Log-Approximable Minimization Problems on Random Inputs

Anders Malmström
Lehrgebiet Mathematische Grundlagen der Informatik
RWTH Aachen, Germany
E-mail: anders@informatik.rwth-aachen.de

Abstract

We extend recent work about the relationship between logically defined classes of NP optimization problems and the asymptotic growth of optimal solutions on random inputs. We consider the syntactic class MIN $F^+\Pi_2(1)$ of minimization problems. Kolaitis and Thakur proved that every problem in this class is log-approximable. We show that for every problem $\mathcal{Q}$ in the class MIN $F^+\Pi_2(1)$ there exist polynomials $g(n)$ and $h(n)$ such the optimal solution of $\mathcal{Q}$ almost surely grows like $\Theta(g(n)+h(n)\cdot\log n)$. Applying this result we show without using any complexity theoretical assumptions that the problem MIN GRAPH COLOURING is not in MIN $F^+\Pi_2(1)$. With the same method we get a similar criterion for each class MIN $F^+\Pi_2(k)$. Using the fact that on a random graph with n nodes the chromatic number is $n/(2\log n)$ almost surely we prove that MIN GRAPH COLOURING is not in MIN $F^+\Pi_2$, resolving a conjecture by Kolaitis and Thakur.

1 Introduction

Many NP-complete decision problems come from combinatorial optimization problems by putting a bound on the cost of a desired solution. Assuming P $\neq$ NP we cannot find polynomial-time algorithms which solve optimization problems exactly. This does not exclude to find nearly optimal solutions, that means solutions whose relative error compared to an optimal solution is smaller than a constant. For many problems their approximation properties are known. Some have efficient approximization algorithms, others like the TSP are as hard to approximate as to solve exactly. It would be interesting to know more about the structural reasons for these differences.

Motivated by Fagin's characterization of NP by existential second-order logic on finite structures [4], Papadimitriou and Yannakakis introduced in [11] methods for defining NP optimization problems using logic. They showed that in some important cases there is a relation between the logical representation of an optimization problem and its approximation properties. They defined the two syntactic classes of optimization problems, MAX SNP and MAX NP, and proved that for all problems $\mathcal{Q}$ in these classes there exists a polynomial-time algorithm that approximates $\mathrm{opt}_{\mathcal{Q}}$ up to a constant factor. Many natural maximization problems like MAX CUT and MAX SAT are contained in MAX NP. Surprisingly the analogue definitions for minimization problems do not lead to a similar result (see [7] and [8] for details). However, Kolaitis and Thakur introduced in [7] a different syntactic class, called MIN $F^+\Pi_1$, and proved that all problems in this class are approximable in polynomial time up to a constant factor. MIN $F^+\Pi_1$ consists of all minimization problems $\mathcal{Q}$ whose input instances $\mathfrak{A}$ are finite structures such that the optimum can be expressed as

$$\mathrm{opt}_{\mathcal{Q}}(\mathfrak{A}) = \min_S\{|S| : (\mathfrak{A}, S) \models \forall\bar{x}\psi(\bar{x}, S)\},$$

where S is a single predicate variable and $\psi(\bar{x}, S)$ is a quantifier-free first-order formula in which all occurrences of S are positive.

In [8] the same authors introduced the syntactic class MIN $F^+\Pi_2$ which is an extension of MIN $F^+\Pi_1$.

Definition 1.1 For every $k \in \mathbb{N}$, the class MIN $F^+\Pi_2(k)$ consists of all minimization problems $\mathcal{Q}$ whose optimum can be expressed as

$$\mathrm{opt}_{\mathcal{Q}}(\mathfrak{A}) = \min_S\{|S| : (\mathfrak{A}, S) \models \forall\bar{x}\exists\bar{y}\psi(\bar{x}, \bar{y}, S)\},$$

where S is a single predicate variable, $\psi(\bar{x}, \bar{y}, S)$ is a quantifier-free formula in disjunctive normal form (DNF), in which all occurrences of S are positive and S occurs at most k times in each disjunct. Moreover, let MIN $F^+\Pi_2 = \bigcup_k$ MIN $F^+\Pi_2(k)$.

Kolaitis and Thakur proved that every problem in MIN $F^+\Pi_2(1)$ is log-approximable. An example for this class is MIN DOMINATING SET. Given a graph, the problem asks for the cardinality of the smallest subset S of vertices such that every vertex either is in S or is adjacent to a vertex in S. In fact MIN DOMINATING SET is complete for MIN $F^+\Pi_2(1)$. It is definable by the expression

$$\mathrm{opt}_{\mathrm{MIN\ DS}}(G) = \min_S\{|S| : (G, S) \models \forall x\exists y(Sx \vee (Exy \wedge Sy))\}.$$

Other complete problems in this class are MIN SET COVER and MIN HITTING SET.

