Completeness in the Polynomial-Time Hierarchy
A Compendium*

Marcus Schaefer
School of CTI
DePaul University
243 S. Wabash Avenue
Chicago, Illinois 60604, USA
schaefer@cs.depaul.edu

Christopher Umans
Computer Science Department
Caltech
1200 East California Boulevard
Pasadena, CA 91125
uumans@cs.caltech.edu

October 5, 2008

Abstract

We present a Garey/Johnson-style list of problems known to be complete for the second and higher levels of the polynomial-time Hierarchy (polynomial hierarchy, or **PH** for short). We also include the best-known hardness of approximation results. The list will be updated as necessary.

Updates

The compendium currently lists more than 80 problems. Latest changes include:

- added [GT26] SUCCINCT k-KING,
- added [GT25] SUCCINCT k-DIAMETER,
- added [GT4] SUCCINCT k-RADIUS at third level,
- added [GT24] MINIMUM VERTEX COLORING DEFINING SET,
- added [GT23] GRAPH SANDWICH PROBLEM FOR Π,
- added [L24] MINIMUM 3SAT DEFINING SET,
- added [L23] $\exists\exists_t!$ 3SAT,
- open problem MEE solved, now [L22],
- open problem THUE NUMBER solved, now [GT22],

- added open problem [O9] THUE CHROMATIC NUMBER,
- added open problem [O8] STRONG CHROMATIC NUMBER,
- added [L21] ∃∃!-3SAT,
- added [GT21] UNIQUE k-LIST COLORABILITY,
- added open problem [O7] THUE NUMBER,
- added [GT20] PEBBLING NUMBER,

1 Introduction

In this paper we have compiled a Garey/Johnson-style list of complete problems in the polynomial-time hierarchy, at the second level and above. For optimization problems, we also include any known hardness of approximation results. This list is based on a thorough, but not infallible, literature search. We should also point out that we have not verified all of the quoted results. We realize that the list is incomplete (and will in all likelihood remain so), but we are planning on regularly updating it, as further problems come to our attention.

Definitions relevant to specific problems are contained in the list below. We briefly review the definition of the polynomial hierarchy (**PH**). **PH** is defined recursively from the classes **P** and **NP** by:

$$\begin{aligned}\mathbf{\Sigma_0^p} &= \mathbf{\Pi_0^p} = \mathbf{P}\\ \mathbf{\Sigma_i^p} &= \mathbf{NP}^{\mathbf{\Sigma_{i-1}^p}}\\ \mathbf{\Pi_i^p} &= \mathbf{coNP}^{\mathbf{\Pi_{i-1}^p}}\end{aligned}$$

where $\mathbf{coNP} = \{\overline{L} : L \in \mathbf{NP}\}$.

In the next three sections we list problems complete for the second level of **PH**, problems complete for the third level of **PH**, and a selection of problems in **PH** whose complexity remains open. We should mention that there are natural problems complete for higher levels in nonclassical logics. Within each section the problems are categorized by area, and individual problems are labeled in Garey/Johnson style (e.g., GT3 for the third graph theory problem). We distinguish optimization problems by an asterisk at the beginning of their label.

2 The Second level

2.1 Logic

[L1] ∀∃**3SAT**

Given: Boolean formula $\varphi(x, y)$ in 3-CNF.

Question: Is it true that $(\forall x)(\exists y)\varphi(x,y)$?

Reference: Stockmeyer [75], Wrathall [88].

Comments: $\mathbf{\Pi_2^p}$-complete. Remains $\mathbf{\Pi_2^p}$-complete if φ is representable by a planar circuit (Gutner [28]). Stockmeyer and Wrathall showed that deciding QSAT_k, the set of true formulas with $k-1$ quantifier alternations beginning with an $\exists$ quantifier, is $\mathbf{\Sigma_k^p}$-complete. Earlier, Meyer and Stockmeyer [56] had shown that QUANTIFIED BOOLEAN FORMULAE, the problem of deciding the truth of quantified Boolean formulas (without restriction on the number of alternations), is **PSPACE**-complete. See MINMAX SAT for the optimization variant.

[L2] NOT-ALL-EQUAL∀∃3SAT

Given: 3-CNF formula $\varphi(x,y)$.

Question: Is it true that for every truth-assignment to x there is a truth-assignment to y such that each clause in $\varphi(x,y)$ contains both a true and a false literal?

Reference: Eiter, Gottlob [21].

Comments: $\mathbf{\Pi_2^p}$-complete.

[*L3] MONOTONE MINIMUM WEIGHT WORD

Given: A Π_1 nondeterministic circuit C that accepts a nonempty monotone set (although C may contain NOT gates) and an integer k. A Π_1 *nondeterministic circuit* is an ordinary Boolean circuit with two sets of inputs x and y. We say that C *accepts an input* x iff $(\forall y)C(x,y)=1$. A *monotone set* is a subset S for which $x \in S$ implies $x' \in S$ for all $x' \succeq x$, where $\succeq$ is the bitwise partial order on bitstrings.

Question: Does C accept an input x with at most k ones?

Reference: Umans [81].

Comments: $\mathbf{\Sigma_2^p}$-complete. Also $\mathbf{\Sigma_2^p}$-hard to approximate to within $n^{1-\epsilon}$, where n is the size of circuit C [81, 78]. The generalized version with m sets of inputs x and $y_1, y_2, \ldots, y_{m-1}$ in which C accepts an input x iff $(\forall y_1)(\exists y_2)(\forall y_3)\ldots C(x, y_1, y_2, \ldots y_{m-1})$ is $\mathbf{\Sigma_m^p}$-complete and $\mathbf{\Sigma_m^p}$-hard to approximate to within $n^{1-\epsilon}$ [83, 78]. Maximization version of MONOTONE MAXIMUM ZEROS.

[*L4] MONOTONE MAXIMUM ZEROS

Given: A Π_1 nondeterministic circuit C that accepts a nonempty monotone set (although C may contain NOT gates) and an integer k. See MONOTONE MINIMUM WEIGHT WORD above for the relevant definitions.

Question: Does C accept an input x with at least k zeros?

Reference: Umans [83].

Comments: $\mathbf{\Sigma_2^p}$-complete. Also $\mathbf{\Sigma_2^p}$-hard to approximate to within $n^{1/8-\epsilon}$, where n is the size of circuit C. The generalized version with m sets of inputs x and $y_1, y_2, \ldots, y_{m-1}$ in which C accepts an input x iff $(\forall y_1)(\exists y_2)(\forall y_3)\ldots C(x, y_1, y_2, \ldots y_{m-1})$ is $\mathbf{\Sigma_m^p}$-complete and $\mathbf{\Sigma_m^p}$-hard to approximate to within $n^{1/8-\epsilon}$. Minimization version of MONOTONE MINIMUM WEIGHT WORD.

[L5] GENERALIZED 3-CNF CONSISTENCY

Given: Two sets A and B of Boolean formulas.

Question: Is there a Boolean formula φ such that $\varphi \wedge \psi$ is satisfiable for all $\psi \in A$, and unsatisfiable for all $\psi \in B$?

Reference: Ko, Tzeng [44].

Comments: $\mathbf{\Sigma_2^p}$-complete. Similar in structure to PATTERN CONSISTENCY, and GRAPH CONSISTENCY.

[*L6] MIN DNF

Given: A DNF formula φ and an integer k. The *size* of a formula is the number of occurrences of literals in the formula.

Question: Is there a DNF formula ψ such that $\psi \equiv \varphi$ and ψ has size at most k?

Reference: Umans [84].

Comments: $\mathbf{\Sigma_2^p}$-complete. Also $\mathbf{\Sigma_2^p}$-hard to approximate to within $n^{1/4-\epsilon}$ (resp., $n^{1/3-\epsilon}$), where n is the size of φ (resp., n is the number of terms in φ) [81, 83, 78]. The variant in which the *size* is the number of terms is also $\mathbf{\Sigma_2^p}$-complete, and $\mathbf{\Sigma_2^p}$-hard to approximate to within the same factors. The problem is also known as $\mathsf{MEE_{DNF}}$, and MIN. If we drop the restriction to DNF formulas, we obtain MEE. The complexity of the variant MINIMAL is not known.

[*L7] IRREDUNDANT

Given: A DNF formula φ and an integer k.

Question: Is there a subset of at most k terms from φ whose disjunction is equivalent to φ?

Reference: Umans [81].

Comments: $\mathbf{\Sigma_2^p}$-complete. Also $\mathbf{\Sigma_2^p}$-hard to approximate to within $n^{1/4-\epsilon}$ (resp., $n^{1/3-\epsilon}$), where n is the number of occurrences of literals in φ (resp., n is the number of terms in φ) [81, 83, 78]. Minimization version of MAXIMUM TERM DELETION. The variant in which φ is a 3-DNF tautology is called MIN DNF TAUTOLOGY and remains $\mathbf{\Sigma_2^p}$-complete [24, 70], and $\mathbf{\Sigma_2^p}$-hard to approximate to within n^{ϵ} [70].

[*L8] MAXIMUM TERM DELETION

Given: A DNF formula φ and an integer k.

Question: Can one delete at least k terms from φ so that the remaining DNF is equivalent to φ?

Reference: Umans [83].

Comments: $\mathbf{\Sigma_2^p}$-complete. Also $\mathbf{\Sigma_2^p}$-hard to approximate to within n^ϵ for some constant $\epsilon > 0$, where n is the number of occurrences of literals in φ [83, 78]. Maximization version of IRREDUNDANT.

[*L9] SHORT CNF

Given: A DNF formula φ and an integer k in unary. The *size* of a formula is the number of occurrences of literals in the formula.

Question: Is there a CNF formula ψ such that $\psi \equiv \varphi$ and ψ has size at most k?

Reference: Schaefer, Umans [70].

Comments: $\mathbf{\Sigma_2^p}$-complete. Also $\mathbf{\Sigma_2^p}$-hard to approximate to within a factor n^ϵ, where n is the size of φ. The problem was proposed by Papadimitriou [61, Problem 17.3.12].

[*L10] SHORTEST IMPLICANT CORE

Given: A DNF formula φ, an implicant C of φ, and an integer k. An *implicant* of φ is a set of literals whose conjunction implies φ. The size of an implicant is its size as a set.

Question: Is there an implicant $C' \subseteq C$ of φ of size at most k?

Reference: Umans [84].

Comments: $\mathbf{\Sigma_2^p}$-complete. Also $\mathbf{\Sigma_2^p}$-hard to approximate to within $n^{1-\epsilon}$, where n is the number of occurrences of literals in φ [81, 78]. Minimization version of MAXIMUM LITERAL DELETION.

[*L11] MAXIMUM LITERAL DELETION

Given: A DNF formula φ, an implicant C of φ, and an integer k. See SHORTEST IMPLICANT CORE above for the relevant definitions.

Question: Is there a subset $D \subseteq C$ of size at least k for which $C' = C \setminus D$ is an implicant of φ?

Reference: Umans [83].

Comments: $\mathbf{\Sigma_2^p}$-complete. Also $\mathbf{\Sigma_2^p}$-hard to approximate to within n^ϵ for some constant $\epsilon > 0$, where n is the number of occurrences of literals in φ [83, 78]. Minimization version of SHORTEST IMPLICANT CORE.

[*L12] SHORTEST IMPLICANT

Given: A Boolean circuit φ, and an integer k.

Question: Is there an implicant C of φ of size at most k? See SHORTEST IMPLICANT CORE above for the relevant definitions.

Reference: Umans [84].

Comments: $\mathbf{\Sigma_2^p}$-complete. Also $\mathbf{\Sigma_2^p}$-hard to approximate to within $n^{1-\epsilon}$, where n is the number of occurrences of literals in φ. The variant in which φ is a Boolean formula remains $\mathbf{\Sigma_2^p}$-complete and $\mathbf{\Sigma_2^p}$-hard to approximate to within the same factor [82]. The variant in which φ is a DNF formula is complete for a class between **coNP** and $\mathbf{\Sigma_2^p}$ called $\mathbf{GC}(\log^2 n, \mathbf{coNP})$ [84], and $\mathbf{GC}(\log^2 n, \mathbf{coNP})$-hard to approximate to within an $(1/3 - \epsilon)\log n$ *additive* factor, where n is the number of terms in φ [83].

[L13] CIRCUIT RESTRICTION

Given: Two circuits C_1 and C_2 on the same set of variables V. Two circuits are *equivalent* if they compute the same truth-table on V. A *restriction* of a circuit is obtained by setting some of the variables to constant values in $\{0, 1\}$.

