Εθνικό και Καποδιστριακό Πανεπιστήμιο
Αθηνών

Τμήμα Μαθηματικών
Μεταπτυχιακό Πρόγραμμα Λογικής και Θεωρίας Αλγορίθμων και Υπολογισμού

Διχοτομίες Πολυπλοκότητας σε Προβλήματα Μέτρησης

Διπλωματική Εργασία
του
Στυλιανού Δεσποτάκη

Επιβλέπων: Άρης Παγουρτζής
Επίκουρος Καθηγητής

Ιούλιος 2012

Η παρούσα Διπλωματική Εργασία
εκπονήθηκε στα πλαίσια των σπουδών
για την απόκτηση του
Μεταπτυχιακού Διπλώματος Ειδίκευσης
στη
Λογική και Θεωρία Αλγορίθμων και Υπολογισμού
που απονέμει το
Τμήμα Μαθηματικών
του
Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών

Εγκρίθηκε την 23η Ιουλίου 2012 από Εξεταστική Επιτροπή
αποτελούμενη από τους:

Ονοματεπώνυμο	**Βαθμίδα**	**Υπογραφή**
1. : Αρ. Παγουρτζής	Επίκ. Καθηγητής	……………………
2. : Ε. Ζάχος	Καθηγητής	……………………
3. : Δ. Φωτάκης	Λέκτορας	……………………

Περίληψη

Στην παρούσα διπλωματική εργασία μελετάμε θεωρήματα διχοτομίας που αφορούν κυρίως προβλήματα μέτρησης. Ένα θεώρημα διχοτομίας για μια κλάση προβλημάτων, στην υπολογιστική πολυπλοκότητα, είναι μια πλήρης ταξινόμηση των προβλημάτων της οικογένειας σε υπολογιστικά εύκολα και σε υπολογιστικά δύσκολα προβλήματα, χωρίς ενδιάμεσα προβλήματα. Εξαιτίας του θεωρήματος του Ladner δε μπορούμε να βρούμε διχοτομία για ολόκληρες τις κλάσεις NP και #P, αλλά παρ' όλα αυτά υπάρχουν μεγάλες υποκλάσεις προβλημάτων που μοντελοποιούν πολλά "φυσικά" προβλήματα και για τις οποίες υπάρχουν θεωρήματα διχοτομίας. Ξεκινάμε με το μοντέλο των προβλημάτων ικανοποίησης περιορισμών ως πρόβλημα απόφασης (CSP) και στη συνέχεια μελετάμε τις κλάσεις των προβλημάτων μέτρησης ομομορφισμών γραφήματος, των προβλημάτων μέτρησης ικανοποίησης περιορισμών (#CSP) και των προβλημάτων Holant. Τέλος, θα κάνουμε μια σύντομη εισαγωγή στους ολογραφικούς αλγορίθμους, που είναι ειδικού τύπου πολυωνυμικοί αλγόριθμοι για προβλήματα μέτρησης.

Λέξεις κλειδιά: Διχοτομία, Πρόβλημα Ικανοποίησης Περιορισμών, Ομομορφισμοί Γραφήματος, Προβλήματα Holant, Ολογραφικοί Αλγόριθμοι

Abstract

In this thesis we study dichotomy theorems mainly of counting problems. A dichotomy theorem for a class of problems, in computational complexity, is a complete classification of the problems of this class in computationally easy and computationally hard problems, without intermediate problems. Due to Ladner's theorem we cannot find a dichotomy for the whole classes NP and #P, however there are large subclasses of NP (#P), that model many "natural" problems, for which dichotomy theorems exist. We begin with the framework of the decision version of constraint satifaction problems (CSP) and then we study the classes of graph homomorphisms problems, of counting constraint satisfaction problems (#CSP) and of Holant problems. Finally, we will make a brief introduction to holographic algorithms, a special type of polynomial-time algorithms for counting problems.

Keywords: Dichotomy, Constraint Satisfaction Problem, Graph Homomorphisms, Holant Problems, Holographic Algorithms

Ευχαριστίες

Θα ήθελα με αφορμή αυτή τη διπλωματική να ευχαριστήσω όλους μου τους καθηγητές για τις γνώσεις που μου πρόσφεραν και για την ουσιαστική επιρροή που είχαν στη μέχρι τώρα πορεία μου.

Πιο συγκεκριμένα, θα ήθελα να ευχαριστήσω τον κ. Παγουρτζή για την καθοδήγησή του όλο αυτό το διάστημα, και τα άλλα δύο μέλη τις τριμελούς επιτροπής, κ. Ζάχο και κ. Φωτάκη. Επίσης θα ήθελα να ευχαριστήσω τους καθηγητές μου κ. Κολλιόπουλο, κ. Θηλυκό και κ. Κουτσουπιά για τη στήριξή τους και τις συμβουλές που μου πρόσφεραν.

Επιπλέον, θα ήθελα να ευχαριστήσω τον Ανδρέα για την καλή συνεργασία που είχαμε καθ' όλη τη διάρκεια της συγγραφής αυτής της εργασίας. Να ευχαριστήσω επίσης, τους συμφοιτητές μου και φίλους μου στο μΠλ∀ και την εργαστηριακή ομάδα του Corelab για τις ευχάριστες και δημιουργικές συζητήσεις που είχαμε.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένειά μου για την αγάπη και τη στήριξη που μου έχει προσφέρει όλα αυτά τα χρόνια.