It is not known whether MIN $F^+\Pi_2$ is in APX, the class of all NP optimization problems that are approximable up to a constant factor. In fact this is unlikely since Lund and Yannakakis showed in [10] that there exists a constant c, such that MIN DOMINATING SET cannot be approximated within ratio $c \log n$ unless NP is contained in DTIME$[\mathrm{n}^{\mathrm{poly}\log \mathrm{n}}]$.

For an overview of NP optimization problems see [3] or [6].

To show that a problem is in a logically defined class is usually done by giving the desired logical formula. On the other hand showing that a specific problem does not belong to a class is more difficult. In [2] Behrendt, Compton and Grädel investigated a new technique. They introduced a probabilistic method that characterizes the growth of the size of expected optimal solutions for problems that belong to MAX Σ_1. They showed that on a random input it grows like $\Theta(p(n))$, where $p(n)$ is a polynomial and n the cardinality of the input. In [5] we established a similar criterion for membership in MIN $F^+\Pi_1$. Using this result we proved that MIN DOMINATING SET is not contained in MIN $F^+\Pi_1$.

In this paper we show an analogue result for MIN $F^+\Pi_2(1)$ which looks different to the hitherto existing criteria.

Theorem 1.2 (Probabilistic criterion for MIN $F^+\Pi_2(1)$**)** *Let $\mathcal{Q}$ be a problem in* MIN $F^+\Pi_2(1)$. *Then there exist polynomials $g(n)$ and $h(n)$ such that almost surely*

$$opt_{\mathcal{Q}}(\mathfrak{A}) = \Theta(g(n) + h(n) \cdot \log n)$$

for a randomly chosen σ-structure $\mathfrak{A}$, where $n = |\mathfrak{A}|$.

As a relevant example we have MIN DOMINATING SET. In [5] we proved that for a random graph G with n vertices and edge probability $1/2$ almost surely $\text{opt}_{\text{MIN DS}}(G)$ is $\log n$.

Applying this result we can show without using any complexity theoretical assumptions that the problem MIN GRAPH COLOURING is not in MIN $F^+\Pi_2(1)$. MIN GRAPH COLOURING is the problem of computing the chromatic number of a graph. With the same method we get a similar criterion for each class MIN $F^+\Pi_2(k)$. Using the fact that on a random graph with n nodes the chromatic number is $n/(2\log n)$ almost surely [1] we can show that MIN GRAPH COLOURING is not in MIN $F^+\Pi_2$, resolving a conjecture by Kolaitis and Thakur [8].

2 Preliminaries

Definition 2.1 An *NP optimization problem* is a tuple $\mathcal{Q} = (\mathcal{I}_{\mathcal{Q}}, \mathcal{F}_{\mathcal{Q}}, f_{\mathcal{Q}}, \text{opt})$ such that

- $\mathcal{I}_{\mathcal{Q}}$ is the set of *input instances.*
- $\mathcal{F}_{\mathcal{Q}}(I)$ is the set of *feasible solutions* for input I. Here 'feasible' means that the size of the elements $S \in \mathcal{F}_{\mathcal{Q}}(I)$ is polynomially bounded in the size of I and that the set $\{(I,S) : S \in \mathcal{F}_{\mathcal{Q}}(I)\}$ is recognizable in polynomial time.
- $f_{\mathcal{Q}} : \{(I,S) : S \in \mathcal{F}_{\mathcal{Q}}(I)\} \to \mathbb{N}$ is a polynomial-time computable function, called the *cost function.*
- $\text{opt}_{\mathcal{Q}}$ is one of the two functions defined below with $\mathcal{I}_{\mathcal{Q}}$ as domain and positive integers as values:

  $$\text{opt}_{\mathcal{Q}}(I) = \min_S f_{\mathcal{Q}}(I,S) \text{ or } \text{opt}_{\mathcal{Q}}(I) = \max_S f_{\mathcal{Q}}(I,S).$$

  In the first case we say $\mathcal{Q}$ is a *minimization problem* and in the second case we say $\mathcal{Q}$ is a *maximization problem.*

For every NP optimization problem $\mathcal{Q}$, the following decision problem is in NP: Given an instance I of $\mathcal{Q}$ and a natural number k, is there a solution $S \in \mathcal{F}_{\mathcal{Q}}(I)$ such that $f_{\mathcal{Q}}(I,S) \leqslant k$ when opt=min, (or $f_{\mathcal{Q}}(I,S) \geqslant k$, when opt = max).

First we give the exact result about the approximation properties of the class Min $F^+\Pi_2(k)$, $k \geqslant 1$.