Question: Is C_1 a restriction of C_2?

Reference: Borchert, Ranjan [7].

Comments: $\mathbf{\Sigma_2^p}$-complete. Three other variants are also $\mathbf{\Sigma_2^p}$-complete: allowing variables to be renamed, allowing variables to be set and renamed, or replacing variables by literals [7]. If variables are renamed bijectively, the problem turns into BOOLEAN ISOMORPHISM which is likely to be intermediary between the first and second level of the hierarchy [1, 8].

[*L14] MINMAX SAT

Given: 3-CNF formula $\varphi(x, y)$ and integer k.

Question: For every truth-assignment to x, is there a truth-assignment to y making at least k clauses in $\varphi(x, y)$ true?

Reference: Meyer, Stockmeyer [56].

Comments: $\mathbf{\Pi_2^p}$-complete. Optimization version of $\forall\exists$3SAT. Let us call $f(\varphi)$ the largest k such that for every x there exists a y making at least k clauses in $\varphi(x, y)$ true. Then there is a $c > 0$ such that approximating $f(\varphi)$ to within a factor of c is $\mathbf{\Pi_2^p}$-hard. This follows from work on debate systems (generalizing the PCP characterization of **NP**) [13, 41] as pointed out in [42]. Ko and Lin [43] showed that the c-approximation problem remains $\mathbf{\Pi_2^p}$-hard if the number of occurrences of each variable is bounded by a constant B (MINMAX SAT B). This result is used in the proof that LONGEST DIRECTED CIRCUIT is $\mathbf{\Pi_2^p}$-complete. Havev, Regev, and Ta-Shma [33] showed that MINMAX SAT B remains $\mathbf{\Pi_2^p}$-complete, even if we know that in positive instances all clauses are true.

[L15] $\exists^*\forall^*$ **SATISFIABILITY IN FOL WITH ONE UNARY FUNCTION**

Given: A first-order formula φ whose quantifier part is of the form $\exists^*\forall^*$, where φ may contain equality and one unary function, but no other constant, function, or relation symbols.

Question: Is there a model for φ?

Reference: Börger, Grädel, Gurevich [9, Theorem 6.4.19].

Comments: $\mathbf{\Sigma_2^p}$-complete; the harder part being membership in $\mathbf{\Sigma_2^p}$. This is a special case of Ramsey's decidability result of the satisfiability problem for $\exists^*\forall^*$ formulas with equality, but no other relation symbols (which is **NEXP**-complete). The following variants of the satisfiability problem are also $\mathbf{\Sigma_2^p}$-complete: quantifier part of the form $\exists^*\forall^*$, and no relation or function symbols except for equality; quantifier part of the form $\exists^*\forall^*$, and at most one unary relation (no function symbols, no equality); quantifier part of the form $\exists^2\forall^*$, and relations of arbitrary arity (no functions, no equality). See [9, Theorem 6.4.7]. Also, see $\exists^*\forall^*$ CNF SATISFIABILITY WITH EQUALITY.

[L16] $\exists^*\forall^*$ **CNF SATISFIABILITY WITH EQUALITY**

Given: A first-order formula φ whose quantifier part is of the form $\exists^*\forall^*$, and whose quantifier-free part is in 3-CNF and may contain equality, but no function, or relation symbols.

Question: Is there a model for φ of cardinality three?

Reference: Pichler [63].

Comments: $\mathbf{\Sigma_2^p}$-complete. Remains $\mathbf{\Sigma_2^p}$-complete if cardinality is any fixed integer at least three. See $\exists^*\forall^*$ SATISFIABILITY IN FOL WITH ONE UNARY FUNCTION.

[L17] CONSTRAINTS OVER PARTIALLY SPECIFIED FUNCTIONS

Given: A set of partially specified Boolean functions $f_1, \ldots, f_n$, and a Boolean formula φ over $f_1, \ldots, f_n$. A *partially specified Boolean function* f is a circuit with three output values: 1, 0, and d (for "don't care").

Question: Can the "don't care" values in $f_1, \ldots, f_n$ be set to 0 and 1 such that φ, when interpreted over the resulting Boolean functions, is always true?

Reference: Sriram, Tandon, Dasgupta, Chakrabarti [73].

Comments: $\mathbf{\Pi_2^p}$-complete.

[L18] $\exists\forall\exists\exists$ **PRESBURGER ARITHMETIC**

Given: A first-order formula φ of Presburger arithmetic, that is, allowing addition and equality, whose quantifier part is of the form $\exists\forall\exists\exists$.

Question: Is φ true in the natural numbers?

Reference: Schöning [72].

Comments: $\mathbf{\Sigma_2^p}$-complete. Truth in Presburger Arithmetic of formulas with prefix $\exists_1\forall_2\ldots\forall_m\exists^3$ is $\mathbf{\Sigma_m^p}$-complete if m is even, and the truth of formulas with prefix $\exists_1\forall_2\ldots\exists_m\exists^3$ is $\mathbf{\Sigma_m^p}$-complete if m is odd. The $\exists\forall$ case is **NP**-complete.

[L19] GRAPH SATISFIABILITY

Given: 3-CNF formula φ. With a formula φ we associate a graph $G(\varphi)$ on the variables and clauses of φ with an edge between a variable and a clause, if the variable occurs in the clause (positively, or negatively). We call φ *graph-satisfiable* if every ψ with $G(\varphi) = G(\psi)$ is satisfiable (i.e. the satisfiability of φ only depends on the graph $G(\varphi)$).

Question: Is φ graph satisfiable?

Reference: Szeider [76, 77].

Comments: $\mathbf{\Pi_2^p}$-complete. For 2-CNF formulas graph satisfiability can be recognized in linear time. Reduction from 2-COLORING EXTENSION.

[L20] ARGUMENT COHERENCE

Given: Digraph (without self-loops) $H = (X, A)$, called an *argument system*. X is the set of *arguments*, and A the set of *attacks*; we say x *attacks* y if $(x, y) \in A$. An argument $x \in X$ is *attacked* by $S \subseteq X$ if $(y, x) \in A$ for some $y \in S$. A set of arguments S is *conflict-free* if no argument in S is attacked by S. An argument $x \in X$ is *acceptable with respect to* S if for every $y \in X$ that attacks x there is a $z \in S$ that attacks y. A set of arguments S is *admissible* if every argument in S is acceptable with respect to S. A *preferred extension* is a maximal admissible set. A *stable extension* S is a conflict free set that attacks every argument in $\overline{S}$. H is *coherent* if every preferred extension is stable.

Question: Is H coherent?

Reference: Dunne, Bench-Capon [17].

Comments: $\mathbf{\Pi_2^p}$-complete. The proof also shows that the question of whether a given argument occurs in every preferred extension is $\mathbf{\Pi_2^p}$-complete as well.

[L21] $\exists\exists!$-3SAT

Given: 3-CNF formula φ. "$\exists!$" is interpreted as "there is exactly one".

Question: Is $\exists x \exists! y \varphi(x, y)$ true?

Reference: Marx [52].

Comments: $\mathbf{\Sigma_2^p}$-complete. Used to show UNIQUE k-LIST COLORABILITY $\mathbf{\Sigma_2^p}$-complete.

[*L22] MINIMUM EQUIVALENT EXPRESSION

Given: A well-formed Boolean formula φ, integer k. The *size* $|\varphi|$ of a formula is the number of occurrences of literals in the formula.

Question: Is there a well-formed Boolean formula ψ for which $\psi \equiv \varphi$, and $|\psi| < k$?

Reference: Buchfuhrer, Umans [10]. Mentioned as an open problem in Garey, Johnson [25].

Comments: $\mathbf{\Sigma_2^p}$-complete under Turing-reductions [10] if all Boolean formulas are over signature $\{\vee, \wedge, \neg\}$; trivially hard for **coNP**, and hard for $\mathbf{P}_{||}^{\mathbf{NP}}$ (**P** with parallel access to **NP**) as shown by Hemaspaandra and Wechsung [35]. MEE_d, the problem restricted to $\{\vee, \wedge, \neg\}$-Boolean formulas of depth at most d is also $\mathbf{\Sigma_2^p}$-complete under Turing reductions for any fixed $d \geq 3$ [10]. Completeness under many-one reductions of MEE and MEE_d is open, as is the original version suggested by Garey, Johnson with Boolean formulas over signature $\{\vee, \wedge, \neg, \rightarrow\}$. Restricted to DNF formulas, the problem is MIN DNF, which is $\mathbf{\Sigma_2^p}$-complete. Also see MINIMAL.

[L23] $\exists\exists!_t$-3SAT

Given: 3-CNF formula $\varphi(x, y)$ with a proper partial assignment over y. A *partial assignment over S* assigns truth-values to a subset of the S-variables. It is *proper* if every clause in φ contains a true literal. An *assignment* assigns truth-values to all variables in the formula. It *respects* a partial assignment, if it agrees with the truth-values of the partial assignment.

Question: Is $\exists x \exists!_t y \varphi(x, y)$ true? That is, is there a partial assignment t' over x so that there is a unique proper assignment of φ which respects t'?

Reference: Hatami, Maserrat [31].

Comments: $\mathbf{\Sigma_2^p}$-complete. Used to show MINIMUM 3SAT DEFINING SET $\mathbf{\Sigma_2^p}$-complete.

[L24] MINIMUM 3SAT DEFINING SET

Given: 3-CNF formula φ, integer k. A *defining set* is a partial assignment of truth-values to variables of φ which has a unique extension to a satisfying assignment of φ. The *size* of a defining set is the number of variables that are assigned truth-values.

Question: Does φ have a defining set of size at most k?

Reference: Hatami, Maserrat [31].

Comments: $\mathbf{\Sigma_2^p}$-complete. Reduction from $\exists\exists_t!$ 3SAT. Used to show MINIMUM VERTEX COLORING DEFINING SET $\mathbf{\Sigma_2^p}$-complete.

2.2 Graph Theory

[GT1] GRAPH CONSISTENCY

Given: Two sets A and B of (finite) graphs.

Question: Is there a graph G such that every graph in A is isomorphic to a subgraph of G, but no graph in B is isomorphic to a subgraph of G?

Reference: Ko, Tzeng [44] (GRAPH RECONSTRUCTION).

Comments: $\mathbf{\Sigma_2^p}$-complete. Similar in structure to PATTERN CONSISTENCY, and GENERALIZED 3-CNF CONSISTENCY.

[*GT2] MINMAX CLIQUE

Given: Graph $G = (V, E)$, a partition $(V_{i,j})_{i \in I, j \in J}$ of V, integer k. For a function $t : I \rightarrow J$ let f_t be the size of the largest clique in G restricted to $\bigcup_{i \in I} V_{i,t(i)}$.

Question: Is $\min_{t \in J^I} f_t(G) \geq k$?

Reference: Ko, Lin [42].

Comments: $\mathbf{\Pi_2^p}$-complete. There is a $c > 0$ such that approximating $f_t(G)$ to within a factor c is $\mathbf{\Pi_2^p}$-hard. Also see MAXMIN VERTEX COVER.

[*GT3] MINMAX CIRCUIT

Given: Graph $G = (V, E)$, a partition $(V_{i,j})_{i \in I, j \in J}$ of V, integer k. For a function $t : I \to J$ let f_t be the length of the longest cycle in G restricted to $\bigcup_{i \in I} V_{i,t(i)}$.

Question: Is $\min_{t \in J^I} f_t(G) \geq k$?

Reference: Ko, Lin [42].

Comments: $\mathbf{\Pi_2^p}$-complete. It is not known whether the c-approximation version of this problem remains $\mathbf{\Pi_2^p}$-complete.

[GT4] DYNAMIC HAMILTONIAN CIRCUIT

Given: Graph $G = (V, E)$, subset B of E. For a subset D of E, define $G_D = (V, E - D)$.

Question: Is it true that for all $D \subseteq B$ with $|D| \leq |B|/2$, G_D has a Hamilton cycle.

Reference: Ko, Lin [42].

Comments: $\mathbf{\Pi_2^p}$-complete.

[*GT5] LONGEST DIRECTED CIRCUIT

Given: Directed graph $G = (V, E)$, and a subset E' of E of *alterable* edges, integer k. For $D \subseteq E$ let G_D be the graph obtained from G by substituting each edge (u, v) in D by its reverse edge (v, u). Define f_D to be the length of the longest cycle in G_D.

