On sketching approximations for symmetric Boolean CSPs

Joanna Boyland\(^*\) Michael Hwang\(^*\) Tarun Prasad\(^*\) Noah Singer\(^*†\)
Santhoshini Velusamy\(^‡\)

May 4, 2022

Abstract

A Boolean maximum constraint satisfaction problem, \textit{Max-CSP}(\(f\)), is specified by a predicate \(f : \{-1, 1\}^k \rightarrow \{0, 1\}\). An \(n\)-variable instance of \textit{Max-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 [CGV20] obtained explicit ratios characterizing the \(\sqrt{n}\)-space streaming approximability of every predicate. For \(k \geq 3\), Chou, Golovnev, Sudan, and Velusamy [CGSV21a] 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 \(\epsilon > 0\), \textit{Max-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 \(\alpha'_{k} = 2^{-(k-1)(1 - k^{-2})} / 2\), we show that for odd \(k \geq 3\), \(\alpha(\text{kAND}) = \alpha'_{k}\), and for even \(k \geq 2\), \(\alpha(\text{kAND}) = 2\alpha'_{k+1}\). Thus, for every \(k\), \(\text{kAND}\) 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 [CGS + 22] implying that \textit{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-\(k\)-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 [CGSV21a] only implies that \(\alpha(f)\) can be computed to arbitrary precision in \(\text{PSPACE}\), and thus closed-form expressions need not have existed \textit{a priori}. Our analyses involve identifying and exploiting structural “saddle-point” properties of this dichotomy.

Separately, for all \textit{threshold} functions, we give optimal “bias-based” approximation algorithms generalizing [CGV20] while simplifying [CGSV21a]. Finally, we investigate the \(\sqrt{n}\)-space \textit{streaming} lower bounds in [CGSV21a], and show that they are \textit{incomplete} for \(3\text{AND}\), i.e., they fail to rule out \((\alpha(3\text{AND}) - \epsilon)\)-approximations in \(o(\sqrt{n})\) space.

\(^*\)Harvard College, Harvard University, Cambridge, MA, USA. Emails: \{jboyland,michaelhwang1,tarumuraliprasad,noahsinger\}@college.harvard.edu.
\(^†\)Supported by the Harvard College Research Program.
\(^‡\)School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA. Supported in part by a Google Ph.D. Fellowship, a Simons Investigator Award to Madhu Sudan, and NSF Awards CCF 1715187 and CCF 2152413. Email: svelusamy@g.harvard.edu.
1 Introduction

In this work, we consider the streaming approximability of various Boolean constraint satisfaction problems, and we begin by defining these terms. See [CGSV21a, §1.1-2] for more details on the definitions.

1.1 Setup: The streaming approximability of Boolean CSPs

1.1.1 Boolean CSPs

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

For Boolean vectors $a = (a_1, \ldots, a_n), b = (b_1, \ldots, b_n) \in \{−1, 1\}^n$, let $a \odot b$ denote their coordinate-wise product $(a_1 b_1, a_2 b_2, \ldots, a_n b_n)$. An assignment $\sigma = (\sigma_1, \ldots, \sigma_n) \in \{−1, 1\}^n$ satisfies $C$ iff $f(b \odot \sigma|_j) = 1$, where $\sigma|_j$ is the $k$-tuple $(\sigma_{j_1}, \ldots, \sigma_{j_k})$ (i.e., $\sigma$ satisfies $C$ iff $f(b_1 \sigma_{j_1}, \ldots, b_k \sigma_{j_k}) = 1$). An instance $\Psi$ of Max-CSP($f$) consists of constraints $C_1, \ldots, C_m$ with non-negative weights $w_1, \ldots, w_m$ where $C_i = (j(i), b(i))$ and $w_i \in \mathbb{R}$ for each $i \in [m]$; the value $\text{val}_\Psi(\sigma)$ of an assignment $\sigma$ to $\Psi$ is the (weighted) fraction of constraints in $\Psi$ satisfied by $\sigma$, i.e., $\text{val}_\Psi(\sigma) \overset{\text{def}}{=} \frac{1}{W} \sum_{i=1}^m w_i \cdot f(b(i) \odot \sigma|_{j(i)})$, where $W = \sum_{i=1}^m w_i$. The value $\text{val}_\Psi$ of an instance $\Psi$ is the maximum value of any assignment $\sigma \in \{−1, 1\}^n$, i.e., $\text{val}_\Psi \overset{\text{def}}{=} \max_{\sigma \in \{−1, 1\}^n} \text{val}_\Psi(\sigma)$.

1.1.2 Approximations to CSPs

For $\alpha \in [0, 1]$, we consider the problem of $\alpha$-approximating Max-CSP($f$). In this problem, the goal of an algorithm $A$ is to, on input an instance $\Psi$, output an estimate $A(\Psi)$ such that with probability at least $\frac{2}{3}$, $\alpha \cdot \text{val}_\Psi \leq A(\Psi) \leq \text{val}_\Psi$. For $\beta < \gamma \in [0, 1]$, we also consider the closely related $(\beta, \gamma)$-Max-CSP($f$). In this problem, the input instance $\Psi$ is promised to either satisfy $\text{val}_\Psi \leq \beta$ or $\text{val}_\Psi \geq \gamma$, 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 Max-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 $\Psi$, 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 [CGSV21a, §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 $\rho(f) \overset{\text{def}}{=} \Pr_{b \sim \text{Unif}\{-1, 1\}^k}[f(b) = 1]$, where $\text{Unif}\{-1, 1\}^k$ 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 $\mathbb{E}_{a \sim \text{Unif}\{-1, 1\}^n}[\text{val}_f(a)] = \Pr_{b \sim \text{Unif}\{-1, 1\}^n}[f(b) = 1] = \rho(f)$. 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 [KKS15] showed that Max-2XOR is approximation-resistant to $o(\sqrt{n})$-space streaming algorithms. Later, Chou, Golovnev, and Velusamy [CGV20], building on earlier work of Guruswami, Velusamy, and Velingker [GVV17], gave an $O(\log n)$-space linear sketching algorithm which $(\frac{4}{5} - \epsilon)$-approximates Max-2AND for every $\epsilon > 0$ and showed that $(\frac{4}{5} + \epsilon)$-approximations require $\Omega(\sqrt{n})$ space, even for streaming algorithms.

In two recent works [CGSV21a; CGSV21b], Chou, Golovnev, Sudan, and Velusamy proved so-called dichotomy theorems for sketching CSPs. In [CGSV21a], 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 [CGSV21b], they extend it to the more general case of CSPs over finite alphabets.\[^1\]

[CGSV21a] 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 [CGSV21a] is the following dichotomy theorem: For any $0 \leq \gamma < \beta \leq 1$, either

1. $(\beta, \gamma)$-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 $f, \gamma, \beta$) 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).

A priori, it may be possible to achieve an $(\alpha(f) + \epsilon)$-approximation with a $o(\sqrt{n})$-space streaming algorithm. But [CGSV21a] 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., [KKS15; CGV20]), and prove several new ones. In particular, [CGSV21a] 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 $\mathbb{E}_{b \sim \mathcal{D}}[b] = 0$ for all $i \in [k]$ (where $[k] \overset{\text{def}}{=} \{1, \ldots, k\}$), then Max-CSP($f$) is streaming approximation-resistant. For symmetric Boolean CSPs, they also prove the converse, and thus give a complete characterization for approximation resistance [CGSV21a, Lemma 2.14]. However, besides Max-2AND, [CGSV21a]

\[^1\]More precisely, [CGSV21a] and [CGSV21b] 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.
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 [CGSV21a] leave unanswered:

1. Can the framework in [CGSV21a] be used to find closed-form sketching approximability ratios \( \alpha(f) \) for approximable problems Max-CSP\((f)\) beyond Max-2AND?

2. As observed in [CGS+22, §1.3], [CGSV21a] 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 [CGSV21a] suffice to resolve the streaming approximability of every function?

4. The optimal (\( \alpha(f) - \epsilon \))-approximation algorithm for Max-CSP\((f)\) in [CGSV21a] requires running a “grid” of \( O(1/\epsilon^2) \) distinguishers for \( (\beta, \gamma) \)-Max-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 \( f_{S,k} : \{-1,1\}^k \rightarrow \{0,1\} \) as the indicator function for the set \( \{b \in \{-1,1\}^k : \text{wt}(b) \in S\} \) (where \( \text{wt}(b) \) denotes the Hamming weight of \( b \)). That is, \( f_{S,k}(x) = 1 \) if and only if \( \text{wt}(x) \in S \). Some well-studied examples of functions in this class include \( k\text{AND} = f_{\{k\},k} \), the threshold functions \( \text{Th}_k^i = f_{\{i,i+1,\ldots,k\},k} \), and “exact weight” functions \( \text{Ex}_k^i = f_{\{i\},k} \).\(^3\)

1.3.1 The sketching approximability of Max-\(k\text{AND}\)

Chou, Golovnev, and Velusamy [CGV20] showed that \( \alpha(2\text{AND}) = \frac{4}{9} \) and \((4/9 + \epsilon)\)-approximation can be ruled out even for \( o(\sqrt{n})\)-space streaming algorithms. For \( k \geq 3 \), while Chou, Golovnev, Velusamy, and Sudan [CGSV21a] give optimal sketching approximation algorithms for Max-\(k\text{AND}\), they do not explicitly analyze the approximation ratio \( \alpha(k\text{AND}) \), and show only that it lies between \( 2^{-k} \) and \( 2^{-(k-1)} \).

In this paper, we analyze the dichotomy theorem in [CGSV21a], and obtain a closed-form expression for the sketching approximability of Max-\(k\text{AND}\) for every \( k \). For odd \( k \geq 3 \), define the constant

\[
\alpha_k' \overset{\text{def}}{=} \left( \frac{(k-1)(k+1)}{4k^2} \right)^{(k-1)/2} = 2^{-(k-1)} \cdot \left( 1 - \frac{1}{k^2} \right)^{(k-1)/2}.
\]  

(1.1)

In Section 4, we prove the following:

---

\(^2\)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 \( b \in \{0,1\} \rightarrow \{-1\}^k \in \{-1,1\}^k \), for consistency with [CGSV21a].

