On Sketching Approximations for Symmetric Boolean CSPs

Joanna Boyland
Harvard College, Harvard University, Cambridge, MA, USA

Michael Hwang
Harvard College, Harvard University, Cambridge, MA, USA

Tarun Prasad
Harvard College, Harvard University, Cambridge, MA, USA

Noah Singer
Department of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA
Harvard College, Harvard University, Cambridge, MA, USA

Santhoshini Velusamy
School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA

Abstract

A Boolean maximum constraint satisfaction problem, $\mathsf{Max\text{-}CSP}(f)$, is specified by a predicate $f : \{-1,1\}^k \to \{0,1\}$. An n-variable instance of $\mathsf{Max\text{-}CSP}(f)$ consists of a list of constraints, each of which applies f to k distinct literals drawn from the n variables. For $k = 2$, Chou, Golovnev, and Velusamy [8] obtained explicit ratios characterizing the $\sqrt{n}$-space *streaming* approximability of every predicate. For $k \geq 3$, Chou, Golovnev, Sudan, and Velusamy [7] proved a general dichotomy theorem for $\sqrt{n}$-space *sketching* algorithms: For every f, there exists $\alpha(f) \in (0,1]$ such that for every $\epsilon > 0$, $\mathsf{Max\text{-}CSP}(f)$ is $(\alpha(f) - \epsilon)$-approximable by an $O(\log n)$-space *linear* sketching algorithm, but $(\alpha(f) + \epsilon)$-approximation sketching algorithms require $\Omega(\sqrt{n})$ space.

In this work, we give closed-form expressions for the sketching approximation ratios of multiple families of *symmetric* Boolean functions. Letting $\alpha'_k = 2^{-(k-1)}(1-k^{-2})^{(k-1)/2}$, we show that for odd $k \geq 3$, $\alpha(k\mathsf{AND}) = \alpha'_k$, and for even $k \geq 2$, $\alpha(k\mathsf{AND}) = 2\alpha'_{k+1}$. Thus, for every k, $k\mathsf{AND}$ can be $(2 - o(1))2^{-k}$-approximated by $O(\log n)$-space *sketching* algorithms; we contrast this with a lower bound of Chou, Golovnev, Sudan, Velingker, and Velusamy [5] implying that *streaming* $(2+\epsilon) \cdot 2^{-k}$-approximations require $\Omega(n)$ space! We also resolve the ratio for the "at-least-$(k-1)$-1's" function for all even k; the "exactly-$\frac{k+1}{2}$-1's" function for odd $k \in \{3, \ldots, 51\}$; and fifteen other functions. We stress here that for general f, the dichotomy theorem in [7] only implies that $\alpha(f)$ can be computed to arbitrary precision in PSPACE, and thus closed-form expressions need not have existed *a priori*. Our analyses involve identifying and exploiting structural "saddle-point" properties of this dichotomy.

Separately, for all *threshold* functions, we give optimal "bias-based" approximation algorithms generalizing [8] while simplifying [7]. Finally, we investigate the $\sqrt{n}$-space *streaming* lower bounds in [7], and show that they are *incomplete* for 3AND, i.e., they fail to rule out $(\alpha(3\mathsf{AND}) - \epsilon)$-approximations in $o(\sqrt{n})$ space.

2012 ACM Subject Classification Theory of computation → Approximation algorithms analysis; Theory of computation → Streaming, sublinear and near linear time algorithms; Theory of computation → Discrete optimization

Keywords and phrases Streaming algorithms, constraint satisfaction problems, approximability

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2022.38

Category APPROX

Related Version *Full Version*: `https://arxiv.org/abs/2112.06319` [3]
Included in fourth author's thesis: `https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37371750` [21]

Supplementary Material *Interactive Resource (Interactive Mathematica notebook)*: https://notebookarchive.org/2022-03-a5vpzhg/

Funding *Noah Singer*: Supported by the Harvard College Research Program, and by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE2140739. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.
Santhoshini Velusamy: Supported in part by a Google Ph.D. Fellowship, a Simons Investigator Award to Madhu Sudan, and NSF Awards CCF 1715187 and CCF 2152413.

Acknowledgements All authors of this paper are or were supervised by Madhu Sudan. We would like to thank him for advice and guidance on both technical and organizational aspects of this project, as well as for many helpful comments on earlier drafts of this paper.

1 Introduction

In this work, we consider the *streaming approximability* of various *Boolean constraint satisfaction problems*, and we begin by defining these terms. See [7, §1.1-2] for more details on the definitions.

1.1 Setup: The streaming approximability of Boolean CSPs

1.1.1 Boolean CSPs

Let $f : \{-1,1\}^k \to \{0,1\}$ be a Boolean function. In an n-variable instance of the problem $\mathsf{Max\text{-}CSP}(f)$, a *constraint* is a pair $C = (\mathbf{b}, \mathbf{j})$, where $\mathbf{j} = (j_1, \dots, j_k) \in [n]^k$ is a k-tuple of distinct *indices*, and $\mathbf{b} = (b_1, \dots, b_k) \in \{-1,1\}^k$ is a *negation pattern.*

For Boolean vectors $\mathbf{a} = (a_1, \dots, a_n), \mathbf{b} = (b_1, \dots, b_n) \in \{-1,1\}^n$, let $\mathbf{a} \odot \mathbf{b}$ denote their *coordinate-wise product* $(a_1b_1, \dots, a_nb_n)$. An *assignment* $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_n) \in \{-1,1\}^n$ satisfies C iff $f(\mathbf{b} \odot \boldsymbol{\sigma}|_{\mathbf{j}}) = 1$, where $\boldsymbol{\sigma}|_{\mathbf{j}}$ is the k-tuple $(\sigma_{j_1}, \dots, \sigma_{j_k})$ (i.e., $\boldsymbol{\sigma}$ satisfies C iff $f(b_1\sigma_{j_1}, \dots, b_k\sigma_{j_k}) = 1$). An *instance* Ψ of $\mathsf{Max\text{-}CSP}(f)$ consists of constraints $C_1, \dots, C_m$ with non-negative weights $w_1, \dots, w_m$ where $C_i = (\mathbf{j}(i), \mathbf{b}(i))$ and $w_i \in \mathbb{R}$ for each $i \in [m]$; the *value* $\mathsf{val}_\Psi(\boldsymbol{\sigma})$ of an assignment $\boldsymbol{\sigma}$ to Ψ is the (weighted) fraction of constraints in Ψ satisfied by $\boldsymbol{\sigma}$, i.e., $\mathsf{val}_\Psi(\boldsymbol{\sigma}) \stackrel{\text{def}}{=} \frac{1}{W} \sum_{i\in[m]} w_i \cdot f(\mathbf{b}(i) \odot \boldsymbol{\sigma}|_{\mathbf{j}(i)})$, where $W = \sum_{i=1}^m w_i$. The *value* val_Ψ of an instance Ψ is the maximum value of any assignment $\boldsymbol{\sigma} \in \{-1,1\}^n$, i.e., $\mathsf{val}_\Psi \stackrel{\text{def}}{=} \max_{\boldsymbol{\sigma}\in\{-1,1\}^n} \mathsf{val}_\Psi(\boldsymbol{\sigma})$.

1.1.2 Approximations to CSPs

For $\alpha \in [0,1]$, we consider the problem of *α-approximating $\mathsf{Max\text{-}CSP}(f)$*. In this problem, the goal of an algorithm $\mathcal{A}$ is to, on input an instance Ψ, output an estimate $\mathcal{A}(\Psi)$ such that with probability at least $\frac{2}{3}$, $\alpha \cdot \mathsf{val}_\Psi \le \mathcal{A}(\Psi) \le \mathsf{val}_\Psi$. For $\beta < \gamma \in [0,1]$, we also consider the closely related (β,γ)-$\mathsf{Max\text{-}CSP}(f)$. In this problem, the input instance Ψ is promised to either satisfy $\mathsf{val}_\Psi \le \beta$ or $\mathsf{val}_\Psi \ge \gamma$, and the goal is to decide which is the case with probability at least $\frac{2}{3}$.

1.1.3 Streaming and sketching algorithms for CSPs

For various Boolean functions f, we consider algorithms which attempt to approximate $\mathsf{Max\text{-}CSP}(f)$ instances in the *(single-pass, insertion-only) space-s streaming setting.* Such algorithms can only use space s (which is ideally small, such as $O(\log n)$, where n is the number of variables in an input instance), and, when given as input a CSP instance Ψ, can only read the list of constraints in a single, left-to-right pass.

We also consider a (seemingly) weak class of streaming algorithms called *sketching algorithms*, where the algorithm's output is determined by an length-s string called a "sketch" produced from the input stream, and the sketch itself has the property that the sketch of the concatenation of two streams can be computed from the sketches of the two component streams. (See [7, §3.3] for a formal definition.) A special case of sketching algorithms are *linear sketches*, where each sketch (i.e., element of $\{0,1\}^s$) encodes an element of a vector space and we perform vector addition to combine two sketches.

1.2 Prior work and motivations

1.2.1 Prior results on streaming and sketching Max-CSP(f)

We first give a brief review of what is already known about the streaming and sketching approximability of Max-CSP(f). For $f : \{-1,1\}^k \to \{0,1\}$, let $\rho(f) \stackrel{\text{def}}{=} \Pr_{\mathbf{b}\sim\mathsf{Unif}(\{-1,1\}^k)}[f(\mathbf{b}) = 1]$, where $\mathsf{Unif}(\{-1,1\}^k)$ denotes the uniform distribution on $\{-1,1\}^k$. For every f, the Max-CSP(f) problem has a trivial $\rho(f)$-approximation algorithm given by simply outputting $\rho(f)$ since $\mathbb{E}_{\mathbf{a}\sim\mathsf{Unif}(\{-1,1\}^n)}[\mathsf{val}_\Psi(\mathbf{a})] = \Pr_{\mathbf{b}\sim\mathsf{Unif}(\{-1,1\}^k)}[f(\mathbf{b}) = 1] = \rho(f)$. We refer to a function f as *approximation-resistant* for some class of algorithms (e.g., streaming or sketching algorithms with some space bound) if it cannot be $(\rho(f)+\epsilon)$-approximated for any constant $\epsilon > 0$. Otherwise, we refer to f as *approximable* for the class of algorithms.

The first two CSPs whose $o(\sqrt{n})$-space streaming approximabilities were resolved were Max-2XOR and Max-2AND. Kapralov, Khanna, and Sudan [18] showed that Max-2XOR is approximation-resistant to $o(\sqrt{n})$-space streaming algorithms. Later, Chou, Golovnev, and Velusamy [8], building on earlier work of Guruswami, Velusamy, and Velingker [12], gave an $O(\log n)$-space linear sketching algorithm which $(\frac{4}{9}-\epsilon)$-approximates Max-2AND for every $\epsilon > 0$ and showed that $(\frac{4}{9}+\epsilon)$-approximations require $\Omega(\sqrt{n})$ space, even for streaming algorithms.

In two recent works [7, 6], Chou, Golovnev, Sudan, and Velusamy proved so-called *dichotomy theorems* for sketching CSPs. In [7], they prove the dichotomy for CSPs over the Boolean alphabet with negations of variables (i.e., the setup we described in Section 1.1.1). In [6], they extend it to the more general case of CSPs over finite alphabets.[1] See [6, §1] and [21] for more general background on CSPs in the streaming setting.

[7] is most relevant for our purposes, as it concerns Boolean CSPs. For a fixed constraint function $f : \{-1,1\}^k \to \{0,1\}$, the main result in [7] is the following *dichotomy theorem*: For any $0 \le \gamma < \beta \le 1$, either

1. (β,γ)-Max-CSP(f) has an $O(\log n)$-space linear sketching algorithm, *or*
2. For all $\epsilon > 0$, sketching algorithms for $(\beta+\epsilon,\gamma-\epsilon)$-Max-CSP($f$) require $\Omega(\sqrt{n})$ space.

Distinguishing whether (1) or (2) applies is equivalent to deciding whether two convex polytopes (which depend on f,γ,β) intersect. We omit a technical statement of this criterion, and instead focus on the following corollary: there exists an $\alpha(f) \in [0,1]$ such that Max-CSP(f) can be $(\alpha(f)-\epsilon)$-approximated by $O(\log n)$-space linear sketches, but not $(\alpha(f)+\epsilon)$-approximated by $o(\sqrt{n})$-space sketches, for all $\epsilon > 0$; furthermore, $\alpha(f)$ equals the solution to an explicit minimization problem, which we describe in Section 2.1 (in the special case where f is symmetric).

[1] More precisely, [7] and [6] both consider the more general case of CSPs defined by *families* of functions of a specific arity. We do not need this generality for the purposes of our paper, and therefore omit it.

A priori, it may be possible to achieve an $(\alpha(f)+\epsilon)$-approximation with a $o(\sqrt{n})$-space *streaming* algorithm. But [7] also extends the lower bound (case 2 of the dichotomy) to cover streaming algorithms when special objects called *padded one-wise pairs* exist. See Section 2.4 below for a definition (again, specialized for symmetric functions). The padded one-wise pair criterion is sufficient to recover all previous streaming approximability results for Boolean functions (i.e., [18, 8]), and prove several new ones. In particular, [7] proves that if $f : \{-1,1\}^k \to \{0,1\}$ has the property that there exists $\mathcal{D} \in \Delta(f^{-1}(1))$ such that $\mathbb{E}_{\mathbf{b}\sim\mathcal{D}}[b_i] = 0$ for all $i \in [k]$ (where $[k] \stackrel{\text{def}}{=} \{1,\dots,k\}$), then $\mathsf{Max\text{-}CSP}(f)$ is streaming approximation-resistant. For symmetric Boolean CSPs, they also prove the converse, and thus give a complete characterization for approximation resistance [7, Lemma 2.14]. However, besides $\mathsf{Max\text{-}2AND}$, [7] does not explicitly analyze the approximation ratio of any CSP that is "approximable", i.e., not approximation resistant.

