Submodularity and the Complexity of Constraint Satisfaction

Andrei Krokhin
Durham University, UK

Three Well-Known Problems

k-SAT: is a given k-CNF formula satisfiable?

$$F = (\neg x \vee y \vee \neg z) \wedge (x \vee y \vee \neg z) \wedge (\neg x \vee \neg y \vee z)$$

Linear Equations: does a given system of linear equations have a solution in the fixed field K?

$$\begin{cases} 2x + 2y + 3z = 1 \\ 3x - 2y - 2z = 0 \\ 5x - y + 10z = 2 \end{cases}$$

Graph k-colouring: given a graph, can its vertices be coloured with k colours so that adjacent vertices are different colour?

Valued Constraints

- D – a fixed finite set with $|D| > 1$;
- $R_D^{(m)} = \{f \mid f : D^m \to \mathbb{Q}_+ \cup \{\infty\}\}$, $R_D = \bigcup_{m=1}^{\infty} R_D^{(m)}$.

Definition 1 *A valued constraint over a set of variables* $V = \{x_1, x_2, \ldots, x_n\}$ *is an expression of the form* $f(\mathbf{x})$ *where*

- $f \in R_D^{(m)}$ *is the constraint (cost) function,*
- $\mathbf{x} = (x_{i_1}, \ldots, x_{i_m})$ *the constraint scope.*

Interpretation: when assigning values to the variables, say $\varphi(x_i) = a_i$, the constraint incurs a cost of $f(a_{i_1}, \ldots, a_{i_m})$.

Valued Constraint Satisfaction Problem

VCSP

Instance: A collection $f_1(\mathbf{x}_1), \ldots, f_q(\mathbf{x}_q)$ of valued constraints over $V = \{x_1, \ldots, x_n\}$, possibly with weights $w_i \in \mathbb{Q}_+$ $(1 \leq i \leq q)$.

Goal: Find an assignment $\phi : V \to D$ that minimises the total cost; in other words, minimise the function $f : D^n \to \mathbb{Q}_+ \cup \{\infty\}$ defined by

$$f(x_1, \ldots, x_n) = \sum_{i=1}^{q} w_i \cdot f_i(\mathbf{x}_i).$$

Special Cases

Let $f(x_1, \ldots, x_n) = \sum_{i=1}^{q} w_i \cdot f_i(\mathbf{x}_i)$ be an instance of VCSP

- If $\mathrm{Im}(f_i) \subseteq \{0, \infty\}$ for all i, we get CSP
  - think "$0 =$ satisfied" — can one satisfy all $f_i(\mathbf{x}_i)$?
- If $\mathrm{Im}(f_i) \subseteq \{0, 1\}$, we get MAX CSP
  - want to satisfy maximum number of $f_i(\mathbf{x}_i)$
  - will use notation $P_D = \{g \in R_D \mid \mathrm{Im}(g) \subseteq \{0, 1\}\}$
- This talk – $\mathrm{Im}(f_i) \subseteq \mathbb{Q}_+$ – no infinite values
  - minimisation of "weakly separable" functions
  - will use notation $Q_D = \{g \in R_D \mid \mathrm{Im}(g) \subseteq \mathbb{Q}_+\}$

Parameterisation of VCSP

For a finite set $\Gamma \subseteq R_D$ (called a constraint language), VCSP(Γ) consists of all VCSP instances in which all constraint functions f_i belong to Γ.

Example 1 *Let* $D = \{0,1\}$ *and let* $\Gamma = \{neq\}$ *where* $neq(x,y) = a$ *if* $x \neq y$ *and* $neq(x,y) = b$ $(> a)$ *otherwise. Then* VCSP(Γ) *is precisely* Max Cut.

Indeed, for a graph $G = (V,E)$ *with* $V = \{x_1,\ldots,x_n\}$*, computing maximum cut is the same as minimising*

$$f(x_1,\ldots,x_n) = \sum_{e=(x_i,x_j)\in E} neq(x_i,x_j).$$

A Complexity Classification Project

How does the complexity of VCSP(Γ) depend on Γ?