Περιεχόμενα

Κεφάλαιο 1

Εισαγωγή

1.1 Βασικοί Ορισμοί

Σκοπός της Υπολογιστικής Πολυπλοκότητας είναι να ταξινομήσει τα διάφορα προβλήματα σε κλάσεις πολυπλοκότητας. Σε αυτή την παράγραφο θα ορίσουμε κάποιες βασικές κλάσεις πολυπλοκότητας και αναγωγών, τα οποία θα χρειαστούμε στη συνέχεια. Για μια πιο αναλυτική εισαγωγή, παραπέμπουμε στα [2, 34].

Όταν περιορίζουμε μια μηχανή Turing (TM) ως προς τον χρόνο που μπορεί να τρέξει, τότε προκύπτει ο παρακάτω ορισμός:

Ορισμός 1.1.1. *Έστω f μια συνάρτηση από το $\mathbb{N}$ στο $\mathbb{N}$. Θα λέμε ότι η TM M λειτουργεί σε χρόνο $f(n)$, αν για κάθε συμβολοσειρά εισόδου x, ο χρόνος που χρειάζεται η M σε είσοδο x είναι το πολύ $f(|x|)$* [1]*. Ας υποθέσουμε ότι μια γλώσσα $L \subseteq (\Sigma - \{\sqcup\})^*$ αποφασίζεται από μια TM η οποία λειτουργεί σε χρόνο $f(n)$. Θα λέμε ότι $L \in \mathrm{TIME}(f(n))$.*

Αν από την άλλη περιορίσουμε τη μηχανή ως προς τον χώρο τον οποίο μπορεί να χρησιμοποιήσει, τότε έχουμε τον παρακάτω ορισμό:

Ορισμός 1.1.2. *Έστω f μια συνάρτηση από το $\mathbb{N}$ στο $\mathbb{N}$. Θα λέμε ότι η TM M λειτουργεί σε χώρο $f(n)$, αν για κάθε συμβολοσειρά εισόδου x, η M χρειάζεται χώρο το πολύ $f(|x|)$. Έστω L μια γλώσσα. Θα λέμε ότι $L \in \mathrm{SPACE}(f(n))$ αν υπάρχει TM που αποφασίζει την L με τη χρήση το πολύ $f(n)$ κελιών στην ταινία εργασίας της.*

Για μη-ντετερμινιστικές μηχανές Turing συμβολίζουμε τις κλάσεις πολυπλοκότητας με περιορισμό στο χρόνο και στο χώρο ως $\mathrm{NTIME}(f(n))$ και $\mathrm{NSPACE}(f(n))$ αντίστοιχα.

Ορίζουμε επίσης τις παρακάτω κλάσεις πολυπλοκότητας:

[1] το $|x|$ συμβολίζει το μήκος της συμβολοσειράς x

$\mathrm{L} = \mathrm{SPACE}(\log n)$
$\mathrm{NL} = \mathrm{NSPACE}(\log n)$
$\mathrm{P} = \bigcup_{k>0} \mathrm{TIME}(n^k)$
$\mathrm{NP} = \bigcup_{k>0} \mathrm{NTIME}(n^k)$
$\mathrm{EXP} = \bigcup_{k>0} \mathrm{TIME}(2^{n^k})$
$\mathrm{PSPACE} = \bigcup_{k>0} \mathrm{SPACE}(n^k)$
$\mathrm{NPSPACE} = \bigcup_{k>0} \mathrm{NSPACE}(n^k)$

Οι παραπάνω κλάσεις συνδέονται μεταξύ τους ως εξής:

$$\mathrm{L} \subseteq \mathrm{NL} \subseteq \mathrm{P} \subseteq \mathrm{NP} \subseteq \mathrm{PSPACE}.$$

Για τους περισσότερες από τους παραπάνω εγκλεισμούς παραμένει ανοιχτό το ερώτημα αν είναι γνήσιοι ή όχι. Το κεντρικό και πιο διάσημο ανοιχτό ερώτημα της υπολογιστικής πολυπλοκότητας είναι η σχέση μεταξύ των κλάσεων P και NP, επειδή για τα περισσότερα προβλήματα που συναντούμε στη βιβλιογραφία, αυτά που μπορούν να λυθούν αποτελεσματικά ανήκουν στο P και αυτά για τα οποία δε γνωρίζουμε αποδοτικούς αλγορίθμους ανήκουν στο NP.

Οι παραπάνω ορισμοί αφορούν προβλήματα απόφασης (γλώσσες). Εμείς θα ασχοληθούμε κυρίως με προβλήματα μέτρησης (συναρτήσεις), όπου ενδιαφερόμαστε για το πλήθος των λύσεων ενός δοσμένου προβλήματος. Η κύρια κλάση για τα προβλήματα μέτρησης είναι η εξής:

Ορισμός 1.1.3 (Valiant [36])**.** *$\#\boldsymbol{P}$ είναι η κλάση όλων των συναρτήσεων $f : \Sigma^* \to \mathbb{N}$ για τις οποίες, για κάθε $x \in \Sigma^*$, $f(x) = \mathrm{acc}_M(x)$ για κάποια NP-TM M, όπου $\mathrm{acc}_M(x)$ είναι το πλήθος των υπολογιστικών μονοπατιών της M που αποδέχονται με είσοδο x.*

Επίσης, η κλάση συναρτήσεων που υπολογίζονται από μια ντετερμινιστική μηχανή Turing πολυωνυμικού χρόνου είναι η FP. Το ανοιχτό ερώτημα για την πολυπλοκότητα των προβλημάτων μέτρησης, αντίστοιχο αυτού του P vs. NP, είναι το εάν ισχύει $\mathrm{FP} = \#\mathrm{P}$ ή $\mathrm{FP} \subset \#\mathrm{P}$, με το δεύτερο να θεωρείται πιο πιθανό.