Theorem 2.2 (Kolaitis, Thakur [8]) *Let $\mathcal{Q}$ be an optimization problem in the class* Min $F^+\Pi_2(k)$, $k \geqslant 1$. *Then there exists a polynomial-time approximation algorithm and a constant c such that for every instance $\mathfrak{A}$ of $\mathcal{Q}$, the algorithm produces a feasible solution on which the cost function takes on a value less than or equal to $c(opt_{\mathcal{Q}}(\mathfrak{A}))^k \cdot \log(|\mathfrak{A}|)$. In particular, every problem in the class* Min $F^+\Pi_2(1)$ *is log-approximable.*

We see that the most important of these classes is Min $F^+\Pi_2(1)$. We show our main result for this class and then we extend it to Min $F^+\Pi_2(k)$ for each k.

With some simple syntactical transformations we get

Lemma 2.3 Min $F^+\Pi_1 \subsetneq$ Min $F^+\Pi_2(1)$.

Moreover we need some definitions and results from the theory of asymptotic probabilities and from the theory of graphs.

As usual, for functions $f, g : \mathbb{N} \rightarrow \mathbb{R}$, we write $f = \Theta(g)$, if there are constants $c, d > 0$ such that $cg(n) \leqslant f(n) \leqslant dg(n)$ for sufficiently large n, and we write $f \sim g$ to indicate that $\lim_{n\rightarrow\infty} f(n)/g(n) = 1$.

Definition 2.4 Let σ be a finite relational vocabulary and $x_1, \ldots, x_k$ be a sequence of distinct variables. A maximal consistent set t of σ-atoms and negated σ-atoms (including equalities and inequalities) in $x_1, \ldots, x_k$ is called an *atomic σ-type in $x_1 \ldots, x_k$*. Since such a set is always finite, we can form in first-order logic the conjunction of the formulae in t; by abuse of notation, we denote this conjunction by $t(x_1, \ldots, x_k)$. On every σ-structure $\mathfrak{A}$, the type t defines the set of realizations $t^{\mathfrak{A}} = \{\bar{u} \in A^k : \mathfrak{A} \models t(\bar{u})\}$, where A is the universe of $\mathfrak{A}$.

A $\varnothing$-type is called an equality type e. So e is a maximal consistent set of equalities $x_i = x_j$ and inequalities $x_i \neq x_j$, where $1 \leqslant i < j \leqslant k$, which defines on every structure $\mathfrak{A}$ the set $e^{\mathfrak{A}} = \{\bar{u} \in A^k : \mathfrak{A} \models e(\bar{u})\}$.

For each atomic type t let r_t be the unique natural number of pairwise distinct components of each tuple of type t.

Theorem 2.5 (Kolaitis, Vardi [9]) *For every formula $\psi(\bar{x})$ in the infinitary logic $L^{\omega}_{\infty\omega}$ there exist atomic types $t_i(\bar{x})$ such that*

$$\mathfrak{A} \models \forall \bar{x}(\psi(\bar{x}) \leftrightarrow \bigvee_i t_i(\bar{x}))$$

for all but an exponentially decreasing fraction of structures $\mathfrak{A}$ when the size of $\mathfrak{A}$ tends to infinity.

Lemma 2.6 *For each $n \in \mathbb{N}$, let $G = (V, E)$ be a* random graph *of the following form: V is the disjoint union of two sets A of cardinality $a(n)$ and B of cardinality $b(n)$, $a(n), b(n)$ non-constant polynomials. A is an independent set and B is a clique of G and the probability for every $(u, v) \in A \times B$ to be an edge is independently a constant p.*
Then almost surely the domination number of G satisfies

$$\gamma(G) \sim \frac{1}{\log \frac{1}{1-p}} \log a(n) = O(\log n)$$

Proof. Let $X_r(G)$ be the number of different dominating sets D of cardinality r. Thus

$$\gamma(G) = \min\{r \in \mathbb{N} : X_r(G) \geqslant 1\}.$$

Without loss of generality $D \subset B$. We estimate the random variable X_r for $r = r(n)$. Define $q = 1 - p$. Let D be an arbitrary subset of B of cardinality $r(n) = \log_{1/q} a(n)$. The probability that D is a dominating set is

$$\left(1 - (1-p)^{r(n)}\right)^{a(n)} = \left(1 - q^{r(n)}\right)^{a(n)} = \left(1 - \frac{1}{a(n)}\right)^{a(n)}$$

and this tends to $\frac{1}{e}$ if n tends to infinity. We infer that $\gamma(G) \leqslant \log_{1/q} a(n)$ almost surely.