Question: Is $l(G) = \min_{D \subseteq E'} f_D \geq k$?

Reference: Ko, Lin [43].

Comments: $\mathbf{\Pi_2^p}$-complete. There is a constant $c > 0$ such that approximating $l(G)$ to within a factor of c is $\mathbf{\Pi_2^p}$-hard.

[GT6] SUCCINCT TOURNAMENT REACHABILITY

Given: Circuit C representing a tournament graph $G = (V, E)$ (i.e., $C(u, v) = 1$ if and only if $(u, v) \in E$), and two vertices s, t. A *tournament graph* has exactly one edge between each pair of vertices.

Question: Is t reachable from s in G?

Reference: Nickelsen, Tantau [79, 60].

Comments: $\mathbf{\Pi_2^p}$-complete. The more interesting part is showing that the problem lies in $\mathbf{\Pi_2^p}$. Remains in $\mathbf{\Pi_2^p}$ for graphs of bounded independence number (instead of tournaments); a generalization of this variant lies in $\mathbf{\Pi_3^p}$, but is not known to be complete. The variant of the tournament problem in which G must be strongly connected is also $\mathbf{\Pi_2^p}$-complete.

[*GT7] SUCCINCT TOURNAMENT DOMINATING SET

Given: Circuit C representing a tournament graph $G = (V, E)$ (i.e., $C(u, v) = 1$ if and only if $(u, v) \in E$), and an integer k. A *tournament graph* has exactly one edge between each pair of vertices.

Question: Does G have a dominating set of size at most k? A *dominating set* is a subset $V' \subseteq V$ such that every vertex is reachable in zero or one steps from V'.

Reference: Umans [83].

Comments: $\mathbf{\Sigma_2^p}$-complete. Also $\mathbf{\Sigma_2^p}$-hard to approximate to within $n^{1/2-\epsilon}$, where n is the size of the circuit C [83, 78]. The nonsuccinct version is considered in [62].

[GT8] 3-COLORING EXTENSION

Given: Graph G.

Question: Can any 3-coloring of the leaves of G be extended to a 3-coloring of all of G?

Reference: Ajtai, Fagin, Stockmeyer [2].

Comments: $\mathbf{\Pi_2^p}$-complete, even if G has maximum degree at most 4. The general version of the problem has two players alternating in k rounds with vertices of degree i being colored in round $i < k$, and all remaining vertices colored in round k. This last player wins, if he can complete a legal coloring. This problem is $\mathbf{\Sigma_k^p}$-complete if k is odd, and $\mathbf{\Pi_2^p}$-complete if k is even, even if the graph has maximum degree at most $\max\{k, 4\}$. Also see 2-COLORING EXTENSION.

[GT9] GENERALIZED GRAPH COLORING

Given: Graphs F, G.

Question: Is there a two-coloring of the vertices of F which does not contain a monochromatic G as a subgraph?

Reference: Rutenburg [64].

Comments: $\mathbf{\Sigma_2^p}$-complete even if G is restricted to be complete. The completeness proof also works for other **coNP**-complete families of graphs, see, for example, the GENERALIZED NODE DELETION problem. For edge colorings compare to ARROWING and STRONG ARROWING.

[*GT10] GENERALIZED NODE DELETION

Given: Graphs F, G, integer k.

Question: Can we remove at most k vertices from F such that the resulting graph does not contain G as a subgraph?

Reference: $\mathbf{\Sigma_2^p}$-completeness is claimed in Rutenburg [64] without proof.

Comments: $\mathbf{\Sigma_2^p}$-complete even if G is restricted to be complete. No nonapproximability results are known.

[GT11] GENERALIZED RAMSEY NUMBER

Given: Graph F, a partial two-coloring of the edges of F, integer k.

Question: Does every two-coloring of F which extends the given two-coloring contain a clique on k vertices.

Reference: Ko, Lin [42]. A proof can also be found in [16].

Comments: $\mathbf{\Pi_2^p}$-complete. See also ARROWING.

[GT12] ARROWING

Given: Graphs F, G, and H.

Question: Does $F \to (G, H)$, i.e., does every edge-coloring of F with colors red and green contain either a red G, or a green H as a subgraph?

Reference: Schaefer [67].

Comments: $\mathbf{\Pi_2^p}$-complete even if G is a fixed tree on at least three vertices, and H a complete graph. The problem is **coNP**-complete for fixed three-connected graphs G and H [12]. If F is a complete graph, then the problem is **NP**-hard [11], but not known to be $\mathbf{\Pi_2^p}$-complete. $K_n \to (K_m, K_\ell)$ is unlikely to be $\mathbf{\Pi_2^p}$-complete, since it lies in $\mathbf{coNP}^{\mathsf{LOGCLIQUE}}$, where LOGCLIQUE is the problem of deciding whether a graph F has a clique of size at least $\log |F|$. This version is particularly interesting since it corresponds to computing Ramsey numbers. Also see STRONG ARROWING, and GENERALIZED RAMSEY NUMBER. The vertex-coloring version of this problem is called GENERALIZED GRAPH COLORING.

[GT13] STRONG ARROWING

Given: Graphs F, G, and H.

Question: Does $F \rightarrowtail (G, H)$, i.e. does every edge-coloring of F with colors red and green contain either a red G, or a green H as an induced subgraph of F?

Reference: Schaefer [67].

Comments: $\mathbf{\Pi_2^p}$-complete if G is a fixed star $K_{1,p}$ ($p \geq 2$), and H a complete graph, or $G = H = K_{1,n}$ (the diagonal case). The noninduced version $F \to (K_{1,n}, K_{1,m})$ is in **P** [12]. Also see ARROWING.

[GT14] 2-COLORING EXTENSION

Given: Graph G, set of vertices S.

Question: Can any 2-coloring of S be extended to a 3-coloring of G?

Reference: Szeider [76].

Comments: $\mathbf{\Pi_2^p}$-complete. Reduction from NAE∀∃3SAT. Used to show GRAPH SATISFIABILITY $\mathbf{\Pi_2^p}$-complete. Also see 3-COLORING EXTENSION.

[GT15] BIPARTITE GRAPH (2, 3)-CHOOSABILITY

Given: Bipartite graph G, function $f : V \to \{2,3\}$. G is called *f-choosable*, if for every assignment of $f(v)$ colors to each node v, one color can be chosen for each node to obtain a *proper coloring*; that is, a coloring in which adjacent vertices have different colors.

Question: Is G f-choosable?

Reference: Attributed to Rubin in Erdős, Rubin, Taylor [22].

Comments: $\mathbf{\Pi_2^p}$-complete. Remains $\mathbf{\Pi_2^p}$-complete if G is restricted to be planar (Gutner [28]). Also see LIST CHROMATIC NUMBER.

[*GT16] LIST CHROMATIC NUMBER

Given: Graph G, integer k. G is called *k-choosable*, if for every assignment of k colors to every node, one color can be chosen for each node to obtain a *proper coloring*; that is, a coloring in which adjacent vertices have different colors. The *list chromatic number*, $\chi_\ell(G)$, also known as the *choice number* of G is the smallest k such that G is k-choosable.

Question: Is $\chi_\ell(G) \leq k$?

Reference: Gutner, Tarsi [29].

Comments: $\mathbf{\Pi_2^p}$-complete for any fixed $k \geq 3$. Reduction from BIPARTITE GRAPH $(2,3)$-CHOOSABILITY. Remains $\mathbf{\Pi_2^p}$-complete if G is bipartite. For $k = 2$, the problem is solvable in polynomial time using a result of Erdős, Rubin, Taylor [22]. Gutner [28] shows that the following planar versions of the problem remain $\mathbf{\Pi_2^p}$-complete: determining whether a planar triangle-free graph is 3-choosable, determining whether a planar graph is 4-choosable, determining whether a union of two forests (on a shared vertex set) is 3-choosable. Also see BIPARTITE GRAPH $(2,3)$-CHOOSABILITY and UNIQUE k-LIST COLORABILITY.

[*GT17] GROUP CHROMATIC NUMBER

Given: Graph $G = (V,E)$, integer k. For a fixed Abelian group A, G is said to be *A-colorable* if for every orientation of the edges of G, and every *edge-labelling* $\phi : E \to A$, there is a vertex-coloring $c : V \to A$, such that $\phi(u,v) \neq c(u) - c(v)$ for all directed edges (u,v) of G. The *group chromatic number* $\chi_g(G)$ is the smallest number ℓ such that G is A-colorable for all Abelian groups of order at least ℓ.

Question: Is $\chi_g(G) \leq k$?

Reference: Král' [46]. Also in Král' and Nejedlý [47].

Comments: $\mathbf{\Pi_2^p}$-complete for any fixed $k \geq 3$. Also see GROUP CHOOSABILITY.

[GT18] GROUP CHOOSABILITY

Given: Graph $G = (V,E)$, integer ℓ. For a fixed Abelian group A, G is said to be *A-ℓ-choosable* if for every orientation of the edges of G, every list assignment $L : V \to \binom{A}{\ell}$, and every *edge-labelling* $\phi : E \to A$, there is a vertex-coloring $c : V \to A$ with $c(u) \in L(u)$, such that $\phi(u,v) \neq c(u) - c(v)$ for all directed edges (u,v) of G.

Question: Is G A-ℓ-choosable?

Reference: Král' and Nejedlý [47].

Comments: $\mathbf{\Pi_2^p}$-complete for any fixed group A of order at least 3 and any fixed $\ell \geq 3$. In particular, it is $\mathbf{\Pi_2^p}$-complete to decide whether G is A-colorable (also in [46]). The problem becomes polynomial-time solvable if $\ell \leq 2$. GROUP CHOOSABILITY generalizes LIST CHROMATIC NUMBER. Also see the closely related GROUP CHROMATIC NUMBER.

[*GT19] CLIQUE COLORING

Given: Graph $G = (V, E)$, integer k. A *k-clique-coloring* is a function $c : V \to \{1, \ldots, k\}$ such that every maximal clique of G contains two vertices of different color.

Question: Does G have a k-clique-coloring?

Reference: Marx [51].

Comments: $\mathbf{\Sigma_2^p}$-complete for any fixed $k \geq 2$. A k-clique-coloring of G is not necessarily a k-clique-coloring of the subgraphs of G. The variant HEREDITARY CLIQUE COLORING, in which the graph and all its induced subgraphs are required to be k-clique colorable turns out to be $\mathbf{\Pi_3^p}$-complete. CLIQUE CHOOSABILITY is another $\mathbf{\Pi_3^p}$-complete variant.

[*GT20] PEBBLING NUMBER

Given: Graph $G = (V, E)$, integer k. Vertices of the graph can contain pebbles. A *pebbling move* along an edge $uv \in E$ removes two pebbles from u and adds one pebble to v. The *pebbling number* $\pi(G)$ is the smallest number k of pebbles such that for all distributions of k pebbles on G and for all target vertices $v \in V$ there is a sequence of pebbling moves that places a pebble on v.

Question: Is $\pi(G) \leq k$?

Reference: Milans, Clark [57].

Comments: $\mathbf{\Pi_2^p}$-complete. Remains $\mathbf{\Pi_2^p}$-complete for a single target vertex which is part of the input. Determining the *optimal pebbling number*, $\hat{\pi}(G)$, the smallest number k of pebbles such that there is a distribution of k pebbles on G such that for every target vertex $v \in V$ there is a sequence of pebbling moves that places a pebble on v, is **NP**-complete. The complexity of deciding $\pi(G) = |V|$ remains open (note that $\pi(G) \geq |V|$).

[*GT21] UNIQUE k-LIST COLORABILITY

Given: Graph $G = (V, E)$, integer k. A *k-list coloring* L assigns k colors to each node of G. The graph is *L-colorable* if there is a proper coloring of the graph such that every vertex v is assigned a color from its list $L(v)$. A graph is *k-list colorable* (or *k-choosable*) if there is a k-list coloring L such that G is L-colorable. A graph is *uniquely k-list colorable* if there is a k-list coloring L such that there is exactly one L-coloring of G.

Question: Is G uniquely k-list colorable?

Reference: Marx [52].

Comments: $\mathbf{\Sigma_2^p}$-complete. Reduction from ∃∃!-3SAT. Remains $\mathbf{\Sigma_2^p}$-complete for $k = 3$ or if each of the lists contains 2 or 3 elements. Can be decided in polynomial time for $k = 2$ (Mahdian and Mahmoodian, see [52]). Also, see LIST CHROMATIC NUMBER.