Για να δώσουμε αργότερα κάποια αποτελέσματα πληρότητας, θα χρησιμοποιήσουμε τις παρακάτω αναγωγές:

Ορισμός 1.1.4. *Θα λέμε ότι ένα πρόβλημα A ανάγεται σε ένα πρόβλημα B κατά Karp-αναγωγή (polynomial-time many one) και θα συμβολίζεται με $A \leq_m^p B$, αν υπάρχει μια συνάρτηση f υπολογίσιμη σε πολυωνυμικό χρόνο, τέτοια ώστε για κάθε x, $x \in A$ αν και μόνο αν $f(x) \in B$.*

Στην περίπτωση συναρτήσεων f_A και f_B, η παραπάνω αναγωγή ορίζεται ως εξής:

$$f_A(x) \leq_m^p f_B \text{ αν } \exists g \in \mathrm{FP}, \forall x\, f_A(x) = g(f_B(x)).$$

Ορισμός 1.1.5. *Θα λέμε ότι ένα πρόβλημα A (ή συνάρτηση) ανάγεται σε ένα πρόβλημα B κατά Cook-αναγωγή (polynomial-time Turing) και θα συμβολίζεται με $A \leq_T^p B$, αν η A μπορεί να υπολογιστεί από μια ντετερμινιστική TM M σε πολυωνυμικό χρόνο με τη χρήση ενός μαντείου (oracle) για τη B. Αν η M καλέσει μία μόνο φορά το μαντείο για τη B, τότε έχουμε την Cook* [1] *αναγωγή, που συμβολίζεται με $A \leq_{[1]-T}^p B$.*

Να σημειώσουμε ότι η Cook-αναγωγή ισχύει όχι μόνο για γλώσσες (προβλήματα), αλλά και για συναρτήσεις. Για τα προβλήματα μέτρησης θα χρησιμοποιήσουμε επίσης και την φειδωλή (parsimonious) αναγωγή:

Ορισμός 1.1.6. *Θα λέμε ότι ένα πρόβλημα μέτρησης $\#A$ ανάγεται σε ένα πρόβλημα μέτρησης $\#B$ με φειδωλή αναγωγή (parsimonious reduction) αν υπάρχει μια συνάρτηση f υπολογίσιμη σε πολυωνυμικό χρόνο τέτοια ώστε, για κάθε x,* $|\{y \mid (x,y) \in A\}| = |\{z \mid (f(x),z) \in B\}|$.

1.2 Το Θεώρημα του Ladner και οι Επιπτώσεις του

Όπως είπαμε ένα από κύρια ανοιχτά προβλήματα είναι μια απόδειξη για το αν ισχύει ή όχι ότι $\mathrm{P} \neq \mathrm{NP}$. Από τη στιγμή που το πρόβλημα είναι ακόμη ανοιχτό, το καλύτερο που μπορούμε να ελπίζουμε είναι να κατατάξουμε κάθε πρόβλημα στο NP να είναι είτε NP-complete είτε να λύνεται στο P. Η παραπάνω πρόταση ισχύει και για τα προβλήματα μέτρησης, λόγω έλλειψης απόδειξης του $\mathrm{FP} \neq \#\mathrm{P}$, μπορούμε να κατατάξουμε κάθε πρόβλημα στο $\#\mathrm{P}$ έτσι ώστε να είναι είτε $\#\mathrm{P}$-complete είτε να λύνεται σε πολυωνυμικό χρόνο;

Και στα δύο παραπάνω ερωτήματα απάντησε αρνητικά ο Ladner με το παρακάτω θεώρημά του [31].

Θεώρημα 1.2.1. *Αν $\mathrm{P} \neq \mathrm{NP}$ τότε υπάρχει μια γλώσσα $L \in \mathrm{NP}$ η οποία δεν ανήκει στο P αλλά ούτε είναι NP-complete.*

Η ακόλουθη απόδειξη μπορεί να βρεθεί στο [2] (με κάποιες λεπτομέρειες να παραλείπονται):

Απόδειξη του θεωρήματος 1.2.1. Έστω M_i η μηχανή Turing που έχει σαν περιγραφή τη δυαδική αναπαράσταση του i. Ας θεωρήσουμε τη γλώσσα $\mathrm{SAT}_H = \{\phi 01^{n^{H(n)}} \mid \phi \in \mathrm{SAT} \wedge n = |\phi|\}$, όπου $H : \mathbb{N} \to \mathbb{N}$ είναι η συνάρτηση που ορίζεται ως εξής:
$H(n)$ είναι ο μικρότερος αριθμός $i < \log \log n$ τέτοιος ώστε για κάθε $x \in \{0,1\}^\star$ με $|x| \leq \log n$, η M_i δίνει ως έξοδο το $\mathrm{SAT}_H(x)$ μετά από το πολύ $i \cdot |x|^i$ βήματα. Αν δεν υπάρχει τέτοιος αριθμός i τότε $H(n) = \log \log n$.

