Solving the Identifying Code Set Problem with Grouped Independent Support*

Anna L.D. Latour
National University of Singapore, Singapore

Arunabha Sen
Arizona State University, Tempe, AZ, USA

Kuldeep S. Meel
National University of Singapore, Singapore

1 Introduction

We introduce an extension of the concept of an *independent support* of a Boolean formula [1]: *grouped independent support (GIS)*. Then, we present a case study of how we can reduce the well-studied and NP-hard constraint optimisation problem of finding a minimal *generalised identifying code set (GICS)* for a network, to finding a minimal GIS. Finding a minimal GIS is computationally harder than finding a minimal GICS. Specifically: checking if a candidate for a GIS is indeed a GIS is in co-NP, while checking if a candidate GICS is indeed a GICS is in P. However, we show that reducing the minimal GICS problem to the minimal GIS problem yields an exponential reduction in the encoding size, compared to the current state of the art in GICS solving [4]. Such a dramatic reduction presents an opportunity to solve the GICS problem for much larger problem instances than what is possible with the current state of the art, *if* we can find an algorithm for finding a minimal GIS that is fast in practice. We introduce and implement such an algorithm, gismo[1], and demonstrate its efficiency.

2 Grouped Independent Support

We use $\sigma := X \mapsto \{0,1\}$ to denote an assignment of truth values to the Boolean variables X, and $\sigma_{\downarrow Y} := Y \mapsto \{0,1\}$ to denote an assignment that is *projected* on a subset $Y \subseteq X$. We denote the set of solutions of a formula F with $Sol(F)$. We also define the *support* of a set of sets $\mathcal{X}$ as $sup\,(\mathcal{X}) := \bigcup_{X \in \mathcal{X}} X$. Let T be a set of items, and let $S \subseteq T$ be a subset. Given a set $\mathcal{C}$ of constraints on sets, we call S *set-minimal w.r.t.* $\mathcal{C}$ if S satisfies all constraints in $\mathcal{C}$ and there exists no proper subset of S that also satisfies all those constraints. We call S a *cardinality-minimal* set if S is minimal, and there exists no $S' \subseteq T$, with $|S'| < |S|$, that is also set-minimal. Now we define the GIS of a Boolean formula as follows:

▶ **Definition 1.** *Given a Boolean formula $F(Z, A)$, with $Z \cap A = \varnothing$, and a partition $\mathcal{G}$ of Z. We refer to an element $G \in \mathcal{G}$ as a* variable group. *The set $\mathcal{I} \subseteq \mathcal{G}$ is a* grouped independent support (GIS) *of $\langle F(Z, A), \mathcal{G}\rangle$ if the following holds:*

$$\forall \sigma_1, \sigma_2 \in Sol(F).\, \left(\left(\sigma_{1\downarrow sup(\mathcal{I})} = \sigma_{2\downarrow sup(\mathcal{I})}\right) \leftrightarrow \left(\sigma_{1\downarrow Z} = \sigma_{2\downarrow Z}\right)\right). \tag{1}$$

Intuitively, Equation (1) says that the truth values that a solution σ assigns to the variables in $sup\,(\mathcal{I})$ uniquely define the truth values that σ assigns to the variables in $Z \setminus sup\,(\mathcal{I})$, if $\mathcal{I}$ is a GIS of $\langle F, \mathcal{G}\rangle$. The auxiliary variables A cannot be part of the GIS.

Observe that $\mathcal{G}$ is in itself a grouped independent support of F. Our algorithm, gismo (inspired by a state-of-the-art independent support algorithm [8]), starts by assuming that all

* Extended abstract of our IJCAI 2023 paper [3]. Preprint: `https://arxiv.org/abs/2306.15693`.

[1] Available open-source at `https://github.com/meelgroup/gismo`.

groups are part of the GIS for F. Then, it iterates over the variable groups in $\mathcal{G}$, and checks for each group if it can be removed from the candidate GIS. If yes, that group is removed, if not, it remains. After considering each group, gismo returns a set-minimal GIS.

In gismo, we use a version of Padoa's Theorem [5] to check if a variable $z_i \in Z$ must be part of an independent support for $F(Z, A)$. Specifically, we construct a formula ψ_i that is unsatisfiable if z_i is not needed in the independent support. If ψ_i is satisfiable for at least one $z_i \in G$, then group G remains in the candidate GIS. Each ψ_i is evaluated at most once.

3 The Generalised Identifying Code Set Problem

To demonstrate the merits of GIS, we present a case study, in the form of a novel modelling and solving method for a generalised version of the *identifying code set (ICS)* [2] problem.