1.2.2 Questions from previous work

In this work, we address several major questions about streaming approximations for Boolean CSPs which Chou, Golovnev, Sudan, and Velusamy [7] leave unanswered:

1. Can the framework in [7] be used to find closed-form sketching approximability ratios $\alpha(f)$ for approximable problems $\mathsf{Max\text{-}CSP}(f)$ beyond $\mathsf{Max\text{-}2AND}$?
2. As observed in [5, §1.3], [7] implies the following "trivial upper bound" on streaming approximability: for all f, $\alpha(f) \leq 2\rho(f)$. How tight is this upper bound?
3. Does the streaming lower bound (the "padded one-wise pair" criterion) in [7] suffice to resolve the streaming approximability of every function?
4. The optimal $(\alpha(f)-\epsilon)$-approximation algorithm for $\mathsf{Max\text{-}CSP}(f)$ in [7] requires running a "grid" of $O(1/\epsilon^2)$ distinguishers for (β,γ)-$\mathsf{Max\text{-}CSP}(f)$ distinguishing problems in parallel. Can we obtain simpler optimal sketching approximations?

1.3 Our results

We study the questions in Section 1.2.2 for *symmetric* Boolean CSPs. Symmetric Boolean functions are those functions that depend only on the Hamming weight of the input, i.e., number of 1's in the input.[2] For a set $S \subseteq [k]$, we define $f_{S,k} : \{-1,1\}^k \to \{0,1\}$ as the indicator function for the set $\{\mathbf{b} \in \{-1,1\}^k : \mathrm{wt}(\mathbf{b}) \in S\}$ (where $\mathrm{wt}(\mathbf{b})$ denotes the Hamming weight of $\mathbf{b}$). That is, $f_{S,k}(\mathbf{x}) = 1$ if and only if $\mathrm{wt}(\mathbf{x}) \in S$. Some well-studied examples of functions in this class include $k\mathsf{AND} = f_{\{k\},k}$, the *threshold functions* $\mathsf{Th}_k^i = f_{\{i,i+1,\dots,k\},k}$, and "exact weight" functions $\mathsf{Ex}_k^i = f_{\{i\},k}$.[3]

---

[2] Note that the inputs are in $\{-1,1\}^k$; we define the Hamming weight as the number of 1's, and not -1's (which is arguably more "natural" under the mapping $b \in \{0,1\} \mapsto (-1)^b \in \{-1,1\}$), for consistency with [7].

[3] By [7, Lemma 2.14], if S contains elements $s \leq \frac{k}{2}$ and $t \geq \frac{k}{2}$, not necessarily distinct, then $f_{S,k}$ supports one-wise independence and is therefore approximation-resistant (even to streaming algorithms). Thus, we focus on the case where all elements of S are either larger than or smaller than $\frac{k}{2}$. Moreover, note that if $S' = \{k-s : s \in S\}$, every instance of $\mathsf{Max\text{-}CSP}(f_{S,k})$ can be viewed as an instance of $\mathsf{Max\text{-}CSP}(f_{S',k})$ with the same value, since for any constraint $C = (\mathbf{b},\mathbf{j})$ and assignment $\boldsymbol{\sigma} \in \{-1,1\}^n$, we have $f_{S,k}(\mathbf{b} \odot \boldsymbol{\sigma}|_{\mathbf{j}}) = f_{S',k}(\mathbf{b} \odot (-\boldsymbol{\sigma})|_{\mathbf{j}})$. Thus, we further narrow our focus to the case where every element of S is larger than $\frac{k}{2}$.

1.3.1 The sketching approximability of Max-kAND

Chou, Golovnev, and Velusamy [8] showed that $\alpha(2\mathsf{AND}) = \frac{4}{9}$ (and $(\frac{4}{9}+\epsilon)$-approximation can be ruled out even for $o(\sqrt{n})$-space streaming algorithms). For $k \geq 3$, while Chou, Golovnev, Velusamy, and Sudan [7] give optimal sketching approximation algorithms for Max-kAND, they do not explicitly analyze the approximation ratio $\alpha(k\mathsf{AND})$, and show only that it lies between 2^{-k} and $2^{-(k-1)}$.

In this paper, we analyze the dichotomy theorem in [7], and obtain a closed-form expression for the sketching approximability of Max-kAND for every k. For odd $k \geq 3$, define the constant

$$\alpha'_k \stackrel{\text{def}}{=} \left(\frac{(k-1)(k+1)}{4k^2}\right)^{(k-1)/2} = 2^{-(k-1)} \cdot \left(1 - \frac{1}{k^2}\right)^{(k-1)/2}. \tag{1}$$

In Section 4, we prove the following:

▶ **Theorem 1.** *For odd $k \geq 3$, $\alpha(k\mathsf{AND}) = \alpha'_k$, and for even $k \geq 2$, $\alpha(k\mathsf{AND}) = 2\alpha'_{k+1}$.*

Since $\rho(k\mathsf{AND}) = 2^{-k}$, Theorem 1 also has the following important corollary:

▶ **Corollary 2.** $\lim_{k\to\infty} \frac{\alpha(k\mathsf{AND})}{2\rho(k\mathsf{AND})} = 1$.

Recall that [7] implies that $\alpha(f) \leq 2\rho(f)$ for all functions f. Indeed, Chou, Golovnev, Sudan, Velusamy, and Velingker [5] show that any function f cannot be $(2\rho(f)+\epsilon)$-approximated even by $o(n)$-space streaming algorithms. On the other hand, in Section 1.3.3 below, we describe simple $O(\log n)$-space sketching algorithms for Max-kAND achieving the optimal ratio from [7]. Thus, as $k \to \infty$, these algorithms achieve an asymptotically optimal approximation ratio even among $o(n)$-space streaming algorithms!

1.3.2 The sketching approximability of other symmetric functions

We also analyze the sketching approximability of a number of other symmetric Boolean functions. Specifically, for the threshold functions Th_k^{k-1} for even k, we show that:

▶ **Theorem 3.** *For even $k \geq 2$, $\alpha(\mathsf{Th}_k^{k-1}) = \frac{k}{2}\alpha'_{k-1}$.*

We prove Theorem 3 in Section 5.1 using techniques similar to our proof of Theorem 1. We also provide partial results for $\mathsf{Ex}_k^{(k+1)/2}$, including closed forms for small k and an asymptotic analysis of $\alpha(\mathsf{Ex}_k^{(k+1)/2})$:

▶ **Theorem 4** (Informal version of Theorem 25)**.** *For odd $k \in \{3, \dots, 51\}$, there is an explicit expression for $\alpha(\mathsf{Ex}_k^{(k+1)/2})$ as a function of k.*

▶ **Theorem 5.** $\lim_{\text{odd } k\to\infty} \frac{\alpha\left(\mathsf{Ex}_k^{(k+1)/2}\right)}{\rho\left(\mathsf{Ex}_k^{(k+1)/2}\right)} = 1$.

We prove Theorems 4 and 5 in Section 5.2. Finally, in Section 5.3, we explicitly resolve fifteen other cases (e.g., $f_{\{2,3\},3}$ and $f_{\{4\},5}$) not covered by Theorems 1, 3, and 4.

1.3.3 Simple approximation algorithms for threshold functions

Chou, Golovnev, and Velusamy's optimal $(\frac{4}{9}-\epsilon)$-approximation for 2AND [8], like Guruswami, Velingker, and Velusamy's earlier $(\frac{2}{5}-\epsilon)$-approximation [12], is based on measuring a quantity called the *bias* of an instance Ψ, denoted $\mathsf{bias}(\Psi)$, which is defined as follows: For each $i \in [n]$,

$\mathsf{diff}_i(\Psi)$ is the difference in total weight between constraints where x_i occurs positively and negatively, and $\mathsf{bias}(\Psi) \stackrel{\text{def}}{=} \frac{1}{km} \sum_{i=1}^n |\mathsf{diff}_i(\Psi)| \in [0,1]$.[4] In the sketching setting, $\mathsf{bias}(\Psi)$ can be estimated using standard ℓ_1-norm sketching algorithms [16, 17].

In Section 7, we give simple optimal bias-based approximation algorithms for threshold functions:

▶ **Theorem 6.** *Let $f_{S,k} = \mathsf{Th}_k^i$ be a threshold function. Then for every $\epsilon > 0$, there exists a piecewise linear function $\gamma : [-1,1] \to [0,1]$ and a constant $\epsilon' > 0$ such that the following is a sketching $(\alpha(f_{S,k}) - \epsilon)$-approximation for Max-CSP$(f_{S,k})$: On input Ψ, compute an estimate $\widehat{b}$ for bias(Ψ) up to a multiplicative $(1 \pm \epsilon')$ error and output $\gamma(\widehat{b})$.*

Our construction generalizes the algorithm in [8] for 2AND to all threshold functions, and is also a simplification, since the [8] algorithm computes a more complicated function of $\widehat{b}$.

For all CSPs whose approximability we resolve in this paper, we apply an analytical technique which we term the "max-min method;" see the discussion in Section 2.3 below. For such CSPs, our algorithm can be extended to solve the problem of outputting an approximately optimal *assignment* (instead of just the value of such an assignment). Indeed, for this problem, we give a simple randomized streaming algorithm using $O(n)$ space and time:

▶ **Theorem 7** (Informal version of Theorem 34)**.** *Let $f_{S,k}$ be a function for which the max-min method applies, such as kAND, or Th_k^{k-1} (for even k). Then there exists a constant $p^* \in [0,1]$ such that following algorithm, on input Ψ, outputs an assignment with expected value at least $\alpha(f_{S,k})\mathsf{val}_\Psi$: Assign variable i to 1 if $\mathsf{diff}_i(\Psi) \geq 0$ and -1 otherwise, and then flip each variable's assignment independently with probability p^*.*

Our algorithm can potentially be derandomized using universal hash families, as in Biswas and Raman's recent derandomization [1] of the Max-2AND algorithm in [8].

1.3.4 Sketching vs. streaming approximability

Theorem 1 implies that $\alpha(3\mathsf{AND}) = \frac{2}{9}$. We prove that the padded one-wise pair criterion of Chou, Golovnev, Sudan, and Velusamy [7] is not sufficient to completely resolve the *streaming* approximability of Max-3AND:

▶ **Theorem 8** (Informal version of Theorem 12 + Observation 13)**.** *The padded one-wise pair criterion in [7] does not rule out a $o(\sqrt{n})$-space streaming $(\frac{2}{9} + \epsilon)$-approximation for 3AND for every $\epsilon > 0$; however, it does rule out such an algorithm for $\epsilon \gtrapprox 0.0141$.*

We state these results formally in Section 2.4 and prove them in Section 6. Separately, Theorem 3 implies that $\alpha(\mathsf{Th}_4^3) = \frac{4}{9}$, and the padded one-wise pair criterion *can* be used to show that $(\frac{4}{9} + \epsilon)$-approximating Max-CSP(Th_4^3) requires $\Omega(\sqrt{n})$ space in the streaming setting (see Observation 22 below).

1.4 Related work

The classical approximability of Max-kAND has been the subject of intense study, both in terms of algorithms [11, 10, 26, 23, 25, 13, 14, 4] and hardness-of-approximation [15, 24, 22, 19, 9, 20], given its intimate connections to k-bit PCPs. Charikar, Makarychev, and

[4] [12, 8] did not normalize by $\frac{1}{kW}$.

Makarychev [4] constructed an $\Omega(k2^{-k})$-approximation to Max-kAND, while Samorodnitsky and Trevisan [20] showed that $k2^{-(k-1)}$-approximations and $(k+1)2^{-k}$-approximations are **NP**- and UG-hard, respectively.

Interestingly, recalling that $\alpha(k\mathsf{AND}) \to 2\rho(k\mathsf{AND}) = 2^{-(k-1)}$ as $k \to \infty$, in the large-k limit our simple randomized algorithm (given in Theorem 7) matches the performance of Trevisan's [23] parallelizable LP-based algorithm for kAND, which (to the best of our knowledge) was the first work on the general kAND problem! The subsequent works [13, 14, 4] superseding [23] use more complex techniques involving semidefinite programming, but are structurally similar to our algorithm in Theorem 7: They all involve "guessing" an assignment $\mathbf{x} \in \mathbb{Z}_2^n$ and then perturbing each bit with constant probability.

2 Our techniques

In this section, we give a more detailed background on the technical aspects of the dichotomy theorem in [7], and explain the novel aspects of our analysis.

2.1 The Chou, Golovnev, Sudan, and Velusamy [7] framework for symmetric functions

In this section, we describe the Chou, Golovnev, Sudan, and Velusamy [7] framework for finding the optimal sketching approximation ratio of a symmetric Boolean function $f_{S,k}$.

Let $\Delta(\{-1,1\}^k)$ denote the space of all distributions on $\{-1,1\}^k$. For a distribution $\mathcal{D} \in \Delta(\{-1,1\}^k)$ and $\mathbf{x} \in \{-1,1\}^k$, we use $\mathcal{D}(\mathbf{x})$ to denote the probability of sampling $\mathbf{x}$ in $\mathcal{D}$. To a distribution $\mathcal{D} \in \Delta(\{-1,1\}^k)$ we associate a *canonical instance* $\Psi_{\mathcal{D}}$ of Max-CSP$(f_{S,k})$ on k variables as follows. Let $\mathbf{j} = (1,\ldots,k)$. For every negation pattern $\mathbf{b} \in \{-1,1\}^k$, $\Psi_{\mathcal{D}}$ contains the constraint $(\mathbf{b},\mathbf{j})$ with weight $\mathcal{D}(\mathbf{b})$.