Μπορούμε να υπολογίσουμε την $H(n)$ αναδρομικά υπολογίζοντας πρώτα τα $H(k)$ για $k \leq \log n$. Ο αλγόριθμος χρειάζεται να προσομοιώσει το πολύ $\log \log n$ μηχανές Turing για κάθε είσοδο με μήκος το πολύ $\log n$ και για το πολύ $\log \log n \cdot (\log n)^{\log \log n}$ βήματα. Στο τέλος κάθε τέτοιας προσομοίωσης πρέπει να ελέγχουμε αν $M_i(x) = \mathrm{SAT}_H(x)$ και για να το κάνουμε αυτό χρειαζόμαστε την τιμή του $H(k)$ για ένα $k \leq \log n$. Εύκολα μπορούμε να δούμε ότι ο συνολικός υπολογισμός του $H(n)$ γίνεται το πολύ σε $O(n^3)$ βήματα. Αυτό μας δείχνει και ότι η συνάρτηση $H(n)$ είναι καλώς ορισμένη.

Τώρα ας υποθέσουμε ότι $\mathrm{SAT}_H \in \mathrm{P}$, και άρα υπάρχει μηχανή Turing M που λύνει το SAT_H σε το πολύ $c \cdot n^c$ βήματα για κάποια σταθερά c. Θα υπάρχει κάποιος αριθμός $i > c$ τέτοιος ώστε $M_i = M$. Εξ' ορισμού, για $n > 2^{2^i}$, $H(n) \leq i$, οπότε $H(n) = O(1)$.

Από την άλλη μεριά, αν $H(n) = O(1)$, τότε η H μπορεί να πάρει μόνο πεπερασμένου πλήθους τιμές, οπότε υπάρχει ένα i τέτοιο ώστε $H(n) = i$, για άπειρα σε πλήθος n. Από αυτό συμπεραίνουμε ότι η M_i λύνει το SAT_H σε χρόνο $i \cdot n^i$, γιατί διαφορετικά αν υπήρχε κάποιο x για το οποίο η M_i δε δίνει σωστό αποτέλεσμα μέσα σε αυτό το χρόνο, τότε για σχεδόν όλα τα $n > 2^{|x|}$ (εκτός ίσως από το 2^{2^i}) θα ίσχυε $H(n) \neq i$.

Στις δύο τελευταίες παραγράφους δείξαμε ότι $\mathrm{SAT}_H \in \mathrm{P}$ αν και μόνο αν $H(n) = O(1)$. Επιπλέον, η συνεπαγωγή της δεύτερης παραγράφου ισχύει ακόμη και αν υποθέσουμε μόνο ότι υπάρχει σταθερά C τέτοια ώστε $H(n) \leq C$ για άπειρα σε πλήθος n, που σημαίνει ότι αν $\mathrm{SAT}_H \notin \mathrm{P}$ τότε το $H(n)$ τείνει στο άπειρο όταν το n τείνει στο άπειρο.

Έστω τώρα ότι $\mathrm{SAT}_H \in \mathrm{P}$: από τα παραπάνω συμπεραίνουμε ότι $H(n) \leq C$ για κάποια σταθερά C, από το οποίο προκύπτει ότι το SAT_H είναι απλά το SAT αν κολλήσουμε σε κάθε στιγμιότυπο το πολύ πολυωνυμικά σε πλήθος 1. Αυτό όμως θα σήμαινε ότι ένας πολυωνυμικός αλγόριθμος για το SAT_H θα μπορούσε να χρησιμοποιηθεί για να λυθεί το SAT σε πολυωνυμικό χρόνο, και άρα θα ίσχυε $\mathrm{P} = \mathrm{NP}$.

Ας υποθέσουμε τέλος ότι το SAT_H είναι NP-complete: οπότε υπάρχει αναγώγή r από το SAT στο SAT_H που τρέχει σε χρόνο $O(n^k)$ για κάποια σταθερά k. Έχουμε ήδη δείξει ότι $\mathrm{SAT}_H \notin \mathrm{P}$, οπότε το $H(n)$ τείνει στο άπειρο. Αφού η r τρέχει σε χρόνο $O(n^k)$, για αρκετά μεγάλα n θα πρέπει να στέλνει στιγμιότυπα του SAT μεγέθους n σε στιγμιότυπα του SAT_H μεγέθους το πολύ $n^H(n)$. Συνεπώς αρκετά μεγάλους τύπους ψ, η αναγωγή r τους στέλνει σε μια συμβολοσειρά του τύπου $\phi 01^{|\phi|^{H(|\phi|)}}$, όπου ϕ ένας τύπος μικρότερος κατά ένα σταθερό πολυωνυμικό παράγοντα. Πιο συγκεκριμένα θα είναι $|\phi| + 1 + |\phi|^{H(|\phi|)} = O(|\psi|^i)$, που σημαίνει ότι $|\phi| = o(|\psi|)$. Εφαρμόζοντας αναδρομικά τη διαδικασία μέχρι να φτάσουμε σε αρκετά μικρό τύπο, έχουμε έναν αλγόριθμο πολυωνυμικού χρόνου για το SAT, κάτι που έρχεται σε αντίθεση με την υπόθεση ότι $\mathrm{P} \neq \mathrm{NP}$. $\square$

Ας σημειώσουμε ότι η συγκεκριμένη τεχνική που είδαμε στην απόδειξη, μπορεί να εφαρμοστεί και σε άλλες κλάσεις και έτσι να προκύψουν τα παρακάτω πορίσματα.

Πόρισμα 1.2.2. *Αν υποθέσουμε ότι* $\mathrm{P} \neq \mathrm{NP}$, *τότε υπάρχει μια άπειρη ιεραρχία από διαχωρισμένες κλάσεις πολυπλοκότητας μεταξύ του* P *και του* NP.