\(^3\)By [CGSV21a, 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 Max-CSP\((f_{S',k})\) can be viewed as an instance of Max-CSP\((f_{S,k})\) with the same value, since for any constraint \( C = (b,1) \) and assignment \( \sigma \in \{-1,1\}^n \), we have \( f_{S,k}(b \oplus \sigma) = f_{S',k}(b \oplus (\neg \sigma)) \). Thus, we further narrow our focus to the case where every element of \( S \) is larger than \( \frac{k}{2} \).
Theorem 1.2. For odd $k \geq 3$, $\alpha(k \text{AND}) = \alpha_k'$, and for even $k \geq 2$, $\alpha(k \text{AND}) = 2\alpha_{k+1}'$.

Since $\rho(k \text{AND}) = 2^{-k}$, Theorem 1.2 also has the following important corollary:

Corollary 1.3. $\lim_{k \to \infty} \frac{\alpha(k \text{AND})}{2^k \rho(k \text{AND})} = 1$.

Recall that [CGSV21a] implies that $\alpha(f) \leq 2\rho(f)$ for all functions $f$. Indeed, Chou, Golovnev, Sudan, Velusamy, and Velingker [CGS+22] 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 $\text{Max}$-$k$-$\text{AND}$ achieving the optimal ratio from [CGSV21a]. 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 $\text{Th}_k^{k-1}$ for even $k$, we show that:

Theorem 1.4. For even $k \geq 2$, $\alpha(\text{Th}_k^{k-1}) = \frac{k}{2} \alpha_k$.

We prove Theorem 1.4 in Section 5.1 using techniques similar to our proof of Theorem 1.2. We also provide partial results for $\text{Ex}_k^{(k+1)/2}$, including closed forms for small $k$ and an asymptotic analysis of $\alpha(\text{Ex}_k^{(k+1)/2})$:

Theorem 1.5 (Informal version of Theorem 5.11). For odd $k \in \{3, \ldots, 51\}$, there is an explicit expression for $\alpha(\text{Ex}_k^{(k+1)/2})$ as a function of $k$.

Theorem 1.6. $\lim_{\text{odd } k \to \infty} \frac{\alpha(\text{Ex}_k^{(k+1)/2})}{\rho(\text{Ex}_k^{(k+1)/2})} = 1$.

We prove Theorems 1.5 and 1.6 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 1.2, 1.4 and 1.5.

1.3.3 Simple approximation algorithms for threshold functions

Chou, Golovnev, and Velusamy’s optimal $(\frac{4}{5} - \epsilon)$-approximation for $2 \text{AND}$ [CGV20], like Gurusswami, Velingker, and Velusamy’s earlier $(\frac{2}{5} - \epsilon)$-approximation [GVV17], is based on measuring a quantity called the bias of an instance $\Psi$, denoted $\text{bias}(\Psi)$, which is defined as follows: For each $i \in [n]$, $\text{diff}_i(\Psi)$ is the difference in total weight between constraints where $x_i$ occurs positively and negatively, and $\text{bias}(\Psi) \overset{\text{def}}{=} \frac{1}{kn} \sum_{i=1}^{n} |\text{diff}_i(\Psi)| \in [0,1]$. In the sketching setting, $\text{bias}(\Psi)$ can be estimated using standard $\ell_1$-norm sketching algorithms [Ind06; KNW10].

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

Theorem 1.7. Let $f_{S,k} = \text{Th}_k^S$ 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 $\text{Max}$-$\text{CSP}(f_{S,k})$: On input $\Psi$, compute an estimate $\hat{b}$ for $\text{bias}(\Psi)$ up to a multiplicative $(1 \pm \epsilon')$ error and output $\gamma(\hat{b})$.

\[\text{[GVV17; CGV20] did not normalize by } \frac{1}{kW}.\]
Our construction generalizes the algorithm in [CGV20] for 2AND to all threshold functions, and is also a simplification, since the [CGV20] algorithm computes a more complicated function of \( \hat{b} \).

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 1.8** (Informal version of Theorem 7.7). Let \( f_{S,k} \) be a function for which the max-min method applies, such as \( k\text{AND} \), or \( Th_{k-1}^k \) (for even \( k \)). Then there exists a constant \( p^* \in [0,1] \) such that following algorithm, on input \( \Psi \), outputs an assignment with expected value at least \( \alpha(f_{S,k})\text{val}_\Psi \):

Assign variable \( i \) to 1 if \( \text{diff}_i(\Psi) \geq 0 \) 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 [BR21] of the Max-2AND algorithm in [CGV20].

### 1.3.4 Sketching vs. streaming approximability

**Theorem 1.2** implies that \( \alpha(3\text{AND}) = \frac{2}{3} \). We prove that the padded one-wise pair criterion of Chou, Golovnev, Sudan, and Velusamy [CGSV21a] is not sufficient to completely resolve the streaming approximability of Max-3AND:

**Theorem 1.9** (Informal version of Theorem 2.12 + Observation 2.13). The padded one-wise pair criterion in [CGSV21a] does not rule out a \( o(\sqrt{n}) \)-space streaming \( (\frac{2}{3} + \epsilon) \)-approximation for 3AND for every \( \epsilon > 0 \); however, it does rule out such an algorithm for \( \epsilon \gtrsim 0.0141 \).

We state these results formally in Section 2.4 and prove them in Section 6. Separately, Theorem 1.4 implies that \( \alpha(Th_3^3) = \frac{4}{7} \), and the padded one-wise pair criterion can be used to show that \( (\frac{4}{7} + \epsilon) \)-approximating Max-CSP(Th_3^3) requires \( \Omega(\sqrt{n}) \) space in the streaming setting (see Observation 5.7 below).

### 1.4 Related work

The classical approximability of Max-\( k \)AND has been the subject of intense study, both in terms of algorithms [GW95; FG95; Zw98; Tr98a; TSSW00; Has04; Has05; CMM09] and hardness-of-approximation [Has01; Tr98b; ST98; ST00; EH08; ST09], given its intimate connections to \( k \)-bit PCPs. Charikar, Makarychev, and Makarychev [CMM09] constructed an \( \Omega(k2^{-k}) \)-approximation to Max-\( k \)AND, while Samorodnitsky and Trevisan [ST09] showed that \( k2^{-(k-1)} \)-approximations and \( (k+1)2^{-k} \)-approximations are \( \text{NP} \)- and \( \text{UG} \)-hard, respectively.

Interestingly, recalling that \( \alpha(k\text{AND}) \to 2\rho(k\text{AND}) = 2^{-(k-1)} \) as \( k \to \infty \), in the large-\( k \) limit our simple randomized algorithm (given in Theorem 1.8) matches the performance of Trevisan’s [Tr98a] parallelizable LP-based algorithm for \( k \)AND, which (to the best of our knowledge) was the first work on the general \( k \)AND problem! The subsequent works [Has04; Has05; CMM09] superseding [Tr98a] use more complex techniques involving semidefinite programming, but are structurally similar to our algorithm in Theorem 1.8: They all involve “guessing” an assignment \( x \in \mathbb{Z}_2^n \) 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 [CGSV21a], and explain the novel aspects of our analysis.

2.1 The Chou, Golovnev, Sudan, and Velusamy [CGSV21a] framework for symmetric functions

In this section, we describe the Chou, Golovnev, Sudan, and Velusamy [CGSV21a] framework for finding the optimal sketching approximation ratio of a symmetric Boolean function \( f \).

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

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

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

\[
\lambda_S(D, p) \overset{\text{def}}{=} \mathbb{E}_{a \sim D, b \sim \text{Bern}(p)^k} [f_{S,k}(a \circ b)] = \mathbb{E}_{b \sim \text{Bern}(p)^k} [\text{val}_{\Psi_D}(b)]
\]

(2.1)

denote the expected value of a “\( p \)-biased symmetric assignment” on \( D \)’s canonical instance. Also, for a symmetric distribution \( D \in \Delta_k \), we define its (scalar) marginal

\[
\mu(D) \overset{\text{def}}{=} \mathbb{E}_{b \sim D} [b_1] = \cdots = \mathbb{E}_{b \sim D} [b_k].
\]

(2.2)

In general, \( \lambda_S \) is linear in \( D \) and degree-\( k \) in \( p \), and \( \mu \) is linear in \( D \). For \( D \in \Delta_k \), we provide explicit formulas for \( \lambda_S \) and \( \mu \) in Section 3.

Roughly, [CGSV21a] states that Max-CSP(\( f_{S,k} \)) is hard to approximate in the sketching setting if there exist distributions \( D_N, D_Y \in \Delta_k \) such that (1) \( \mu(D_N) = \mu(D_Y) \) and (2) \( D_Y \)’s canonical instance is highly satisfied by the trivial (all-ones) assignment but (3) \( D_N \)’s canonical instance is not well-satisfied by any “biased symmetric assignment”. To be precise, for \( D \in \Delta(\{-1,1\}^k) \), let

\[
\beta_S(D) \overset{\text{def}}{=} \sup_{p \in [0,1]} \lambda_S(D, p) \quad \text{and} \quad \gamma_S(D) \overset{\text{def}}{=} \lambda_S(D, 1),
\]

(2.3)

and define

\[
\alpha(f_{S,k}) \overset{\text{def}}{=} \inf_{D_N, D_Y \in \Delta_k : \mu(D_N) = \mu(D_Y)} \left( \frac{\beta_S(D_N)}{\gamma_S(D_Y)} \right).
\]

(2.4)