▶ **Definition 2.** *Given an undirected, loop-free graph $\Gamma := (V, E)$ with nodes V and edges E, and a $D \subseteq V$. We define the* signature *of $U \subseteq V$ as the following tuple: $s_U := \langle S_U^0, S_U^1 \rangle$, where $S_U^0 := U \cap D$ and $S_U^1 := N_1^+(U) \cap D$. Here, $N_1^+(U)$ is the* closed 1-neighbourhood *of U, which contains U and all direct neighbours of nodes in U: $N_1^+(U) := U \cup \bigcup_{u \in U} N_1(u)$.*

▶ **Definition 3.** *Given a graph $\Gamma := (V, E)$, a* maximum identifiable set size $1 \leq k \leq |V|$*, and $D \subseteq V$. We call D a* generalised identifying code set (GICS) *of $\langle \Gamma, k \rangle$ if, for all $U, W \subseteq V$ with $|U| \leq k$, $|W| \leq k$ and $U \neq W$, we have $s_U \neq s_W$. Hence, if D is a GICS of $\langle \Gamma, k \rangle$, then the signatures of all subsets of V with cardinality at most k are unique. Given a $\langle \Gamma, k \rangle$, the* GICS problem *asks to find a $D \subseteq V$ such that D is a GICS of $\langle \Gamma, k \rangle$, and $|D|$ is minimised.*

The current state of the art in ICS solving supports $k = 1$ only, and encodes the problem as an *integer-linear program (ILP)*, employing an off-the-shelf *mixed-integer problem (MIP)* solver to find a cardinality-minimal D [4]. However, naively extending this ILP encoding to support our *GICS* setting (with $k \geq 1$) would cause the number of linear constraints in the encoding to grow as as $O\left(\binom{|V|}{k}^2\right)$. Instead, we create a Boolean formula $F_k(X \cup Y, A) = F_{\text{signature}} \wedge F_{\text{cardinality},k}$ (the possibly empty set of auxiliary variables A originates from the translation of the cardinality constraint into CNF), with

$$F_{\text{signature}} := \bigwedge_{v \in V} \left(y_v \leftrightarrow \bigvee_{u \in N_1^+(v)} x_u \right) \quad \text{and} \quad F_{\text{cardinality},k} := \sum_{v \in V} x_v \leq k. \tag{2}$$

In conjunctive normal form (CNF), F_k has $O(k \cdot |V| + |E|)$ clauses [6, 7], *e.g.*, we obtain a model that grows linearly with the problem size. We now make the following claim:

▶ **Lemma 4.** *Given a loop-free, undirected network $\Gamma := (V, E)$ on nodes V and edges E, a maximum identifiable set size $0 < k \leq |V|$, and given a GIS of $F_k(X \cup Y, A)$ $\mathcal{I} \subseteq \mathcal{G}$ with groups $\mathcal{G} := \{G_v := \{x_v, y_v\}\}$. The set $D := \{v \in V \mid G_v \in \mathcal{I}\}$ is a GICS of Γ.*

4 Experimental Results

In our experiments on 50 networks of different types and 9 different values of k, we compared the performance of gismo to that of the ILP method, pbpbs, that is based on the state of the art [4]. Overall, gismo solved 289/450 problem instances, while pbpbs solved only 36/450. We also find that gismo can solve instances that are up to 40× larger than the instances that can be solved by pbpbs, and can do so up to 520× faster in terms of median running time. In the majority of instances that were solved by both methods, the set-minimal solution from gismo was at most 10% larger than the cardinality-minimal solution from pbpbs.

Acknowledgements

This work was supported in part by National Research Foundation Singapore under its NRF Fellowship Programme [NRF-NRFFAI1-2019-0004], Ministry of Education Singapore Tier 2 grant MOE-T2EP20121-0011, and Ministry of Education Singapore Tier 1 Grant [R-252-000-B59-114]. The computational work for this article was performed on resources of the National Supercomputing Centre, Singapore `www.nscc.sg`. One of the authors (Sen) acknowledges the support of the National University of Singapore during his sabbatical there.

References

1 Supratik Chakraborty, Daniel J. Fremont, Kuldeep S. Meel, Sanjit A. Seshia, and Moshe Y. Vardi. Distribution-aware sampling and weighted model counting for SAT. In *AAAI*, pages 1722–1730. AAAI Press, 2014.

2 Mark G. Karpovsky, Krishnendu Chakrabarty, and Lev B. Levitin. On a new class of codes for identifying vertices in graphs. *IEEE Trans. Inf. Theory*, 44(2):599–611, 1998.

3 Anna L.D. Latour, Arunabha Sen, and Kuldeep S. Meel. Solving the identifying code set problem with grouped independent support. In *Proceedings of the 32nd International Joint Conference on Artificial Intelligence (IJCAI23)*, 8 2023.

4 Malhar Padhee, Reetam Sen Biswas, Anamitra Pal, Kaustav Basu, and Arunabha Sen. Identifying unique power system signatures for determining vulnerability of critical power system assets. *SIGMETRICS Perform. Evaluation Rev.*, 47(4):8–11, 2020.

5 A. Padoa. Essai d'une théorie algébrique des nombres entiers, précédé d'une introduction logique à une theorie déductive quelconque. *Bibliothèque du Congrès international de philosophie*, 3:309–365, 1901.

6 Tobias Philipp and Peter Steinke. PBLib — A library for encoding Pseudo-Boolean constraints into CNF. In *SAT*, volume 9340 of *Lecture Notes in Computer Science*, pages 9–16. Springer, 2015.

7 Carsten Sinz. Towards an optimal CNF encoding of Boolean cardinality constraints. In *CP*, volume 3709 of *Lecture Notes in Computer Science*, pages 827–831. Springer, 2005.

8 Mate Soos and Kuldeep S. Meel. Arjun: An efficient independent support computation technique and its applications to counting and sampling. In *ICCAD*, page 71. ACM, 2022.