- Sets Γ vary enormously
- Dichotomic tendency: either tractable or **NP**-hard
- Goal: identify all the tractable cases
- Goal: find a unified explanation of the tractability
- Goal: identify seeds of hardness/intractability
- Want: BIG PICTURE
- A lot of activity, powerful theory, strong results

Important Technicality: Core

Definition 2 *A constraint language* Γ *is a core if, for any* $a \in D$*, there is an instance* I *of* $\mathrm{VCSP}(\Gamma)$ *such that each optimal solution to* I *assigns* a *to some variable.*

Intuition: if Γ is not a core then there is $a \in D$ such that each instance of $\mathrm{VCSP}(\Gamma)$ has an optimal solution not involving a, so $\mathrm{VCSP}(\Gamma)$ reduces to a similar problem over a smaller domain.

Example 2 *For* $|D| = 2$, Γ *is not a core iff there is* $a \in D$ *such that* $f(a, \ldots, a) \leq f(x_1, \ldots, x_n)$ *for all* $f \in \Gamma$*.*
In this case $\mathrm{VCSP}(\Gamma)$ *is trivial.*

The Boolean Case: Submodularity!

Let $D = \{0, 1\}$. A function $f : D^n \to \mathbb{Q}_+$ is submodular iff

$$f(\mathbf{a} \vee \mathbf{b}) + f(\mathbf{a} \wedge \mathbf{b}) \leq f(\mathbf{a}) + f(\mathbf{b}) \text{ for all } \mathbf{a}, \mathbf{b} \in D^n.$$

Clearly, if Γ consists of submodular functions then VCSP(Γ) is tractable (because SFM is tractable).

Theorem 1 (Cohen, Cooper, Jeavons, AK '06)
Let $D = \{0, 1\}$ and let $\Gamma \subseteq Q_D$ be a core.
If each $f \in \Gamma$ is submodular then VCSP(Γ) *is tractable.*
Otherwise, VCSP(Γ) *is* **NP**-*hard.*

More Submodularity

Submodularity can be extended to any finite D with a fixed total order (to define $\vee$ and $\wedge$). Again, if Γ consists of submodular functions then $\mathrm{VCSP}(\Gamma)$ is tractable.

Theorem 2 (Jonsson, Klasson, AK '06)
Let $|D| = 3$ and let $\Gamma \subseteq P_D$ be a core. If there is a total order ϱ on D such that each $f \in \Gamma$ is submodular wrt ϱ then $\mathrm{VCSP}(\Gamma)$ is tractable. Otherwise, it is **NP***-hard.*

Theorem 3 (Kolmogorov, Živný '11)
Let D be any finite set and let $P_D^{(1)} \subseteq \Gamma \subseteq Q_D$.
If there is a total order ϱ on D such that each $f \in \Gamma$ is submodular wrt ϱ then $\mathrm{VCSP}(\Gamma)$ is tractable. Otherwise, it is **NP***-hard.*

Submodularity-Like Conditions

Modify $\vee$ and $\wedge$, using some additional structure on D

- Bisubmodularity/Directed Submodularity (Qi)
  - $D = \{-1, 0, 1, \}$ with order $-1 > 0 < 1$
  - $1 \vee_0 -1 = -1 \vee_0 1 = 0$ and $x \vee_0 y = \max(x, y)$ o/w
  - $1 \wedge_0 -1 = -1 \wedge_0 1 = 0$ and $x \wedge_0 y = \min(x, y)$ o/w
- $L^\natural$-convexity (Murota)
- Submodularity on a tree (Kolmogorov)
- Submodularity on a lattice/poset (Topkis)
- Submodularity in a bush (Madeup)

Multimorphisms

Definition 3 *A tuple* $\mathbf{F} = \langle F_1, \ldots, F_k \rangle$ *of operations* $F_i : D^m \to D$ *is called a multimorphism (MM) of* $f \in R_D^{(n)}$ *if, for all* $\mathbf{a}_1, \ldots, \mathbf{a}_m \in D^n$,

$$\frac{1}{k} \sum_{i=1}^{k} f(F_i(\mathbf{a}_1, \ldots, \mathbf{a}_m)) \leq \frac{1}{m} \sum_{j=1}^{m} f(\mathbf{a}_j).$$

In this case, one also says that $\mathbf{F}$ *improves* f.