For every symmetric function \( f_{S,k} \), [CGSV21a] proves that \( \alpha(f_{S,k}) \) is the optimal sketching approximation ratio for Max-CSP(\( f_{S,k} \)).
Theorem 2.5 (Combines [CGSV21a, 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 a linear sketching $(\alpha(f_{S,k}) - \epsilon)$-approximation to $\text{Max-CSP}(f_{S,k})$ in $O(\log n)$ space, but any sketching $(\alpha(f_{S,k}) + \epsilon)$-approximation to $\text{Max-CSP}(f_{S,k})$ 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 Eq. (2.4). Instead, [CGSV21a] 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 Eq. (2.2)). Instead, we must consider a vector marginal $\mu(\mathcal{D}) = (\mu_1, \ldots, \mu_k)$ with $i$-th component $\mu_i = \mathbb{E}_{b \sim \mathcal{D}}[b_i]$; correspondingly, $\mathcal{D}_N$ and $\mathcal{D}_Y$ are required to satisfy the constraint $\mu(\mathcal{D}_N) = \mu(\mathcal{D}_Y)$. These issues motivate our focus on symmetric functions in this paper. Since we need to consider only symmetric distributions in Eq. (2.4), $\mathcal{D}_Y$ 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 $\alpha(2\text{AND}) = \frac{4}{9}$, Chou, Golovnev, Sudan, and Velusamy [CGSV21a, Example 1] use the following reformulation of the optimization problem on the right hand side of Eq. (2.4). For a symmetric function $f_{S,k}$ and $\mu \in [-1,1]$, let

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

(2.6)

then

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

(2.7)

The optimization problem on the right-hand side of Eq. (2.7) appears simpler than that of Eq. (2.4) 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 $2\text{AND} = f_{\{2\},2}$, Chou, Golovnev, Sudan, and Velusamy [CGSV21a] show that $\gamma_{\{2\},2}(\mu)$ is an explicit linear function of $\mu$; 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 $\mu$ (see Lemma 3.3 below), we do not know how to find closed forms for $\beta_{S,k}(\mu)$ even for $3\text{AND}$. Thus, in this work we introduce a different formulation of the optimization problem:

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

(2.8)

This reformulation is valid because

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

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 Eq. (2.8) 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 [CGSV21a] even in the 2AND case. Let $\mathcal{D}_N^*$ minimize the right-hand side of Eq. (2.8), 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 Eq. (2.8), and applying the max-min inequality, we get

$$\alpha(f_{S,k}) = \inf_{\mathcal{D}_N \in \Delta_k} \sup_{p \in [0,1]} \left( \frac{\lambda_S(\mathcal{D}_N, p)}{\gamma_{k_S}(\mu(\mathcal{D}_N))} \right) \geq \sup_{p \in [0,1]} \inf_{\mathcal{D}_N \in \Delta_k} \left( \frac{\lambda_S(\mathcal{D}_N, p)}{\gamma_{k_S}(\mu(\mathcal{D}_N))} \right)$$

$$\geq \inf_{\mathcal{D}_N \in \Delta_k} \left( \frac{\lambda_S(\mathcal{D}_N, p^*)}{\gamma_{k_S}(\mu(\mathcal{D}_N))} \right). \quad (2.9)$$

Given $p^*$, the right-hand side of Eq. (2.9) 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 Eq. (2.9) equals $\alpha(f_{S,k})$; that is, $(\mathcal{D}_N^*, p^*)$ is a saddle point of $\frac{\lambda_S(\mathcal{D}_N, p)}{\gamma_{k_S}(\mu(\mathcal{D}_N))}$. \footnote{This term comes from the optimization literature; such points are also said to satisfy the “strong max-min property” (see, e.g., [BV04, 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.}

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 $\frac{\lambda_S(\mathcal{D}_N^*, p^*)}{\gamma_{k_S}(\mu(\mathcal{D}_N^*)))}$ 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 $\alpha(k\text{AND})$ for odd $k$ (i.e., the odd case of Theorem 1.2), we guess $\mathcal{D}_N^* \in \Delta_k$ and $p^* = k+1$ (by using Mathematica for small cases), and then check the saddle-point and maximization conditions in two separate lemmas (Lemmas 4.1 and 4.2, respectively). Then, we show that $\alpha(k\text{AND}) = \alpha_k^*$, by analyzing the right hand side of the appropriate instantiation of Eq. (2.9). We use similar techniques for $k\text{AND}$ for even $k$ (also Theorem 1.2) and for various other cases in Sections 5.1 to 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 $f_{\{3\},4}$.

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 Max-CSP($f_{S,k}$) where $f_{S,k} = \Theta_k^1$ is a threshold function (see Theorem 7.7).

2.4 Streaming lower bounds

Chou, Golovnev, Sudan, and Velusamy [CGSV21a] 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 $(\gamma, \beta)$-Max-CSP($f$) for streaming algorithms:

**Definition 2.10** (Padded one-wise pairs, [CGSV21a, §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
In this section, we give explicit formulas for the quantities $\mu$, $\lambda$, and $\gamma$. For any $k$, let $\epsilon \in [k]$, let $\epsilon_{i,k} \equiv -1 + \frac{2i}{k}$.

Lemma 3.1. For any $D \in \Delta_k$,

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

Proof of Lemma 3.1. By definition (Eq. (2.2)), $\mu(D) = \mathbb{E}_{b \sim \mathbb{D}}[b]$. We use linearity of expectation; the contribution of weight-$i$ vectors to $\mu(D)$ is $D(i) \cdot \frac{1}{k}(i \cdot 1 + (k - i) \cdot (-1)) = \epsilon_{i,k} D(i)$. □
Lemma 3.2. For any $D \in \Delta_k$ and $p \in [0, 1]$, we have

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

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

Proof. By linearity of expectation and symmetry, it suffices to fix $s$ and $i$ and calculate, given a fixed string $a = (a_1, \ldots, a_k)$ of Hamming weight $i$ and a random string $b = (b_1, \ldots, b_k) \sim \text{Bern}(p)^k$, the probability of the event $\text{wt}(a \odot b) = s$.

Let $A = \text{supp}(a) = \{ t \in [k] : a_t = 1 \}$ and similarly $B = \text{supp}(b)$. We have $|A| = i$ and $s = \text{wt}(a \odot b) = |A \cap B| + |([k] \setminus A) \cap ([k] \setminus B)]$.

Let $j = |A \cap B|$, and consider cases based on $j$.

Given fixed $j$, we must have $|A \cap B| = j$ and $|([k] \setminus A) \cap ([k] \setminus B)] = s - j$. Thus if $j$ satisfies $j \leq i, s - j \leq k - i, j \geq 0, j \leq s$, we have $\binom{i}{j}$ choices for $A \cap B$ and $\binom{k-i}{s-j}$ choices for $([k] \setminus A) \cap ([k] \setminus B)]$; together, these completely determine $B$. Moreover $\text{wt}(b) = |B| = |B \cap A| + |B \cap ([k] \setminus A)| = j + (k - i) - (s - j) = k - s - i + 2j$, yielding the desired formula.

Lemma 3.3. 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 $\min \left\{ \frac{1+\mu}{1+\epsilon_{s,k}}, 1, \frac{1-\mu}{1-\epsilon_{t,k}} \right\}$).

Proof. For $\mu \in [-1, 1]$, in (Eq. (2.6)) we defined

$$
\gamma_{S,k}(\mu) = \sup_{D_Y \in \Delta_k, \mu(D_Y) = \mu} \gamma_S(D_Y),
$$

where by Eqs. (2.1) and (2.3), $\gamma_S(D_Y) = \sum_{i \in S} D_Y(i)$. For $D_Y \in \Delta_k$, let $\text{supp}(D_Y) = \{ i \in [k] : D_Y(i) > 0 \}$. We handle cases based on $\mu$.

Case 1: $\mu \in [-1, \epsilon_{s,k}]$. Our strategy is to reduce to the case $\text{supp}(D_Y) \subseteq \{0, s\}$ while preserving the marginal $\mu$ and (non-strictly) increasing the value of $\gamma_S$.

Consider the following operation on a distribution $D_Y \in \Delta_k$: For $u < v < w \in [k]$, increase $D_Y(u)$ by $D_Y(v) \frac{w-u}{w-v}$, increase $D_Y(w)$ by $D_Y(v) \frac{w-u}{w-v}$, and set $D_Y(v)$ to zero. Note that this results in a new distribution with the same marginal, and

$$
D_Y(v) \frac{w-v}{w-u} \epsilon_u,k + D_Y(v) \frac{v-u}{w-u} \epsilon_{w,k} = D_Y(v) \epsilon_{v,k}.
$$

Given an initial distribution $D_Y$, we can apply this operation to zero out $D_Y(v)$ for $v \in \{1, \ldots, s-1\}$ by redistributing to $D_Y(0)$ and $D_Y(s)$, preserving the marginal and only increasing the value of $\gamma_S$ (since $v \notin S$ while $s \in S$). Similarly, we can redistribute $D_Y(v)$ to $D_Y(t)$ and $D_Y(k)$ when
v ∈ \{t + 1, \ldots, k - 1\}, and to \(D_Y(s)\) and \(D_Y(t)\) when \(v \in \{s + 1, \ldots, t - 1\}\). Thus, we need only consider the case supp(\(D\)) ⊆ \{0, s, t, k\}. We assume for simplicity that 0, s, t, k are distinct.

By definition of \(\epsilon_{i,k}\) we have

\[
\mu(D) = -D_Y(0) + D_Y(s) \left(-1 + \frac{2s}{k}\right) + D_Y(t) \left(-1 + \frac{2t}{k}\right) + D_Y(k) \leq -1 + \frac{2s}{k}
\]

(by assumption for this case). Substituting \(D_Y(s) = 1 - D_Y(0) - D_Y(t) - D_Y(k)\) and multiplying through by \(\frac{k}{2}\), we have

\[
kD_Y(k) - sD_Y(0) - sD_Y(t) - sD_Y(k) + tD_Y(t) \leq 0;
\]

defining \(\delta = D_Y(t)\left(\frac{k}{s} - 1\right) + D_Y(k)\left(\frac{k}{s} - 1\right)\), we can rearrange to get \(D_Y(0) \geq \delta\). Then given \(D_Y\), we can zero out \(D_Y(t)\) and \(D_Y(k)\), decrease \(D_Y(0)\) by \(\delta\), and correspondingly increase \(D_Y(s)\) by \(D_Y(t) + D_Y(k) + \delta\). This preserves the marginal since

\[
(\delta + D_Y(t) + D_Y(k)) \epsilon_{s,k} = -\delta + D_Y(t) \epsilon_{t,k} + D_Y(k)
\]

and can only increase \(\gamma_S\).