[*GT22] THUE NUMBER

Given: A graph $G = (V, E)$, integer k. A word w is *square-free* (or *non-repetitive*) if there are no u, v, w such that $w = uvvw$ (with v not the empty word). A *non-repetitive* k-edge coloring of G is a k-edge coloring of G such that for any path in G, the sequence of colors along the path is square-free. The smallest k such that G has a non-repetitive k-edge coloring is called the *Thue number* of G.

Question: Is the Thue number of G at most k?

Reference: Manin [50].

Comments: $\mathbf{\Sigma_2^p}$-complete. Deciding whether a given edge coloring is non-repetitive is **coNP**-complete. If we only have to avoid non-repetitive sequences up to a certain length, the problem is **NP**-complete. Thue number was first defined in Alon, Grytczuk, Hauszczak, Riordan [4]. Named after Axel Thue who proved that there are infinite square-free words. Also see THUE CHROMATIC NUMBER (open problems).

[*GT23] GRAPH SANDWICH PROBLEM FOR Π

Given: Graphs F, F' so that $F \subseteq F'$.

Question: Is there a graph G satisfying Π so that $F \subseteq G \subseteq F'$?

Reference: Schaefer [69].

Comments: $\mathbf{\Sigma_2^p}$-complete for the property of being P_k-free where $k = \Theta(|V(G)|^{1/2})$. Open whether there is a natural property Π, such as being well-covered, for which problem is $\mathbf{\Sigma_2^p}$-complete.

[GT24] MINIMUM VERTEX COLORING DEFINING SET

Given: Graph G, integer k. A *defining set* for a vertex coloring is a partial vertex coloring which has a unique extension to a legal vertex coloring of G. The *size* of a defining set is the number of vertices colored.

Question: Does G have a vertex coloring defining set of size at most k?

Reference: Hatami, Maserrat [31].

Comments: $\mathbf{\Sigma_2^p}$-complete for vertex 3-colorings. Reduction from MINIMUM 3SAT DEFINING SET. For a discussion on the relationship to the forcing chromatic number, see [30].

[*GT25] SUCCINCT k-DIAMETER

Given: Circuit C representing a directed graph $G = (V, E)$ (i.e., $C(u, v) = 1$ if and only if $(u, v) \in E$). The *diameter* of a directed graph is the largest distance between any two vertices of the graph. The *distance* between two vertices is the length of a smallest directed path between the vertices.

Question: Does G have diameter at most k?

Reference: Hemaspaandra, Hemaspaandra, Tantau, Watanabe [34].

Comments: $\mathbf{\Sigma_2^p}$-complete for any fixed $k \geq 2$. Remains $\mathbf{\Sigma_2^p}$-complete for tournaments (directed graphs for which there is exactly one edge between any two vertices) and undirected graphs [80]. Also see SUCCINCT k-DIAMETER and SUCCINCT k-RADIUS.

[*GT26] SUCCINCT k-KING

Given: Circuit C representing a directed graph $G = (V, E)$ (i.e., $C(u, v) = 1$ if and only if $(u, v) \in E$), integer k. A vertex is a *k-king* is every vertex in the graph can be reached by a directed path of length at most k.

Question: Does G contain a k-king?

Reference: Hemaspaandra, Hemaspaandra, Tantau, Watanabe [34].

Comments: $\mathbf{\Pi_2^p}$-complete for any fixed $k \geq 2$. Remains $\mathbf{\Pi_2^p}$-complete for tournaments (directed graphs for which there is exactly one edge between any two vertices). Also see SUCCINCT k-KING and SUCCINCT k-DIAMETER.

2.3 Sets and Partitions

[*SP1] SUCCINCT SET COVER

Given: A collection $S = \{\varphi_1, \varphi_2, \ldots, \varphi_m\}$ of 3-DNF formulas on n variables, and an integer k.

Question: Is there a subset $S' \subseteq S$ of size at most k for which $\vee_{\varphi \in S'} \varphi \equiv 1$?

Reference: Umans [81].

Comments: $\mathbf{\Sigma_2^p}$-complete. Also $\mathbf{\Sigma_2^p}$-hard to approximate to within $n^{1-\epsilon}$, where n is the number of occurrences of literals in $\varphi_1, \varphi_2, \ldots, \varphi_m$ [81, 78]. The restriction in which all the ϕ_i except ϕ_1 are single literals, and ϕ_1 evaluates to 1 on at least 1/2 of the domain remains $\mathbf{\Sigma_2^p}$-complete and $\mathbf{\Sigma_2^p}$-hard to approximate to within the same factor. This restriction can be seen as a succinct version of RICH HYPERGRAPH COVER [83], whose complexity was considered in [62].

[SP2] GENERALIZED SUBSET SUM

Given: Two vectors u and v of integers, and an integer t.

Question: Is $(\exists x)(\forall y)[ux + vy \neq t]$ true, where the variables x and y are binary vectors of the same length as u and v?

Reference: Berman, Karpinski, Larmore, Plandowski, Rytter [6].

Comments: $\mathbf{\Sigma_2^p}$-complete. Used to show FULLY COMPRESSED TWO-DIMENSIONAL PATTERN MATCHING $\mathbf{\Pi_2^p}$-complete.

[*SP3] MAXMIN VERTEX COVER

Given: Graph $G = (V, E)$, a partition $(V_{i,j})_{i \in I, j \in J}$ of V, integer k. For a function $t : I \to J$ let f_t be the size of a smallest vertex cover of G restricted to $\bigcup_{i \in I} V_{i,t(i)}$.

Question: Is $\max_{t \in J^I} f_t(G) \leq k$?

Reference: Ko, Lin [42].

Comments: $\mathbf{\Pi_2^p}$-completeness follows from $\mathbf{\Pi_2^p}$-completeness of MINMAX CLIQUE using the standard transformation between vertex covers and cliques. The nonapproximability result for MINMAX CLIQUE does not carry over, and no nonapproximability results are currently known.

[*SP4] MINMAX THREE DIMENSIONAL MATCHING

Given: Set W, partition $(W_{i,j})_{i \in I, j \in J}$ of W, set S of three-element subsets of W, and an integer k. Call a set $S' \subseteq S$ a *matching in* $W' \subseteq W$, if the sets in S' are mutually disjoint subsets of W'. For a function $t : I \to J$ let $f_t(W)$ be the size $|S'|$ of a largest matching S' in $\bigcup_{i \in I} W_{i,t(i)}$.

Question: Is $\min_{t \in J^I} f_t(W) \geq k$?

Reference: Ko, Lin [42].

Comments: $\mathbf{\Pi_2^p}$-complete; reduction from MINMAX SAT YB. There is a $c > 0$ such that approximating $\min_{t \in J^I} f_t(W)$ to within a factor c is $\mathbf{\Pi_2^p}$-hard.

[SP5] ∀∃ THREE DIMENSIONAL MATCHING

Given: Three disjoint sets X_1, X_2, X_3 of the same cardinality, and two disjoint subsets M_1 and M_2 of $X_1 \times X_2 \times X_3$. A *matching* of X_1, X_2, X_3 is a set $S \subseteq X_1 \times X_2 \times X_3$ of size $|X_1| = |X_2| = |X_3|$ such that the components of the elements of S contain all the elements of $X_1 \cup X_2 \cup X_3$.

Question: For any subset S_1 of M_1, is there a subset S_2 of M_2 such that $S_1 \cup S_2$ is a matching?

Reference: McLoughlin [55].

Comments: Used to show $\mathbf{\Pi_2^p}$-completeness of COVERING RADIUS. A gap version of this problem remains $\mathbf{\Pi_2^p}$-hard which implies $\mathbf{\Pi_2^p}$-hardness of approximation for COVERING RADIUS [27].

2.4 Algebra and Number Theory

[AN1] INTEGER EXPRESSION INEQUIVALENCE

Given: Two integer expressions e_1, and e_2 built from binary numbers with operators $+$, and $\cup$. For an integer expression e define $L(e) = \{n\}$ if e is the binary representation of n, $L(e + f) = \{n + m : n \in L(e), m \in L(f)\}$, and $L(e \cup f) = L(e) \cup L(f)$.

Question: Is $L(e_1) \neq L(e_2)$?

Reference: Stockmeyer, Meyer [25]. The result appears in a 1973 conference paper by Stockmeyer and Meyer, and a 1976 paper by Stockmeyer.

Comments: $\mathbf{\Sigma_2^p}$-complete. Probably the first natural problem to be shown $\mathbf{\Sigma_2^p}$-complete. The subset problem $L(e_1) \subseteq L(e_2)$ is $\mathbf{\Pi_2^p}$-complete. The same is true for expressions represented in the general hierarchy input language (GHIL) which according to Wagner [87] was shown by Bentley, Ottmann, and Widmayer (1983). Huynh [39] observes that his result that 1 LETTER TERMINAL ALPHABET GRAMMAR INEQUIVALENCE is $\mathbf{\Sigma_2^p}$-complete implies that deciding the inequivalence of integer expressions over a unary alphabet with operations $\cup$, $\cdot$, ^{2}, and * is also $\mathbf{\Sigma_2^p}$-complete. See INTEGER EXPRESSION CONNECTEDNESS, and INTEGER EXPRESSION COMPONENT LENGTH.

[AN2] INTEGER EXPRESSION CONNECTEDNESS

Given: An integer expression e built from binary numbers with operators $+$, and $\cup$. See INTEGER EXPRESSION INEQUIVALENCEabove for the definition of an integer expression. A set of integers S is called *connected* if for every $x, z \in S$ and any y, if $x < y < z$ then $y \in S$.

Question: Is $L(e)$ connected?

Reference: Wagner [87].

Comments: $\mathbf{\Pi_2^p}$-complete. The result also holds if the input is specified using the general hierarchic input language (GHIL).

[*AN3] BOOLEAN EXPRESSION COMPONENT LENGTH

Given: A Boolean formula φ, integer k. If φ has n Boolean input variables $x_1, \dots, x_n$ we let $L(\varphi) = \{x_1 \cdots x_n : \varphi(x_1, \dots, x_n)\}$ interpreting the binary vector as a natural number. A set of numbers L is called *connected* if for every $x, z \in L$ and any y, if $x < y < z$ then $y \in L$. A maximal connected subset of a set is called a *component*.

Question: Does $L(\varphi)$ have a component of size at least k?

Reference: Wagner [87].

Comments: $\mathbf{\Sigma_2^p}$-complete. For integer expressions the problem is $\mathbf{\Sigma_3^p}$-complete (INTEGER EXPRESSION COMPONENT LENGTH). No nonapproximability results are known.

[AN4] BOUNDED EIGENVECTOR

Given: $n \times n$ integer matrix M, eigenvalue λ of M, subset $I \subseteq \{1, \dots, n\}$, rational number y.

Question: Is there an eigenvector $\mathbf{x} = (x_1, \dots, x_n)$ (for λ) such that $x_1 = y$, $|x_i| \leq c$ (for some fixed c), and $\mathbf{x}$ has maximal ℓ_2-norm among vectors identical to $\mathbf{x}$ on I?

Reference: Eiter, Gottlob [21].

Comments: $\mathbf{\Sigma_2^p}$-complete for any fixed $c \geq 1$, and $y = 0$.

[AN5] SEMILINEAR SET EQUIVALENCE

Given: Finite sets $C_i, P_i, C'_i, P'_i \subseteq \mathbb{N}^k$ $(1 \leq i \leq n)$. Let $L(C, P) = \{c + \sum_{p \in P} \lambda_p p : c \in C, p \in P, \lambda \in \mathbb{N}\}$, and $SL(C_1, \dots, C_n; P_1, \dots P_n) = \bigcup_{i=1}^n L(C_i, P_i)$. Sets of the form L are called *linear*, sets of the form SL *semilinear*.

Question: Is $SL(C_1, \ldots, C_n; P_1, \ldots P_n) = SL(C'_1, \ldots, C'_n; P'_1, \ldots P'_n)$?

Reference: Huynh [40].

Comments: $\mathbf{\Pi_2^p}$-complete, even for $k = 1$.

2.5 Automata and Languages

[AL1] PATTERN CONSISTENCY

Given: Two sets A and B of strings over $\{0, 1\}$. A *pattern* is a string over $\{0, 1\}$ and a set of variables. The language $L(p)$ associated with a pattern p is the set of strings that can be obtained from p by substituting all variables in p by strings over $\{0, 1\}$.