Απόδειξη. Θα εφαρμόσουμε τη τεχνική της απόδειξης του θεωρήματος 1.2.1 για να δειξουμε ότι υπάρχει γλώσσα που δεν ανήκει στο P, ούτε η $\mathrm{SAT}_H =_{\mathrm{def}} \mathrm{SAT}_{H_0}$ ανάγεται σε αυτήν:

Ας θεωρήσουμε τη γλώσσα $\mathrm{SAT}_{H_1} = \{\phi 01^{n^{H(n)}} \mid \phi \in \mathrm{SAT}_{H_0} \wedge n = |\phi|\}$, όπου H είναι η συνάρτηση που ορίσαμε στην απόδειξη του 1.2.1. Η SAT_{H_1} είναι όχι μόνο στο NP, αλλά ανάγεται και στην SAT_{H_0} (αφού H είναι πολυωνυμικά υπολογίσιμη). Από την άλλη μεριά, χρησιμοποιώντας τα ίδια επιχειρήματα όπως πριν, αν $\mathrm{SAT}_{H_1} \in \mathrm{P}$ θα προέκυπτε ότι $\mathrm{SAT}_{H_0} \in \mathrm{P}$. Επίσης μια αναγωγή $\mathrm{SAT}_{H_0} \leq_p \mathrm{SAT}H_1$, θα μας έδινε έναν πολυωνυμικό αλγόριθμο για την SAT_{H_0}.

Ας θεωρήσουμε τώρα την ακόλουθη οικογένεια γλωσσών: $\mathrm{SAT}_{H_i} = \{\phi 01^{n^{H(n)}} \mid \phi \in \mathrm{SAT}_{H_{i-1}} \wedge n = |\phi|\}$. Οι κλάσεις πολυπλοκότητας που ορίζονται από τις αντίστοιχες κλειστότητες (closures) κατά Karp, αποτελούν μια άπειρη ιεραρχία κλάσεων μεταξύ του P και του NP. □

Πόρισμα 1.2.3. *Αν* $\mathrm{FP} \neq \#\mathrm{P}$ *τότε υπάρχει μια άπειρη ιεραρχία διαχωρισμένων κλάσεων πολυπλοκότητας μεταξύ του* FP *και του* #P.

Εκτός από τα τεχνητά προβλήματα που περιγράψαμε στις παραπάνω αποδειξεις (SAT_H), υπάρχουν και κάποιοι φυσικοί υποψήφιοι που πιστεύουμε πως δεν είναι ούτε NP-complete, και ούτε γνωρίζουμε αν ανήκουν στο P. Τέτοια προβλήματα είναι το πρόβλημα ισομορφισμού γραφημάτων και το πρόβλημα της παραγοντοποίησης ως πρόβλημα απόφασης, δηλαδή δοσμένου ενός ακεραίου n και ενός ακεραίου M με $1 \leq M \leq n$, το ερώτημα είναι αν ο n έχει διαιρέτη d με $1 < d < M$. Για τα προβλήματα μέτρησης, μπορούμε να θεωρήσουμε αντίστοιχα το πρόβλημα του υπολογισμού του πλήθους των διαιρετών ενός ακεραίου.

Από την άλλη μεριά, η πλειοψηφία των προβλημάτων απόφασης (μέτρησης) που γνωρίζουμε και που προκύπτουν φυσικά, είναι είτε πολύ εύκολα, δηλαδή ανήκουν στο P(FP), είτε πολύ δύσκολα, δηλαδή είναι NP-complete(#P-complete). Από τη στιγμή που δε μπορούμε να αποδείξουμε ότι αυτό ισχύει για όλα τα προβλήματα στο NP(#P), θα θέλαμε να βρούμε υποκλάσεις που περιέχουν όσο το δυνατόν περισσότερα προβλήματα, τέτοιες ώστε τα προβλήματα που περιέχονται σε αυτές να είναι είτε πολύ εύκολα, είτε πολύ δύσκολα. Μια τέτοιου είδους κλάση θα πρέπει να είναι αρκετά περιορισμένη ώστε θεώρημα αντίστοιχο με αυτό του Ladner να μην ισχύει για αυτή την κλάση, αλλά παράλληλα θα θέλαμε να είναι και όσο το δυνατόν πιο μεγάλη γίνεται. Έτσι θα είχαμε πετύχει να καθορίσουμε πλήρως την πολυπλοκότητα των προβλημάτων που περιέχονται σε αυτή την κλάση. Τέτοιου είδους θεωρήματα στην υπολογιστική πολυπλοκότητα, όπου μια οικογένεια προβλημάτων αποδεικνύεται ότι περιέχει είτε πολύ εύκολα, είτε πολύ δύσκολα προβλήματα και τίποτα ενδιάμεσα, ονομάζονται θεωρήματα διχοτομίας (dichotomy theorems).

1.3 Παραδείγματα Διχοτομιών στην Υπολογιστική Πολυπλοκότητα

Ένα από τα πρώτα θεωρήματα διχοτομίας στην υπολογιστική πολυπλοκότητα οφείλεται στους Jaeger, Vertigan και Welsh [26] και αφορά το πολυώνυμο Tutte ενός γραφήματος.

Ορισμός 1.3.1. *Το πολυώνυμο Tutte ενός γραφήματος G είναι το:*

$$T(G;x,y) = \sum_{A \subseteq E} (x-1)^{\kappa(V,A)-\kappa(V,E)} (y-1)^{|A|-(|V|-\kappa(V,A))}$$

όπου $\kappa(V,A) =$ το πλήθος των συνεκτικών συνιστωσών του γραφήματος (V,A).