We say a distribution $\mathcal{D} \in \Delta(\{-1,1\}^k)$ is *symmetric* if all vectors of equal Hamming weight are equiprobable, i.e., for every $\mathbf{x},\mathbf{y} \in \{-1,1\}^k$ such that $\mathrm{wt}(\mathbf{x}) = \mathrm{wt}(\mathbf{y})$, $\mathcal{D}(\mathbf{x}) = \mathcal{D}(\mathbf{y})$. Let $\Delta_k \subseteq \Delta(\{-1,1\}^k)$ denote the set of all symmetric distributions on $\{-1,1\}^k$. Given $\mathcal{D} \in \Delta_k$, let $\mathcal{D}\langle i\rangle \overset{\text{def}}{=} \sum_{\mathbf{x}\in\{-1,1\}^k:\mathrm{wt}(\mathbf{x})=i} \mathcal{D}(\mathbf{x})$ denote the total probability mass on vectors of Hamming weight i. Note that any vector $(\mathcal{D}\langle 0\rangle,\ldots,\mathcal{D}\langle k\rangle)$ of nonnegative values summing to 1 uniquely determines a distribution $\mathcal{D} \in \Delta_k$; we write $\mathcal{D} = (\mathcal{D}\langle 0\rangle,\ldots,\mathcal{D}\langle k\rangle)$ for notational convenience.

Let $\mathsf{Bern}(p)$ represent a random variable which is 1 with probability p and -1 with probability $1-p$. For $\mathcal{D} \in \Delta(\{-1,1\})^k$ and $p \in [0,1]$, let

$$\lambda_S(\mathcal{D},p) \overset{\text{def}}{=} \mathop{\mathbb{E}}_{\mathbf{a}\sim\mathcal{D},\mathbf{b}\sim\mathsf{Bern}(p)^k}[f_{S,k}(\mathbf{a}\odot\mathbf{b})] = \mathop{\mathbb{E}}_{\mathbf{b}\sim\mathsf{Bern}(p)^k}[\mathsf{val}_{\Psi_{\mathcal{D}}}(\mathbf{b})] \tag{2}$$

denote the expected value of a "p-biased symmetric assignment" on $\mathcal{D}$'s canonical instance. Also, for a symmetric distribution $\mathcal{D} \in \Delta_k$, we define its (scalar) *marginal*

$$\mu(\mathcal{D}) \overset{\text{def}}{=} \mathop{\mathbb{E}}_{\mathbf{b}\sim\mathcal{D}}[b_1] = \cdots = \mathop{\mathbb{E}}_{\mathbf{b}\sim\mathcal{D}}[b_k]. \tag{3}$$

In general, λ_S is linear in $\mathcal{D}$ and degree-k in p, and μ is linear in $\mathcal{D}$. For $\mathcal{D} \in \Delta_k$, we provide explicit formulas for λ_S and μ in Section 3.

Roughly, [7] states that Max-CSP$(f_{S,k})$ is hard to approximate in the sketching setting if there exist distributions $\mathcal{D}_N,\mathcal{D}_Y \in \Delta_k$ such that (1) $\mu(\mathcal{D}_N) = \mu(\mathcal{D}_Y)$ and (2) $\mathcal{D}_Y$'s canonical instance is highly satisfied by the trivial (all-ones) assignment but (3) $\mathcal{D}_N$'s canonical instance is not well-satisfied by any "biased symmetric assignment". To be precise, for $\mathcal{D} \in \Delta(\{-1,1\}^k)$, let

$$\beta_S(\mathcal{D}) \stackrel{\text{def}}{=} \sup_{p\in[0,1]} \lambda_S(\mathcal{D},p) \text{ and } \gamma_S(\mathcal{D}) \stackrel{\text{def}}{=} \lambda_S(\mathcal{D},1), \tag{4}$$

and define

$$\alpha(f_{S,k}) \stackrel{\text{def}}{=} \inf_{\mathcal{D}_N,\mathcal{D}_Y\in\Delta_k:\ \mu(\mathcal{D}_N)=\mu(\mathcal{D}_Y)} \left(\frac{\beta_S(\mathcal{D}_N)}{\gamma_S(\mathcal{D}_Y)}\right). \tag{5}$$

For every symmetric function $f_{S,k}$, [7] proves that $\alpha(f_{S,k})$ is the optimal sketching approximation ratio for Max-CSP$(f_{S,k})$:

▶ **Theorem 9** (Combines [7, Theorem 2.10 and Lemma 2.14]). *Let $f_{S,k} : \{-1,1\}^k \to \{0,1\}$ be a symmetric function. Then for every $\epsilon > 0$, there is an linear sketching $(\alpha(f_{S,k}) - \epsilon)$-approximation to Max-CSP$(f_{S,k})$ in $O(\log n)$ space, but any sketching $(\alpha(f_{S,k}) + \epsilon)$-approximation to Max-CSP$(f_{S,k})$ requires $\Omega(\sqrt{n})$ space.*

▶ Remark. In the general case where $f : \{-1,1\}^k \to \{0,1\}$ is not symmetric, the approximability of f is no longer characterized by Equation (5). Instead, [7] requires taking an infimum over *all* (not necessarily symmetric) distributions $\mathcal{D}_N, \mathcal{D}_Y \in \Delta(\{-1,1\})^k$. Moreover, a general distribution $\mathcal{D} \in \Delta(\{-1,1\})^k$ no longer has a single scalar marginal (as in Equation (3)). Instead, we must consider a *vector* marginal $\boldsymbol{\mu}(\mathcal{D}) = (\mu_1,\ldots,\mu_k)$ with i-th component $\mu_i = \mathbb{E}_{\mathbf{b}\sim\mathcal{D}}[b_i]$; correspondingly, $\mathcal{D}_N$ and $\mathcal{D}_Y$ are required to satisfy the constraint $\boldsymbol{\mu}(\mathcal{D}_N) = \boldsymbol{\mu}(\mathcal{D}_Y)$. These issues motivate our focus on *symmetric* functions in this paper. Since we need to consider only symmetric distributions in Equation (5), $\mathcal{D}_Y$ and $\mathcal{D}_N$ are each parameterized by $k+1$ variables (as opposed to 2^k variables), and there is a single linear equality constraint (as opposed to k constraints).

2.2 Formulations of the optimization problem

In order to show that $\alpha(\text{2AND}) = \frac{4}{9}$, Chou, Golovnev, Sudan, and Velusamy [7, Example 1] use the following reformulation of the optimization problem on the right hand side of Equation (5). For a symmetric function $f_{S,k}$ and $\mu \in [-1,1]$, let

$$\beta_{S,k}(\mu) = \inf_{\mathcal{D}_N\in\Delta_k:\ \mu(\mathcal{D}_N)=\mu} \beta_S(\mathcal{D}_N) \text{ and } \gamma_{S,k}(\mu) = \sup_{\mathcal{D}_Y\in\Delta_k:\ \mu(\mathcal{D}_Y)=\mu} \gamma_S(\mathcal{D}_Y); \tag{6}$$

then

$$\alpha(f_{S,k}) = \inf_{\mu\in[-1,1]} \left(\frac{\beta_{S,k}(\mu)}{\gamma_{S,k}(\mu)}\right). \tag{7}$$

The optimization problem on the right-hand side of Equation (7) appears simpler than that of Equation (5) because it is univariate, but there is a hidden difficulty: Finding an explicit solution requires giving explicit formulas for $\beta_{S,k}(\mu)$ and $\gamma_{S,k}(\mu)$. In the case of $\text{2AND} = f_{\{2\},2}$, Chou, Golovnev, Sudan, and Velusamy [7] show that $\gamma_{\{2\},2}(\mu)$ is an explicit linear function of μ; maximize the quadratic $\lambda_{\{2\}}(\mathcal{D}_N, p)$ over $p \in [0,1]$ to find $\beta_{\{2\}}(\mathcal{D}_N)$; and then minimize $\beta_{\{2\}}(\mathcal{D}_N)$ given $\mu(\mathcal{D}_N) = \mu$ to find $\beta_{\{2\},2}(\mu)$. However, while for general symmetric functions $f_{S,k}$ we can describe $\gamma_{S,k}(\mu)$ as an explicit piecewise linear function of μ (see Lemma 16 below), we do not know how to find closed forms for $\beta_{S,k}(\mu)$ even for 3AND. Thus, in this work we introduce a different formulation of the optimization problem:

$$\alpha(f_{S,k}) = \inf_{\mathcal{D}_N\in\Delta_k} \left(\frac{\beta_S(\mathcal{D}_N)}{\gamma_{S,k}(\mu(\mathcal{D}_N))}\right). \tag{8}$$

This reformulation is valid because

$$\alpha(f_{S,k}) = \inf_{\mu \in [-1,1], \mathcal{D}_N \in \Delta_k:\ \mu(\mathcal{D}_N)=\mu} \left(\frac{\beta_S(\mathcal{D}_N)}{\gamma_{S,k}(\mu)} \right) = \inf_{\mathcal{D}_N \in \Delta_k} \left(\frac{\beta_S(\mathcal{D}_N)}{\gamma_{S,k}(\mu(\mathcal{D}_N))} \right).$$

We view optimizing directly over $\mathcal{D}_N \in \Delta_k$ as an important conceptual switch. In particular, our formulation emphasizes the calculation of $\beta_S(\mathcal{D}_N)$ as the centrally difficult feature, yet we can still take advantage of the relative simplicity of calculating $\gamma_{S,k}(\mu)$.

2.3 Our contribution: The max-min method

A priori, solving the optimization problem on the right-hand side of Equation (8) still requires calculating $\beta_S(\mathcal{D}_N)$, which involves maximizing a degree-k polynomial. To get around this difficulty, we have made a key discovery, which was not noticed by Chou, Golovnev, Sudan, and Velusamy [7] even in the 2AND case. Let $\mathcal{D}_N^*$ minimize the right-hand side of Equation (8), and p^* maximize $\lambda_S(\mathcal{D}_N^*, \cdot)$. After substituting $\beta_S(\mathcal{D}) = \sup_{p \in [0,1]} \lambda_S(\mathcal{D}, p)$ in Equation (8), and applying the max-min inequality, we get

$$\begin{aligned} \alpha(f_{S,k}) = \inf_{\mathcal{D}_N \in \Delta_k} \sup_{p \in [0,1]} \left(\frac{\lambda_S(\mathcal{D}_N, p)}{\gamma_{S,k}(\mu(\mathcal{D}_N))} \right) &\geq \sup_{p \in [0,1]} \inf_{\mathcal{D}_N \in \Delta_k} \left(\frac{\lambda_S(\mathcal{D}_N, p)}{\gamma_{S,k}(\mu(\mathcal{D}_N))} \right) \\ &\geq \inf_{\mathcal{D}_N \in \Delta_k} \left(\frac{\lambda_S(\mathcal{D}_N, p^*)}{\gamma_{S,k}(\mu(\mathcal{D}_N))} \right). \end{aligned} \tag{9}$$

Given p^*, the right-hand side of Equation (9) is relatively easy to calculate, being a ratio of a linear and piecewise linear function of $\mathcal{D}_N$. Our discovery is that, in a wide variety of cases, the quantity on the right-hand side of Equation (9) *equals* $\alpha(f_{S,k})$; that is, $(\mathcal{D}_N^*, p^*)$ is a *saddle point* of $\frac{\lambda_S(\mathcal{D}_N,p)}{\gamma_{S,k}(\mu(\mathcal{D}_N))}$.[5]

This yields a novel technique, which we call the "max-min method", for finding a closed form for $\alpha(f_{S,k})$. First, we guess $\mathcal{D}_N^*$ and p^*, and then, we show analytically that $\frac{\lambda_S(\mathcal{D}_N,p)}{\gamma_{S,k}(\mu(\mathcal{D}_N))}$ has a saddle point at $(\mathcal{D}_N^*, p^*)$ and that $\lambda_S(\mathcal{D}_N, p)$ is maximized at p^*. These imply that $\frac{\lambda_S(\mathcal{D}_N^*,p^*)}{\gamma_{S,k}(\mu(\mathcal{D}_N^*))}$ is a lower and upper bound on $\alpha(f_{S,k})$, respectively. For instance, in Section 4, in order to give a closed form for $\alpha(k\mathsf{AND})$ for odd k (i.e., the odd case of Theorem 1), we guess $\mathcal{D}_N^*\langle \frac{k+1}{2} \rangle = 1$ and $p^* = \frac{k+1}{2k}$ (by using Mathematica for small cases), and then check the saddle-point and maximization conditions in two separate lemmas (Lemmas 17 and 18, respectively). Then, we show that $\alpha(k\mathsf{AND}) = \alpha'_k$ by analyzing the right hand side of the appropriate instantiation of Equation (9). We use similar techniques for kAND for even k (also Theorem 1) and for various other cases in Sections 5.1–5.3.

In all of these cases, the $\mathcal{D}_N^*$ we construct is supported on at most two distinct Hamming weights, which is the property which makes finding $\mathcal{D}_N^*$ tractable (using computer assistance). However, this technique is not a "silver bullet": it is not the case that the sketching approximability of every symmetric Boolean CSP can be exactly calculated by finding the optimal $\mathcal{D}_N^*$ supported on two elements and using the max-min method. Indeed, (as mentioned in Section 5.3) we verify using computer assistance that this is not the case for $f_{\{3\},4}$.

