DNF SPARSIFICATION AND A FASTER DETERMINISTIC COUNTING ALGORITHM

Parikshit Gopalan, Raghu Meka, and Omer Reingold

Abstract. Given a DNF formula $f$ on $n$ variables, the two natural size measures are the number of terms or size $s(f)$ and the maximum width of a term $w(f)$. It is folklore that small DNF formulas can be made narrow: if a formula has $m$ terms, it can be $\varepsilon$-approximated by a formula with width $\log(m/\varepsilon)$. We prove a converse, showing that narrow formulas can be sparsified. More precisely, any width $w$ DNF irrespective of its size can be $\varepsilon$-approximated by a width $w$ DNF with at most $(w \log(1/\varepsilon))^{O(w)}$ terms.

We combine our sparsification result with the work of Luby & Velickovic (1991, Algorithmica 16(4/5):415–433, 1996) to give a faster deterministic algorithm for approximately counting the number of satisfying solutions to a DNF. Given a formula on $n$ variables with poly($n$) terms, we give a deterministic $n^{\tilde{O}(\log \log(n))}$ time algorithm that computes an additive $\varepsilon$ approximation to the fraction of satisfying assignments of $f$ for $\varepsilon = 1/\text{poly}(\log n)$. The previous best result due to Luby and Velickovic from nearly two decades ago had a run time of $n^{\exp(O(\sqrt{\log \log n})}$ (Luby & Velickovic 1991, in Algorithmica 16(4/5):415–433, 1996).

Keywords. Complexity, pseudorandomness, derandomization, DNF formulae.

Subject classification. 68Q25, 68Q87.

1. Introduction

A natural way to represent a Boolean function $f : \{0, 1\}^n \rightarrow \{0, 1\}$ is to write it as a CNF or DNF formula. The class of functions that
admit compact representations of this form (aka polynomial size CNF and DNF formulae) is central to Boolean function analysis, computational complexity, and machine learning.

Given a DNF formula $f$ on $n$ variables, the two natural size measures are the number of terms or size $s(f)$ and the maximum width of a term $w(f)$. The analogous measures for a CNF are the number of clauses and clause width. It is folklore that every DNF formula $f$ with $m$ terms can be $\varepsilon$-approximated by another DNF $g$ where $s(g) \leq m$ and $w(g) \leq \log(m/\varepsilon)$, regardless of $w(f)$. The formula $g$ is a sparsification of $f$ obtained by simply discarding all terms of width larger than $\log(m/\varepsilon)$. In other words, small DNF formulas can be made narrow. An analogous statement can be derived for CNFs.

In this work, we show the reverse connection: narrow formulae can be made small. Indeed, we prove the existence of a strong form of approximation known as sandwiching approximations which are important in pseudorandomness. In this work, we only consider approximators which are also Boolean functions.

**Definition 1.1.** Let $f : \{0, 1\}^n \rightarrow \{0, 1\}$. We say that functions $f_u, f_\ell : \{0, 1\}^n \rightarrow \{0, 1\}$ are $\varepsilon$-sandwiching approximators for $f$ if $f_\ell(x) \leq f(x) \leq f_u(x)$ for every $x \in \{0, 1\}^n$, and

$$
\Pr_{x \in \{0, 1\}^n} [f_\ell(x) \neq f(x)] = \Pr_{x \in \{0, 1\}^n} [(f_\ell(x) = 0) \land (f(x) = 1)] \leq \varepsilon,
$$

$$
\Pr_{x \in \{0, 1\}^n} [f_u(x) \neq f(x)] = \Pr_{x \in \{0, 1\}^n} [(f_u(x) = 1) \land (f(x) = 0)] \leq \varepsilon.
$$

Our main result is the existence of $\varepsilon$-sandwiching approximators for arbitrary width $w$ DNFs using small-width $w$ DNFs where the number of clauses depends only on $w$ and $\varepsilon$.

**Theorem 1.2.** For every width-$w$ DNF formula $f$ and every $\varepsilon > 0$, there exist DNF formulae $f_\ell, f_u$ each of width $w$ and size at most ($w \log(1/\varepsilon))O(w)$ which are $\varepsilon$-sandwiching approximators for $f$.

Our result is proved by a sparsification procedure for DNF formulae which uses the notion of quasi-sunflowers due to Rossman (2010). The best previously known result along these lines