Thus, it suffices to only consider the case supp(\(D_Y\)) ⊆ \{0, s\}. This uniquely determines \(D_Y\) (because \(\mu\) is fixed); we have \(D_Y(0) = \frac{\epsilon_{s,k} - \mu}{\epsilon_{s,k} + 1}\) and \(D_Y(s) = \frac{\mu - \epsilon_{s,k}}{\epsilon_{s,k} - \epsilon_{t,k}}\), yielding the desired value of \(\gamma_S\).

**Case 2:** \(\mu \in [\epsilon_{s,k}, \epsilon_{t,k}]\). We simply construct \(D_Y\) with \(D_Y(s) = \frac{\epsilon_{s,k} - \mu}{\epsilon_{s,k} - \epsilon_{t,k}}\) and \(D_Y(t) = \frac{\mu - \epsilon_{s,k}}{\epsilon_{s,k} - \epsilon_{t,k}}\); we have \(\mu(D_Y) = \mu\) and \(\gamma_S(D_Y) = 1\).

**Case 3:** \(\mu \in [\epsilon_{t,k}, 1]\). Following the symmetric logic to Case 1, we consider \(D_Y\) supported on \{t, k\} and set \(D_Y(t) = \frac{1 - \mu}{1 - \epsilon_{t,k}}\) and \(D_Y(k) = \frac{\mu - \epsilon_{t,k}}{1 - \epsilon_{t,k}}\), yielding \(\mu(D_Y) = \mu\) and \(\gamma_S(D_Y) = D_Y(t)\).

### 4 Analysis of \(\alpha(k\text{AND})\)

In this section, we prove Theorem 1.2 (on the sketching approximability of Max-kAND). Recall that in Eq. (1.1), we defined

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

Theorem 1.2 follows immediately from the following two lemmas:

**Lemma 4.1.** For all odd \(k \geq 3\), \(\alpha(k\text{AND}) \leq \alpha'_k\). For all even \(k \geq 2\), \(\alpha(k\text{AND}) \leq 2\alpha'_{k+1}\).

**Lemma 4.2.** For all odd \(k \geq 3\), \(\alpha(k\text{AND}) \geq \alpha'_k\). For all even \(k \geq 2\), \(\alpha(k\text{AND}) \geq 2\alpha'_{k+1}\).

To begin, we give explicit formulas for \(\gamma_{\{k\},k}(\mu(D))\) and \(\lambda_{\{k\}}(D,p)\). Note that the smallest element of \(\{k\}\) is \(k\), and \(\epsilon_{k,k} = 1\). Thus, for \(D \in \Delta_k\), we have by Lemmas 3.1 and 3.3 that

\[
\gamma_{\{k\},k}(\mu(D)) = \frac{1 + \sum_{i=0}^{k} (-1 + \frac{2i}{k}) D(i)}{2} = \sum_{i=0}^{k} \frac{i}{k} D(i). \quad (4.3)
\]

Similarly, we can apply Lemma 3.2 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 \(\binom{k-1}{s-i} = 1\). Thus, for \(q = 1 - p\), we have

\[
\lambda_{\{k\}}(D,p) = \sum_{i=0}^{k} q^{k-i} p^{i} D(i) \quad (4.4)
\]

Now, we prove Lemma 4.1 directly:
Proof of Lemma 4.1. Consider the case where \( k \) is odd. Define \( D^*_N \) by \( D^*_N(\ell+1) = 1 \) and let \( p^* = \frac{1}{2} + \frac{1}{2k} \). Since

\[
\alpha(k \text{AND}) \leq \frac{\beta(k)(D^*_N)}{\gamma(k,k(\mu(D^*_N)))} \quad \text{and} \quad \beta(k)(D^*_N) = \sup_{p \in [0,1]} \lambda_k(D^*_N, p),
\]

by Eqs. (2.3) and (2.8), respectively, it suffices to check that \( p^* \) maximizes \( \lambda_k(D^*_N, \cdot) \) and

\[
\frac{\lambda_k(D^*_N, p^*)}{\gamma(k,k(\mu(D^*_N)))} = \alpha'_k.
\]

Indeed, by Eq. (4.4),

\[
\lambda_k(D^*_N, p) = (1 - p)^{(k-1)/2}p^{(k+1)/2}.
\]

To show \( p^* \) maximizes \( \lambda_k(D^*_N, \cdot) \), 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, \( \lambda_k(D^*_N, \cdot) \) has critical points only at 0, 1, and \( p^* \), and it is maximized at \( p^* \) since it vanishes at 0 and 1. Finally, by Eqs. (4.3) and (4.4) and the definition of \( \alpha'_k \),

\[
\frac{\lambda_k(D^*_N, p^*)}{\gamma(k,k(\mu(D^*_N)))} = \frac{\left( \frac{1}{2} - \frac{1}{2k} \right)^{(k-1)/2} \left( \frac{1}{2} + \frac{1}{2k} \right)^{(k+1)/2}}{\frac{1}{2} \left( 1 + \frac{1}{k} \right)} = \alpha'_k,
\]

as desired.

Similarly, consider the case where \( k \) is even; here, we define \( D^*_N \) by \( D^*_N(\ell) = \frac{(\ell+1)^2}{(\ell)^2+(\ell+1)^2} \) and \( D^*_N(\ell+1) = \frac{(\ell)^2}{(\ell)^2+(\ell+1)^2} \), and set \( p^* = \frac{1}{2} + \frac{1}{2(k+1)} \). Using Eq. (4.4) to calculate the derivative of \( \lambda_k(D^*_N, \cdot) \) yields

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

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

so \( \lambda_k(D^*_N, \cdot) \) has critical points at 0, 1, \( \frac{1}{2} + \frac{k}{2} \), and \( p^* \); \( p^* \) is the only critical point in the interval [0, 1] for which \( \lambda_k(D^*_N, \cdot) \) is positive, and hence its maximum. Finally, it can be verified algebraically using Eqs. (4.3) and (4.4) that \( \frac{\lambda_k(D^*_N, p^*)}{\gamma(k,k(\mu(D^*_N)))} = 2\alpha'_{k+1} \), as desired.

We prove Lemma 4.2 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 Appendix A:

**Proposition 4.5.** Let \( f : \mathbb{R}^n \to \mathbb{R} \) be defined by the equation \( f(x) = \frac{a^T x}{b^T x} \) for some \( a, b \in \mathbb{R}^n_{\geq 0} \). For every \( y(1), \ldots, y(r) \in \mathbb{R}^n_{\geq 0} \), and every \( x = \sum_{i=1}^r \alpha_i y(i) \) with each \( x_i \geq 0 \), we have \( f(x) \geq \min_y f(y(i)) \). In particular, taking \( r = n \) and \( y(1), \ldots, y(n) \) as the standard basis for \( \mathbb{R}^n \), for every \( x \in \mathbb{R}^n_{\geq 0} \), we have \( f(x) \geq \min_i \frac{a_i}{b_i} \).
Proof of Lemma 4.2. First, suppose $k \geq 3$ is odd. Set $p^* = \frac{1}{2} + \frac{1}{2k} = \frac{k+1}{2k}$. We want to show that

$$\alpha'_k \leq \inf_{\mathcal{D} \in \Delta_k} \frac{\lambda_{\{k\}}(\mathcal{D}, p^*)}{\gamma_{\{k\},k}(\mu(\mathcal{D}))}$$

(max-min inequality, i.e., Eq. (2.9))

$$= \inf_{\mathcal{D} \in \Delta_k} \frac{\sum_{i=0}^{k} (1 - p^*)^{k-i}(p^*)^i \mathcal{D}(i)}{\sum_{i=0}^{k} \frac{i}{k} \mathcal{D}(i)}.$$  

(Eqs. (4.3) and (4.4))

By Proposition 4.5, 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 $\alpha'_k$, 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

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

(4.6)

When $i = \frac{k+1}{2}$ or $\frac{k-1}{2}$, we have equality in Eq. (4.6). We extend to the other values of $i$ by induction. Indeed, when $i \geq \frac{k+1}{2}$, then “i satisfies Eq. (4.6)” implies “$i + 1$ satisfies Eq. (4.6)” because $ri \geq i + 1$, and when $i \leq \frac{k-1}{2}$, then “i satisfies Eq. (4.6)” implies “$i - 1$ satisfies Eq. (4.6)” because $\frac{1}{i} \geq i - 1$.

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 Eq. (4.6) can be derived:

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

and these inequalities follow from the same inductive argument. \qed

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

5.1 $\text{Th}_k^{k-1}$ for even $k$

In this subsection, we prove Theorem 1.4 (on the sketching approximability of $\text{Th}_k^{k-1}$ for even $k \geq 2$). It is necessary and sufficient to prove the following two lemmas:

Lemma 5.1. For all even $k \geq 2$, $\alpha(\text{Th}_k^{k-1}) \leq \frac{k}{2} \alpha'_{k-1}$.

Lemma 5.2. For all even $k \geq 2$, $\alpha(\text{Th}_k^{k-1}) \geq \frac{k}{2} \alpha'_{k-1}$.

Firstly, we give explicit formulas for $\gamma_{\{k-1\},k}$ and $\lambda_{\{k-1\},k}$. We have $\text{Th}_k^{k-1} = f_{\{k-1\},k}$, and $\epsilon_{k-1,k} = -1 + \frac{2(k-1)}{k} = 1 - \frac{2}{k}$. Thus, Lemmas 3.1 and 3.3 give

$$\gamma_{\{k-1\},k}(\mu(\mathcal{D})) = \min \left\{ \frac{1 + \sum_{i=0}^{k} (-1 + \frac{2i}{k}) \mathcal{D}(i)}{2 - \frac{2}{k}}, 1 \right\} = \min \left\{ \sum_{i=0}^{k} \frac{i}{k-1} \mathcal{D}(i), 1 \right\}. \quad (5.3)$$

Next, we calculate $\lambda_{\{k-1\},k}(\mathcal{D}, p)$ with Lemma 3.2. Let $q = 1-p$, and let us examine the coefficient on $\mathcal{D}(i)$. $s = k$ contributes $q^{k-1}p^k$. 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 \( i q^{k-i+1} p^{i-1} \) for \( j = i - 1 \). Thus, altogether we can write

\[
\lambda_{(k-1,k)}(D, p) = \sum_{i=0}^{k} q^{k-i-1} p^{i-1} ((k-i)p^2 + pq + iq^2) D(i).
\]  

(5.4)

Now, we prove Lemmas 5.1 and 5.2.