[5] This term comes from the optimization literature; such points are also said to satisfy the "strong max-min property" (see, e.g., [2, pp. 115, 238]). The saddle-point property is guaranteed by von Neumann's minimax theorem for functions which are concave and convex in the first and second arguments, respectively, but this theorem and the generalizations we are aware of do not apply even to 3AND.

Finally, we remark that the saddle-point property is precisely what defines the value p^* required for our simple classical algorithm for outputting approximately optimal assignments for $\mathsf{Max\text{-}CSP}(f_{S,k})$ where $f_{S,k} = \mathsf{Th}_k^i$ is a threshold function (see Theorem 34).

2.4 Streaming lower bounds

Chou, Golovnev, Sudan, and Velusamy [7] also define the following condition on pairs $(\mathcal{D}_N, \mathcal{D}_Y)$, stronger than $\mu(\mathcal{D}_N) = \mu(\mathcal{D}_Y)$, which implies hardness of (γ,β)-$\mathsf{Max\text{-}CSP}(f)$ for *streaming* algorithms:

▶ **Definition 10** (Padded one-wise pairs, [7, §2.3] (symmetric case)). *A pair of distributions $(\mathcal{D}_Y, \mathcal{D}_N) \in \Delta_k$ forms a* padded one-wise pair *if there exists $\tau \in [0,1]$ and distributions $\mathcal{D}_0, \mathcal{D}'_Y, \mathcal{D}'_N \in \Delta_k$ such that (1) $\mu(\mathcal{D}'_Y) = \mu(\mathcal{D}'_N) = 0$ and (2) $\mathcal{D}_Y = \tau\mathcal{D}_0 + (1-\tau)\mathcal{D}'_Y$ and $\mathcal{D}_N = \tau\mathcal{D}_0 + (1-\tau)\mathcal{D}'_N$.*

▶ **Theorem 11** (Streaming lower bound for padded one-wise pairs, [7, Theorem 2.11] (symmetric case)). *Let $(\mathcal{D}_Y, \mathcal{D}_N)$ be a padded one-wise pair. Then for every $\epsilon > 0$, $(\beta_S(\mathcal{D}_Y)+\epsilon, \gamma_S(\mathcal{D}_N)-\epsilon)$-$\mathsf{Max\text{-}CSP}(f)$ requires $\Omega(\sqrt{n})$ space in the streaming setting.*

We prove that Theorem 11 fails to rule out streaming $(\frac{2}{9}+\epsilon)$-approximations to $\mathsf{Max\text{-}3AND}$ in the following sense:

▶ **Theorem 12.** *There is no infinite sequence $(\mathcal{D}_Y^{(1)}, \mathcal{D}_N^{(1)}), (\mathcal{D}_Y^{(2)}, \mathcal{D}_N^{(2)}), \ldots$ of padded one-wise pairs on Δ_3 such that*

$$\lim_{t\to\infty} \frac{\beta_{\{3\}}(\mathcal{D}_N^{(t)})}{\gamma_{\{3\}}(\mathcal{D}_Y^{(t)})} = \frac{2}{9}.$$

Theorem 12 is proven formally in Section 6; we give a proof outline in Appendix A.

Yet we still can achieve decent bounds using padded one-wise pairs:

▶ **Observation 13.** *The padded one-wise pair $\mathcal{D}_N = (0, 0.45, 0.45, 0.1), \mathcal{D}_Y = (0.45, 0, 0, 0.55)$ (discovered by numerical search)* does *prove a streaming approximability upper bound of $\approx .2362$ for $\mathsf{3AND}$, which is still quite close to $\alpha(\mathsf{3AND}) = \frac{2}{9}$.*

3 Formulas for μ, λ_S, and $\gamma_{S,k}$

In this section, we give explicit formulas for the quantities $\mu(\mathcal{D})$, $\lambda_S(\mathcal{D}, p)$, and $\gamma_{S,k}(\mu)$ (defined in Equations (2), (3), and (6), respectively) which will be used throughout the rest of the paper. For $i \in [k]$, let $\epsilon_{i,k} \stackrel{\text{def}}{=} -1 + \frac{2i}{k}$.

▶ **Lemma 14.** *For any $\mathcal{D} \in \Delta_k$,*

$$\mu(\mathcal{D}) = \sum_{i=0}^{k} \epsilon_{i,k}\, \mathcal{D}\langle i\rangle.$$

Proof of Lemma 14. By definition (Equation (3)), $\mu(\mathcal{D}) = \mathbb{E}_{\mathbf{b}\sim\mathcal{D}}[b_1]$. We use linearity of expectation; the contribution of weight-i vectors to $\mu(\mathcal{D})$ is $\mathcal{D}\langle i\rangle \cdot \frac{1}{k}(i \cdot 1 + (k-i)\cdot(-1)) = \epsilon_{i,k}\, \mathcal{D}\langle i\rangle$. ◀

▶ **Lemma 15.** *For any* $\mathcal{D} \in \Delta_k$ *and* $p \in [0,1]$*, we have*

$$\lambda_S(\mathcal{D}, p) = \sum_{s \in S} \sum_{i=0}^{k} \left(\sum_{j=\max\{0, s-(k-i)\}}^{\min\{i,s\}} \binom{i}{j} \binom{k-i}{s-j} q^{s+i-2j} p^{k-s-i+2j} \right) \mathcal{D}\langle i \rangle$$

where $q \stackrel{\text{def}}{=} 1 - p$*.*

The proof of Lemma 15 is given in the full version [3].

▶ **Lemma 16.** *Let* $S \subseteq [k]$*, and let* s *be its smallest element and* t *its largest element (they need not be distinct). Then for* $\mu \in [-1, 1]$*,*

$$\gamma_{S,k}(\mu) = \begin{cases} \frac{1+\mu}{1+\epsilon_{s,k}} & \mu \in [-1, \epsilon_{s,k}) \\ 1 & \mu \in [\epsilon_{s,k}, \epsilon_{t,k}] \\ \frac{1-\mu}{1-\epsilon_{t,k}} & \mu \in (\epsilon_{t,k}, 1] \end{cases}$$

(which also equals $\min\left\{ \frac{1+\mu}{1+\epsilon_{s,k}}, 1, \frac{1-\mu}{1-\epsilon_{t,k}} \right\}$*).*

The proof of Lemma 16 is given in the full version [3].

4 Analysis of $\alpha(k\mathsf{AND})$

In this section, we prove Theorem 1 (on the sketching approximability of Max-kAND). Recall that in Equation (1), we defined

$$\alpha'_k = \left(\frac{(k-1)(k+1)}{4k^2} \right)^{(k-1)/2}.$$

Theorem 1 follows immediately from the following two lemmas:

▶ **Lemma 17.** *For all odd* $k \geq 3$*,* $\alpha(k\mathsf{AND}) \leq \alpha'_k$*. For all even* $k \geq 2$*,* $\alpha(k\mathsf{AND}) \leq 2\alpha'_{k+1}$*.*

▶ **Lemma 18.** *For all odd* $k \geq 3$*,* $\alpha(k\mathsf{AND}) \geq \alpha'_k$*. For all even* $k \geq 2$*,* $\alpha(k\mathsf{AND}) \geq 2\alpha'_{k+1}$*.*

To begin, we give explicit formulas for $\gamma_{\{k\},k}(\mu(\mathcal{D}))$ and $\lambda_{\{k\}}(\mathcal{D}, p)$. Note that the smallest element of $\{k\}$ is k, and $\epsilon_{k,k} = 1$. Thus, for $\mathcal{D} \in \Delta_k$, we have by Lemmas 14 and 16 that

$$\gamma_{\{k\},k}(\mu(\mathcal{D})) = \frac{1 + \sum_{i=0}^{k} (-1 + \frac{2i}{k}) \mathcal{D}\langle i \rangle}{2} = \sum_{i=0}^{k} \frac{i}{k} \mathcal{D}\langle i \rangle. \tag{10}$$

Similarly, we can apply Lemma 15 with $s = k$; for each $i \in \{0\} \cup [k]$, $\max\{0, s - (k-i)\} = \min\{i, k\} = i$, so we need only consider $j = i$, and then $\binom{i}{j} = \binom{k-i}{s-j} = 1$. Thus, for $q = 1 - p$, we have

$$\lambda_{\{k\}}(\mathcal{D}, p) = \sum_{i=0}^{k} q^{k-i} p^i \mathcal{D}\langle i \rangle \tag{11}$$

The proof of Lemma 17 is given in Appendix A. We also prove Lemma 18 in Appendix A using the max-min method. We rely on the following proposition which is a simple inequality for optimizing ratios of linear functions, which we prove in the full version [3]:

▶ **Proposition 19.** *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be defined by the equation* $f(\mathbf{x}) = \frac{\mathbf{a} \cdot \mathbf{x}}{\mathbf{b} \cdot \mathbf{x}}$ *for some* $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n_{\geq 0}$*. For every* $\mathbf{y}(1), \ldots, \mathbf{y}(r) \in \mathbb{R}^n_{\geq 0}$*, and every* $\mathbf{x} = \sum_{i=1}^{r} \alpha_i \mathbf{y}(i)$ *with each* $x_i \geq 0$*, we have* $f(\mathbf{x}) \geq \min_i f(\mathbf{y}(i))$*. In particular, taking* $r = n$ *and* $\mathbf{y}(1), \ldots, \mathbf{y}(n)$ *as the standard basis for* $\mathbb{R}^n$*, for every* $\mathbf{x} \in \mathbb{R}^n_{\geq 0}$*, we have* $f(\mathbf{x}) \geq \min_i \frac{a_i}{b_i}$*.*

5 Further analyses of $\alpha(f)$ for symmetric Boolean functions f

5.1 Th_k^{k-1} for even k

In this subsection, we prove Theorem 3 (on the sketching approximability of Th_k^{k-1} for even $k \geq 2$). It is necessary and sufficient to prove the following two lemmas:

▶ **Lemma 20.** *For all even $k \geq 2$, $\alpha(\mathsf{Th}_k^{k-1}) \leq \frac{k}{2}\alpha'_{k-1}$.*

▶ **Lemma 21.** *For all even $k \geq 2$, $\alpha(\mathsf{Th}_k^{k-1}) \geq \frac{k}{2}\alpha'_{k-1}$.*

Firstly, we give explicit formulas for $\gamma_{\{k-1,k\},k}$ and $\lambda_{\{k-1,k\}}$. We have $\mathsf{Th}_k^{k-1} = f_{\{k-1,k\},k}$, and $\epsilon_{k-1,k} = -1 + \frac{2(k-1)}{k} = 1 - \frac{2}{k}$. Thus, Lemmas 14 and 16 give

$$\gamma_{\{k-1,k\},k}(\mu(\mathcal{D})) = \min\left\{\frac{1+\sum_{i=0}^{k}(-1+\frac{2i}{k})\,\mathcal{D}\langle i\rangle}{2-\frac{2}{k}}, 1\right\} = \min\left\{\sum_{i=0}^{k}\frac{i}{k-1}\,\mathcal{D}\langle i\rangle, 1\right\}. \quad (12)$$

Next, we calculate $\lambda_{\{k-1,k\}}(\mathcal{D},p)$ with Lemma 15. Let $q = 1-p$, and let us examine the coefficient on $\mathcal{D}\langle i\rangle$. $s = k$ contributes $q^{k-i}p^k$. In the case $i \leq k-1$, $s = k-1$ contributes $(k-i)q^{k-i-1}p^{i+1}$ for $j = i$, and in the case $i \geq 1$, $s = k-1$ contributes $iq^{k-i+1}p^{i-1}$ for $j = i-1$. Thus, altogether we can write

$$\lambda_{\{k-1,k\}}(\mathcal{D},p) = \sum_{i=0}^{k} q^{k-i-1}p^{i-1}\left((k-i)p^2 + pq + iq^2\right)\mathcal{D}\langle i\rangle. \quad (13)$$

The proofs of Lemmas 20 and 21 are given in Appendix A.