Το πολυώνυμο Tutte ορίζεται για κάθε μη κατευθυνόμενο γράφημα και δίνει πληροφορίες για το πως ακριβώς είναι συνεκτικό το γράφημα. Μερικά παραδείγματα που μπρούν να βρεθούν στο βιβλιο του Welsh [40] είναι τα εξής:

- το $T(G;1,1)$ υπολογίζει το πλήθος των δενδροπαραγόντων (spanning trees) του γραφήματος G.
- το $T(G;2,1)$ υπολογίζει το πλήθος των δασών (forests) στο G
- $T(G;1,2)$ υπολογίζει το πλήθος των υποσυνόλων από ακμές που είναι συνεκτικά και καλύπτουν το G.
- το $T(G;2,0)$ υπολογίζει το πλήθος των ακυκλικών προσανατολισμών του G.
- Το χρωματικό πολυώνυμο $P(G;\lambda)$ ενός γραφήματος G με n κορυφές, m ακμές και k συνεκτικές συνιστώσες δίνεται από τον τύπο: $P(G;\lambda) = (-1)^{n-k}\lambda^k T(G;1-\lambda,0)$.
  Αν θεωρήσουμε το λ έναν θετικό ακέραιο, τότε το $P(G;\lambda)$ υπολογίζει το πλήθος των λ-χρωματισμών του G.

Για σταθερά x, y το θεώρημα διχοτομίας των Jaeger, Vertigan και Welsh είναι πάνω στο ακόλουθο πρόβλημα:

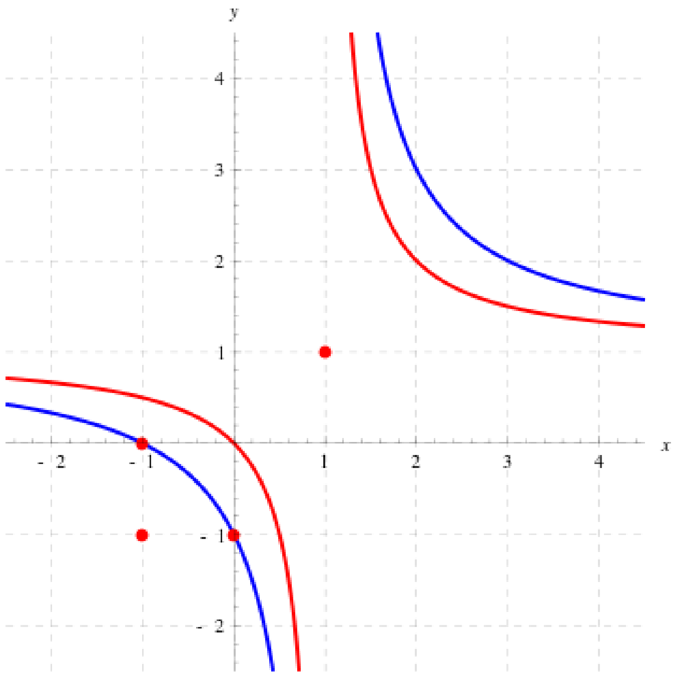
- $\textsc{Tutte}(x,y)$:

– Είσοδος: Ένα γράφημα G.

– Ζητούμενο: $T(G; x, y)$.

Πιο συγκεκριμένα, απέδειξαν ότι το πρόβλημα υπολογισμού του πολυωνύμου Tutte ενός γραφήματος στο σημείο (x, y) του επιπέδου, είναι #P-hard εκτός εάν $(x-1)(y-1) = 1$ ή όταν το (x, y) ισούται με $(1,1), (-1,-1), (0,-1), (-1,0), (i,-i), (-i,i), (j,j^2)$ ή (j^2,j), όπου $j = e^{2\pi i/3}$.

Επίσης οι Vertigan και Welsh [39] έδειξαν ότι αν περιοριστούμε στην περίπτωση των επίπεδων γραφημάτων, τότε και τα σημεία στην υπερβολή $(x-1)(y-1) = 2$ οδηγούν σε υπολογισμό του πολυωνύμου σε πολυωνυμικό χρόνο.



Στο παραπάνω σχήμα, κάθε σημείο (x, y) του επιπέδου αντιστοιχεί σε ένα πρόβλημα TUTTE(x, y). Σε κάθε κόκκινο σημείο, το πρόβλημα λύνεται σε πολυωνυμικό χρόνο. Σε κάθε μπλε σημείο το πρόβλημα είναι #P-hard γενικά, αλλά υπολογίζεται σε πολυωνυμικό χρόνο για επίπεδα γραφήματα. Σε κάθε άλλο σημείο, το πρόβλημα είναι #P-hard ακόμη και για επίπεδα γραφήματα.

Ένα άλλο θεώρημα διχοτομίας που βγήκε το 1990 ήταν για το πρόβλημα H-Χρωματισμού. Ένας H-Χρωματισμός ενός γραφήματος G είναι απλά ένας ομομορφισμός από το G στο H.

Για παράδειγμα, αν $H = K_3$ τότε ο K_3-Χρωματισμός είναι απλά το πρόβλημα 3-χρωματισμού ενός γραφήματος, και πιο γενικά αν $H = K_q$ τότε έχουμε το πρόβλημα του k-χρωματισμού.

