing all edges of $\mathbf{G}$ colored with an element of $[m+1, \ldots, k]$
- discarding the coloring, orientation, and multiplicity of the remaining edges

we end up with G.

Notation: $\mathbf{G}' = G$



Nondeterministic testing

Definition (Lovász-Vesztergombi)

$\mathcal{P}$ is non-deterministically (ND-) testable if there exist

- integers $k \geq m$,
- a testable property $\mathcal{Q}$ of k-colored directed graphs (witness)

such that

$$\mathcal{P} = \{ \mathbf{G}' \mid \mathbf{G} \in \mathcal{Q} \}$$

Nondeterministic testing

Previous work:

- Lovász-Vesztergombi('13): ND-testability equivalent to testability
  - non-effective proof using graph limits
- Gishboliner-Shapira('14):

$$q_{\mathcal{P}}(\varepsilon) \le \mathrm{tf}(\mathrm{poly}(q_{\mathcal{Q}}(\varepsilon/2)))$$

  if $\mathcal{P}$ is ND-testable with witness $\mathcal{Q}$
  - $\mathrm{tf}(t)$: the exponential tower of 2s of height t
  - effective proof using Szemerédi's Regularity Lemma

Nondeterministic testing

Definition (Non-deterministically testable parameter)

f is ND-testable if there are $k \geq m$ and a testable k-colored directed graph parameter g such that

$$f(G) = \max_{\mathbf{G}'=G} g(\mathbf{G}).$$

Theorem (Karpinski and Markó)

Upper bounds for sample complexity in ND-testing:

- *simple graph parameters:* $q_f(\varepsilon) \leq \exp^{(3)}(cq_g^2(\varepsilon/2))$
- r*-uniform hypergraph parameters:* $q_f(\varepsilon) \leq \exp^{(4(r-1)+1)}(c_{r,k}q_g(\varepsilon)/\varepsilon)$
- *in both cases: property testing bounds have same magnitude*

Further research

Further research

- Improve sample bounds for GSE and ND-testing, characterize efficiently testable graph parameters and properties
- Testability of free energies - improve existing bounds, investigate infinite state models such as the n-vector spin model
- Nondeterministic testing with additional global conditions - analogous equivalence?
- Algorithmic version of the Hypergraph Regularity Lemma

Thank You!