Now let $r(n) = (1-\varepsilon)\log_{1/q} a(n)$ for any $\varepsilon > 0$. There are $\binom{b(n)}{(1-\varepsilon)\log_{1/q} a(n)}$ different choices for D in B. Define the indicator random variable X^D by

$$X_r^D(G) := \begin{cases} 1 & : \quad D \text{ is a dominating set of } G \\ 0 & : \quad \text{otherwise} \end{cases}$$

Obviously, $X_r = \sum_D X_r^D$. By linearity of expectation it follows that

$$\begin{aligned} E(X_r) &= \sum_D E(X_r^D) = \binom{b(n)}{(1-\varepsilon)\log_{1/q} a(n)} \left(1 - \frac{1}{a(n)^{(1-\varepsilon)}}\right)^{a(n)} \\ &= \binom{b(n)}{(1-\varepsilon)\log_{1/q} a(n)} \left(1 - \frac{1}{a(n)^{(1-\varepsilon)}}\right)^{a(n)^{(1-\varepsilon)} \cdot \frac{a(n)}{a(n)^{(1-\varepsilon)}}} \end{aligned}$$

Because $\lim_{n\to\infty}(1 - n^{-s})^{n^s} = e^{-1}$ and $\binom{n}{m} \leqslant n^m$ we can infer that

$$\begin{aligned} \lim_{n\to\infty} E(X_r) &\leqslant \lim_{n\to\infty} b(n)^{(1-\varepsilon)\log_{1/q} a(n)} \left(\frac{1}{e}\right)^{a(n)^{\varepsilon}} \\ &\leqslant \lim_{n\to\infty} e^{\ln b(n) \cdot (1-\varepsilon)\log_{1/q} a(n) - a(n)^{\varepsilon}} = 0 \end{aligned}$$

By Markov's inequality, $P[X_r \geqslant 1] \leqslant E(X_r)$, so almost surely
$\gamma(G) > (1-\varepsilon)\log_{1/q} a(n)$ and thus $\gamma(G) \sim \log_{1/q} a(n)$ almost surely. ∎

3 Main Result

In this section we establish a necessary criterion for the membership of a problem $\mathcal{Q}$ in the class MIN $F^+\Pi_2(1)$. We will give a probabilistic estimation of $\text{opt}_\mathcal{Q}$ on a random structure $\mathfrak{A}$. First we need a normal form for the first-order formula that defines $\mathcal{Q}$.

Let $\mathcal{Q}$ be an optimization problem whose input instances are finite structures over a fixed vocabulary σ.

Lemma 3.1 (Normal form for MIN $F^+\Pi_2(k)$**)** *Let $\mathcal{Q} \in$ MIN $F^+\Pi_2(k)$, then there exist a universal first-order formula $\gamma(\bar{x})$ and an existential first-order formula $\varphi(\bar{x}, \bar{z})$ such that for all $\mathfrak{A}$*

$$opt_\mathcal{Q}(\mathfrak{A}) = \min_S\{|S| : (\mathfrak{A}, S) \models \forall\bar{x}[\gamma(\bar{x}) \rightarrow \exists\bar{z}_1 \ldots \exists\bar{z}_k(\varphi(\bar{x}, \bar{z}) \wedge \bigwedge_{j=1}^{k} S\bar{z}_j)]\},$$

and $\bar{z}_j$ is of length m.

PROOF. According to Definition 1.1, the optimum of $\mathcal{Q}$ can be expressed for all input structures $\mathfrak{A}$ as:

$$\text{opt}_\mathcal{Q}(\mathfrak{A}) = \min_S\{|S| : (\mathfrak{A}, S) \models \forall\bar{x}\exists\bar{y}\psi(\bar{x}, \bar{y}, S)\},$$

where S is of arity m and $\psi(\bar{x}, \bar{y}, S)$ is a quantifier-free DNF formula in which all occurrences of S are positive and S occurs at most k times in each disjunct. Define the disjunction of all subformulae in which S does not occur by $\omega(\bar{x}, \bar{y})$. Without loss of generality, we can assume that in all other disjuncts S occurs exactly k times. So let us make the following logical deductions:

$$\begin{aligned}
\forall\bar{x}\exists\bar{y}\psi(\bar{x}, \bar{y}, S) &\equiv \forall\bar{x}\exists\bar{y}[\omega(\bar{x}, \bar{y}) \vee \bigvee_i \psi_i(\bar{x}, \bar{y}, S)] \\
&\equiv \forall\bar{x}\exists\bar{y}[\omega(\bar{x}, \bar{y}) \vee \bigvee_i (\varphi_i(\bar{x}, \bar{y}) \wedge \bigwedge_{j=1}^{k} S\bar{y}_{i_j})] \\
&\equiv \forall\bar{x}\exists\bar{y}\exists\bar{z}_1 \ldots \exists\bar{z}_k[\omega(\bar{x}, \bar{y}) \vee \bigvee_i (\varphi_i(\bar{x}, \bar{y}) \wedge \bigwedge_{j=1}^{k} \bar{z}_j = \bar{y}_{i_j} \wedge \bigwedge_{j=1}^{k} S\bar{z}_j)] \\
&\equiv \forall\bar{x}\exists\bar{y}\exists\bar{z}_1 \ldots \exists\bar{z}_k[\omega(\bar{x}, \bar{y}) \vee (\bigvee_i (\varphi_i(\bar{x}, \bar{y}) \wedge \bigwedge_{j=1}^{k} \bar{z}_j = \bar{y}_{i_j}) \wedge \bigwedge_{j=1}^{k} S\bar{z}_j)] \\
&\equiv \forall\bar{x}[(\forall\bar{y}\neg\omega(\bar{x}, \bar{y})) \rightarrow \exists\bar{z}_1 \ldots \exists\bar{z}_k(\exists\bar{y}\bigvee_i (\varphi_i(\bar{x}, \bar{y}) \wedge \bigwedge_{j=1}^{k} \bar{z}_j = \bar{y}_{i_j}) \\
&\qquad \wedge \bigwedge_{j=1}^{k} S\bar{z}_j))]
\end{aligned}$$