- $f \in Q_{\{0,1\}}$ is submodular iff f has MM $\langle \min, \max \rangle$.
- $f \in Q_{\{-1,0,1\}}$ is bisubmodular iff f has MM $\langle \wedge_0, \vee_0 \rangle$.
- $f \in Q_{Z_p}$ is $L^\natural$-convex iff f has MM $\langle \lfloor \frac{x+y}{2} \rfloor, \lceil \frac{x+y}{2} \rceil \rangle$.

1-Defect Chain MM

Let $\leq$ be a total order on D. A 1-defect chain is obtained from $\leq$ by removing one pair (a, b) such that $a \prec b$.

A pair of operations $\langle \sqcup, \sqcap \rangle$ is a 1-defect chain MM if

- $x \sqcap y = \min(x, y)$ and $x \sqcup y = \max(x, y)$ whenever $\{x, y\} \neq \{a, b\}$
- $a \sqcap b < a \sqcup b$ and $\{a \sqcap b, a \sqcup b\} \cap \{a, b\} = \emptyset$

Bisubmodularity: $D = \{0 < 1 < -1\}$ and $(a, b) = (1, -1)$

Theorem 4 (Jonsson, Kuivinen, Thapper '11)
Let $|D| = 4$ and let $\Gamma \subseteq P_D$ be a core. If Γ is submodular on some chain or has 1-defect chain MM then VCSP(Γ) *tractable. Otherwise, it is* **NP***-hard.*

Generalisation: Fractional Polymorphisms

Definition 4 *For* $1 \leq i \leq k$*, let* $0 \leq \alpha_i \leq 1$, $\sum_{i=1}^{k} \alpha_i = 1$. *A tuple* $\mathbf{F} = \langle(\alpha_1, F_1), \ldots, (\alpha_k, F_k)\rangle$ *of pairs with* $F_i : D^m \to D$ *is called a fractional polymorphism (FP) of a function* $f \in R_D^{(n)}$ *if, for all* $\mathbf{a}_1, \ldots, \mathbf{a}_m \in D^n$.

$$\sum_{i=1}^{k} \alpha_i \cdot f(F_i(\mathbf{a}_1, \ldots, \mathbf{a}_m)) \leq \frac{1}{m} \sum_{j=1}^{m} f(\mathbf{a}_j)$$

In this case, one also says that $\mathbf{F}$ *improves* f.

- Each MM is an FP (with all $\alpha_i = 1/k$)
- If $\mathbf{F}$ improves each function in Γ then it also improves each instance $f = \sum_{i=1}^{q} w_i \cdot f_i(\mathbf{x}_i)$ of VCSP(Γ).

Example: α-Bisubmodularity

Recall bisubmodularity MM:

- $D = \{-1, 0, 1, \}$ with order $-1 > 0 < 1$
- $1 \vee_0 -1 = -1 \vee_0 1 = 0$ and $x \vee_0 y = \max(x, y)$ o/w
- $1 \wedge_0 -1 = -1 \wedge_0 1 = 0$ and $x \wedge_0 y = \min(x, y)$ o/w

Can also define

- $1 \vee_1 -1 = -1 \vee_1 1 = 1$ and $x \vee_0 y = \max(x, y)$ o/w

Definition 5 *For $0 < \alpha \leq 1$, a function $f \in Q_D$ is called α-bisubmodular if it has FP $\langle(\frac{1-\alpha}{2}, \vee_1), (\frac{\alpha}{2}, \vee_0), (\frac{1}{2}, \wedge_0)\rangle$, i.e.*

$$(1-\alpha) \cdot f(\mathbf{a} \vee_1 \mathbf{b}) + \alpha \cdot f(\mathbf{a} \vee_0 \mathbf{b}) + f(\mathbf{a} \wedge_0 \mathbf{b}) \leq f(\mathbf{a}) + f(\mathbf{b}).$$

FPs in Control of Complexity

Theorem 5 (Cohen, Cooper, Jeavons '06)
Let $\Gamma_1, \Gamma_2 \subseteq Q_D$ *be finite. If each FP of* Γ_1 *is an FP of* Γ_2 *then* $\mathrm{VCSP}(\Gamma_2)$ *poly-time reduces to* $\mathrm{VCSP}(\Gamma_1)$*.*

Corollary 1 *If* $\Gamma_1, \Gamma_2 \subseteq Q_D$ *are finite and have exactly the same FPs then* $\mathrm{VCSP}(\Gamma_1)$ *and* $\mathrm{VCSP}(\Gamma_2)$ *are equivalent.*

- Actually, FPs control expressive power of Γ
- Classification can definitely be stated in terms of FPs
- Which FPs guarantee tractability?