**Proof of Lemma 5.1.** As in the proof of Lemma 4.1, it suffices to construct \( D_N^* \) and \( p^* \) such that \( p^* \) maximizes \( \lambda_{(k-1,k)}(D_N^*, \cdot) \) and \( \frac{\lambda_{(k-1,k)}(D_N^*, p^*)}{\gamma_{(k-1,k),k}(\mu(D_N))} = \frac{1}{2} \alpha'_{k-1} \).

We again let \( p^* = \frac{1}{2} + \frac{1}{2(k-1)} \), but define \( D_N^* \) by \( D_N^*(\frac{1}{2}) = \frac{(\frac{1}{2})^2}{(\frac{1}{2})^2 + (\frac{k}{2} - 1)^2} \) and \( D_N^*(\frac{k}{2} + 1) = \frac{(\frac{k}{2} - 1)^2}{(\frac{k}{2})^2 + (\frac{k}{2} - 1)^2} \). By Eq. (5.4), the derivative of \( \lambda_{(k-1,k)}(D_N^*, \cdot) \) is now

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

where \( \xi(p) \) is the cubic

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

Thus, \( \lambda_{(k-1,k)} \)'s critical points on the interval \([0, 1]\) are 0, 1, \( p^* \) and any roots of \( \xi \) in this interval. We claim that \( \xi \) 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

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

(5.5)

whose coefficients are cubic in \( k \). It can be verified by calculating the roots of each coefficient of \( x \) in Eq. (5.5) that all coefficients are positive for \( k \geq 6 \). Thus, Eq. (5.5) cannot have roots for positive \( x \), a contradiction. Hence \( \lambda_{(k-1,k)}(D_N^*, \cdot) \) is maximized at \( p^* \). Finally, it can be verified that \( \frac{\lambda_{(k-1,k)}(D_N^*, p^*)}{\gamma_{(k-1,k),k}(\mu(D_N))} = \frac{1}{2} \alpha'_{k-1} \), as desired.

**Proof of Lemma 5.2.** Define \( p^* = \frac{1}{2} + \frac{1}{2(k-1)} \). Following the proof of Lemma 4.2 and using the lower bound \( \gamma_{(k-1,k),k}(\mu(D_N)) \leq \sum_{i=0}^{k} \frac{1}{k-1} D_N(i) \), it suffices to prove that

\[
\frac{k}{2} \alpha'_{k-1} \leq \inf_{D_N \in \Delta_k} \frac{\sum_{i=0}^{k} (1 - p^*)^{k-i-1}(p^*)^i - (k-i)(p^*)^i + (i+1-p^*)^2) D_N(i)}{\sum_{i=0}^{k} \frac{i}{k-1} D_N(i)}
\]

for which by Proposition 4.5, it in turn suffices to prove that for each \( i \in \{0\} \cup [k] \),

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

(5.6)
We again observe that \( \alpha_{k-1} = (1 - p^r)^{k/2-1}(p^r)^{k/2-1} \), define \( r = \frac{p^r}{1-p^r} = \frac{k}{k-2} \), and factor out \((1 - p^r)^{k-1}\), which simplifies our desired inequality to

\[
\frac{1}{2} \cdot \frac{1}{r} \cdot \frac{k-2}{k-1} (i + r + (k - i)r^2) \geq i. \tag{5.6}
\]

for each \( i \in \{0\} \cup \{k\} \). Again, we assume \( k \geq 6 \) WLOG; the bases cases \( i = k/2 - 1, k/2 \) can be verified directly, and we proceed by induction. If Eq. (5.6) holds for \( i \), and we seek to prove it for \( i + 1 \), it suffices to cross-multiply and instead prove the inequality

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

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. \( \square \)

**Observation 5.7.** For \( \text{Th}^3_k \) the optimal \( D^*_N \) = \((0,0,\frac{4}{15},\frac{1}{5},0) \) does participate in a padded one-wise pair with \( D^*_Y = (\frac{4}{15},0,0,\frac{11}{15},0) \) (given by \( D_0 = (0,0,0,1,0) \), \( \tau = \frac{1}{5} \), \( D'_N = (0,0,1,0,0) \), and \( D'_Y = (\frac{4}{15},0,0,\frac{8}{15},0) \)) so we can rule out streaming \((\frac{4}{9} + \epsilon)\)-approximations to Max-CSP\( \left( \text{Th}^3_k \right) \) in \( o(\sqrt{n}) \) space.

### 5.2 \( E_n^{(k+1)/2} \) for (small) odd \( k \)

In this section, we prove bounds on the sketching approximability of \( E_n^{(k+1)/2} \) for odd \( k \in \{3, \ldots, 51\} \). Define \( D_{0,k} \in \Delta_k \) by \( D_{0,k}(0) = \frac{k-1}{2k} \) and \( D_{0,k}(k) = \frac{k+1}{2k} \). We prove the following two lemmas:

**Lemma 5.8.** For all odd \( k \geq 3 \), \( \alpha(E_n^{(k+1)/2}) \leq \lambda_{\left(\frac{k+1}{2}\right)}(D_{0,k}, p'_k) \), where \( p'_k \) \( \overset{\text{def}}{=} \frac{3k-k^2+\sqrt{4k+k^2-2k^4+k^4}}{4k} \).

**Lemma 5.9.** The following holds for all odd \( k \in \{3, \ldots, 51\} \). For all \( p \in [0,1] \), the expression \( \frac{\lambda_{\left(\frac{k+1}{2}\right)}(1-p)}{\gamma_{\left(\frac{k+1}{2}\right),k}(\mu(\cdot))} \) is minimized at \( D_{0,k} \).

We begin by writing an explicit formula for \( \lambda_{\left(\frac{k+1}{2}\right)} \). **Lemma 3.2** gives

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

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 \( D_{0,k} \), we get:

\[
\lambda_{\left(\frac{k+1}{2}\right)}(D_{0,k}, p) = \left( \frac{k}{k+1/2} \right) \left( \frac{k-1}{2k} (1-p)^{(k+1)/2} p^{(k-1)/2} + \frac{k+1}{2k} (1-p)^{(k-1)/2} p^{(k+1)/2} \right). \tag{5.10}
\]

By **Lemmas 3.1 and 3.3**, \( \gamma_{\left(\frac{k+1}{2}\right),k}(\mu(D_{0,k})) = \gamma_{\left(\frac{k+1}{2}\right),k}(\frac{1}{k}) = 1 \). Thus, **Lemmas 5.8 and 5.9** together imply the following theorem:
Theorem 5.11. For odd \( k \in \{3, \ldots, 51\} \),
\[
\alpha(\Ex_k^{(k+1)/2}) = \left( \frac{k}{k+1} \right) \frac{k-1}{2k} (1-p'k^{(k+1)/2})(p_k^{(k-1)/2} + \frac{k+1}{2k} (1-p'k^{(k-1)/2})(p_k^{(k+1)/2})
\]
where \( p_k' = \frac{2k-k^2+\sqrt{4k+k^2-2k^3+k^4}}{4k} \) as in Lemma 5.8.

Recall that \( \rho(f_{(k+1)/2,k}) = \left( \frac{k}{k+1} \right) 2^{-k} \). Although we currently lack a lower bound on \( \alpha(\Ex_k^{(k+1)/2}) \) for large odd \( k \), the upper bound from Lemma 5.8 suffices to prove Theorem 1.6, i.e., it can be verified that
\[
\lim_{k \text{ odd} \to \infty} \frac{(\frac{k}{k+1}) (\frac{k-1}{2k} (1-p'k^{(k+1)/2})(p_k^{(k-1)/2} + \frac{k+1}{2k} (1-p'k^{(k-1)/2})(p_k^{(k+1)/2})}{\rho(\Ex_k^{(k+1)/2})} = 1.
\]

We remark that for \( \Ex_k^{(k+1)/2} \), our lower bound (Lemma 5.9) is stronger than what we were able to prove for \( k \\text{AND} \) (Lemma 4.2) and \( \Th_k^{k-1} \) (Lemma 5.2) because the inequality holds regardless of \( p \). This is fortunate for us, as the optimal \( p^* \) from Lemma 5.8 is rather messy.\(^6\) It remains to prove Lemmas 5.8 and 5.9.

Proof of Lemma 5.8. Taking the derivative with respect to \( p \) of Eq. (5.10) yields
\[
\frac{d}{dp} \left[ \lambda_{(\frac{k}{k+1})}(\D_{0,k},p) \right] = -\frac{1}{4k} \left( \frac{k}{k+1} \right)^{2} (kp^2 + (2k^2 - 6k)p + (-k^2 + 2k - 1)),
\]
where \( q = 1-p \). Thus, \( \lambda_{(\frac{k}{k+1})}(\D_{0,k}, \cdot) \) has critical points at \( p = 0, 1, p_k' \), and \( 3k-k^2-\sqrt{4k+k^2-2k^3+k^4} \).

This last value is nonpositive for all \( k \geq 0 \) (since \( 3k-k^2-\sqrt{4k+k^2-2k^3+k^4} = -4k(k-1)^2 \)). \( \square \)

The proof of our lower bound (Lemma 5.9) is slightly different than those of our earlier lower bounds (i.e., Lemmas 4.2 and 5.2) in the following sense. For \( i \in \{0 \cup [k]\} \), let \( D_i \in \Delta_k \) be defined by \( D_i(i) = 1 \). For \( k \\text{AND} \) (Lemma 4.2), we used the fact that \( \frac{\lambda_{(k)}(\cdot,p^*)}{\gamma_{(k)}(\cdot,k)} \) is a ratio of linear functions, and thus using Proposition 4.5, it is sufficient to verify the lower bound at \( D_0, \ldots, D_k \).