■

We will now consider the behaviour of $\mathrm{opt}_{\mathcal{Q}}(\mathfrak{A})$ on a randomly chosen σ-structure $\mathfrak{A}$. Let $\Omega(n)$ be the probability space of all σ-structures over universe $\{0,\ldots,n-1\}$ with uniform probability distribution. So $\mathrm{opt}_{\mathcal{Q}}$ is a random variable on $\Omega(n)$.

We show the result for the lowest class MIN $F^+\Pi_2(1)$ and we will generalize it to MIN $F^+\Pi_2(k)$ for an arbitrary k in the next section.

Theorem 3.2 *Let $\mathcal{Q}$ be a problem in* MIN $F^+\Pi_2(1)$. *Then there exist polynomials $g(n)$ and $h(n)$ such that almost surely*

$$opt_{\mathcal{Q}}(\mathfrak{A}) = \Theta(g(n) + h(n)\cdot\log n)$$

for a randomly chosen σ-structure $\mathfrak{A}$, where $n = |\mathfrak{A}|$.

PROOF. By lemma 3.1 we can express the optimum of every $\mathcal{Q} \in$ MIN $F^+\Pi_2(1)$ as:

$$\mathrm{opt}_{\mathcal{Q}}(\mathfrak{A}) = \min_S\{|S| : (\mathfrak{A}, S) \models \forall\bar{x}[\gamma(\bar{x}) \rightarrow \exists\bar{z}(\varphi(\bar{x},\bar{z}) \wedge S\bar{z})]\}.$$

By theorem 2.5 there exists a finite number of atomic types $u_i(\bar{x})$ and $t_j(\bar{x},\bar{z})$ such that

$$\mathfrak{A} \models \forall\bar{x}(\gamma(\bar{x}) \leftrightarrow \bigvee_i u_i(\bar{x})) \text{ and } \mathfrak{A} \models \forall\bar{x}\forall\bar{z}(\varphi(\bar{x},\bar{z}) \leftrightarrow \bigvee_j t_j(\bar{x},\bar{z}))$$

for all but an exponentially decreasing fraction of structures $\mathfrak{A}$. So almost surely

$$\mathrm{opt}_{\mathcal{Q}}(\mathfrak{A}) = \min_S\{|S| : (\mathfrak{A}, S) \models \forall\bar{x}[\bigvee_i u_i(\bar{x}) \rightarrow \exists\bar{z}(\bigvee_j t_j(\bar{x},\bar{z}) \wedge S\bar{z})]\}.$$

For each i there exists at least one j such that t_j is an extension type of u_i. That means we can decompose the formula $t_j(\bar{x},\bar{z})$ in the following way.

$$t_j(\bar{x},\bar{z}) = u_i(\bar{x}) \wedge v_j(\bar{z}) \wedge \alpha_j(\bar{x},\bar{z}) \wedge \beta_j(\bar{x},\bar{z}),$$

where v_j is the type of $\bar{z}$, α_j contains only equalities $x_k = z_l$ and inequalities $x_k \neq z_l$ and β_j contains only atoms and negated atoms (no equalities and inequalities) $R\bar{y}$ where $\bar{y}$ intersects $\bar{x}$ and $\bar{z}$. In this case we call (i,j) feasible. Let F be the finite set of all feasible pairs (i,j). We can infer that

$$\mathrm{opt}_{\mathcal{Q}}(\mathfrak{A}) = \min_S\{|S| : (\mathfrak{A}, S) \models \forall\bar{x} \bigvee_{(i,j)\in F} [u_i(\bar{x}) \rightarrow \exists\bar{z}t_j(\bar{x},\bar{z}) \wedge S\bar{z}]\}.$$