▶ **Observation 22.** *For Th_4^3 the optimal $\mathcal{D}_N^* = (0,0,\frac{4}{5},\frac{1}{5},0)$ does participate in a padded one-wise pair with $\mathcal{D}_Y^* = (\frac{4}{15},0,0,\frac{11}{15},0)$ (given by $\mathcal{D}_0 = (0,0,0,1,0)$, $\tau = \frac{1}{5}$, $\mathcal{D}'_N = (0,0,1,0,0)$, and $\mathcal{D}'_Y = (\frac{4}{15},0,0,\frac{8}{15},0)$) so we can rule out* streaming *$(\frac{4}{9}+\epsilon)$-approximations to Max-CSP(Th_4^3) in $o(\sqrt{n})$ space.*

5.2 $\mathsf{Ex}_k^{(k+1)/2}$ for (small) odd k

In this section, we prove bounds on the sketching approximability of $\mathsf{Ex}_k^{(k+1)/2}$ for odd $k \in \{3,\ldots,51\}$. Define $\mathcal{D}_{0,k} \in \Delta_k$ by $\mathcal{D}_{0,k}\langle 0\rangle = \frac{k-1}{2k}$ and $\mathcal{D}_{0,k}\langle k\rangle = \frac{k+1}{2k}$. We prove the following two lemmas:

▶ **Lemma 23.** *For all odd $k \geq 3$, $\alpha(\mathsf{Ex}_k^{(k+1)/2}) \leq \lambda_{\{\frac{k+1}{2}\}}(\mathcal{D}_{0,k}, p'_k)$, where $p'_k \stackrel{\text{def}}{=} \frac{3k-k^2+\sqrt{4k+k^2-2k^3+k^4}}{4k}$.*

▶ **Lemma 24.** *The following holds for all odd $k \in \{3,\ldots,51\}$. For all $p \in [0,1]$, the expression $\frac{\lambda_{\{\frac{k+1}{2}\}}(\cdot,p)}{\gamma_{\{\frac{k+1}{2}\},k}(\mu(\cdot))}$ is minimized at $\mathcal{D}_{0,k}$.*

We begin by writing an explicit formula for $\lambda_{\{\frac{k+1}{2}\}}$. Lemma 15 gives

$$\lambda_{\{\frac{k+1}{2}\}}(\mathcal{D},p) = \sum_{i=0}^{k}\left(\sum_{j=\max\{0,i-\frac{k-1}{2}\}}^{\min\{i,\frac{k+1}{2}\}}\binom{i}{j}\binom{k}{\frac{k+1}{2}-j}(1-p)^{(k+1)/2+i-2j}p^{(k-1)/2-i+2j}\right)\mathcal{D}\langle i\rangle.$$

For $i \leq \frac{k-1}{2}$, the sum over j goes from 0 to i, and for $i \geq \frac{k+1}{2}$, it goes from $i - \frac{k-1}{2}$ to $\frac{k+1}{2}$. Thus, plugging in $\mathcal{D}_{0,k}$, we get:

$$\lambda_{\{\frac{k+1}{2}\}}(\mathcal{D}_{0,k},p) = \binom{k}{\frac{k+1}{2}}\left(\frac{k-1}{2k}(1-p)^{(k+1)/2}p^{(k-1)/2} + \frac{k+1}{2k}(1-p)^{(k-1)/2}p^{(k+1)/2}\right). \quad (14)$$

By Lemmas 14 and 16, $\gamma_{\{\frac{k+1}{2}\},k}(\mu(\mathcal{D}_{0,k})) = \gamma_{\{\frac{k+1}{2}\},k}(\frac{1}{k}) = 1$. Thus, Lemmas 23 and 24 together imply the following theorem:

▶ **Theorem 25.** *For odd* $k \in \{3, \ldots, 51\}$,

$$\alpha(\mathsf{Ex}_k^{(k+1)/2}) = \binom{k}{\frac{k+1}{2}} \left(\frac{k-1}{2k}(1-p_k')^{(k+1)/2}(p_k')^{(k-1)/2} + \frac{k+1}{2k}(1-p_k')^{(k-1)/2}(p_k')^{(k+1)/2} \right),$$

where $p_k' = \frac{3k-k^2+\sqrt{4k+k^2-2k^3+k^4}}{4k}$ *as in Lemma 23.*

Recall that $\rho(f_{(k+1)/2,k}) = \binom{k}{\frac{k+1}{2}} 2^{-k}$. Although we currently lack a lower bound on $\alpha(\mathsf{Ex}_k^{(k+1)/2})$ for large odd k, the upper bound from Lemma 23 suffices to prove Theorem 5, i.e., it can be verified that

$$\lim_{k \text{ odd} \to \infty} \frac{\binom{k}{\frac{k+1}{2}} \left(\frac{k-1}{2k}(1-p_k')^{(k+1)/2}(p_k')^{(k-1)/2} + \frac{k+1}{2k}(1-p_k')^{(k-1)/2}(p_k')^{(k+1)/2} \right)}{\rho(\mathsf{Ex}_k^{(k+1)/2})} = 1.$$

We remark that for $\mathsf{Ex}_k^{(k+1)/2}$, our lower bound (Lemma 24) is *stronger* than what we were able to prove for $k\mathsf{AND}$ (Lemma 18) and Th_k^{k-1} (Lemma 21) because the inequality holds regardless of p. This is fortunate for us, as the optimal p^* from Lemma 23 is rather messy.[6] The proofs of Lemmas 23 and 24 are given in the full version [3].

5.3 More symmetric functions

In Table 1 below, we list four more symmetric Boolean functions (beyond $k\mathsf{AND}$, Th_k^{k-1}, and $\mathsf{Ex}_k^{(k+1)/2}$) whose sketching approximability we have analytically resolved using the "max-min method". These values were calculated using two functions in the Mathematica code, `estimateAlpha` – which numerically or symbolically estimates the $\mathcal{D}_N$, with a given support, which minimizes α – and `testMinMax` – which, given a particular $\mathcal{D}_N$, calculates p^* for that $\mathcal{D}_N$ and checks analytically whether lower-bounding by evaluating λ_S at p^* proves that $\mathcal{D}_N$ is minimal.

Table 1 Symmetric functions for which we have analytically calculated exact α values using the "max-min method". For a polynomial $P : \mathbb{R} \to \mathbb{R}$ with a *unique* positive real root, let $\text{root}_{\mathbb{R}}(p)$ denote that root, and define the polynomials $P_1(z) = -72 + 4890z - 108999z^2 + 800000z^3$, $P_2(z) = -908 + 5021z - 9001z^2 + 5158z^3$, $P_3(z) = -60 + 5745z - 183426z^2 + 1953125z^3$, $P_4(z) = -344 + 1770z - 3102z^2 + 1811z^3$. (We note that in the $f_{\{4\},5}$ and $f_{\{4,5\},5}$ calculations, we were required to check equality of roots numerically (to high precision) instead of analytically).

S	k	α	$\mathcal{D}_N^*$
$\{2,3\}$	3	$\frac{1}{2} + \frac{\sqrt{3}}{18} \approx 0.5962$	$(0, \frac{1}{2}, 0, \frac{1}{2})$
$\{4,5\}$	5	$8\,\text{root}_{\mathbb{R}}(P_1) \approx 0.2831$	$(0, 0, 1-\text{root}_{\mathbb{R}}(P_2), \text{root}_{\mathbb{R}}(P_2), 0, 0)$
$\{4\}$	5	$8\,\text{root}_{\mathbb{R}}(P_3) \approx 0.2394$	$(0, 0, 1-\text{root}_{\mathbb{R}}(P_4), \text{root}_{\mathbb{R}}(P_4), 0, 0)$
$\{3,4,5\}$	5	$\frac{1}{2} + \frac{3\sqrt{5}}{125} \approx 0.5537$	$(0, \frac{1}{2}, 0, 0, 0, \frac{1}{2})$

We remark that two of the cases in Table 1 (as well as $k\mathsf{AND}$), the optimal $\mathcal{D}_N$ is rational and supported on two coordinates. However, in the other two cases in Table 1, the optimal $\mathcal{D}_N$ involves roots of a cubic.

[6] The analogous statement is false for e.g. $3\mathsf{AND}$, where we had $\mathcal{D}_N^* = (0,0,1,0)$, but at $p = \frac{3}{4}$,

$$\frac{\lambda_{\{3\}}((0, \frac{1}{2}, \frac{1}{2}, 0), \frac{3}{4})}{\gamma_{\{3\},3}(\mu(0, \frac{1}{2}, \frac{1}{2}, 0))} = \frac{3}{16} \leq \frac{27}{128} = \frac{\lambda_{\{3\}}((0,0,1,0), \frac{3}{4})}{\gamma_{\{3\},3}(\mu(0,0,1,0))}.$$

In Section 5.2, we showed that $\mathcal{D}_N^*$ defined by $\mathcal{D}_N^*\langle 0\rangle = \frac{k-1}{2k}$ and $\mathcal{D}_N^*\langle k\rangle = \frac{k+1}{2k}$ is optimal for $\mathsf{Ex}_k^{(k+1)/2}$ for odd $k \in \{3,\dots,51\}$. Using the same $\mathcal{D}_N^*$, we are also able to resolve 11 other cases in which S is "close to" $\{\frac{k+1}{2}\}$; for instance, $S = \{5,6\},\{5,6,7\},\{5,7\}$ for $k = 9$. (We have omitted the values of α and $\mathcal{D}_N$ because they are defined using the roots of polynomials of degree up to 8.)

In all previously-mentioned cases, the condition "$\mathcal{D}_N^*$ has support size 2" was helpful, as it makes the optimization problem over $\mathcal{D}_N^*$ essentially univariate; however, we have confirmed analytically in two other cases ($S = \{3\}, k = 4$ and $S = \{3,5\}, k = 5$) that "max-min method on distributions with support size two" does not suffice for tight bounds on α (see `testDistsWithSupportSize2` in the Mathematica code). However, using the max-min method with $\mathcal{D}_N$ supported on two levels still achieves decent (but not tight) bounds on α. For $S = \{3\}, k = 4$, using $\mathcal{D}_N = (\frac{1}{4},0,0,0,\frac{3}{4})$, we get the bounds $\alpha(f_{\{3\},4}) \in [0.3209, 0.3295]$ (the difference being 2.67%). For $S = \{3,5\}, k = 5$, using $\mathcal{D}_N = (\frac{1}{4},0,0,0,\frac{3}{4},0)$, we get $\alpha(f_{\{3,5\},5}) \in [0.3416, 0.3635]$ (the difference being 6.42%).

Finally, we have also analyzed cases where we get numerical solutions which are very close to tight, but we lack analytical solutions because they likely involve roots of high-degree polynomials. For instance, in the case $S = \{4,5,6\}, k = 6$, setting $\mathcal{D}_N = (0,0,0,0.930013,0,0,0.069987)$ gives $\alpha(f_{\{4,5,6\},6}) \in [0.44409972, 0.44409973]$, differing only by 0.000003%. (We conjecture here that $\alpha = \frac{4}{9}$.) For $S = \{6,7,8\}, k = 8$, using $\mathcal{D}_N = (0,0,0,0,0.699501,0.300499)$, we get the bounds $\alpha(f_{\{6,7,8\},8}) \in [0.20848, 0.20854]$ (the difference being 0.02%).[7]

6 Incompleteness of streaming lower bounds: Proving Theorem 12

In this section, we prove Theorem 12, showing that the streaming lower bounds from [7] (Theorem 11) cannot characterize the *streaming* approximability of 3AND.

▶ **Lemma 26.** *For $\mathcal{D} \in \Delta_3$, the expression*

$$\frac{\lambda_{\{3\}}(\mathcal{D}, \frac{1}{3}\mathcal{D}\langle 1\rangle + \frac{2}{3}\mathcal{D}\langle 2\rangle + \mathcal{D}\langle 3\rangle)}{\gamma_{\{3\},3}(\mu(\mathcal{D}))}$$

is minimized uniquely at $\mathcal{D} = (0,0,1,0)$, with value $\frac{2}{9}$.

Proof. Letting $p = \frac{1}{3}\mathcal{D}\langle 1\rangle + \frac{2}{3}\mathcal{D}\langle 2\rangle + \mathcal{D}\langle 3\rangle$ and $q = 1-p$, by Lemmas 14–16 the expression expands to

$$\frac{\mathcal{D}\langle 0\rangle\, p^3 + \mathcal{D}\langle 1\rangle\, p^2(1-p) + \mathcal{D}\langle 2\rangle\, p(1-p)^2 + \mathcal{D}\langle 3\rangle\,(1-p)^3}{\frac{1}{2}(1 - \mathcal{D}\langle 0\rangle - \frac{1}{3}\mathcal{D}\langle 1\rangle + \frac{1}{3}\mathcal{D}\langle 2\rangle + \mathcal{D}\langle 3\rangle)}.$$

The expression's minimum, and its uniqueness, are confirmed analytically in the Mathematica code. ◀

▶ **Lemma 27.** *Let X be a compact topological space, $Y \subseteq X$ a closed subspace, Z a topological space, and $f : X \to Z$ a continuous map. Let $x^* \in X, z^* \in Z$ be such that $f^{-1}(z^*) = \{x^*\}$. Let $\{x_i\}_{i\in\mathbb{N}}$ be a sequence of points in Y such that $\{f(x_i)\}_{i\in\mathbb{N}}$ converges to z^*. Then $x^* \in Y$.*

[7] Interestingly, in this latter case, we get bounds differing by 2.12% using $\mathcal{D}_N = (0,0,0,0,\frac{9}{13},\frac{4}{13},0,0,0)$ in an attempt to continue the pattern from $f_{\{7,8\},8}$ and $f_{\{8\},8}$ (where we set $\mathcal{D}_N^* = (0,0,0,0,\frac{16}{25},\frac{9}{25},0,0,0)$ and $(0,0,0,0,\frac{25}{41},\frac{16}{41},0,0,0)$ in Section 5.1 and Section 4, respectively).

Proof. By compactness of X, there is a subsequence $\{x_{j_i}\}_{i\in\mathbb{N}}$ which converges to a limit $\widetilde{x}$. By closure, $\widetilde{x} \in Y$. By continuity, $f(\widetilde{x}) = z^*$, so $\widetilde{x} = x^*$. ◀

Finally, the proof of Theorem 12 is given in Appendix A.

7 Simple sketching algorithms for threshold functions

The main goal of this section is to prove Theorem 6, giving a simple "bias-based" sketching algorithm for threshold functions Th_k^i. Given an instance Ψ of $\mathsf{Max\text{-}CSP}(\mathsf{Th}_k^i)$, for $i \in [n]$, let $\mathsf{diff}_i(\Psi)$ denote the total weight of clauses in which x_i appears positively minus the weight of those in which it appears negatively; that is, if Ψ consists of clauses $(\mathbf{b}(1), \mathbf{j}(1)), \ldots, (\mathbf{b}(m), \mathbf{j}(m))$ with weights $w_1, \ldots, w_m$, then

$$\mathsf{diff}_i(\Psi) \stackrel{\text{def}}{=} \sum_{\ell\in[m] \text{ s.t. } j(\ell)_t = i \text{ for some } t\in[k]} b(\ell)_t w_\ell.$$

Let $\mathsf{bias}(\Psi) \stackrel{\text{def}}{=} \frac{1}{kW}\sum_{i=1}^n |\mathsf{diff}_i(\Psi)|$, where $W = \sum_{\ell=1}^m w_\ell$ is the total weight in Ψ.