Question: Is there a pattern p such that $A \subseteq L(p) \subseteq \overline{B}$.

Reference: Ko, Tzeng [44].

Comments: $\mathbf{\Sigma_2^p}$-complete. Similar in structure to GRAPH CONSISTENCY, and GENERALIZED 3-CNF CONSISTENCY.

[AL2] FULLY COMPRESSED TWO-DIMENSIONAL PATTERN MATCHING

Given: Two images succinctly represented by straight-line programs. One image is called the pattern, the other the text. A *straight-line program* is a sequence of instructions of types $A \leftarrow B \oslash C$ (put image B next to image C if images have same height), and $A \leftarrow B \ominus C$ (put image B on top of image C if images have same width). Terminal symbols are 0 and 1.

Question: Is the pattern contained in the text (as a subrectangle)?

Reference: Berman, Karpinski, Larmore, Plandowski, Rytter [6].

Comments: $\mathbf{\Sigma_2^p}$-complete. Reduction from GENERALIZED SUBSET SUM. The fully compressed pattern matching problem for strings (one-dimensional patterns) can be solved in polynomial time (see [65] for a survey on compressed pattern matching).

[AL3] 1LTA GRAMMAR INEQUIVALENCE

Given: Two context-free grammars G_1 and G_2 over a 1-letter terminal alphabet. Let $L(G)$ be the language generated by a grammar G.

Question: Is $L(G_1) \neq L(G_2)$?

Reference: Huynh [39].

Comments: $\mathbf{\Sigma_2^p}$-complete. Reduction from INTEGER EXPRESSION INEQUIVALENCE. The more difficult part here is showing that the problem lies in $\mathbf{\Sigma_2^p}$ by using a variant of Parikh's theorem. The result has consequences for a unary variant of INTEGER EXPRESSION INEQUIVALENCE.

[AL4] POMSET LANGUAGE CONTAINMENT

Given: Two POMSETS P, Q. A POMSET (partially ordered multiset) is a directed acyclic graph (V, E) whose vertices have labels in Σ. The language $L(P)$ associated with a POMSET P is the set of words of length $n = |V|$ over Σ that corresponds to a permutation of the vertices in V which is consistent with the partial order generated by (V, E).

Question: Is $L(P) \subseteq L(Q)$?

Reference: Feigenbaum, Kahn, Lund [23].

Comments: $\mathbf{\Pi_2^p}$-complete. The language membership problem is **NP**-complete, and determining the size of $L(P)$ is span-P complete. Determining whether $L(P) = L(Q)$ obviously lies in $\mathbf{\Pi_2^p}$, and Feigenbaum, Kahn, and Lund showed that it is at least as hard as GRAPH ISOMORPHISM.

[AL5] STAR-FREE REG. EXPRESSION W/ INTERLEAVING EQUIVALENCE

Given: Two regular expressions e_1, e_2 using union, concatenation, and interleaving. For two words $x, y \in \{0,1\}^*$ the operation $|$ of interleaving x and y results in the set $x|y$ containing all words $x_1 y_1 \ldots x_k y_k$ such that $x = x_1 \ldots x_k$, and $y = y_1 \ldots y_k$, where the y_i can have any length (including zero).

Question: Are e_1 and e_2 equivalent, i.e., do they describe the same set of words?

Reference: Mayer, Stockmeyer [54].

Comments: $\mathbf{\Pi_2^p}$-complete. The proof is based on Stockmeyer's INTEGER EXPRESSION INEQUIVALENCE result. There are many versions of the regular expression problem. The standard version has union, concatenation, and Kleene star, and it is **PSPACE**-complete [25, AL9]. Adding interleaving, or intersection (Hunt, 1973; according to [54]) makes it exponential space-complete. Removing both the Kleene star and interleaving gives an **NP**-complete problem (Hunt, Stockmeyer and Mayer, 1973; according to [54]).

[AL6] TRIE$_2$

Given: A sequence Π of patterns of length n and an integer k. A *pattern* is a string in $\{0, 1, *\}^*$. A *call* is a string over $\{0, 1\}$ ($*$ matches both 0 and 1). A *TRIE T* is an ordered rooted tree (i.e. the order of a depth-first search traversal is specified) whose edges have labels in $\{1, 2, \ldots\} \times \{0, 1, *\}$. A *TRIE T for* Π is a TRIE which has as many leaves as Π has patterns. Furthermore if τ is the set of labels along the path to the jth leaf reached in the fixed depth-first search traversal of T, then τ needs to be equal to the jth pattern in Π (for all j). For a call c let $m(c, T)$ be the number of edges that a depth-first traversal of T will visit (we do not continue along an edge whose label is not consistent with c). Intuitively $m(c, T)$ is the number of matches performed by the TRIE to find all patterns in Π matching c.

Question: Is there a call c such that $m(c, T) \geq k$ for all TRIEs T for Π.

Reference: Lin [49].

Comments: $\mathbf{\Pi_2^p}$-complete. There is a constant $0 < c < 1$ for which approximating $f(\Pi) = \min_T \max_c m(c, T)$ to within a factor of c is $\mathbf{\Pi_2^p}$-hard. For a fixed TRIE the problem is **NP**-complete (Dawson, Ramakrishnan, Ramakrishnan, Swift, 1994; according to [49]).

[AL7] BOOLEAN ALGEBRA UNIFICATION

Given: Two terms ϕ and ψ over a Boolean algebra (operations $+$, $\times$, $\neg$ and constants 0, 1) with free constants.

Question: Can ϕ and ψ be unified; that is, is there a substitution σ of the free constants by terms of the Boolean algebra such that $\sigma(\phi)$ and $\sigma(\psi)$ are congruent in the Boolean algebra?

Reference: Baader [5].

Comments: $\mathbf{\Pi_2^p}$-complete. Is **NP**-complete, if free constants are not allowed.

[AL8] SIMPLE XPATH CONTAINMENT

Given: Simple XPath expressions P_1 and P_2. The application of a simple XPath expression P to an XML document results in a set of nodes (of the XML document). We write $P_1 \subseteq P_2$ if for all XML documents the nodes returned by P_1 are contained in the set of nodes returned by P_2. For precise definitions see [14] and references mentioned there.

Question: Does $P_1 \subseteq P_2$? hold?

Reference: Deutsch, Tannen [14].

Comments: $\mathbf{\Pi_2^p}$-complete, as are several variants of the problem.

[AL9] TRACE MONOID PRESENTATION

Given: Two trace monoids $M = M(A, D)$, and $M' = M(A, D')$ such that $D \subseteq D'$. A *trace monoid* $M(A, D)$ is a set of traces, that is, the quotient set $A^*/\{ab = ba | (a, b) \notin D\}$ of equivalence classes of words over the (finite) alphabet A, where two words are equivalent if one can be transformed into the other by repeatedly transposing pairs of letters (a, b) not in D. The *dependence relation* D is required to be reflexive and symmetric. A *trace replacement system* for a trace monoid $M = M(A, D)$ is a subset R of $M \times M$. An element (l, r) or R is considered as a rewriting rule $l \Rightarrow r$ over M. R is called *complete* if it is Noetherian (no infinite chains), and confluent.

Question: Is there a finite, complete trace replacement system R such that $M/R = M'$?

Reference: Diekert, Ochmański, Reinhardt [15].

Comments: $\mathbf{\Sigma_2^p}$-complete. The paper also shows that a similar question about semi-commutation systems is equivalent, and therefore also $\mathbf{\Sigma_2^p}$-complete.

[AL10] PLANAR NET DEADLOCK

Given: A nondeterministic finite automaton A, and an integer n. We construct a planar cellular automaton by placing n^2 copies of A on the n^2 grid points of an $n \times n$ square grid. Neighboring automata communicate by sending and receiving messages. A *deadlock* occurs if a group of automata permanently enter a waiting state (that is, they wait to receive a message which never arrives). An *input* to the network is a binary string of length n whose ith bit is sent to the ith automaton in the first row.

Question: Does the cellular automaton enter a deadlock in at most n time steps, for any possible input?

Reference: Durand, Fabret [18].

Comments: $\mathbf{\Pi_2^p}$-complete. Reduction from FINITE TILING EXTENSION. Recognizing whether a network enters a deadlock in at most n steps for a given input is, of course, **NP**-complete.

2.6 Databases

[D1] MONOTONIC RELATIONAL EXPRESSION CONTAINMENT

Given: Two *monotonic* relational expressions e_1, and e_2, i.e. only using operators select, project, join, and union. We write $\nu_D(e)$ to denote the extension of the relational expression e for a particular database state D.

Question: Is e_1 contained in e_2; that is, is it true that $\nu_D(e_1) \subseteq \nu_D(e_2)$ for all database states D?

Reference: Sagiv, Yannakakis [66].

Comments: $\mathbf{\Pi_2^p}$-complete. Implies that testing equivalence of monotonic relational expressions is also $\mathbf{\Pi_2^p}$-complete.

[D2] RESTRICTED RELATIONAL EXPRESSION CONTAINMENT (INEQ)

Given: Two *restricted* relational expressions e_1, and e_2, i.e. only using operators select, project, and join. The select conditions are allowed to contain inequalities ($\leq$, $<$, $\neq$). We write $\nu_D(e)$ to denote the extension of the relational expression e for a particular database state D.

Question: Is e_1 contained in e_2; that is, is it true that $\nu_D(e_1) \subseteq \nu_D(e_2)$ for all database states D?

Reference: van der Meyden [85].

Comments: $\mathbf{\Pi_2^p}$-complete. Remains $\mathbf{\Pi_2^p}$-complete if only one type of inequality ($\leq$, $<$, or $\neq$) is allowed in the select conditions [85]. It also remains $\mathbf{\Pi_2^p}$-complete, if the expressions are assumed to be *safe* (only variables that occur as arguments of relations can appear in inequalities), and certain other conditions (see [45]). Becomes **coNP**-complete, if all relations are unary. Without inequalities, the problem is **NP**-complete.

[D3] DISJUNCTIVE DATABASE LITERAL INFERENCE

Given: A disjunctive database D, and a literal w. A *disjunctive database* is a collection of formulas of the form $a_1 \vee \ldots \vee a_n \leftarrow b_1 \wedge \ldots \wedge b_k \wedge \overline{b}_{k+1} \wedge \ldots \wedge \overline{b}_m$ where the a_i and b_j are variables. There are different notions of $D \models w$ depending on the semantics chosen.

Question: Does $D \models w$?

Reference: Eiter, Gottlob [20].

Comments: $\mathbf{\Pi_2^p}$-complete for the following semantics: (Extended) Generalized Closed World Assumption, Extended Closed World Assumption, Iterated Closed World Assumption, Perfect Model Semantics, and Disjunctive Stable Semantics. It remains $\mathbf{\Pi_2^p}$-complete in all these cases, if the formulas of the disjunctive database do not contain negation, and there are no integrity clauses (i.e. $n > 0$ in all formulas). The problem is $\mathbf{\Pi_2^p}$-hard, and in $\mathbf{P}^{\mathbf{\Pi_2^p}[O(\log n)]}$ for the Careful Closed World Assumption.

[D4] DISJUNCTIVE DATABASE MODEL EXISTENCE

Given: A disjunctive database D, and a literal w. A *disjunctive database* is a collection of formulas of the form $a_1 \vee \ldots \vee a_n \leftarrow b_1 \wedge \ldots \wedge b_k \wedge \overline{b}_{k+1} \wedge \ldots \wedge \overline{b}_m$ where the a_i and b_j are variables. There are different semantics for what it means to be a model of D.

Question: Is there a model for D?

Reference: Eiter, Gottlob [20].

Comments: $\mathbf{\Pi_2^p}$-complete for Perfect Model Semantics, and Disjunctive Stable Semantics.

2.7 Games and Puzzles

[GP1] STRONG NASH EQUILIBRIUM

Given: A game $\mathcal{G}$ in graphical normal form. A *game* $\mathcal{G}$ consists of a set P of players, and, for every player, a set of neighbors $N(p) \subseteq P - \{p\}$, a set of actions $A(p)$, and a utility function $u_p : \times_{x \in N(p) \cup \{p\}} A(x) \to \mathbb{R}$. The game is in *graphical normal form*, if the utility function of each player is represented as a table. For a collection of players $P' \subseteq P$, an element of $\times_{p \in P'} A(p)$ is called a *strategy*. A strategy is *global*, if $P' = P$. A global strategy x is called a *strong Nash equilibrium* if there is no collection of players P' for whom there is a strategy $y \in \times_{p \in P'} A(p)$ which would strictly increase all of their gains; i.e. for all $p \in P'$ we would have $u_p(x) < u_p(x|y)$, where by $x|y$ we denote the strategy which on P' agrees with y, and with x otherwise.