Let us consider the special case where

$$\mathrm{opt}_{\mathcal{Q}}(\mathfrak{A}) = \min_S\{|S| : (\mathfrak{A}, S) \models \forall\bar{x}[u(\bar{x}) \rightarrow \exists\bar{z}t(\bar{x},\bar{z}) \wedge S\bar{z}]\}.$$

and $t(\bar{x},\bar{z}) = u(\bar{x}) \wedge v(\bar{z}) \wedge \alpha(\bar{x},\bar{z}) \wedge \beta(\bar{x},\bar{z})$. The optimum of $\mathcal{Q}$ is the cardinality of the smallest subset $S \subset v^{\mathfrak{A}}$ such that for every $\bar{x} \in u^{\mathfrak{A}}$ there exists a $\bar{z} \in S$ such that $t(\bar{x},\bar{z})$ becomes TRUE. Let r_u and r_v be the natural numbers from definition 2.4, the number of pairwise distinct components of $\bar{x}$ respectively $\bar{z}$. From these r_v pairwise distinct components of $\bar{z}$ let m be the number of components that are equal to some component of $\bar{x}$. We get that $m \leqslant \min\{r_u, r_v\}$.

Each value of this m components defines a subset $u_l^{\mathfrak{A}}$ of $u^{\mathfrak{A}}$. So $u^{\mathfrak{A}}$ is the disjoint union of $n(n-1)\dots(n-m+1)$ sets $u_l^{\mathfrak{A}}$, and the sets

$v_l^{\mathfrak{A}} = \{\bar{z} : \exists \bar{x} \in u_l^{\mathfrak{A}} \wedge (\bar{x},\bar{z}) \in t^{\mathfrak{A}}\}$ are disjoint too.

We have reduced our problem to finding the minimal subset $S_l \subset v_l^{\mathfrak{A}}$ such that for every $\bar{x} \in u_l^{\mathfrak{A}}$ there exists a $\bar{z} \in S_l$ such that $t(\bar{x},\bar{z})$ becomes TRUE. But this is the same problem as finding a minimal dominating set of the graph $G_l = (V_l, E_l)$ which is defined in the following way. Consider V_l as the disjoint union of $u_l^{\mathfrak{A}}$ and $v_l^{\mathfrak{A}}$, and let $(a,b) \in E_l$ if and only if $(a,b) \in v_l^{\mathfrak{A}} \times v_l^{\mathfrak{A}}$ or $(a,b) \in u_l^{\mathfrak{A}} \times v_l^{\mathfrak{A}} \wedge (a,b) \in t^{\mathfrak{A}}$.

Almost surely the size of each $u_l^{\mathfrak{A}}$ is $u_l(n) = O(n^{r_u - m})$ and the size of $v_l^{\mathfrak{A}}$ is $v_l(n) = O(n^{r_v - m})$.

First we consider the case where $m < \min\{r_u, r_v\}$. The probability that randomly chosen $\bar{x} \in u_l^{\mathfrak{A}}$ and $\bar{z} \in v_l^{\mathfrak{A}}$ are in $t^{\mathfrak{A}}$ is a constant p.

With lemma 2.6 we get that $\gamma(G_l) \sim O(\log u_l(n))$ almost surely.

We can infer that almost surely

$$\begin{aligned} \mathrm{opt}_{\mathcal{Q}}(\mathfrak{A}) = \mathrm{opt}_{\mathcal{Q}}(\dot{\bigcup_l} G_l) &= \sum_l \gamma(G_l) \\ &\sim \sum_l O(\log u_l(n)) = O(n^m) \cdot O(\log n^{r_u - m}) \end{aligned}$$

If $m = \min\{r_u, r_v\}$ then

$$\mathrm{opt}_{\mathcal{Q}}(\mathfrak{A}) = \min\{|u^{\mathfrak{A}}|, |v^{\mathfrak{A}}|\}$$

Let us go back to the general case. For each feasible pair (i,j) we find a polynomial $g_{i,j}(n)$ such that the optimum of $\mathcal{Q}$ on $\mathfrak{A}$ restricted to the types u_i and t_j grows asymptotically like $g_{i,j}(n) \cdot \log n$ or like $g_{i,j}(n)$. This yields that the optimum almost surely grows like the finite sum of polynomials and polynomials multiplied by $\log n$ and we get the assertion. ■

4 Generalization and Application

We can extend this criterion to the whole class MIN $F^+\Pi_2$.

Theorem 4.1 (Probabilistic criterion for MIN $F^+\Pi_2(k)$) *Let $\mathcal{Q}$ be a problem in* MIN $F^+\Pi_2(k)$*. Then there exist polynomials $g_i(n)$, $0 \leqslant i \leqslant k$, such that almost surely*

$$opt_{\mathcal{Q}}(\mathfrak{A}) = \Theta\left(g_0(n) + \sum_{i=1}^{k} g_i(n) \cdot \log^{\frac{1}{i}} n\right)$$

for a randomly chosen σ-structure $\mathfrak{A}$, where $n = |\mathfrak{A}|$.