1-Approximate Polymorphisms

Definition 6 *For* $1 \leq i \leq k$, *let* $0 \leq \alpha_i \leq 1$, $\sum_{i=1}^{k} \alpha_i = 1$. *A tuple* $\langle(\alpha_1, F_1), \ldots, (\alpha_k, F_k)\rangle$ *with* $F_i : D^m \to \mathrm{Distr}(D)$ *is called a 1-approximate polymorphism (1-AP) of a function* $f \in Q_D^{(n)}$ *if, for all* $\mathbf{a}_1, \ldots, \mathbf{a}_m \in D^n$,

$$\mathbb{E}[f(F_i(\mathbf{a}_1, \ldots, \mathbf{a}_m))] \leq \max\{f(\mathbf{a}_1), \ldots, f(\mathbf{a}_m)\}.$$

- Each FP is a 1-AP, since, for functions $F_i : D_m \to D$,

$$\sum_{i=1}^{k} \alpha_i \cdot f(F_i(\ldots)) \leq \frac{1}{m} \sum_{j=1}^{m} f(\mathbf{a}_j) \leq \max_j \{f(\mathbf{a}_j)\}$$

Raghavendra's Dichotomy Theorem

Theorem 6 (Raghavendra' 08)
Let Γ *be a core. Assume that, for each* $\tau > 0$, *there is*

- $\mathbf{F}_\tau$ *– a 1-AP for each function in* VCSP(Γ) *such that in each* $F_i \in \mathbf{F}_\tau$, *each coordinate "has influence* $\leq \tau$*".*

Then VCSP(Γ) *is tractable. Otherwise, it is* **UGC***-hard.*

This (kind of) finishes our classification project, but

1. Can the tractability condition be made more tangible? Simple (binary) MMs or FPs instead of many 1-APs?
2. Can one replace **UGC**-hard by **NP**-hardness?
3. Is MAX CUT the only seed of hardness in VCSP?

Answers for the 3-Element Case

Theorem 7 (Huber, AK, Powell '12)
Let $|D| = 3$ *and let* $\Gamma \subseteq Q_D$ *be a core. If there is a renaming of elements of* D *into* $-1, 0, 1$ *such that*

- Γ *is submodular wrt* $-1 < 0 < 1$ *or*
- Γ *is* α*-bisubmodular for some* $0 < \alpha \leq 1$

then VCSP(Γ) *is tractable. Otherwise,* VCSP(Γ) *can express* Max Cut*, and hence is* **NP***-hard.*

- Tractability follows from Raghavendra's result, the above FPs easily generate the right 1-APs.
- We show how to express Max Cut (hardness part).

Conclusion / Open Problems

1. VCSP: valued constraint satisfaction problem
   - Minimisation of "weakly separable" functions
   - Want: complete complexity classification
   - Dichotomy via 1-APs. Tangible small cute FPs ?
   - Max Cut: the ultimate baddie ?
2. Tractability results in the value oracle model ?
   - FPFM: function minimisation with a given nice FP
   - Submodularity on lattices [AK, Larose; Kuivinen]
   - α-bisubmodular functions ?
   - k-submodular functions ? [Huber, Kolmogorov]

Expressive Power

A set $\Gamma \subseteq Q_D$ can express a function $g \in Q_D^{(n)}$ if there is an instance $f(x_1, \ldots, x_n, y_1, \ldots, y_m) = \sum_{i=1}^{q} w_i \cdot f_i(\mathbf{x}_i)$ of VCSP(Γ) such that

$$g(x_1, \ldots, x_n) = \min_{y_1, \ldots, y_m} f(x_1, \ldots, x_n, y_1, \ldots, y_m) + const.$$

Easy: Γ can express $g \Rightarrow \text{VCSP}(\Gamma) \simeq \text{VCSP}(\Gamma \cup \{g\})$.

Theorem 8 (Cohen, Cooper, Jeavons' 06)
For any finite $\Gamma \subseteq Q_D$ and $g \in Q_D$,

- *either Γ can express g, or*
- *there is an FP of Γ which is not FP of g.*

Which Functions are α-Bisubmodular?