Question: Does $\mathcal{G}$ have a strong Nash equilibrium?

Reference: Gottlob, Greco, Scarello [26].

Comments: $\mathbf{\Sigma_2^p}$-complete. Strong Nash equilibria generalize the notion of pure Nash equilibria whose definition is similar, but instead of arbitrary collection of players only requires local optimality for singleton sets of players. Deciding the existence of pure Nash equilibria is **NP**-complete.

[GP2] FINITE TILING EXTENSION

Given: A finite set C of c colors (including a blank color), a tile set $T \subseteq C^4$, an integer n. We say the four sides of the tile (t, r, b, l) are colored t (top), r (right), b (bottom), l (left). In a *tiling* of the plane tiles cannot be rotated or reversed. In a *legal* tiling, any two adjacent tiles must meet in the same color.

Question: Is there a legally tiled row R of n tiles which cannot be extended to a legal tiling of an $n \times n$ square such that R is the first row of that square and the square is surrounded by blank tiles?

Reference: Durand, Fabret [18]. Also see van Emde Boas [86]. The finite tiling variant of the problem, in which we ask whether an $n \times n$ square can be tiled using the tiles in T is **NP**-complete. This result is attributed to many different authors in different sources, including Lewis (in [86]), Garey, Johnson, and Papadimitriou (in [25]). The ideas for the reduction go back to papers by Robinson, Wang, and Berger (see [86]). The same reduction gives the $\mathbf{\Sigma_2^p}$-completeness result for FINITE TILING EXTENSION. However, it seems that Durand and Fabret were the first authors to make this observation explicitly in print (they actually attribute the result to van Emde Boas).

Comments: $\mathbf{\Sigma_2^p}$-complete. There are many versions of the tiling problem; for a detailed discussion see the survey by van Emde Boas [86]. Durand and Fabret use the tiling problem to show that PLANAR NET DEADLOCK is $\mathbf{\Pi_2^p}$-complete.

2.8 Coding and Cryptography

[*CC1] COVERING RADIUS

Given: A linear code, given by a binary parity-check matrix H of dimensions $m \times n$, integer r. The code associated with H is the set $C = \{\mathbf{x} : \mathbf{x}H^t = \mathbf{0}\}$. The *covering radius* of the code C is $\rho = \max_{\mathbf{x} \in \{0,1\}^n} \min_{\mathbf{c} \in C} d(\mathbf{x}, \mathbf{c})$, where $d(\mathbf{x}, \mathbf{c})$ is the Hamming distance between $\mathbf{x}$ and $\mathbf{c}$.

Question: Is $\rho \leq r$?

Reference: McLoughlin [55].

Comments: $\mathbf{\Pi_2^p}$-complete. Reduction from $\forall\exists$ THREE DIMENSIONAL MATCHING. $\mathbf{\Pi_2^p}$-hard to approximate to within some constant factor $c < 2$; however it can be approximated to within a factor of 2 in **AM** [27].

[*CC2] IDENTIFYING LINEAR CODE

Given: A linear code, given by a binary parity-check matrix H of dimensions $m \times n$, integer r. The code associated with H is the set $C = \{\mathbf{x} : \mathbf{x}H^t = \mathbf{0}\}$. C is called *r-identifying*, if the sets $B_r(\mathbf{x}) \cap C$ are all nonempty and pairwise different for $x \in \{0,1\}^n$.

Question: Is C an r-identifying code?

Reference: Honkala, Lobstein [36].

Comments: $\mathbf{\Pi_2^p}$-complete. Reduction from $\forall\exists$ THREE DIMENSIONAL MATCHING and COVERING RADIUS. A code is called *r-locating-dominating*, if the sets $B_r(\mathbf{x}) \cap C$ are all nonempty and pairwise different for $x \notin C$. Deciding, whether C is r-locating-dominating is also $\mathbf{\Pi_2^p}$-complete.

2.9 Miscellaneous

[*M1] MINIMUM BLOCK ENCODER AND DECODER

Given: Directed graph G, integers p, q, k_1, and k_2. G is a *DIF*, that is, it is a strongly connected graph whose edges are labeled with 0 and 1 such that every vertex has at most one outgoing edge of each label. Define $S(G)$ to be the set of all binary strings that can be obtained by following a directed path in G. A DIF G is called *block-feasible*, if there is a set $C \subseteq \{0,1\}^q$ of size at least 2^p whose closure is contained in $S(G)$; that is, there are 2^p codewords fulfilling the constraints described by G. A circuit D computing an injective function $\{0,1\}^p \to C$ is called an *encoder*, a circuit E computing an injective function $C \to \{0,1\}^p$ is called an *decoder*. The *size* of a circuit is the number of gates in the circuit.

Promise: G is block-feasible.

Question: Is there a decoder D of size at most k_1, and an encoder of size at most k_2?

Reference: Stockmeyer, Modha [74].

Comments: $\mathbf{\Sigma_2^p}$-complete under randomized reduction. MINIMUM BLOCK DECODER is $\mathbf{\Sigma_3^p}$-complete, and the complexity of MINIMUM BLOCK ENCODER is open.

[M2] PETRI NET MARKING EQUIVALENCE

Given: Petri nets (N_1, M_1), (N_2, M_2) which share the same set of places. The two nets are called *marking equivalent* if they have the same set of reachable markings. The Petri nets are assumed to be sinkless and normal, or conflict-free.

Question: Are (N_1, M_1) and (N_2, M_2) marking equivalent?

Reference: Howell, Rosier [37], and Howell, Rosier, Yen [38].

Comments: The general problem is undecidable (Rabin), but it is $\mathbf{\Pi_2^p}$-complete if the Petri nets are sinkless and normal [38], or conflict-free [37]. The problem remains complete if instead of equivalence we ask for containment.

[M3] CONSTRAINT RANKING

Given: A regular set $X \subseteq \{0,1\}^m$, given by a finite automaton, called *attested surface set*, and a collection of constraints $\{C_1, \ldots, C_n\}$. A *constraint* C of an attested surface set X is a function from X to the natural numbers (also computed by a finite automaton). A *ranking* of the constraints is an ordering $\vec{C}$ of $\{C_1, \ldots, C_n\}$. An element x of X is consistent with a ranking $(C_{i_1}, \ldots, C_{i_n})$ if $(C_{i_1}(x), \ldots, C_{i_n}(x)) \leq_{\text{lex}} (C_{i_1}(y), \ldots, C_{i_n}(y))$ for all $y \in \{0,1\}^m$, where $\leq_{lex}$ is the lexicographical ordering.

Question: Is there an $x \in X$ consistent with some ranking $\vec{C}$ of C?

Reference: Eisner [19].

Comments: $\mathbf{\Sigma_2^p}$-complete. Learning-theory problem from phonology.

3 The Third Level

3.1 Graph Theory

[*GT1] PATH VC DIMENSION

Given: A graph $G = (V, E)$, and a integer k. Let $VC_{\text{path}}(G)$ be the size of the largest set $X \subseteq V$ which is shattered by subpaths of G, i.e. such that for each $S \subseteq X$ there is a subpath of G containing all vertices in S, but no vertex of $X \setminus S$.

Question: Is $VC_{\text{path}}(G) \geq k$?

Reference: Schaefer [68].

Comments: Special case of the GRAPH VC DIMENSION problem defined for types of subgraphs of a given graph. Introduced by Kranakis, *et al.* [48] building on an idea of Haussler and Welzl [32]. The problem is also $\mathbf{\Sigma_3^p}$-complete for cycles instead of paths [68]. All other cases investigated so far turn out to be in **P** (stars, neighborhoods), or **NP**-complete (trees, connected sets) [48]. Also see VC DIMENSION and Q-ARY VC DIMENSION. No nonapproximability results are known.

[*GT2] CLIQUE CHOOSABILITY

Given: Graph $G = (V, E)$, integer k. A *k-list assignment* assigns a list $L(v)$ of k colors to every vertex v of G. A *k-clique-list-coloring* chooses for every vertex v a color from $L(v)$ such that every maximal clique of G contains two vertices of different color. The graph is *k-clique-choosable* if there is a k-clique-list-coloring for every k-list assignment.

Question: Is G a k-clique-choosable graph?

Reference: Marx [51].

Comments: $\mathbf{\Pi_3^p}$-complete for any fixed $k \geq 2$. The colorability version, CLIQUE COLORING is $\mathbf{\Sigma_2^p}$-complete. Also see HEREDITARY CLIQUE COLORING.

[*GT3] HEREDITARY CLIQUE COLORING

Given: Graph $G = (V, E)$, integer k. A *k-clique-coloring* is a function $c : V \to \{1, \ldots, k\}$ such that every maximal clique of G contains two vertices of different color. G is *hereditarily k-clique-colorable* if there it has a k-coloring which is a k-clique-coloring for all induced subgraphs of G.

Question: Does G have a hereditary k-clique-coloring?

Reference: Marx [51].

Comments: $\mathbf{\Pi_3^p}$-complete for any fixed $k \geq 3$. The complexity of the case $k = 2$ remains open. Also see CLIQUE COLORING and CLIQUE CHOOSABILITY.

[*GT4] SUCCINCT k-RADIUS

Given: Circuit C representing a directed graph $G = (V, E)$ (i.e., $C(u, v) = 1$ if and only if $(u, v) \in E$), integer k. The *r-neighborhood* of a vertex is the set of vertices that are reachable from the vertex by a path of length at most r. The *radius* of a directed graph is the radius r of the smallest r-neighborhood that contains all of G.

Question: Does G have radius at most k?

Reference: Hemaspaandra, Hemaspaandra, Tantau, Watanabe [34].

Comments: $\mathbf{\Sigma_3^p}$-complete for any fixed $k \geq 2$. Not known to remain $\mathbf{\Sigma_3^p}$-complete for tournaments (directed graphs for which there is exactly one edge between any two vertices). For undirected graphs, the problem is also $\mathbf{\Sigma_3^p}$-complete and becomes $\mathbf{\Sigma_2^p}$-complete for $k = 1$ [80]. Also see SUCCINCT k-DIAMETER and SUCCINCT k-KING.

3.2 Sets and Partitions

[*SP1] VC DIMENSION

Given: A collection $\mathcal{C}$ of subsets of a finite set U, represented succinctly by a Boolean circuit C such that $C(i, x) = 1$ if and only if element x is in the i-th set S_i, and an integer k.

Question: Is $VC(\mathcal{C}) \geq k$, i.e. is there a set $X \subseteq U$ of size at least k, such that for every $S \subseteq X$ there is an i such that $S = S_i \cap X$?

Reference: Schaefer [71].

Comments: $\mathbf{\Sigma_3^p}$-complete. Also $\mathbf{\Sigma_3^p}$-hard to approximate to within a factor of $2 - \epsilon$, but can be approximated to within a factor of 2 in **AM** [58]. If $\mathcal{C}$ is represented nonsuccinctly by a matrix the problem is **LOGNP**-complete as shown by Papadimitriou and Yannakakis [62]. Also see PATH VC DIMENSION and Q-ARY VC DIMENSION.

[*SP2] Q-ARY VC DIMENSION

Given: A collection $\mathcal{C}$ of vectors in $\{1, 2, \ldots q\}^U$, where U is a finite set, represented succinctly by a Boolean circuit C such that $C(i, x)$ is the x-th element of the i-th vector, and an integer k.

Question: Is $VC_q(\mathcal{C}) \geq k$, i.e. is there a set $X \subseteq U$ of size at least k, such that $\{(v_x)_{x \in X} | v \in \mathcal{C}\} = \{1, 2, \ldots q\}^X$?

Reference: Mossel, Umans [59].

Comments: $\mathbf{\Sigma_3^p}$-complete. Also $\mathbf{\Sigma_3^p}$-hard to approximate to within a factor of $q - \epsilon$, but can be approximated to within a factor of q in **AM**. Also see PATH VC DIMENSION and VC DIMENSION.

3.3 Algebra and Number Theory

[*AN1] INTEGER EXPRESSION COMPONENT LENGTH

Given: An integer expression e built from binary numbers with operators $+$, and $\cup$, and a number k. For an integer expression e define $L(e) = \{n\}$, if e is the binary representation of n, $L(e + f) = \{n + m : n \in L(e), m \in L(f)\}$, and $L(e \cup f) = L(e) \cup L(f)$. A set of numbers L is called *connected* if for every $x, z \in L$ and any y, if $x < y < z$ then $y \in L$. A maximal connected subset of a set is called a *component*.