For \( \Th_k^{k-1} \) (Lemma 5.2), \( \frac{\lambda_{(k-1,k)}(\cdot,p^*)}{\gamma_{(k-1,k)}(\cdot,k)} \) is not a ratio of linear functions, because the denominator \( \gamma_{(k-1,k)}(\mu(D)) = \min\{\sum_{i=0}^{k} i^{-1} \D(i), 1\} \) is not linear over \( \Delta_k \). However, we managed to carry out the proof by upper-bounding the denominator with the linear function \( \gamma'(D) = \sum_{i=0}^{k} i^{-1} \D(i) \), and then invoking Proposition 4.5 (again, to show that it suffices to verify the lower bound at \( D_0, \ldots, D_k \)).

For \( \Ex_k^{(k+1)/2} \), we show that it suffices to verify the lower bound on a larger (but still finite) set of distributions.

Proof of Lemma 5.9. Recalling that \( \epsilon_{(k+1)/2,k} = \frac{1}{k} \), let \( \Delta_k^+ = \{ D \in \Delta_k : \mu(D) \leq \frac{1}{k}\} \) and \( \Delta_k^- = \{ D \in \Delta_k : \mu(D) \geq \frac{1}{k}\} \). Note that \( \Delta_k^+ \cup \Delta_k^- = \Delta_k \), and restricted to either \( \Delta_k^+ \) or \( \Delta_k^- \), \( \frac{\lambda_{(k-1,k)}(\mu(\cdot))}{\gamma_{(k-1,k)}(\mu(\cdot))} \) is linear and thus we can apply Proposition 4.5 to \( \frac{\lambda_{(k-1,k)}(\cdot,p^*)}{\gamma_{(k-1,k)}(\cdot,k)} \).

\(^6\)The analogous statement is false for e.g. \( 3\AND \), where we had \( D_3 = (0,0,1,0) \), but at \( p = \frac{3}{4} \),
\[
\frac{\lambda_{(3)}((0,\frac{3}{4},0,\frac{3}{4}),(0,\frac{3}{2},0),0))}{\gamma_{(3)}(\mu(0,\frac{3}{4},0,\frac{3}{4}),(0,\frac{3}{2},0),0))} = \frac{3}{16} \leq \frac{27}{128} = \frac{\lambda_{(3)}((0,0,1,0),(0,\frac{3}{2},0),0))}{\gamma_{(3)}(\mu(0,0,1,0),(0,\frac{3}{2},0),0))}.
\]
Let $D_{i,j} \in \Delta_k$, for $i < \frac{k+1}{2}$, $j > \frac{k+1}{2}$, be defined by $D_{i,j}(i) = \frac{2i-(k+1)}{2j-n}$ and $D_{i,j}(j) = \frac{(k+1)-2i}{2j-n}$. Note that $\mu(D_{i,j}) = \frac{1}{k}$ for each $i,j$. We claim that $\{D_i\}_{i \leq \frac{k+1}{2}} \cup \{D_{i,j}\}$ are the extreme points of $\Delta^*_k$, or more precisely, that every distribution $D \in \Delta^*_k$ can be represented as a convex combination of these distributions. Indeed, this follows constructively from the procedure which, given a distribution $D$, subtracts from each $D(i)$ for $i < \frac{k+1}{2}$ (adding to the coefficient of the corresponding $D_i$) until the marginal of the (renormalized) distribution is $\frac{1}{k}$, and then subtracts from pairs $D(i), D(j)$ with $i < \frac{k+1}{2}$ and $j > \frac{k+1}{2}$, adding it to the coefficient of the appropriate $D_{i,j}$) until $D$ vanishes (i.e., $D(i)$ is zero for all $i \in \{0\} \cup \{k\}$). Similarly, every distribution $D \in \Delta^*_k$ can be represented as a convex combination of the distributions $\{D_{i,j}\}_{i \geq \frac{k+1}{2}} \cup \{D_{i,j}\}$. Thus, by Proposition 4.5, it is sufficient to verify that

$$\frac{\lambda_{\frac{k+1}{2}}(D,p)}{\gamma_{\frac{k+1}{2}}(\mu(D))} \geq \frac{\lambda_{\frac{k+1}{2}}(D^*_N,p)}{\gamma_{\frac{k+1}{2}}(\mu(D^*_N))}$$

for each $D \in \{D_i\} \cup \{D_{i,j}\}$. Treating $p$ as a variable, for each odd $k \in \{3, \ldots, 51\}$ we produce a list of $O(k^2)$ degree-$k$ polynomial inequalities in $p$ which we verify using Mathematica.

5.3 More symmetric functions

In Table 1 below, we list four more symmetric Boolean functions (beyond $k$AND, $Th_k^{k-1}$, and $E_k^{(k+1)/2}$) 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 $D_N$, with a given support, which minimizes $\alpha$ — and testMinMax — which, given a particular $D_N$, calculates $p^*$ for that $D_N$ and checks analytically whether lower-bounding by evaluating $\lambda_S$ at $p^*$ proves that $D_N$ is minimal.

<table>
<thead>
<tr>
<th>$S$</th>
<th>$k$</th>
<th>$\alpha$</th>
<th>$D^*_N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>{2, 3}</td>
<td>3</td>
<td>$\frac{1}{2} + \frac{\sqrt{3}}{18} \approx 0.5962$</td>
<td>$(0, \frac{1}{2}, 0, \frac{1}{2})$</td>
</tr>
<tr>
<td>{4, 5}</td>
<td>5</td>
<td>8 root$_R$($P_1$) $\approx 0.2831$</td>
<td>$(0, 0, 1 - \text{root}_R(P_2), \text{root}_R(P_2), 0, 0)$</td>
</tr>
<tr>
<td>{4}</td>
<td>5</td>
<td>8 root$_R$($P_3$) $\approx 0.2394$</td>
<td>$(0, 0, 1 - \text{root}_R(P_4), \text{root}_R(P_4), 0, 0)$</td>
</tr>
<tr>
<td>{3, 4, 5}</td>
<td>5</td>
<td>$\frac{1}{2} + \frac{3\sqrt{3}}{125} \approx 0.5537$</td>
<td>$(0, \frac{1}{2}, 0, 0, 0, \frac{1}{2})$</td>
</tr>
</tbody>
</table>

Table 1: Symmetric functions for which we have analytically calculated exact $\alpha$ values using the “max-min method”. For a polynomial $P : \mathbb{R} \to \mathbb{R}$ with a unique positive real root, let root$_R(p)$ 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).

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

In Section 5.2, we showed that $D^*_N$ defined by $D^*_N(0) = \frac{k+1}{2k}$ and $D^*_N(k) = \frac{k+1}{2k}$ is optimal for $E_k^{(k+1)/2}$ for odd $k \in \{3, \ldots, 51\}$. Using the same $D^*_N$, 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 $\alpha$ and $D_N$ because they are defined using the roots of polynomials of degree up to 8.)

In all previously-mentioned cases, the condition “$D^*_N$ has support size 2” was helpful, as it makes the optimization problem over $D^*_N$ 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 $\alpha$ (see testDistsWithSupportSize2 in the Mathematica code). However, using the max-min method with $D_N$ supported on two levels still achieves decent (but not tight) bounds on $\alpha$. For $S = \{3\}, k = 4$, using $D_N = \{\frac{1}{2}, 0, 0, 0, \frac{2}{3}\}$, 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 $D_N = \{\frac{1}{5}, 0, 0, 0, \frac{3}{7}, 0\}$, 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 $D_N = (0, 0, 0, 0.930013, 0, 0, 0.069987)$ gives $\alpha(f_{\{4, 5, 6\}, 6}) \in [0.44409972, 0.44409973]$, differing only by 0.000003%. (We conjecture here that $\alpha = \frac{4}{5}$.) For $S = \{6, 7, 8\}, k = 8$, using $D_N = (0, 0, 0, 0, 0.699501, 0.300499)$, we get the bounds $\alpha(f_{\{6, 7, 8\}, 8}) \in [0.20848, 0.20854]$ (the difference being 0.02%).

### 6 Incompleteness of streaming lower bounds: Proving Theorem 2.12

In this section, we prove Theorem 2.12, showing that the streaming lower bounds from [CGSV21a] (Theorem 2.11) cannot characterize the streaming approximability of 3AND.

**Lemma 6.1.** For $D \in \Delta_3$, the expression

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

is minimized uniquely at $D = (0, 0, 1, 0)$, with value $\frac{2}{3}$.

**Proof.** Letting $p = \frac{1}{3}D(1) + \frac{2}{3}D(2) + D(3)$ and $q = 1 - p$, by Lemmas 3.1 to 3.3 the expression expands to

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

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

**Lemma 6.2.** 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 $x^* \in Y$.

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

Finally, we have:

---

7Interestingly, in this latter case, we get bounds differing by 2.12% using $D_N = (0, 0, 0, 0, 0.930013, 0, 0, 0.069987)$ in an attempt to continue the pattern from $f_{\{7,8\}, 8}$ and $f_{\{8\}, 8}$ (where we set $D_N = (0, 0, 0, 0.44409972, 0, 0, 0)$ and $(0, 0, 0, 0, 0.699501, 0.300499)$ in Section 5.1 and Section 4, respectively).
Proof of Theorem 2.12. By Lemma 6.1, \( \frac{\beta_{\{3\}}(\mathcal{D}_N)}{\gamma_{\{3\}}(\mathcal{D}_Y)} \) is minimized uniquely at \( \mathcal{D}^*_N = (0, 0, 1, 0) \). By Lemma 3.1 we have \( \mu(\mathcal{D}^*_Y) = \frac{1}{3} \), and by inspection from the proof of Lemma 3.3 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 \( \frac{2}{3} \) using topological properties. View a distribution \( \mathcal{D} \in \Delta_3 \) as the vector \( (\mathcal{D}(0), \mathcal{D}(1), \mathcal{D}(2), \mathcal{D}(3)) \in \mathbb{R}^4 \). 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 \( P \subset M \) 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., \( \lambda_3 \)) 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 \( \frac{2}{3} \) as \( i \to \infty \), since \( M \) is compact and \( P \) is closed, Lemmas 6.1 and 6.2 imply that \( (\mathcal{D}_N^{(i)}, \mathcal{D}_Y^{(i)}) \in P, \) a contradiction.

\[
\square
\]

7 Simple sketching algorithms for threshold functions

The main goal of this section is to prove Theorem 1.7, giving a simple “bias-based” sketching algorithm for threshold functions \( \text{Th}_k \). Given an instance \( \Psi \) of Max-CSP(\( \text{Th}_k \)), for \( i \in [n] \), let \( \text{diff}_i(\Psi) \) 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 \( \Psi \) consists of clauses \( (b(1), j(1)), \ldots, (b(m), j(m)) \) with weights \( w_1, \ldots, w_m \), then
\[
\text{diff}_i(\Psi) \overset{\text{def}}{=} \sum_{\ell \in [m]} b(\ell)_i w_\ell.
\]

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

Let \( S = \{i, \ldots, k\} \) so that \( \text{Th}_k^i = f_{S,k} \). Recall the definitions of \( \beta_{S,k}(\mu) \) and \( \gamma_{S,k}(\mu) \) from Eq. (2.7). Our simple algorithm for Max-CSP(\( \text{Th}_k^i \)) relies on the following two lemmas, which we prove below:

Lemma 7.1. \( \text{val}_\Psi \leq \gamma_{S,k}(\text{bias}(\Psi)) \).