Οι Hell και Nesetril [25] απέδειξαν ότι το πρόβλημα του H-Χρωματισμού είναι στο P αν το γράφημα H είναι διμερές, ενώ είναι NP-complete σε κάθε άλλη περίπτωση.

Πολλά πρόσφατα θεωρήματα διχοτομίας αφορούν προβλήματα από τη στατιστική φυσική. Στην στατιστική φυσική μοντελοποιούν συστήματα spin σε γραφήματα, και θέλουν να υπολογίσουν τη συνάρτηση διαμέρισης (partition function) ενός συστήματος spin ένα γράφημα. Ένα από τα πιο μελετημένα μοντέλα στη στατιστική φυσική είναι το μοντέλο Ising, που πρωτοεισήχθη τη δεκαετία του 1920 από τους Lenz και Ising για να μελετήσουν τον φερρομαγνητισμό (ferromagnetism). Ένα στιγμιότυπο του μοντέλου δίνεται από ένα σύνολο από n τοποθεσίες, ένα σύνολο από ενέργειες αλληλεπίδρασης V_{ij} για κάθε ζεύγος i, j από τοποθεσίες, μια ένταση μγνητικού πεδίου B, και μια αντίστροφη θερμοκρασία (inverse temperature) β. Κατάσταση του συστήματος που ορίζεται από αυτές τις παραμέτρους είναι είναι μία από τις 2^n πιθανές αναθέσεις σ από ± 1 spins σε κάθε τοποθεσία. Η ενέργεια μιας κατάστασης σ συμβολίζεται με $H(\sigma)$ και ορίζεται ως:

$$H(\sigma) = -\sum_{\{i,j\}} V_{ij}\sigma_i\sigma_j - B\sum_k \sigma_k.$$

Το ενδιαφέρον κομμάτι σε αυτό το άθροισμα είναι ο πρώτος όρος, που αποτελείται από τη συνεισφορά από τα ζεύγη τοποθεσιών. Η σημασία αυτής της έκφρασης έρχεται από την κατανομή Gibbs, σύμφωνα με την οποία η πιθανότητα το σύστημα να βρίσκεται σε μια κατάσταση σ είναι ανάλογη της ποσότητας $\exp(-\beta H(\sigma))$. Από αυτό συμπεραίνουμε ότι η πιθανότητα μιας κατάστασης σ είναι $1/Z \times \exp(-\beta H(\sigma))$, όπου ο παράγοντας κανονικοποίησης Z, που καλείται συνάρτηση διαμέρισης του συστήματος, είναι ο

$$Z = \sum_{\sigma \in \{-1,1\}^n} \exp(-\beta H(\sigma)).$$

Να σημειώσουμε εδώ ότι ο υπολογισμός της συνάρτησης διαμέρισης του μοντέλου Ising ισοδυναμεί με το πρόβλημα $\textsc{Tutte}(x, y)$ με x, y τέτοια ώστε $(x-1)(y-1) = 2$. ¨Αλλα σημαντικά μοντέλα της στατιστικής φυσικής είναι το μοντέλο Hard-Core (που είναι ισοδύναμο με τον υπολογισμό των ανεξάρτητων συνόλων σε ένα γράφημα), το μοντέλο Potts και το μοντέλο Antiferromagnetic Potts (που είναι ισοδύναμο με τον υπολογισμό του πλήθους των χρωματισμών ενός γραφήματος). Τα θεωρήματα διχοτομίας που έχουμε αφορούν ειδικές περιπτώσεις αυτών των μοντέλων και λένε ότι η συνάρτηση διαμέρισης μπορεί να προσεγγιστεί (FPRAS) κάτω από κάποιες τιμές μιας παραμέτρου του συστήματος. Για το μοντέλο Ising η παράμετρος αυτή είναι η β.

Αυτό που είναι πιο αξιοσημείωτο με αυτές τις διχοτομίες είναι ότι συμπίπτουν με τις μεταβάσεις φάσεις (phase transitions) που έχουν αποδειχθεί από τους φυσικούς.

Έτσι εάν η παράμετρος είναι κάτω από μια κρίσιμη τιμή, τότε το πρόβλημα είναι εύκολο, ενώ όταν η παράμετρος υπερβεί αυτή τη τιμή, τότε το πρόβλημα γίνεται #P-hard, και επιπλέον το σύστημα αλλάζει φάση.

Κεφάλαιο 2

Το Πρόβλημα Ικανοποίησης Περιορισμών (CSP) ως Πρόβλημα Απόφασης

Σκοπός αυτού του κεφαλαίου είναι να εισάγουμε το πρόβλημα ικανοποίησης περιορισμών (constraint satisfaction problem ή απλούστερα CSP) ως πρόβλημα απόφασης και να παρουσιάσουμε ορισμένα από τα πιο σημαντικά αποτελέσματα γύρω από τη θεωρία του.

Η προέλευση του προβλήματος οφείλεται στον Ugo Montanari και συγκεκριμένα σε ένα άρθρο του το 1974 [33] που ασχολείται με το πρόβλημα για εφαρμογές επεξεργασίας εικόνας.

Για μια πρώτη διαίσθηση μπορούμε να σκεφτόμαστε το CSP σαν ένα γενικό ομογενές πλαίσιο που μοντελοποιεί πολλά γνωστά συνδυαστικά προβλήματα. Παραδείγματα τέτοιων προβλημάτων είναι γραφοθεωρητικά (χρωματισμός γραφήματος, κάλυψη κορυφών, ...), το SAT, προβλήματα βάσεων δεδομένων, τεχνητής νοημοσύνης, επιχειρησιακής έρευνας, βελτιστοποίησης και πολλά άλλα.