Question: Does $L(e)$ have a component of size at least k?

Reference: Wagner [87].

Comments: $\mathbf{\Sigma_3^p}$-complete. The result also holds if using the general hierarchic input language (GHIL) for specifying the input. If the set of integers is specified by a Boolean formula, the problem is $\mathbf{\Sigma_2^p}$-complete (BOOLEAN EXPRESSION COMPONENT LENGTH). See INTEGER EXPRESSION INEQUIVALENCE, INTEGER EXPRESSION CONNECTEDNESS.

3.4 Miscellaneous

[*M1] MINIMUM BLOCK DECODER

Given: Directed graph G, integers p, q, and k. G is a *DIF*, that is, it is a strongly connected graph whose edges are labeled with 0 and 1 such that every vertex has at most one outgoing edge of each label. Define $S(G)$ to be the set of all binary strings that can be obtained by following a directed path in G. A DIF G is called *block-feasible*, if there is a set $C \subseteq \{0,1\}^q$ of size at least 2^p whose closure is contained in $S(G)$; that is, there are 2^p codewords fulfilling the constraints described by G. A circuit D computing an injective function $\{0,1\}^p \to C$ is called an *encoder*, a circuit E computing an injective function $C \to \{0,1\}^p$ is called an *decoder*. The *size* of a circuit is the number of gates in the circuit.

Promise: G is block-feasible.

Question: Is there a decoder D of size at most k?

Reference: Stockmeyer, Modha [74].

Comments: $\mathbf{\Sigma_3^p}$-complete. Remains $\mathbf{\Sigma_3^p}$-complete if $p > \alpha q$ for any $\alpha < 1$ and G has finite memory (from a certain length onward the acceptance of each string by G ends in a unique vertex only depending on the string). MINIMUM BLOCK ENCODER AND DECODER lies in $\mathbf{\Sigma_2^p}$, and is hard for $\mathbf{\Sigma_2^p}$ under randomized reductions. The complexity of the variant MINIMUM BLOCK ENCODER is open.

4 Open Problems

[O1] RAMSEY

Given: Finite Graphs G, and H.

Question: Does $K_n \to (G,H)$, i.e. does every edge-coloring of K_n with colors red and green contain either a red G, or a green H as a subgraph.

Comments: The problem is **NP**-hard [11], but not known to be $\mathbf{\Pi_2^p}$-complete. Also see ARROWING.

[*O2] MINIMUM EQUIVALENT EXPRESSION

Comments: Solved for $\{\vee, \wedge, \neg\}$-Boolean formulas, open over signature $\{\vee, \wedge, \neg, \to\}$. See [L22].

[*O3] MINIMAL

Given: A well-formed Boolean formula φ. The *size* $|\varphi|$ of a formula is the number of occurrences of literals in the formula.

Question: There is no well-formed Boolean formula ψ such that $\psi \equiv \varphi$ and $|\psi| < |\varphi|$.

Reference: Meyer, Stockmeyer [56].

Comments: **coNP**-hard [35], and in $\mathbf{\Pi_2^p}$. Also see MEE, and MIN DNF.

[*O4] MINIMUM BLOCK ENCODER

Given: Directed graph G, integers p, q, and k. G is a *DIF*, that is, it is a strongly connected graph whose edges are labeled with 0 and 1 such that every vertex has at most one outgoing edge of each label. Define $S(G)$ to be the set of all binary strings that can be obtained by following a directed path in G. A DIF G is called *block-feasible*, if there is a set $C \subseteq \{0,1\}^q$ of size at least 2^p whose closure is contained in $S(G)$; that is, there are 2^p codewords fulfilling the constraints described by G. A circuit D computing an injective function $\{0,1\}^p \to C$ is called an *encoder*, a circuit E computing an injective function $C \to \{0,1\}^p$ is called an *decoder*. The *size* of a circuit is the number of gates in the circuit.

Promise: G is block-feasible.

Question: Is there an encoder E of size at most k?

Reference: Stockmeyer, Modha [74].

Comments: Lies in $\mathbf{\Sigma_2^p}$, and is **NP**-hard. The similar MINIMUM BLOCK DECODER problem is $\mathbf{\Sigma_3^p}$-complete. Finding an encoder and a decoder of total size at most k also lies in $\mathbf{\Sigma_2^p}$, with complexity open. Putting separate bounds on the size of decoder and encoder leads to MINIMUM BLOCK ENCODER AND DECODER which is $\mathbf{\Sigma_2^p}$-complete under randomized reducibility.

[O5] POMSET LANGUAGE EQUALITY

Given: Two POMSETS P, Q. A POMSET (partially ordered multiset) is a directed acyclic graph (V,E) whose vertices have labels in Σ. The language $L(P)$ associated with a POMSET P is the set of words of length $n = |V|$ over Σ that corresponds to a permutation of the vertices in V which is consistent with the partial order generated by (V,E).

Question: Is $L(P) = L(Q)$?

Reference: Feigenbaum, Kahn, Lund [23].

Comments: In $\mathbf{\Pi_2^p}$, and at least as hard as GRAPH ISOMORPHISM. See POMSET LANGUAGE CONTAINMENT.

[O6] DISJUNCTIVE DATABASE FORMULA INFERENCE

Given: A disjunctive database D, and a formula φ. A *disjunctive database* is a collection of formulas of the form $a_1 \vee \ldots \vee a_n \leftarrow b_1 \wedge \ldots \wedge b_k \wedge \overline{b}_{k+1} \wedge \ldots \wedge \overline{b}_m$ where the a_i and b_j are variables. There are different notions of $D \models w$ depending on the semantics chosen.

Question: Does $D \models w$?

Reference: Eiter, Gottlob [20].

Comments: The problem is $\mathbf{\Pi_2^p}$-hard, and lies in $\mathbf{P}^{\mathbf{\Pi_2^p}[O(\log n)]}$ for the Generalized Closed World Assumption, and the Careful Closed World Assumption. If φ is a literal, it is known to be $\mathbf{\Pi_2^p}$-complete in the Generalized Closed World Assumption semantics.

[*O7] THUE NUMBER

Comments: Solved. See [GT22].

[*O8] STRONG CHROMATIC NUMBER

Given: A graph $G = (V, E)$, integer k. If k divides $|G|$ we call G *strongly k-colorable* if for every partition of V into pairwise disjoint sets of size k there is a proper coloring of G such that every color occurs exactly once in each set of the partition. If k does not divide G we add at most k isolated vertices to G so it does. The *strong chromatic number* of G is the smallest k such that G is strongly k-colorable.

Question: Is the strong chromatic number of G at most k?

Reference: The strong chromatic number was defined by Alon [3].

Comments: In $\mathbf{\Pi_2^p}$. Alon points out that in case the graph has bounded degree and the partition is given it can be decided in polynomial time whether a strong coloring exists using Beck's effective version of the Lovasz Local Lemma.

[*O9] THUE CHROMATIC NUMBER

Given: A graph $G = (V, E)$, integer k. A word w is *square-free* (or *non-repetitive*) if there are no u, v, w such that $w = uvvw$ (with v not the empty word). A *non-repetitive k-(vertex) coloring* of G is a k-coloring of G such that for any path in G, the sequence of colors along the path is square-free. The smallest k such that G has a non-repetitive k-coloring is called the *Thue chromatic number* of G.

Question: Is the Thue chromatic number of G at most k?

Reference: The Thue chromatic number first defined in Alon, Grytczuk, Hauszczak, Riordan [4].

Comments: In $\mathbf{\Sigma_2^p}$. Given a 4-coloring of a graph, it is **coNP**-complete to decide whether it is non-repetitive [53]. Named after Axel Thue who proved that there are infinite square-free words. Also see THUE NUMBER.

Acknowledgment. Many thanks to Lane Hemaspaandra for many helpful suggestions. Thanks to Nikolay Pelov for pointing out a mistake in the statement of GENERALIZED SUBSET SUM.

References

[1] M. Agrawal and T. Thierauf. The Boolean Isomorphism Problem. In *37th Annual Symposium on Foundations of Computer Science*, pages 422–430, Burlington, Vermont, 14–16 Oct. 1996. IEEE.

[2] M. Ajtai, R. Fagin, and L. J. Stockmeyer. The closure of monadic NP. *Journal of Computer and System Sciences*, 60(3):660–716, 2000.

[3] N. Alon. The strong chromatic number of a graph. *Random Structures Algorithms*, 3(1):1–7, 1992.

[4] N. Alon, J. Grytczuk, M. Hauszczak, and O. Riordan. Nonrepetitive colorings of graphs. In M. Káronski, J. Spencer, and A. Ruciński, editors, *Proceedings of the 10th International Conference on Random Structures and Algorithms (RS&A-01)*, volume 21, 3–4 of *Random Structures & Algorithms*, pages 336–346, Danvers, MA, Aug. 6–10 2002. Wiley Periodicals.

[5] F. Baader. On the complexity of Boolean unification. *Information Processing Letters*, 67(4):215–220, 31 Aug. 1998.

[6] P. Berman, M. Karpinski, L. L. Larmore, W. Plandowski, and W. Rytter. On the complexity of pattern matching for highly compressed two-dimensional texts. In A. Apostolico and J. Hein, editors, *Combinatorial Pattern Matching, 8th Annual Symposium*, volume 1264 of *Lecture Notes in Computer Science*, pages 40–51, Aarhus, Denmark, 30 June–2 July 1997. Springer.

[7] B. Borchert and D. Ranjan. The circuit subfunction relations are $\sum_2^p$-complete. Research Report MPI-I-93-121, Max-Planck-Institut fr Informatik, Im Stadtwald, D-66123 Saarbrcken, Germany, May 1993.

[8] B. Borchert, D. Ranjan, and F. Stephan. On the computational complexity of some classical equivalence relations on boolean functions. *MST: Mathematical Systems Theory*, 31, 1998.

[9] E. Börger, E. Grädel, and Y. Gurevich. *The Classical Decision Problem*. Springer Verlag, Berlin, 1997.

[10] D. Buchfuhrer and C. Umans. The complexity of boolean formula minimization. In L. Aceto, I. Damgrd, L. A. Goldberg, M. M. Halldrsson, A. Inglfsdttir, and I. Walukiewicz, editors, *ICALP*, volume 5125 of *Lecture Notes in Computer Science*, pages 24–35. Springer, 2008.

[11] S. A. Burr. Some undecidable problems involving the edge-coloring and vertex-coloring of graphs. *Discrete Mathematics*, 50:171–177, 1984.

[12] S. A. Burr. On the Computational Complexity of Ramsey-Type Problems. In Nešetřil and Rödl, editors, *Mathematics of Ramsey Theory*. Springer-Verlag, 1990.

[13] A. Condon, J. Feigenbaum, C. Lund, and P. Shor. Probabilistically checkable debate systems and approximation algorithms for PSPACE-hard functions. In *Proc. 25th Ann. ACM Symp. on Theory of Computing*, pages 305–314, 1993.

[14] A. Deutsch and V. Tannen. Containment and integrity constraints for xpath fragments. In M. Lenzerini, D. Nardi, W. Nutt, and D. Suciu, editors, *Proceedings of the 8th International Workshop on Knowledge Representation meets Databases (KRDB 2001), Rome, Italy, September 15, 2001*, number 45 in CEUR Workshop Proceedings, 2001.

[15] V. Diekert, E. Ochmański, and K. Reinhardt. On confluent semi-commutations: decidability and complexity results. *Information and Computation*, 110:164–182, 1994.

[16] D.-Z. Du and K.-I. Ko. *Theory of Computational Complexity*. Wiley, 2000.

[17] P. E. Dunne and T. J. M. Bench-Capon. Coherence in finite argument systems. *AIJ: Artificial Intelligence*, 141(1):187–203, 2002.

[18] B. Durand and A.-C. Fabret. On the complexity of deadlock detection in families of planar nets. *Theoretical Computer Science*, 215(1–2):225–237, 1999.

[19] J. Eisner. Easy and hard constraint ranking in optimality theory. In *Finite-State Phonology, Proceedings of the 5th SIGPHON*, 2000.

[20] T. Eiter and G. Gottlob. Complexity aspects of various semantics for disjuntive databases. In *Symposium on Principles of Database Systems*, pages 158–167, 1993.