Let $f \in Q^{(n)}_{\{-1,0,1\}}$. Say that f is submodular in each orthant if, for any $a_1, a_2, \ldots, a_n \in \{-1, 1\}$, the restriction of f to $\prod_{i=1}^{n} \{0, a_i\}$ is submodular.

Let $U(f)$ denote the set of all unary functions of the form $g(x) = f(b_1, x, \ldots, b_l, x, \ldots, xb_n)$. A function $g \in F^{(1)}_D$ is α-bisubmodular if $(1 + \alpha) \cdot g(0) \leq \alpha \cdot g(1) + g(-1)$.

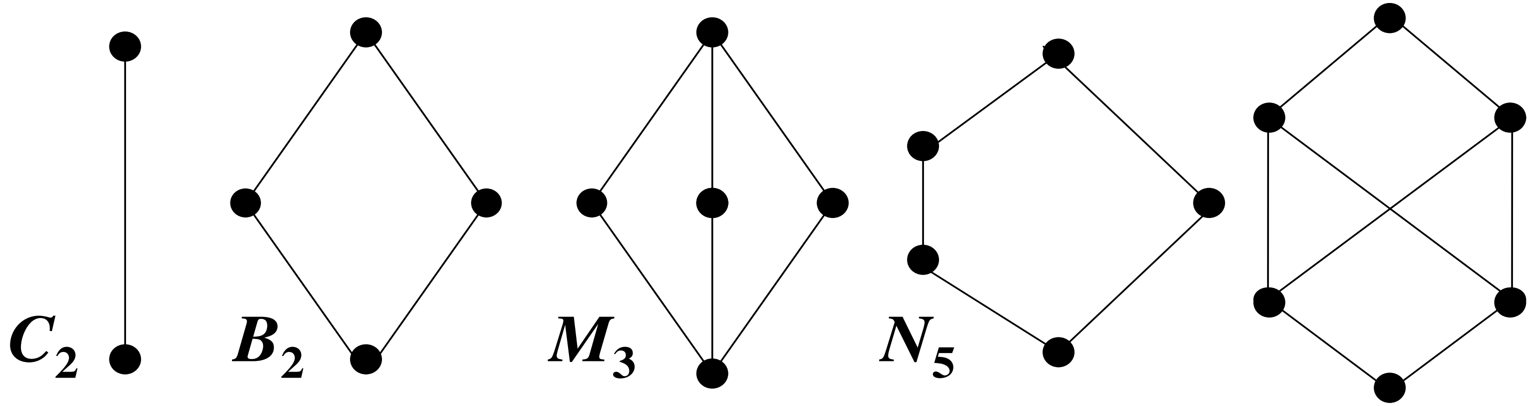
Lemma 1 (Huber, AK, Powell '12)
For any $f \in Q_{\{-1,0,1\}}$, f is α-bisubmodular iff

1. *f is submodular in each orthant, and*
2. *each function in $U(f)$ is α-bisubmodular*

Lattices

A lattice $\mathcal{L}$ is a partial order in which any $a, b \in \mathcal{L}$ have

- a least common upper bound (join) $a \sqcup b$, and
- a greatest common lower bound (meet) $a \sqcap b$



A distributive lattice is one representable by subsets of a set (or, equivalently, containing neither M_3 nor N_5).

Submodularity on lattices

Definition 7 *Let $\mathcal{L}$ be a lattice on a finite set D. A function $f : D^n \to \mathbb{Q}$ is called submodular on $\mathcal{L}$ if*

$$f(\mathbf{a}) + f(\mathbf{b}) \geq f(\mathbf{a} \sqcup \mathbf{b}) + f(\mathbf{a} \sqcap \mathbf{b}) \text{ for all } \mathbf{a}, \mathbf{b} \in D^n.$$

Problem 1 *Fix a finite lattice $\mathcal{L}$ and let $\mathrm{SFM}(\mathcal{L})$ be the problem of minimising a given n-ary submodular function on $\mathcal{L}$. Is there an algorithm solving $\mathrm{SFM}(\mathcal{L})$ in polynomial time in n (in the oracle value model)?*

NB. True for the two-element lattice C_2 (Grötschel et al.).