Let $S = \{i, \ldots, k\}$ so that $\mathsf{Th}_k^i = f_{S,k}$. Recall the definitions of $\beta_{S,k}(\mu)$ and $\gamma_{S,k}(\mu)$ from Equation (7). Our simple algorithm for $\mathsf{Max\text{-}CSP}(\mathsf{Th}_k^i)$ relies on the following two lemmas, which we prove below:

▶ **Lemma 28.** *$\mathsf{val}_\Psi \le \gamma_{S,k}(\mathsf{bias}(\Psi))$.*

▶ **Lemma 29.** *$\mathsf{val}_\Psi \ge \beta_{S,k}(\mathsf{bias}(\Psi))$.*

Together, these two lemmas imply that outputting $\alpha(\mathsf{Th}_k^i)\cdot\gamma_{S,k}(\mathsf{bias}(\Psi))$ gives an $\alpha(\mathsf{Th}_k^i)$-approximation to $\mathsf{Max\text{-}CSP}(\mathsf{Th}_k^i)$, since $\alpha(\mathsf{Th}_k^i) = \inf_{\mu\in[-1,1]} \frac{\beta_{S,k}(\mu)}{\gamma_{S,k}(\mu)}$ (Equation (7)). We can implement this as a small-space sketching algorithm (up to an arbitrarily small constant $\epsilon > 0$ in the approximation ratio) because $\mathsf{bias}(\Psi)$ is measurable using ℓ_1-sketching algorithms (as used also in [12, 8, 7]) and $\gamma_{S,k}(\cdot)$ is piecewise linear:

▶ **Theorem 30** ([16, 17])**.** *For every $\epsilon > 0$, there exists an $O(\log n/\epsilon^2)$-space randomized sketching algorithm for the following problem: The input is a stream S of updates of the form $(i, v) \in [n] \times \{-\mathrm{poly}(n), \ldots, \mathrm{poly}(n)\}$, and the goal is to estimate the ℓ_1-norm of the vector $x \in [n]^n$ defined by $x_i = \sum_{(i,v)\in S} v$, up to a multiplicative factor of $1 \pm \epsilon$.*

▶ **Corollary 31.** *For $f : \{-1,1\}^k \to \{0,1\}$ and every $\epsilon > 0$, there exists an $O(\log n/\epsilon^2)$-space randomized sketching algorithm for the following problem: The input is an instance Ψ of $\mathsf{Max\text{-}CSP}(\mathsf{Th}_k^i)$ (given as a stream of constraints), and the goal is to estimate $\mathsf{bias}(\Psi)$ up to a multiplicative factor of $1 \pm \epsilon$.*

Proof. Invoke the ℓ_1-norm sketching algorithm from Theorem 30 as follows: On each input constraint $(\mathbf{b} = (b_1, \ldots, b_k), \mathbf{j} = (j_1, \ldots, j_k))$ with weight w, insert the updates $(j_1, wb_1), \ldots, (j_k, wb_k)$ into the stream (and normalize appropriately). ◀

Theorem 6 then follows from Lemmas 16, 28, and 29 and Corollary 31; we include a formal proof in Appendix A for completeness.

To prove Lemmas 28 and 29, we require a bit more setup. Adapting notation from [7, §4.2], given an instance Ψ of $\mathsf{Max\text{-}CSP}(\mathsf{Th}_k^i)$ and a "negation pattern" $\mathbf{a} = (a_1, \ldots, a_n) \in \{-1,1\}^n$ for the variables, let $\Psi^{\mathbf{a}}$ be the instance which results from Ψ by "flipping" the variables according to $\mathbf{a}$ (formally, each constraint $(\mathbf{b}, \mathbf{j})$ is replaced with $(\mathbf{b} \odot \mathbf{a}|_{\mathbf{j}}, \mathbf{j})$). We summarize the useful properties of this operation in the following claim:

▶ **Proposition 32.** *Let Ψ be an instance of Max-CSP(Th_k^i) and $\mathbf{a} = (a_1, \ldots, a_n) \in \{-1,1\}^n$. Then:*

i *For each $i \in [n]$, $\mathsf{diff}_i(\Psi^{\mathbf{a}}) = a_i \mathsf{diff}_i(\Psi)$.*
ii *$\mathsf{bias}(\Psi) = \mathsf{bias}(\Psi^{\mathbf{a}})$.*
iii *For any $\boldsymbol{\sigma} \in \{-1,1\}^n$, $\mathsf{val}_{\Psi^{\mathbf{a}}}(\boldsymbol{\sigma}) = \mathsf{val}_{\Psi}(\mathbf{a} \odot \boldsymbol{\sigma})$.*
iv *$\mathsf{val}_{\Psi^{\mathbf{a}}} = \mathsf{val}_{\Psi}$.*

The proof of Proposition 32 is given in the full version [3].

Also, given an instance Ψ, we define its "symmetrized canonical distribution" $\mathcal{D}_\Psi^{\mathrm{sym}} \in \Delta_k$ to be the distribution obtained by sampling a constraint at random from Ψ and outputting its "randomly permuted negation pattern". Formally, let S_k denote the set of permutations $[k] \to [k]$. For a vector $\mathbf{b} = (b_1, \ldots, b_k) \in \{-1,1\}^k$ and a permutation $\boldsymbol{\pi} \in \mathsf{S}_k$, let $\boldsymbol{\pi}(\mathbf{b}) = (b_{\boldsymbol{\pi}(1)}, \ldots, b_{\boldsymbol{\pi}(k)})$. Let $C(i) = (\mathbf{b}(i), \mathbf{j}(i))$ denote the i-th constraint of Ψ, with weight w_i, and let $W = \sum_{i=1}^m w_i$ be the total weight. To sample a random vector from $\mathcal{D}_\Psi^{\mathrm{sym}}$, we sample $i \in [m]$ with probability w_i/W, sample a permutation $\boldsymbol{\pi} \sim \mathsf{Unif}(\mathsf{S}_k)$, and output $\boldsymbol{\pi}(\mathbf{b}(i))$. The useful properties of $\mathcal{D}_\Psi^{\mathrm{sym}}$ are summarized in the following claim:

▶ **Proposition 33.** *Let Ψ be an instance of Max-CSP(Th_k^i). Then:*

i *For any $p \in [0,1]$, $\mathbb{E}_{\mathbf{a} \sim \mathsf{Bern}(p)^n}[\mathsf{val}_\Psi(\mathbf{a})] = \lambda_S(\mathcal{D}_\Psi^{\mathrm{sym}}, p)$.*
ii *$\mu(\mathcal{D}_\Psi^{\mathrm{sym}}) = \frac{1}{kW} \sum_{i=1}^n \mathsf{diff}_i(\Psi) \leq \mathsf{bias}(\Psi)$.*
iii *If $\mathsf{diff}_i(\Psi) \geq 0$ for all $i \in [n]$, then $\mu(\mathcal{D}_\Psi^{\mathrm{sym}}) = \mathsf{bias}(\Psi)$.*

The proof of Proposition 33 is given in the full version [3]. The proofs of Lemmas 28 and 29 are given in Appendix A.

Finally, we state another consequence of Lemma 28 – a simple randomized, $O(n)$-time-and-space streaming algorithm for *outputting* approximately-optimal assignments when the max-min method applies.

▶ **Theorem 34.** *Let Th_k^i be a threshold function and $p^* \in [0,1]$ be such that the max-min method applies, i.e.,*

$$\alpha(\mathsf{Th}_k^i) = \inf_{\mathcal{D}_N \in \Delta_k} \left(\frac{\lambda_S(\mathcal{D}_N, p^*)}{\gamma_{S,k}(\mu(\mathcal{D}_N))} \right).$$

Then the following algorithm, on input Ψ, outputs an assignment with expected value at least $\alpha(\mathsf{Th}_k^i) \cdot \mathsf{val}_\Psi$: Assign every variable to 1 if $\mathsf{diff}_i(\Psi) \geq 0$, and -1 otherwise, and then flip each variable's assignment independently with probability p^*.

The proof of Theorem 34 is given in the full version [3].

Discussion

In this paper, we introduce the max-min method and use it to resolve the streaming approximability of a wide variety of symmetric Boolean CSPs (including infinite families such as Max-kAND for all k, and Th_k^{k-1} for all even k). However, these techniques are in a sense "ad hoc" since we use computer assistance to guess the optimal solution for our optimization problem. We leave the question of whether the max-min method can be applied to determine the sketching approximability for all symmetric Boolean CSPs as an interesting open problem.

Separately, we also establish that the techniques developed in [7] are not sufficient to characterize the *streaming* approximability of all CSPs. Indeed, we show that their streaming lower bound based on "padded one-wise pairs" cannot match the approximation ratio of their optimal sketching algorithm for Max-3AND. While we believe that no $o(\sqrt{n})$-space streaming algorithm can beat their sketching algorithm for Max-3AND, proving this will require new techniques.

References

1 Arindam Biswas and Venkatesh Raman. Sublinear-Space Approximation Algorithms for Max r-SAT. In *Computing and Combinatorics (COCOON 2021, Tainan, Taiwan, October 24-26, 2021)*, volume 13025 of *LNCS*, pages 124–136. Springer, Cham, 2021.

2 Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, first edition, March 2004. doi:10.1017/CBO9780511804441.

3 Joanna Boyland, Michael Hwang, Tarun Prasad, Noah Singer, and Santhoshini Velusamy. Sketching approximations for (some) symmetric Boolean CSPs: Closed-form ratios and simple algorithms, February 2022. arXiv:2112.06319.

4 Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Near-optimal algorithms for maximum constraint satisfaction problems. *ACM Transactions on Algorithms*, 5(3):1–14, July 2009. Conference version in SODA 2007. doi:10.1145/1541885.1541893.

5 Chi-Ning Chou, Alexander Golovnev, Madhu Sudan, Ameya Velingker, and Santhoshini Velusamy. Linear Space Streaming Lower Bounds for Approximating CSPs. In *Proceedings of the 54th Annual ACM Symposium on Theory of Computing (STOC 2022, Rome, Italy, June 20-24, 2022)*, 2022. To appear.

6 Chi-Ning Chou, Alexander Golovnev, Madhu Sudan, and Santhoshini Velusamy. Approximability of all finite CSPs with linear sketches. In *Proceedings of the 62nd Annual IEEE Symposium on Foundations of Computer Science (FOCS 2021, Denver, CO, USA, February 7-10, 2022)*. IEEE Computer Society, 2021. doi:10.1109/FOCS52979.2021.00117.

7 Chi-Ning Chou, Alexander Golovnev, Madhu Sudan, and Santhoshini Velusamy. Approximability of all Boolean CSPs with linear sketches, February 2022. arXiv:2102.12351v7.

8 Chi-Ning Chou, Alexander Golovnev, and Santhoshini Velusamy. Optimal Streaming Approximations for all Boolean Max-2CSPs and Max-kSAT. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS 2020, Virtual, November 16-19, 2020)*, pages 330–341. IEEE Computer Society, November 2020. doi:10.1109/FOCS46700.2020.00039.

9 Lars Engebretsen and Jonas Holmerin. More efficient queries in PCPs for NP and improved approximation hardness of maximum CSP. *Random Structures and Algorithms*, 33(4):497–514, December 2008. Conference version in STACS 2005. doi:10.1002/rsa.20226.

10 Uriel Feige and Michel X. Goemans. Approximating the value of two prover proof systems, with applications to MAX 2SAT and MAX DICUT. In *Proceedings of the 3rd Israel Symposium on the Theory of Computing and Systems (ISTCS 2003, January 4-6, 1995)*, pages 182–189. IEEE Computer Society, 1995. doi:10.1109/ISTCS.1995.377033.

11 Michel X. Goemans and David P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM*, 42(6):1115–1145, November 1995. Conference version in STOC 1994. doi:10.1145/227683.227684.

12 Venkatesan Guruswami, Ameya Velingker, and Santhoshini Velusamy. Streaming Complexity of Approximating Max 2CSP and Max Acyclic Subgraph. In Klaus Jansen, José D. P. Rolim, David Williamson, and Santosh S. Vempala, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX 2017, Berkeley, CA, USA, August 16-18, 2017)*, volume 81 of *LIPIcs*, pages 8:1–8:19. Schloss Dagstuhl — Leibniz-Zentrum für Informatik, August 2017. doi:10.4230/LIPIcs.APPROX-RANDOM.2017.8.

13 Gustav Hast. Approximating Max kCSP Using Random Restrictions. In Klaus Jansen, Sanjeev Khanna, José D. P. Rolim, and Dana Ron, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX 2004, Cambridge, MA, USA, August 22-24, 2004)*, volume 3122 of *LNCS*, pages 151–162. Springer, 2004. doi:10.1007/978-3-540-27821-4_14.

14 Gustav Hast. Approximating Max kCSP – Outperforming a Random Assignment with Almost a Linear Factor. In Luís Caires, Giuseppe F. Italiano, Luís Monteiro, Catuscia Palamidessi, and Moti Yung, editors, *Automata, Languages and Programming (ICALP 2005, July 11-15, 2005)*, volume 3580 of *LNCS*, pages 956–968. Springer, 2005. doi:10.1007/11523468_77.

15 Johan Håstad. Some optimal inapproximability results. *Journal of the ACM*, 48(4):798–859, 2001. doi:10.1145/502090.502098.

16 Piotr Indyk. Stable distributions, pseudorandom generators, embeddings, and data stream computation. *Journal of the ACM*, 53(3):307–323, May 2006. Conference version in FOCS 2000. doi:10.1145/1147954.1147955.