[21] T. Eiter and G. Gottlob. Note on the complexity of some eigenvector problems. Technical Report CD-TR 95/89, Christian Doppler Laboratory for Expert Systems, TU Vienna, Dec. 1995.

[22] P. Erdős, A. L. Rubin, and H. Taylor. Choosability in graphs. *Congressus Numerantium*, 26:125–157, 1979.

[23] J. Feigenbaum, J. A. Kahn, and C. Lund. Complexity Results for POMSET Languages. *SIAM Journal on Discrete Mathematics*, 6:432–442, 1993.

[24] L. Fortnow and J. Killian, April 2002. Personal communication.

[25] M. R. Garey and D. S. Johnson. *Computers and Intractability.* Freeman, San Francisco, 1979.

[26] G. Gottlob, G. Greco, and F. Scarcello. Pure nash equilibria: Hard and easy games. In *Proceedings of TARK 2003*, 2003.

[27] V. Guruswami, D. Micciancio, and O. Regev. The comlexity of the covering radius problem on lattices and code. In *Proceedings of the 19th annual IEEE conference on computational complexity - CCC '04*, pages 161–173, Amherst, Massachusetts, USA, June 2004. IEEE.

[28] S. Gutner. The complexity of planar graph choosability. *Discrete Mathematics*, 159:119–130, 1996.

[29] S. Gutner and M. Tarsi. Some results on $(a : b)$-choosability. Unpublished manuscript, 1992.

[30] F. Harary, W. Slany, and O. Verbitsky. On the computational complexity of the forcing chromatic number. *SIAM J. Comput.*, 37(1):1–19, 2007.

[31] H. Hatami and H. Maserrat. On the computational complexity of defining sets. *Discrete Appl. Math.*, 149(1-3):101–110, 2005.

[32] D. Haussler and E. Welzl. Epsilon-nets and simplex range queries. *Discrete and Computational Geometry*, 2:127–151, 1987.

[33] I. Haviv, O. Regev, and A. Ta-Shma. On the hardness of satisfiability with bounded occurrences in the polynomial-time hierarchy. *Theory of Computing*, 3(1):45–60, 2007.

[34] E. Hemaspaandra, L. A. Hemaspaandra, T. Tantau, and O. Watanabe. On the complexity of kings. In E. Csuhaj-Varjú and Z. Ésik, editors, *FCT*, volume 4639 of *Lecture Notes in Computer Science*, pages 328–340. Springer, 2007.

[35] E. Hemaspaandra and G. Wechsung. The Minimization Problem for Boolean Formulas. In *38th Annual Symposium on Foundations of Computer Science*, pages 575–584, Miami Beach, Florida, 20–22 Oct. 1997. IEEE.

[36] I. Honkala and A. Lobstein. On identifying codes in binary hamming spaces. *Journal of Combinatorial Theory*, 99:232–243, 2002.

[37] R. R. Howell and L. E. Rosier. Completeness results for conflict-free vector replacement systems. *Journal of Computer and System Sciences*, 37(3):349–366, Dec. 1988.

[38] R. R. Howell, L. E. Rosier, and H.-C. Yen. Normal and sinkless petri nets. *Journal of Computer and System Sciences*, 46:1–26, 1993.

[39] D. T. Huynh. Deciding the Inequivalence of Context-free Grammars with 1-Letter Terminal Alphabet is Σ_2^p-complete. *Theoretical Computer Science*, 33(2-3):305–326, Oct. 1984.

[40] T.-D. Huynh. The Complexity of Semilinear Sets. *Journal of Information Processing and Cybernetics*, 18(6):291–338, 1982.

[41] M. A. Kiwi, C. Lund, A. Russell, D. A. Spielman, and R. Sundaram. Alternation in interaction. In *Structure in Complexity Theory Conference*, pages 294–303, 1994.

[42] K.-I. Ko and C.-L. Lin. On the complexity of min-max optimization problems and their approximation. In D.-Z. Du and P. M. Pardalos, editors, *Minimax and Applications*, pages 219–239. Kluwer Academic Publishers, Boston, 1995.

[43] K.-I. Ko and C.-L. Lin. On the longest circuit in an alterable digraph. *Journal of Global Optimization*, 7:279–295, 1995.

[44] K.-I. Ko and W.-G. Tzeng. Three Σ_2^p-complete Problems in Computational Learning Theory. *Computational Complexity*, 1:269–310, 1991.

[45] P. G. Kolaitis, D. L. Martin, and M. N. Thakur. On the complexity of the containment problem for conjunctive queries with built-in predicates. In *PODS '98. Proceedings of the ACM Symposium on Principles of Database Systems*, pages 197–204. ACM Press, 1998.

[46] D. Král'. Group coloring is π_2^p-complete. Technical Report 2003-166, Institut teoretické informatiky (ITI), Universita Karlova v Praze, 2003. Available at `http://iti.mff.cuni.cz/series/files/iti166ps.zip`.

[47] D. Král' and P. Nejedlý. Group coloring and list group coloring are π_2^p-complete. In *Mathematical Foundations of Computer Science 2004: 29th International Symposium*, pages 274– 286. Springer-Verlag, 2004.

[48] E. Kranakis, D. Krizanc, B. Ruf, J. Urrutia, and G. Woeginger. The VC-dimension of set systems defined by graphs. *Discrete Applied Mathematics and Combinatorial Operations Research and Computer Science*, 77:237–257, 1997.

[49] C.-L. Lin. Optimizing TRIEs for Ordered Pattern Matching is Π_2^p-Complete. In *Proceedings of the 10th Annual Conference on Structure in Complexity Theory (SCTC '95)*, pages 238–245, Los Alamitos, CA, USA, June 1995. IEEE Computer Society Press.

[50] F. Manin. The complexity of nonrepetitive edge coloring of graphs. *arXiv.org*, arXiv:0709.4497, 2007.

[51] D. Marx. Complexity of clique coloring and related problems, 2004. manuscript.

[52] D. Marx. Complexity of unique list colorability, 2005. manuscript.

[53] D. Marx and M. Schaefer. The Complexity of Nonrepetitive Coloring. To appear in Discrete Applied Mathematics, 2008.

[54] A. J. Mayer and L. J. Stockmeyer. The complexity of word problems - this time with interleaving. *Information and Computation*, 115:293–311, 1994.

[55] A. McLoughlin. The complexity of computing the covering radius of a code. *IEEE Transactions on Information Theory*, 30:800–804, 1984.

[56] A. R. Meyer and L. J. Stockmeyer. The equivalence problem for regular expressions with squaring requires exponential space. In *13th Annual Symposium on Switching and Automata Theory*, pages 125–129, The University of Maryland, 25–27 Oct. 1972. IEEE.

[57] K. Milans and B. Clark. The complexity of graph pebbling. Manuscript.

[58] E. Mossel and C. Umans. On the complexity of approximating the VC dimension. In *Proceedings of the 16th Annual Conference on Computational Complexity*, pages 220–225. IEEE Computer Society Press, 2001.

[59] E. Mossel and C. Umans. On the complexity of approximating the VC dimension. Full version of [58]. Submitted, 2002.

[60] A. Nickelsen and T. Tantau. On reachability in graphs with bounded independence number. In *8th Annual International Conference, COCOON*, pages 554–563, 2002.

[61] C. H. Papadimitriou. *Computational Complexity*. Addison-Wesley, New York, 1994.

[62] C. H. Papadimitriou and M. Yannakakis. On limited nondeterminism and the complexity of the V–C dimension. *Journal of Computer and System Sciences*, 53(2):161–70, 1996.

[63] R. Pichler. On the complexity of equational problems in CNF over a finite domain. In P. Baumgartner and H. Zhang, editors, *Proc. 3rd Int. Workshop on First-Order Theorem Proving (FTP 2000)*, number 5/2000 in Fachberichte Informatik der Universität Koblenz-Landau, pages 182–193, St. Andrews, Scotland, UK, 2000. available at http://www.uni-koblenz.de/fb4/publikationen/gelbereihe/RR-5-2000/.

[64] V. Rutenburg. Complexity of generalized graph coloring. In J. Gruska, B. Rovan, and J. Wiedermann, editors, *Mathematical Foundations of Computer Science, 12th Symposium*, volume 233 of *lncs*, pages 573–581, Bratislava, Czechoslovakia, 25–29 Aug. 1986. Springer.

[65] W. Rytter. Algorithms on compressed strings and arrays. In *Conference on Current Trends in Theory and Practice of Informatics*, pages 48–65, 1999.

[66] Y. Sagiv and M. Yannakakis. Equivalences Among Relational Expressions with the Union and Difference Operators. *Journal of the Association for Computing Machinery*, 27(4):633–655, Oct. 1980.

[67] M. Schaefer. Graph Ramsey theory and the polynomial hierarchy. In *Proceedings of the 31st Annual ACM Symposium on Theory of Computing*, pages 592–601. ACM, 1999.

[68] M. Schaefer. Deciding the Vapnik-Cervonenkis dimension is Σ_3^P-complete, ii. Technical Report TR00-006, DePaul University, Oct. 2000.

[69] M. Schaefer. The graph sandwich problem for a **conp** property. Technical Report TR06-011, DePaul University, 2006.

[70] M. Schaefer and C. Umans. Σ_2^p-completeness of MIN DNF TAUTOLOGY. Manuscript, Apr. 2002.

[71] M. Schäfer. Deciding the Vapnik-Cervonenkis dimension is Σ_3^P-complete. *Journal of Computer and System Sciences*, 58:177–182, 1999.

[72] U. Schöning. Complexity of Presburger arithmetic with fixed quantifier dimension. *Theory of Computing Systems*, 30(4):423–428, July/Aug. 1997.

[73] S. Sriram, R. Tandon, P. Dasgupta, and P. P. Chakrabarti. Symbolic verification of boolean constraints over partially specified functions. In *Proceedings of ISCAS (IEEE International Symposium on Circuits and Systems) 2001*, pages 113–116, 2001.

[74] L. Stockmeyer and D. S. Modha. Links between complexity theory and constrained block coding. *IEEE Transactions on Information Theory*, 48(1):59–88, 2002.

[75] L. J. Stockmeyer. The polynomial-time hierarchy. *Theoretical Computer Science*, 3(1):1–22, Oct. 1976.

[76] S. Szeider. Generalizations of matched cnf formulas, 2002. Fifth International Symposium on the Theory and Applications of Satisfiability Testing.

[77] S. Szeider. Generalizations of matched CNF formulas. *Ann. Math. Artif. Intell.*, 43(1-4):223–238, 2005. Theory and applications of satisfiability testing.

[78] A. Ta-Shma, C. Umans, and D. Zuckerman. Loss-less condensers, unbalanced expanders, and extractors. In *Proceedings of the 33rd Annual ACM Symposium on Theory of Computing (STOC)*, pages 143–152. ACM, 2001.

[79] T. Tantau. A note on the complexity of the reachability problem for tournaments. Technical Report TR01-092, ECCC, Nov. 2001.

[80] T. Tantau. Complexity of the undirected radius and diameter problems for succinctly represented graphs. Technical Report SIIM-TR-A-08-02, Universität zu Lübeck, 2008.

[81] C. Umans. Hardness of Approximating Σ_2^p Minimization Problems. In *IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 465–474, 1999.

[82] C. Umans. On the complexity and inapproximability of shortest implicant problems. In *Automata, Languages and Programming, 26th International Colloquium (ICALP 99)*, pages 687–696, July 1999.

[83] C. Umans. *Approximability and Completeness in the Polynomial Hierarchy.* PhD thesis, U.C. Berkeley, 2000.

[84] C. Umans. The Minimum Equivalent DNF Problem and Shortest Implicants. *Journal of Computer and System Sciences*, 63(4):597–611, Dec. 2001.

[85] R. van der Meyden. The complexity of querying infinite data about linearly ordered domains. *Journal of Computer and System Sciences*, 54(1):113–135, 1997.

[86] P. van Emde Boas. The convenience of tilings. In A. Sorbi, editor, *Complexity, Logic, and Recursion Theory*, pages 331–363. Marcel Dekker, Inc., 1997.

[87] K. W. Wagner. The Complexity of Combinatorial Problems with Succinct Input Representation. *Acta Informatica*, 23(3):325–356, 1986.

[88] C. Wrathall. Complete sets and the polynomial-time hierarchy. *Theoretical Computer Science*, 3(1):23–33, Oct. 1976.