First we need a generalization of lemma 2.6 to $k+1$-hypergraphs. A $k+1$-hyperdominating set is a set of vertices such that for every vertex there exist k vertices in this set and these $k+1$ vertices form a $k+1$-hyperedge. The hyperdomination number is the cardinality of the smallest hyperdominating set.

Lemma 4.2 *For each $n \in \mathbb{N}$, let $H = (V, E)$ be a random $k+1$-hypergraph of the following form: V is the disjoint union of the sets A and $B_1, \ldots, B_k$ of cardinalities $a(n), b_1(n), \ldots, b_k(n)$, non-constant polynomials. A is an independent set, $B = \bigcup_{i=1}^{k} B_i$ is a $k+1$-hyperclique of H and every tuple $(u, v_1, \ldots, v_k) \in A \times B_1 \times \ldots \times B_k$ is independently a $k+1$-hyperedge with constant probability p.*
Then almost surely the $k+1$-hyperdomination number of H satisfies

$$\tilde{\gamma}(H) \sim O(\log^{\frac{1}{k}} n)$$

Proof. Let $X_r(H)$ be the number of different hyperdominating sets D of cardinality r. Thus

$$\tilde{\gamma}(H) = \min\{r \in \mathbb{N} : X_r(H) \geqslant 1\}.$$

Without loss of generality $D \subset B$. We estimate the random variable X_r for $r = r(n)$. Define $q = 1 - p$. Let D_i be an arbitrary subset of B_i of cardinality $\tilde{r}(n) = \log_{1/q}^{\frac{1}{k}} a(n)$. Thus $r(n) = k \cdot \tilde{r}(n)$. The probability that $D = \bigcup_{i=1}^{k} D_i$ is a hyperdominating set is

$$\left(1 - (1-p)^{r(n)^k}\right)^{a(n)} = \left(1 - q^{r(n)^k}\right)^{a(n)} = \left(1 - \frac{1}{a(n)}\right)^{a(n)}$$

and this tends to $\frac{1}{e}$ if n tends to infinity. We infer that $\tilde{\gamma}(H) \leqslant k \log_{1/q}^{\frac{1}{k}} a(n)$ almost surely.

Now let $\tilde{r}(n) = ((1-\varepsilon)\log_{1/q} a(n))^{\frac{1}{k}}$ for any $\varepsilon > 0$. There are $\prod_{i=1}^{k} \binom{b_i(n)}{\tilde{r}(n)}$ different choices for D in B. Define the indicator random variable X^D by

$$X_r^D(G) := \begin{cases} 1 & : \quad D \text{ is a dominating set of } G \\ 0 & : \quad \text{otherwise} \end{cases}$$

Obviously $X_r = \sum_D X_r^D$. By linearity of expectation it follows that

$$\begin{aligned} E(X_r) &= \sum_D E(X_r^D) = \left(\prod_{i=1}^{k} \binom{b_i(n)}{\tilde{r}(n)}\right)\left(1 - \frac{1}{a(n)^{(1-\varepsilon)}}\right)^{a(n)} \\ &= \left(\prod_{i=1}^{k} \binom{b_i(n)}{\tilde{r}(n)}\right)\left(1 - \frac{1}{a(n)^{(1-\varepsilon)}}\right)^{a(n)^{(1-\varepsilon)} \cdot \frac{a(n)}{a(n)^{(1-\varepsilon)}}} \end{aligned}$$

Because $\lim_{n\to\infty}(1-n^{-s})^{n^s} = e^{-1}$ and $\binom{n}{m} \leqslant n^m$ we can infer that

$$\begin{aligned}
\lim_{n\to\infty} E(X_r) &\leqslant \lim_{n\to\infty} \left(\prod_{i=1}^{k} b_i(n)^{\tilde{r}(n)}\right)\left(\frac{1}{e}\right)^{a(n)^\varepsilon} \leqslant \lim_{n\to\infty} b(n)^{k\tilde{r}(n)}\left(\frac{1}{e}\right)^{a(n)^\varepsilon} \\
&\leqslant \lim_{n\to\infty} e^{\ln b(n)\cdot k\tilde{r}(n) - a(n)^\varepsilon} = 0
\end{aligned}$$

By Markov's inequality, $P[X_r \geqslant 1] \leqslant E(X_r)$, so almost surely $\tilde{\gamma}(H) > k((1-\varepsilon)\log_{1/q} a(n))^{\frac{1}{k}}$ and thus $\tilde{\gamma}(H) \sim k \log_{1/q}^{\frac{1}{k}} a(n)$ almost surely. ■

PROOF. [of Theorem 4.1] By lemma 3.1 we can express the optimum of every $\mathcal{Q} \in \text{MIN } F^+\Pi_2(k)$ as:

$$\text{opt}_{\mathcal{Q}}(\mathfrak{A}) = \min_S\{|S| : (\mathfrak{A}, S) \models \forall \bar{x}[\gamma(\bar{x}) \rightarrow \exists \bar{z}_1 \ldots \exists \bar{z}_k(\varphi(\bar{x}, \bar{z}) \wedge \bigwedge_{i=1}^{k} S\bar{z}_i)]\}.$$