17 Daniel M. Kane, Jelani Nelson, and David P. Woodruff. On the Exact Space Complexity of Sketching and Streaming Small Norms. In *Proceedings of the 2010 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2010, Austin, TX, USA, January 17-19, 2010)*, pages 1161–1178. Society for Industrial and Applied Mathematics, 2010. doi:10.1137/1.9781611973075.93.

18 Michael Kapralov, Sanjeev Khanna, and Madhu Sudan. Streaming lower bounds for approximating MAX-CUT. In *Proceedings of the 26th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2015, San Diego, California, USA, January 4-6, 2015)*, pages 1263–1282. Society for Industrial and Applied Mathematics, January 2015. doi:10.1137/1.9781611973730.84.

19 Alex Samorodnitsky and Luca Trevisan. A PCP characterization of NP with optimal amortized query complexity. In *Proceedings of the 32nd Annual ACM Symposium on Theory of Computing (STOC 2000, Portland, OR, USA, May 21-23, 2000)*, pages 191–199, Portland, Oregon, United States, 2000. Association for Computing Machinery. doi:10.1145/335305.335329.

20 Alex Samorodnitsky and Luca Trevisan. Gowers Uniformity, Influence of Variables, and PCPs. *SIAM Journal on Computing*, 39(1):323–360, January 2009. Conference version in STOC 2006. doi:10.1137/070681612.

21 Noah Singer. *On Streaming Approximation Algorithms for Constraint Satisfaction Problems*. Bachelor's thesis, Harvard University, Cambridge, MA, March 2022.

22 Madhu Sudan and Luca Trevisan. Probabilistically checkable proofs with low amortized query complexity. In *Proceedings of the 39th Annual Symposium on Foundations of Computer Science (SFCS 1998, Palo Alto, CA, USA, November 8-11, 1998)*, pages 18–27. IEEE Computer Society, 1998. doi:10.1109/SFCS.1998.743425.

23 Luca Trevisan. Parallel Approximation Algorithms by Positive Linear Programming. *Algorithmica*, 21(1):72–88, May 1998. doi:10.1007/PL00009209.

24 Luca Trevisan. Recycling queries in PCPs and in linearity tests. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing (STOC 1998, Dallas, Texas, USA, May 24-26, 1998)*, pages 299–308. Association for Computing Machinery, 1998. doi:10.1145/276698.276769.

25 Luca Trevisan, Gregory B. Sorkin, Madhu Sudan, and David P. Williamson. Gadgets, Approximation, and Linear Programming. *SIAM Journal on Computing*, 29(6):2074–2097, January 2000. Conference version in FOCS 1996. doi:10.1137/S0097539797328847.

26 Uri Zwick. Approximation algorithms for constraint satisfaction problems involving at most three variables per constraint. In *Proceedings of the 9th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 1998, San Francisco, CA, USA, January 25-27, 1998)*, pages 201–210. Association for Computing Machinery, 1998. doi:10.5555/314613.314701.

A Miscellaneous technical proofs

Proof outline of Theorem 12. As discussed in Section 2.3, since $k = 3$ is odd, to prove Theorem 1 we show, using the max-min method, that $\mathcal{D}_N^* = (0,0,1,0)$ minimizes $\frac{\beta_{\{3\}}(\cdot)}{\gamma_{\{3\},3}(\mu(\cdot))}$. We can show that the corresponding $\gamma_{\{3\},3}$ value is achieved by $\mathcal{D}_Y^* = (\frac{1}{3},0,0,\frac{2}{3})$. In particular, $(\mathcal{D}_N^*, \mathcal{D}_Y^*)$ are not a padded one-wise pair.

We can show that the minimizer of $\gamma_{\{3\}}$ for a particular μ is in general unique. Hence, it suffices to furthermore show that $\mathcal{D}_N^*$ is the *unique* minimizer of $\frac{\beta_{\{3\}}(\cdot)}{\gamma_{\{3\},3}(\mu(\cdot))}$. For this purpose, the max-min method is not sufficient because $\frac{\lambda_{\{3\}}(\cdot,p^*)}{\gamma_{\{3\},3}(\mu(\cdot))}$ is not uniquely minimized at $\mathcal{D}_N^*$ (where we chose $p^* = \frac{2}{3}$). Intuitively, this is because p^* is not a good enough estimate for the maximizer of $\lambda_{\{3\}}(\mathcal{D}_N, \cdot)$. To remedy this, we observe that $\lambda_{\{3\}}((1,0,0,0),\cdot), \lambda_{\{3\}}((0,1,0,0),\cdot)$, $\lambda_{\{3\}}((0,0,1,0),\cdot)$ and $\lambda_{\{3\}}((0,0,0,1),\cdot)$ are minimized at $0, \frac{1}{3}, \frac{2}{3}$, and 1, respectively. Hence, we instead lower-bound $\lambda_{\{3\}}(\mathcal{D}_N,\cdot)$ by evaluating at $\frac{1}{3}\mathcal{D}_N\langle 1\rangle + \frac{2}{3}\mathcal{D}_N\langle 2\rangle + \mathcal{D}_N\langle 3\rangle$, which does suffice to prove the uniqueness of $\mathcal{D}_N^*$. The theorem then follows from continuity arguments. ◀

Proof of Lemma 17. Consider the case where k is odd. Define $\mathcal{D}_N^*$ by $\mathcal{D}_N^*\langle\frac{k+1}{2}\rangle = 1$ and let $p^* = \frac{1}{2} + \frac{1}{2k}$. Since

$$\alpha(k\mathsf{AND}) \le \frac{\beta_{\{k\}}(\mathcal{D}_N^*)}{\gamma_{\{k\},k}(\mu(\mathcal{D}_N^*))} \text{ and } \beta_{\{k\}}(\mathcal{D}_N) = \sup_{p\in[0,1]} \lambda_{\{k\}}(\mathcal{D}_N^*, p),$$

by Equations (4) and (8), respectively, it suffices to check that p^* maximizes $\lambda_{\{k\}}(\mathcal{D}_N^*,\cdot)$ and

$$\frac{\lambda_{\{k\}}(\mathcal{D}_N^*, p^*)}{\gamma_{\{k\},k}(\mu(\mathcal{D}_N^*))} = \alpha'_k.$$

Indeed, by Equation (11),

$$\lambda_{\{k\}}(\mathcal{D}_N^*, p) = (1-p)^{(k-1)/2}p^{(k+1)/2}.$$

To show p^* maximizes $\lambda_{\{k\}}(\mathcal{D}_N^*,\cdot)$, we calculate its derivative:

$$\frac{d}{dp}\left[(1-p)^{(k-1)/2}p^{(k+1)/2}\right] = -(1-p)^{(k-3)/2}p^{(k-1)/2}\left(kp - \frac{k+1}{2}\right),$$

which has zeros only at $0, 1$, and p^*. Thus, $\lambda_{\{k\}}(\mathcal{D}_N^*,\cdot)$ has critical points only at $0, 1$, and p^*, and it is maximized at p^* since it vanishes at 0 and 1. Finally, by Equations (10) and (11) and the definition of α'_k,

$$\frac{\lambda_{\{k\}}(\mathcal{D}_N^*, p^*)}{\gamma_{\{k\},k}(\mu(\mathcal{D}_N^*))} = \frac{\left(\frac{1}{2}-\frac{1}{2k}\right)^{(k-1)/2}\left(\frac{1}{2}+\frac{1}{2k}\right)^{(k+1)/2}}{\frac{1}{2}\left(1+\frac{1}{k}\right)} = \alpha'_k,$$

as desired.

Similarly, consider the case where k is even; here, we define $\mathcal{D}_N^*$ by $\mathcal{D}_N^*\langle\frac{k}{2}\rangle = \frac{\left(\frac{k}{2}+1\right)^2}{\left(\frac{k}{2}\right)^2+\left(\frac{k}{2}+1\right)^2}$ and $\mathcal{D}_N^*\langle\frac{k}{2}+1\rangle = \frac{\left(\frac{k}{2}\right)^2}{\left(\frac{k}{2}\right)^2+\left(\frac{k}{2}+1\right)^2}$, and set $p^* = \frac{1}{2} + \frac{1}{2(k+1)}$. Using Equation (11) to calculate the derivative of $\lambda_{\{k\}}(\mathcal{D}_N^*,\cdot)$ yields

$$\frac{d}{dp}\left[\frac{\left(\frac{k}{2}+1\right)^2}{\left(\frac{k}{2}\right)^2+\left(\frac{k}{2}+1\right)^2}(1-p)^{k/2}p^{k/2} + \frac{\left(\frac{k}{2}\right)^2}{\left(\frac{k}{2}\right)^2+\left(\frac{k}{2}+1\right)^2}(1-p)^{k/2-1}p^{k/2+1}\right]$$
$$= -\frac{k}{2+2k+2k^2}(1-p)^{k/2-2}p^{k/2-1}\left(\frac{k}{2}+1-2p\right)\left((k+1)p - \left(\frac{k}{2}+1\right)\right),$$

so $\lambda_{\{k\}}(\mathcal{D}_N^*,\cdot)$ has critical points at $0, 1, \frac{1}{2}+\frac{k}{4}$. and p^*; p^* is the only critical point in the interval $[0,1]$ for which $\lambda_{\{k\}}(\mathcal{D}_N^*,\cdot)$ is positive, and hence is its maximum. Finally, it can be verified algebraically using Equations (10) and (11) that $\frac{\lambda_{\{k\}}(\mathcal{D}_N^*,p^*)}{\gamma_{\{k\},k}(\mu(\mathcal{D}_N^*))} = 2\alpha'_{k+1}$, as desired. ◀

Proof of Lemma 18. First, suppose $k \geq 3$ is odd. Set $p^* = \frac{1}{2} + \frac{1}{2k} = \frac{k+1}{2k}$. We want to show that

$$\alpha'_k \leq \inf_{\mathcal{D}_N \in \Delta_k} \frac{\lambda_{\{k\}}(\mathcal{D}_N, p^*)}{\gamma_{\{k\},k}(\mu(\mathcal{D}_N))} \qquad \text{(max-min inequality, i.e., Equation (9))}$$
$$= \inf_{\mathcal{D}_N \in \Delta_k} \frac{\sum_{i=0}^k (1-p^*)^{k-i}(p^*)^i \, \mathcal{D}_N\langle i\rangle}{\sum_{i=0}^k \frac{i}{k} \, \mathcal{D}_N\langle i\rangle}. \qquad \text{(Equations (10) and (11))}$$

By Proposition 19, it suffices to check that

$$\forall i \in \{0\} \cup [k], \quad (1-p^*)^{k-i}(p^*)^i \geq \alpha'_k \cdot \frac{i}{k}.$$

By definition of α'_k, we have that $\alpha'_k = (1-p^*)^{(k-1)/2}(p^*)^{(k-1)/2}$. Defining $r = \frac{p^*}{1-p^*} = \frac{k+1}{k-1}$ (so that $p^* = r(1-p^*)$), factoring out $(1-p^*)^k$, and simplifying, we can rewrite our desired inequality as

$$\forall i \in \{0\} \cup [k], \quad \frac{1}{2}(k-1)r^{i-\frac{k-1}{2}} \geq i. \tag{15}$$

When $i = \frac{k+1}{2}$ or $\frac{k-1}{2}$, we have equality in Equation (15). We extend to the other values of i by induction. Indeed, when $i \geq \frac{k+1}{2}$, then "i satisfies Equation (15)" implies "$i+1$ satisfies Equation (15)" because $ri \geq i+1$, and when $i \leq \frac{k-1}{2}$, then "i satisfies Equation (15)" implies "$i-1$ satisfies Equation (15)" because $\frac{1}{r}i \geq i-1$.

Similarly, in the case where $k \geq 2$ is even, we set $p^* = \frac{1}{2} + \frac{1}{2(k+1)}$ and $r = \frac{p^*}{1-p^*} = \frac{k+2}{k}$. In this case, for $i \in \{0\} \cup [k]$ the following analogue of Equation (15) can be derived:

$$\forall i \in \{0\} \cup [k], \quad \frac{1}{2}kr^{i-\frac{k}{2}} \geq i,$$

and these inequalities follow from the same inductive argument. ◀

Proof of Lemma 20. As in the proof of Lemma 17, it suffices to construct $\mathcal{D}_N^*$ and p^* such that p^* maximizes $\lambda_{\{k-1,k\}}(\mathcal{D}_N^*, \cdot)$ and $\frac{\lambda_{\{k-1,k\}}(\mathcal{D}_N^*, p^*)}{\gamma_{\{k-1,k\},k}(\mu(\mathcal{D}_N^*))} = \frac{k}{2}\alpha'_{k-1}$.

We again let $p^* = \frac{1}{2} + \frac{1}{2(k-1)}$, but define $\mathcal{D}_N^*$ by $\mathcal{D}_N^*\langle \frac{k}{2}\rangle = \frac{\left(\frac{k}{2}\right)^2}{\left(\frac{k}{2}\right)^2 + \left(\frac{k}{2}-1\right)^2}$ and $\mathcal{D}_N^*\langle \frac{k}{2}+1\rangle = \frac{\left(\frac{k}{2}-1\right)^2}{\left(\frac{k}{2}\right)^2 + \left(\frac{k}{2}-1\right)^2}$. By Equation (13), the derivative of $\lambda_{\{k-1,k\}}(\mathcal{D}_N^*, \cdot)$ is now

$$\frac{d}{dp}\Bigg[\frac{\left(\frac{k}{2}\right)^2}{\left(\frac{k}{2}\right)^2 + \left(\frac{k}{2}-1\right)^2}(1-p)^{k/2-1}p^{k/2-1}\left(\frac{k}{2}p^2 + pq + \frac{k}{2}q^2\right) +$$
$$\frac{\left(\frac{k}{2}-1\right)^2}{\left(\frac{k}{2}\right)^2 + \left(\frac{k}{2}-1\right)^2}(1-p)^{k/2-2}p^{k/2}\left(\left(\frac{k}{2}-1\right)p^2 + pq + \left(\frac{k}{2}+1\right)q^2\right)\Bigg]$$
$$= -\frac{1}{8(k^2-2k+2)}(1-p)^{k/2-3}p^{k/2-2}(-k + (2(k-1)p)\xi(p),$$

where $\xi(p)$ is the cubic

$$\xi(p) = -8k(k-1)p^3 + 2(k^3+k^2+6k-12)p^2 - 2(k^3-4)p + k^2(k-2).$$

Thus, $\lambda_{\{k-1,k\}}$'s critical points on the interval $[0,1]$ are $0, 1, p^*$ and any roots of ξ in this interval. We claim that ξ has no additional roots in the interval $(0,1)$. This can be verified directly by calculating roots for $k = 2, 4$, so assume WLOG $k \geq 6$.