Lemma 7.2. \( \text{val}_\Psi \geq \beta_{S,k}(\text{bias}(\Psi)) \).

Together, these two lemmas imply that outputting \( \alpha(\text{Th}_k^i) \cdot \gamma_{S,k}(\text{bias}(\Psi)) \) gives an \( \alpha(\text{Th}_k^i) \)-approximation to Max-CSP(\( \text{Th}_k^i \)), since \( \alpha(\text{Th}_k^i) = \inf_{\mu \in [-1,1]} \frac{\beta_{S,k}(\mu)}{\gamma_{S,k}(\mu)} \) (Eq. (2.7)). We can implement this as a small-space sketching algorithm (up to an arbitrarily small constant \( \epsilon > 0 \) in the approximation ratio) because \( \text{bias}(\Psi) \) is measurable using \( \ell_1 \)-sketching algorithms (as used also in [GVV17; CGV20; CGSV21a]) and \( \gamma_{S,k}(\cdot) \) is piecewise linear.
Theorem 7.3 ([Ind06; KNW10]). 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 $(i, v) \in [n] \times \{-\poly(n), \ldots, \poly(n)\}$, and the goal is to estimate the $\ell_1$-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 7.4. 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 $\Psi$ of Max-CSP($\text{Th}_k^i$) (given as a stream of constraints), and the goal is to estimate $\text{bias}(\Psi)$ up to a multiplicative factor of $1 \pm \epsilon$.

Proof. Invoke the $\ell_1$-norm sketching algorithm from Theorem 7.3 as follows: On each input constraint $(b = (b_1, \ldots, b_k), j = (j_1, \ldots, j_k))$ with weight $w$, insert the updates $(j_1, wb_1), \ldots, (j_k, wb_k)$ into the stream (and normalize appropriately).

Theorem 1.7 then follows from Lemmas 7.1 and 7.2 and Corollary 7.4; we include a formal proof in Appendix A for completeness.

To prove Lemmas 7.1 and 7.2, we require a bit more setup. Adapting notation from [CGSV21a, §4.2], given an instance $\Psi$ of Max-CSP($\text{Th}_k^i$) and a “negation pattern” $a = (a_1, \ldots, a_n) \in \{-1, 1\}^n$ for the variables, let $\Psi^a$ be the instance which results from $\Psi$ by “flipping” the variables according to $a$ (formally, each constraint $(b, j)$ is replaced with $(b \circ a_j, j)$). We summarize the useful properties of this operation in the following claim:

Proposition 7.5. Let $\Psi$ be an instance of Max-CSP($\text{Th}_k^i$) and $a = (a_1, \ldots, a_n) \in \{-1, 1\}^n$. Then:

1. For each $i \in [n]$, $\text{diff}_i(\Psi^a) = a_i \text{diff}_i(\Psi)$.
2. $\text{bias}(\Psi) = \text{bias}(\Psi^a)$.
3. For any $\sigma \in \{-1, 1\}^n$, $\text{val}_{\Psi^a}(\sigma) = \text{val}_{\Psi}(a \odot \sigma)$.
4. $\text{val}_{\Psi^a} = \text{val}_{\Psi}$.

Proof. For Item i, we have

$$\text{diff}_i(\Psi^a) = \sum_{\ell \in [m] \text{ s.t. } j(\ell) = i \text{ for some } t \in [k]} a_j(\ell), b(\ell)t w_\ell \quad \text{(definition of diff}_i)$$

$$= a_i \sum_{\ell \in [m] \text{ s.t. } j(\ell) = i \text{ for some } t \in [k]} b(\ell)t w_\ell$$

$$= a_i \text{diff}_i(\Psi). \quad \text{(definition of diff}_i)$$

Item ii follows immediately from Item i and the definition $\text{bias}(\Psi) = \frac{1}{nW} \sum_{i=1}^n |\text{diff}_i(\Psi)|$. For Item iii, we have

$$\text{val}_{\Psi^a}(\sigma) = \frac{1}{W} \sum_{i \in [m]} w_i \text{Th}_k^i((b(i) \odot a|_{j(i)}) \odot \sigma|_{j(i)}) \quad \text{(definitions of } \Psi^a \text{ and val)}$$

$$= \frac{1}{W} \sum_{i \in [m]} w_i \text{Th}_k^i(b(i) \odot (\sigma \odot a)|_{j(i)})$$

$$= \text{val}_{\Psi}(\sigma \odot a). \quad \text{(definition of val)}$$

Finally, Item iv follows from Item iii and the fact that $\{\sigma : \sigma \in \{-1, 1\}^n\} = \{\sigma \odot a : \sigma \in \{-1, 1\}^n\}$:

$$\text{val}_{\Psi^a} = \max_{\sigma \in \{-1, 1\}^n} \text{val}_{\Psi^a}(\sigma) = \max_{\sigma \in \{-1, 1\}^n} \text{val}_{\Psi}(\sigma \odot a) = \max_{\sigma \in \{-1, 1\}^n} \text{val}_{\Psi}(\sigma) = \text{val}_{\Psi}. \quad \square$$
Thus, by linearity of expectation, we have \( \mu(D^\text{sym}_\Psi) \) to be the distribution obtained by sampling a constraint at random from \( \Psi \) and outputting its “randomly permuted negation pattern”. Formally, let \( S_k \) denote the set of permutations \([k] \to [k] \). For a vector \( b = (b_1, \ldots, b_k) \in \{-1, 1\}^k \) and a permutation \( \pi \in S_k \), let \( \pi(b) = (b_{\pi(1)}, \ldots, b_{\pi(k)}) \). Let \( C(i) = (b(i), j(i)) \) denote the \( i \)-th constraint of \( \Psi \), with weight \( w_i \), and let \( W = \sum_{i=1}^m w_i \) be the total weight. To sample a random vector from \( D^\text{sym}_\Psi \), we sample \( i \in [m] \) with probability \( \frac{w_i}{W} \), sample a permutation \( \pi \sim \text{Unif}(S_k) \), and output \( \pi(b(i)) \). The useful properties of \( D^\text{sym}_\Psi \) are summarized in the following claim:

**Proposition 7.6.** Let \( \Psi \) be an instance of Max-CSP(\( \text{Th}_k^1 \)). Then:

1. For any \( p \in [0, 1] \), \( \mathbb{E}_{a \sim \text{Bern}(p)^n} [\text{val}_\Psi(a)] = \lambda_S(D^\text{sym}_\Psi, p) \).

2. \( \mu(D^\text{sym}_\Psi) = \frac{1}{kW} \sum_{i=1}^n \text{diff}_i(\Psi) \leq \text{bias}(\Psi) \).

3. If \( \text{diff}_i(\Psi) \geq 0 \) for all \( i \in [n] \), then \( \mu(D^\text{sym}_\Psi) = \text{bias}(\Psi) \).

**Proof.** We begin with **Item i.** Fix a constraint \((b, j)\). We make two observations. Firstly, if we sample \( a \sim \text{Bern}(p)^n \), the distribution of \( a_j \) is identical to \( \text{Bern}(p)^k \). Secondly, for a fixed vector \( b \in \{-1, 1\}^k \), if we sample \( a \sim \text{Bern}(p)^k \) and \( \pi \sim \text{Unif}(S_k) \), the distributions of \( \pi(b) \odot a \) and \( \pi(b \odot a) \) are identical. Thus, we have

\[
\mathbb{E}_{a \sim \text{Bern}(p)^n} [\text{Th}_k^i(b(i) \odot a)] = \mathbb{E}_{a \sim \text{Bern}(p)^k} [\text{Th}_k^i(b \odot a)] \quad \text{(first observation)}
\]

\[
= \mathbb{E}_{a \sim \text{Bern}(p)^k, \pi \sim \text{Unif}(S_k)} [\text{Th}_k^i(\pi(b \odot a))] \quad \text{(symmetry of Th}_k^i)\]

\[
= \mathbb{E}_{a \sim \text{Bern}(p)^k, \pi \sim \text{Unif}(S_k)} [\text{Th}_k^i(\pi(b) \odot a)]. \quad \text{(second observation)}
\]

Thus, by linearity of expectation, we have \( \mathbb{E}_{a \sim \text{Bern}(p)^n} [\text{val}_\Psi(a)] = \mathbb{E}_{a \sim \text{Bern}(p)^k, b \sim D^\text{sym}_\Psi} [\text{Th}_k^i(b \odot a)] = \lambda_S(D^\text{sym}_\Psi, p) \), as desired.