Analogue to the proof of theorem 3.2 we decompose the formula into atomic σ-types and consider the special case where

$$\text{opt}_{\mathcal{Q}}(\mathfrak{A}) = \min_S\{|S| : (\mathfrak{A}, S) \models \forall \bar{x}[u(\bar{x}) \rightarrow \exists \bar{z}_1 \ldots \exists \bar{z}_k t(\bar{x}, \bar{z}) \wedge \bigwedge_{i=1}^{k} S\bar{z}_i]\}$$

and $t(\bar{x}, \bar{z}) = t(\bar{x}, \bar{z}_1, \ldots, \bar{z}_k) = u(\bar{x}) \wedge v_1(\bar{z}_1) \wedge \ldots \wedge v_k(\bar{z}_k) \wedge \alpha(\bar{x}, \bar{z}_1, \ldots, \bar{z}_k) \wedge \beta(\bar{x}, \bar{z}_1, \ldots, \bar{z}_k)$ on almost all structures.

The optimum of $\mathcal{Q}$ is the cardinality of the smallest subset $S \subset \bigcup_{i=1}^{k} v_i^{\mathfrak{A}}$ such that for every $\bar{x} \in u^{\mathfrak{A}}$ there exist $\bar{z}_1, \ldots, \bar{z}_k \in S$ such that $t(\bar{x}, \bar{z}_1, \ldots, \bar{z}_k)$ becomes TRUE.

Similar to the proof of theorem 3.2 we reduce $\mathcal{Q}$ to the hyperdominating set problem. In the case that α contains only inequalities we can apply lemma 4.2 directly and we get that the optimum is $O(\log^{\frac{1}{k}} n)$ almost surely. In the other cases we have to split up the problem into subproblems of the desired form. ■

This theorem leads us to the following conclusion.

Proposition 4.3 MIN GRAPH COLOURING, MIN CLIQUE PARTITION *are not in* MIN $F^+\Pi_2$.

PROOF. By [1] the chromatic number $\chi(G)$ of a graph satisfies $\chi(G) \sim n/2\log n$ almost surely. MIN CLIQUE PARTITION is the same problem as MIN GRAPH COLOURING on the complementary graph. ■

With the same arguments as in [5] we can show that the criterion remains true if we use more powerful logics than first-order for example fixed point logic.

References

[1] N. Alon and J. Spencer. *The Probabilistic Method.* Wiley, 1991.

[2] Th. Behrendt, K. Compton, and E. Grädel. Optimization problems: Expressibility, approximation properties, and expected asymptotic growth of optimal solutions. In E. Börger, G. Jäger, H. Kleine Büning, S. Martini, and M.M. Richter, editors, *Computer Science Logic, 6th Workshop, CSL '92, San Miniato 1992, Selected Papers*, volume 702 of *LNCS*, pages 43–60. Springer-Verlag, 1993.

[3] P. Crescenzi and V. Kann. A compendium of NP optimization problems. unpublished, 1994.

[4] R. Fagin. Generalized first–order spectra and polynomial–time recognizable sets. In R. M. Karp, editor, *Complexity of Computation, SIAM-AMS Proceedings, Vol. 7*, pages 43–73, 1974.

[5] E. Grädel and A. Malmström. Approximable minimization problems and optimal solutions on random input. In E. Börger, Y. Gurevich, and K. Meinke, editors, *Computer Science Logic, 7th Workshop, CSL '93, Swansea 1993, Selected Papers*, volume 832 of *LNCS*, pages 139–149. Springer-Verlag, 1994.

[6] V. Kann. *On the Approximability of NP-complete Optimization Problems.* PhD thesis, Royal Institute of Technology, Stockholm, 1992.

[7] Ph. Kolaitis and M. Thakur. Logical definability of NP-optimization problems. Technical Report CRL-90-48, University of California, Santa Cruz, October 1990.

[8] Ph. Kolaitis and M. Thakur. Approximation properties of NP minimization classes. In *Proc. 6th IEEE Symp. on Structure in Complexity Theory*, pages 353–366, 1991.

[9] Ph. Kolaitis and M. Vardi. Infinitary logics and 0–1 laws. *Information and Computation*, 98:258–294, 1992.

[10] C. Lund and M. Yannakakis. On the hardness of approximating minimization problems. In *Proc. 25th ACM Symp. on Theory of Computing*, pages 286–293, New York, 1993. ACM.

[11] C. Papadimitriou and M. Yannakakis. Optimization, approximization and complexity classes. *Journal of Computer and System Sciences*, 43:425–440, 1991.