Suppose $\xi(p) = 0$ for some $p \in (0,1)$, and let $x = \frac{1}{p} - 1 \in (0, \infty)$. Then $p = \frac{1}{1+x}$; plugging this in for p and multiplying through by $(x+1)^3$ gives the new cubic

$$(k^3 - 2k^2)x^3 + (k^3 - 6k^2 + 8)x^2 + (k^3 - 4k^2 + 12k - 8)x + (k^3 - 8k^2 + 20k - 16) = 0 \quad (16)$$

whose coefficients are cubic in k. It can be verified by calculating the roots of each coefficient of x in Equation (16) that all coefficients are positive for $k \geq 6$. Thus, Equation (16) cannot have roots for positive x, a contradiction. Hence $\lambda_{\{k-1,k\}}(\mathcal{D}_N^*, \cdot)$ is maximized at p^*. Finally, it can be verified that $\frac{\lambda_{\{k-1,k\}}(\mathcal{D}_N^*, p^*)}{\gamma_{\{k-1,k\},k}(\mu(\mathcal{D}_N^*))} = \frac{k}{2}\alpha'_{k-1}$, as desired. ◀

Proof of Lemma 21. Define $p^* = \frac{1}{2} + \frac{1}{2(k-1)}$. Following the proof of Lemma 18 and using the lower bound $\gamma_{\{k-1,k\},k}(\mu(\mathcal{D}_N)) \leq \sum_{i=0}^{k} \frac{i}{k-1}\mathcal{D}_N\langle i\rangle$, it suffices to show that

$$\frac{k}{2}\alpha'_{k-1} \leq \inf_{\mathcal{D}_N \in \Delta_k} \frac{\sum_{i=0}^{k}(1-p^*)^{k-i-1}(p^*)^{i-1}((k-i)(p^*)^2 + p^*(1-p^*) + i(1-p^*)^2)\,\mathcal{D}_N\langle i\rangle}{\sum_{i=0}^{k}\frac{i}{k-1}\mathcal{D}_N\langle i\rangle}$$

for which by Proposition 19, it in turn suffices to prove that for each $i \in \{0\} \cup [k]$,

$$\frac{k}{2}\alpha'_{k-1}\frac{i}{k-1} \leq (1-p^*)^{k-i-1}(p^*)^{i-1}((k-i)(p^*)^2 + p^*(1-p^*) + i(1-p^*)^2).$$

We again observe that $\alpha'_{k-1} = (1-p^*)^{k/2-1}(p^*)^{k/2-1}$, define $r = \frac{p^*}{1-p^*} = \frac{k}{k-2}$, and factor out $(1-p^*)^{k-1}$, which simplifies our desired inequality to

$$\frac{1}{2}r^{i-\frac{k}{2}-1} \cdot \frac{k-2}{k-1}\left(i + r + (k-i)r^2\right) \geq i. \quad (17)$$

for each $i \in \{0\} \cup [k]$. Again, we assume $k \geq 6$ WLOG; the bases cases $i = \frac{k}{2} - 1, \frac{k}{2}$ can be verified directly, and we proceed by induction. If Equation (17) holds for i, and we seek to prove it for $i+1$, it suffices to cross-multiply and instead prove the inequality

$$r(i + 1 + r + (k-(i+1))r^2)i \geq (i+1)(i + r + (k-i)r^2),$$

which simplifies to

$$(k-2i)(k-1)(k^2 - 4i - 4) \leq 0,$$

which holds whenever $\frac{k}{2} \leq i \leq \frac{k^2-4}{4}$ (and $\frac{k^2-4}{4} \geq k$ for all $k \geq 6$). The other direction (where $i \leq \frac{k}{2} - 1$ and we induct downwards) is similar. ◀

Proof of Theorem 6. To get an $(\alpha - \epsilon)$-approximation to val_Ψ, let $\delta > 0$ be small enough such that $\frac{1-\delta}{1+\delta}\alpha(\mathsf{Th}_k^i) \geq \alpha(\mathsf{Th}_k^i) - \epsilon$. We claim that calculating an estimate $\widehat{b}$ for $\mathsf{bias}(\Psi)$ (using Corollary 31) up to a multiplicative δ factor and outputting $\widehat{v} = \alpha(\mathsf{Th}_k^i)\gamma_{S,k}(\frac{\widehat{b}}{1+\delta})$ is sufficient.

Indeed, suppose $\widehat{b} \in [(1-\delta)\mathsf{bias}(\Psi), (1+\delta)\mathsf{bias}(\Psi)]$; then $\frac{\widehat{b}}{1+\delta} \in [\frac{1-\delta}{1+\delta}\mathsf{bias}(\Psi), \mathsf{bias}(\Psi)]$. Now we observe

$$\begin{aligned}
\gamma_{S,k}\left(\frac{\widehat{b}}{1+\delta}\right) &\geq \gamma_{S,k}\left(\frac{1-\delta}{1+\delta}\mathsf{bias}(\Psi)\right) && \text{(monotonicity of } \gamma_{S,k}) \\
&= \min\left\{\frac{1 + \frac{1-\delta}{1+\delta}\mathsf{bias}(\Psi)}{1+\epsilon_{s,k}}, 1\right\} && \text{(Lemma 16)} \\
&\geq \frac{1-\delta}{1+\delta}\min\left\{\frac{1+\mathsf{bias}(\Psi)}{1+\epsilon_{s,k}}, 1\right\} && (\delta > 0) \\
&= \frac{1-\delta}{1+\delta}\gamma_{S,k}(\mathsf{bias}(\Psi)). && \text{(Lemma 16)}
\end{aligned}$$

Then we conclude

$$\begin{aligned}(\alpha(\mathsf{Th}_k^i)-\epsilon)\mathsf{val}_\Psi &\le (\alpha(\mathsf{Th}_k^i)-\epsilon)\gamma_{S,k}(\mathsf{bias}(\Psi)) && \text{(Lemma 28)}\\ &\le \alpha(\mathsf{Th}_k^i)\cdot\frac{1-\delta}{1+\delta}\gamma_{S,k}(\mathsf{bias}(\Psi)) && \text{(assumption on } \delta)\\ &\le \widehat{v} && \text{(our observation)}\\ &\le \alpha(\mathsf{Th}_k^i)\gamma_{S,k}(\mathsf{bias}(\Psi)) && \text{(monotonicity of } \gamma_{S,k})\\ &\le \beta_{S,k}(\mathsf{bias}(\Psi)) && \text{(Equation (7))}\\ &\le \mathsf{val}_\Psi, && \text{(Lemma 29)}\end{aligned}$$

as desired. ◀

Proof of Theorem 12. By Lemma 26, $\frac{\beta_{\{3\}}(\mathcal{D}_N)}{\gamma_{\{3\},3}(\mu(\mathcal{D}_N))}$ is minimized *uniquely* at $\mathcal{D}_N^* = (0,0,1,0)$. By Lemma 14 we have $\mu(\mathcal{D}_N^*) = \frac{1}{3}$, and by inspection from the proof of Lemma 16 below, $\gamma_{\{3\}}(\mathcal{D}_Y)$ with $\mu(\mathcal{D}_Y) = \frac{1}{3}$ is uniquely minimized by $\mathcal{D}_Y^* = (\frac{1}{3},0,0,\frac{2}{3})$.

Finally, we rule out the possibility of an infinite sequence of padded one-wise pairs which achieve ratios arbitrarily close to $\frac{2}{9}$ using topological properties. View a distribution $\mathcal{D} \in \Delta_3$ as the vector $(\mathcal{D}\langle 0\rangle, \mathcal{D}\langle 1\rangle, \mathcal{D}\langle 2\rangle, \mathcal{D}\langle 3\rangle) \in \mathbb{R}^4$. Let $D \subset \mathbb{R}^4$ denote the set of such distributions. Let $M \subset D \times D \subset \mathbb{R}^8$ denote the subset of pairs of distributions with matching marginals, and let $M' \subset M$ denote the subset of pairs with uniform marginals and $P \subset M$ the subset of padded one-wise pairs. D, M, M', and P are compact (under the Euclidean topology); indeed, D, M, and M' are bounded and defined by a finite collection of linear equalities and strict inequalities, and letting $M' \subset M$ denote the subset of pairs of distributions with matching *uniform* marginals, P is the image of the compact set $[0,1] \times D \times M' \subset \mathbb{R}^{13}$ under the continuous map $\tau \times \mathcal{D}_0 \times (\mathcal{D}'_Y, \mathcal{D}'_N) \mapsto (\tau\mathcal{D}_0 + (1-\tau)\mathcal{D}'_Y, \tau\mathcal{D}_0 + (1-\tau)\mathcal{D}'_N)$. Hence, P is closed.

Now the function

$$\alpha : M \to \mathbb{R} \cup \{\infty\} : (\mathcal{D}_N, \mathcal{D}_Y) \mapsto \frac{\beta_{\{3\}}(\mathcal{D}_N)}{\gamma_{\{3\}}(\mathcal{D}_Y)}$$

is continuous, since a ratio of continuous functions is continuous, and $\beta_{\{3\}}$ is a single-variable supremum of a continuous function (i.e., λ_S) over a compact interval, which is in general continuous in the remaining variables. Thus, if there were a sequence of padded one-wise pairs $\{(\mathcal{D}_N^{(i)}, \mathcal{D}_Y^{(i)}) \in P\}_{i\in\mathbb{N}}$ such that $\alpha(\mathcal{D}_N^{(i)}, \mathcal{D}_Y^{(i)})$ converges to $\frac{2}{9}$ as $i \to \infty$, since M is compact and P is closed, Lemmas 26 and 27 imply that $(\mathcal{D}_N^*, \mathcal{D}_Y^*) \in P$, a contradiction. ◀

Proof of Lemma 28. Let $\mathbf{opt} \in \{-1,1\}^n$ denote the optimal assignment for Ψ. Then

$$\begin{aligned}\mathsf{val}_\Psi &= \mathsf{val}_\Psi(\mathbf{opt}) && \text{(definition of } \mathbf{opt})\\ &= \mathsf{val}_{\Psi^{\mathbf{opt}}}(1^n) && \text{(Item iii of Proposition 32)}\\ &= \lambda_S(\mathcal{D}^{\mathrm{sym}}_{\Psi^{\mathbf{opt}}}, 1) && \text{(Item i of Proposition 33 with } p=1)\\ &= \gamma_S(\mathcal{D}^{\mathrm{sym}}_{\Psi^{\mathbf{opt}}}) && \text{(definition of } \gamma_S\text{, Equation (4))}\\ &\le \gamma_{S,k}(\mu(\mathcal{D}^{\mathrm{sym}}_{\Psi^{\mathbf{opt}}})) && \text{(definition of } \gamma_{S,k}\text{, Equation (6))}\\ &\le \gamma_{S,k}(\mathsf{bias}(\Psi^{\mathbf{opt}})) && \text{(Item ii of Proposition 33 and monotonicity of } \gamma_{S,k})\\ &= \gamma_{S,k}(\mathsf{bias}(\Psi)), && \text{(Item ii of Proposition 32)}\end{aligned}$$

as desired. ◀

Proof of Lemma 29. Let $\mathbf{maj} \in \{-1,1\}^n$ denote the assignment assigning x_i to 1 if $\mathsf{diff}_i(\Psi) \geq 0$ and -1 otherwise. Now

$$\begin{aligned}
\mathsf{val}_\Psi &= \mathsf{val}_{\Psi^{\mathbf{maj}}} && \text{(Item iv of Proposition 32)}\\
&\geq \sup_{p\in[0,1]} \left(\mathop{\mathbb{E}}_{\mathbf{a}\sim\mathsf{Bern}(p)^n} [\mathsf{val}_{\Psi^{\mathbf{maj}}}(\mathbf{a})] \right) && \text{(probabilistic method)}\\
&= \sup_{p\in[0,1]} (\lambda_S(\mathcal{D}^{\mathrm{sym}}_{\Psi^{\mathbf{maj}}}, p)) && \text{(Item i of Proposition 33)}\\
&\geq \beta_S(\mathcal{D}^{\mathrm{sym}}_{\Psi^{\mathbf{maj}}}) && \text{(definition of } \beta_S\text{, Equation (4))}\\
&\geq \beta_{S,k}(\mu(\mathcal{D}^{\mathrm{sym}}_{\Psi^{\mathbf{maj}}})) && \text{(definition of } \beta_{S,k}\text{, Equation (6))}\\
&= \beta_{S,k}(\mathsf{bias}(\Psi^{\mathbf{maj}})) && \text{(Item iii of Proposition 33)}\\
&= \beta_{S,k}(\mathsf{bias}(\Psi)), && \text{(Item ii of Proposition 32)}
\end{aligned}$$

as desired. ◀