For **Item ii**, we have

\[
\frac{1}{kW} \sum_{i=1}^n \text{diff}_i(\Psi) = \frac{1}{kW} \sum_{i=1}^m w_i \sum_{t=1}^k b(\ell)_t \quad \text{(definition of diff}_i(\Psi))
\]

\[
= \frac{1}{W} \sum_{i=1}^m w_i \mathbb{E}_{\pi \sim S_k} [\pi(b)_1] \quad \text{(where } \pi(b) = (\pi(b)_1, \ldots, \pi(b)_k))
\]

\[
= \mathbb{E}_{b \sim D^\text{sym}_\Psi} [b_1] \quad \text{(definition of } D^\text{sym}_\Psi)
\]

\[
= \mu(D^\text{sym}_\Psi). \quad \text{(definition of } \mu)
\]

Finally, **Item iii** follows immediately from **Item ii** and the definition \( \text{bias}(\Psi) = \frac{1}{kW} \sum_{i=1}^n \text{diff}_i(\Psi) \).

Now, we are equipped to prove the lemmas:
Proof of Lemma 7.1. Let \( \text{opt} \in \{-1, 1\}^n \) denote the optimal assignment for \( \Psi \). Then

\[
\text{val}_\Psi = \text{val}_\Psi(\text{opt}) \quad \text{(definition of } \text{opt})
\]

\[
= \text{val}_{\Psi, \text{opt}}(1^n) \quad \text{(Item iii of Proposition 7.5)}
\]

\[
= \lambda_S(D_{\Psi, \text{opt}}^\text{sym}, 1) \quad \text{(Item i of Proposition 7.6 with } p = 1)
\]

\[
= \gamma_S(D_{\Psi, \text{opt}}^\text{sym}) \quad \text{(definition of } \gamma_S \text{, Eq. (2.3))}
\]

\[
\leq \gamma_{S,k}(\mu(D_{\Psi, \text{opt}}^\text{sym})) \quad \text{(definition of } \gamma_{S,k} \text{, Eq. (2.6))}
\]

\[
\leq \gamma_{S,k}(\text{bias}(\Psi^{\text{opt}})) \quad \text{(Item ii of Proposition 7.6 and monotonicity of } \gamma_{S,k})
\]

\[
= \gamma_{S,k}(\text{bias}(\Psi)) \quad \text{(Item ii of Proposition 7.5)}
\]

as desired.

Proof of Lemma 7.2. Let \( \text{maj} \in \{-1, 1\}^n \) denote the assignment assigning \( x_i \) to 1 if \( \text{diff}_i(\Psi) \geq 0 \) and \(-1\) otherwise. Now

\[
\text{val}_\Psi = \text{val}_{\Psi, \text{maj}} \quad \text{(Item iv of Proposition 7.5)}
\]

\[
\geq \sup_{p \in [0,1]} \left( \mathbb{E}_{a \sim \text{Bern}(p)^n} \left[ \text{val}_{\Psi, \text{maj}}(a) \right] \right) \quad \text{(probabilistic method)}
\]

\[
= \sup_{p \in [0,1]} (\lambda_S(D_{\Psi, \text{maj}}^\text{sym}, p)) \quad \text{(Item i of Proposition 7.6)}
\]

\[
\geq \beta_S(D_{\Psi, \text{maj}}^\text{sym}) \quad \text{(definition of } \beta_S \text{, Eq. (2.3))}
\]

\[
\geq \beta_{S,k}(\mu(D_{\Psi, \text{maj}}^\text{sym})) \quad \text{(definition of } \beta_{S,k} \text{, Eq. (2.6))}
\]

\[
= \beta_{S,k}(\text{bias}(\Psi^{\text{maj}})) \quad \text{(Item iii of Proposition 7.6)}
\]

\[
= \beta_{S,k}(\text{bias}(\Psi)) \quad \text{(Item ii of Proposition 7.5)}
\]

as desired.

Finally, we state another consequence of Lemma 7.1 — a simple randomized, \( O(n) \)-time-and-space streaming algorithm for outputting approximately-optimal assignments when the max-min method applies.

**Theorem 7.7.** Let \( \text{Th}_k^j \) be a threshold function and \( p^* \in [0,1] \) be such that the max-min method applies, i.e.,

\[
\alpha(\text{Th}_k^j) = \inf_{D_N \in \Delta_k} \left( \frac{\lambda_S(D_N, p^*)}{\gamma_{S,k}(\mu(D_N))} \right).
\]

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

**Proof.** Let \( p^* \) be as in the theorem statement, and define \( \text{maj} \) as in the proof of Lemma 7.2. We output the assignment \( \text{maj} \odot a \) for \( a \sim \text{Bern}(p^*)^n \), and our goal is to show that its expected value is at least \( \alpha(\text{Th}_k^j) \cdot \text{val}_\Psi \).

Our assumption that the max-min method applies asserts in particular that

\[
\lambda_S(D_{\Psi, \text{maj}}^\text{sym}, p^*) \geq \alpha(\text{Th}_k^j) \gamma_{S,k}(\mu(D_{\Psi, \text{maj}}^\text{sym})). \quad (7.8)
\]
Thus our expected output value is

\[
E_{a \sim \text{Bern}(p^*)} [\text{val}_\Psi(\text{maj} \odot a)] = E_{a \sim \text{Bern}(p^*)} [\text{val}_\Psi(\text{maj} a)] = \lambda_S(D^\text{sym}_\Psi \text{maj}, p^*) \geq \alpha(\text{Th}_k^i) \gamma_{S,k} \mu(D^\text{sym}_\Psi \text{maj})) \geq \alpha(\text{Th}_k^i) \gamma_{S,k} \text{bias}(\Psi)) \geq \alpha(\text{Th}_k^i) \text{val}_\Psi,
\]

as desired.

\[\square\]

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 \text{Max-}k\text{AND} for all \(k\), and \(\text{Th}_k^{k-1}\) 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 [CGSV21a] 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 \text{Max-3AND}. While we believe that no \(o(\sqrt{n})\)-space streaming algorithm can beat their sketching algorithm for \text{Max-3AND}, proving this will require new techniques.
References


A Miscellaneous technical proofs

Proof of Proposition 4.5. Firstly, we show that it suffices WLOG to take the special case where \( r = n \) and \( y(1), \ldots, y(n) \) is the standard basis for \( \mathbb{R}^n \). Indeed, assume the special case and note that for a general case, we can let \( a' = (a \cdot y(1), \ldots, a \cdot y(r)) \), \( b' = (b \cdot y(1), \ldots, b \cdot y(r)) \), \( x' = (x_1, \ldots, x_r) \), and let \( y'(1), \ldots, y'(r) \) be the standard basis for \( \mathbb{R}^r \). Then \( x' = \sum_{i=1}^{r} a_i y'(i) \) and

\[
    f(x) = \frac{\sum_{i=1}^{r} (a \cdot y(i)) \alpha_i}{\sum_{i=1}^{r} (b \cdot y(i)) \alpha_i} = \frac{a' \cdot x'}{b' \cdot x'} \geq \min_{i \in [r]} \frac{a' \cdot y'(i)}{b' \cdot y'(i)} = \min_{i \in [r]} \frac{a \cdot y(i)}{b \cdot y(i)}.
\]

Now we prove the special case: Assume \( r = n \) and \( y(1), \ldots, y(n) \) is the standard basis for \( \mathbb{R}^n \). We have \( f(y(i)) = \frac{a_i}{b_i} \). Assume WLOG that \( f(y(1)) = \min \{ f(y(i)) : i \in [n] \} \), i.e., \( \frac{a_i}{b_i} \leq \frac{a_i}{b_i} \) for all \( i \in [n] \). Then \( a_i \geq \frac{a_i b_i}{b_i} \) for all \( i \in [n] \), so

\[
    a \cdot x \geq \sum_{i=1}^{n} \frac{a_i b_i}{b_i} \alpha_i = a_1 \frac{b_1}{b_1} (b \cdot x).
\]

Hence

\[
    f(x) = \frac{a \cdot x}{b \cdot x} \geq \frac{a_1}{b_1} = f(y(1)),
\]

as desired. \( \square \)

Proof of Theorem 1.7. To get an \((\alpha - \epsilon)\)-approximation to \( \text{val}_\Psi \), let \( \delta > 0 \) be small enough such that \( \frac{1-\delta}{1+\delta} \alpha(Th^1_k) \geq \alpha(Th^1_k) - \epsilon \). We claim that calculating an estimate \( \hat{b} \) for \( \text{bias}(\Psi) \) (using Corollary 7.4) up to a multiplicative \( \delta \) factor and outputting \( \hat{v} = \alpha(Th^1_k) \gamma_{S,k}(\frac{\hat{b}}{1+\delta}) \) is sufficient.

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

\[
    \gamma_{S,k} \left( \frac{\hat{b}}{1+\delta} \right) \geq \gamma_{S,k} \left( \frac{1-\delta}{1+\delta} \text{bias}(\Psi) \right) \quad \text{(monotonicity of} \gamma_{S,k})
\]

\[
    = \min \left\{ \frac{1 + \frac{\delta}{1+\delta} \text{bias}(\Psi)}{1 + \epsilon_{s,k}}, 1 \right\} \quad \text{(Lemma 3.3)}
\]

\[
    \geq \frac{1-\delta}{1+\delta} \min \left\{ \frac{1 + \text{bias}(\Psi)}{1 + \epsilon_{s,k}}, 1 \right\} \quad (\delta > 0)
\]

\[
    = \frac{1-\delta}{1+\delta} \gamma_{S,k}(\text{bias}(\Psi)). \quad \text{(Lemma 3.3)}
\]

Then we conclude

\[
    (\alpha(Th^1_k) - \epsilon)\text{val}_\Psi \leq (\alpha(Th^1_k) - \epsilon) \gamma_{S,k}(\text{bias}(\Psi)) \quad \text{(Lemma 7.1)}
\]

\[
    \leq \alpha(Th^1_k) \cdot \frac{1-\delta}{1+\delta} \gamma_{S,k}(\text{bias}(\Psi)) \quad \text{(assumption on} \ \delta)\]

\[
    \leq \hat{v} \quad \text{(our observation)}
\]

\[
    \leq \alpha(Th^1_k) \gamma_{S,k}(\text{bias}(\Psi)) \quad \text{(monotonicity of} \gamma_{S,k})
\]

\[
    \leq \beta_{S,k}(\text{bias}(\Psi)) \quad \text{(Eq. (2.7))}
\]

\[
    \leq \text{val}_\Psi, \quad \text{(Lemma 7.2)}
\]

as desired. \( \square \)