2.1 Ορισμός του Προβλήματος Απόφασης

Αρχικά θα εισάγουμε κάποιους βασικούς ορισμούς.

Ορισμός 2.1.1. *Πεδίο (domain) ονομάζεται ένα πεπερασμένο σύνολο D από στοιχεία, που συνήθως θα συμβολίζεται ως $D = \{0, 1, \cdots, n-1\}$.*

Ορισμός 2.1.2. *Ο πληθικός αριθμός (cardinality) ενός πεδίου D θα συμβολίζεται ως $|D| = n$.*

Ορισμός 2.1.3. *Μια k-δική σχέση (relation) R στο πεδίο D θα είναι ένα υποσύνολο $R \subseteq D^k$.*

Για παράδειγμα αν έχουμε ως πεδίο το σύνολο $D = \{0, 1, 2\}$, μια διμελής σχέση στο D είναι η $L = \{(0, 0), (0, 1), (0, 2), (1, 1), (1, 2), (2, 2)\}$.

Ορισμός 2.1.4. *Ένας k-δικός περιορισμός είναι ένα κατηγόρημα εφαρμοζόμενο σε ένα διάνυσμα k μεταβλητών.*

Για παράδειγμα τα $C_1(x, y) = (x = y)$, $C_2(x, y) = (x \leq y)$ και $C_3(x, y, z) = (x \neq y \neq z \neq x)$ είναι περιορισμοί.

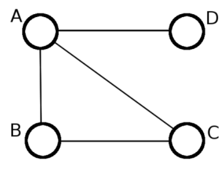
Από μια σχέση R μπορούμε να κατασκευάσουμε έναν περιορισμό $R(\vec{x})$ τέτοιο ώστε το R να αντιστοιχεί στο σύνολο λύσεων του $R(\vec{x})$. Αν για παράδειγμα θεωρήσουμε τη σχέση L που αναφέραμε προηγουμένως για το πεδίο $D = \{0, 1, 2\}$, τότε ο περιορισμός $L(x, y)$ αντιστοιχεί στον περιορισμό $x \leq y$ για το D.

Επίσης να παρατηρήσουμε ότι μια σχέση R μπορεί να οδηγήσει σε περισσότερους του ενός περιορισμούς. Για παράδειγμα οι $R(x, y)$ και $R(y, x)$ είναι διαφορετικοί περιορισμοί.

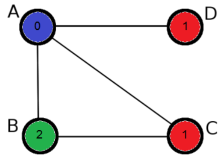
Ήρθε η ώρα να ορίσουμε το πρόβλημα CSP. Ως είσοδος δίνεται ένας τύπος ϕ, ο οποίος είναι πεπερασμένη σύζευξη από περιορισμούς στο πεδίο D. Το ζητούμενο είναι αν όλοι οι περιορισμοί είναι ικανοποιήσιμοι, δηλαδή αν υπάρχει αποτίμηση στις μεταβλητές έτσι ώστε να ικανοποιούνται όλοι οι περιορισμοί.

Για παράδειγμα αν έχουμε τον τύπο $(a = b) \wedge (b \neq d) \wedge (d < c) \wedge (b < c)$ στο πεδίο $D = \{0, 1, 2, 3\}$, τότε η αποτίμηση $a = 2, b = 2, c = 3, d = 2$ δεν είναι λύση γιατί δεν ικανοποιείται ο δεύτερος περιορισμός. Η αποτίμηση όμως $a = 1, b = 1, c = 3, d = 2$ αποτελεί λύση, άρα ο τύπος είναι ικανοποιήσιμος.

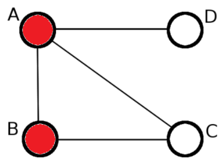
Ας δούμε τώρα μερικά παραδείγματα προβλημάτων που μοντελοποιούνται από το πλαίσιο του CSP. Το πιο απλό ίσως παράδειγμα είναι αυτό του 3-χρωματισμού ενός γραφήματος. Σ'αυτό το πρόβλημα δίνεται ένα μη κατευθυνόμενο γράφημα και μας ρωτάνε αν μπορούμε να χρωματίσουμε τις κορυφές του με τρία χρώματα έτσι ώστε οποιεσδήποτε δύο κορυφές ενώνονται με ακμή να έχουν διαφορετικό χρώμα. Θα θεωρήσουμε ως πεδίο D το $\{0, 1, 2\}$, ένα στοιχείο για κάθε χρώμα και ως μεταβλητές τις κορυφές του γραφήματος. Για παράδειγμα στο γράφημα της εικόνας θα έχουμε μεταβλητές το σύνολο $V = \{a, b, c, d\}$.



Τέλος όλοι οι περιορισμοί θα αντιστοιχούν στη σχέση $\neq$ και θα υπάρχει ένας για κάθε ακμή. Στο παράδειγμά λοιπόν ο τύπος που θα προκύψει ως είσοδος θα είναι ο $(a \neq b) \wedge (a \neq c) \wedge (b \neq c) \wedge (a \neq d)$ και μια λύση είναι η $a = 0, b = 2, c = d = 1$, άρα το γράφημα είναι 3-χρωματίσιμο.



Ένα άλλο παράδειγμα προβλήματος που μοντελοποιείται ως CSP είναι η k-κάλυψη κορυφών. Σ'αυτό το πρόβλημα δίνεται πάλι ως είσοδος ένα γράφημα και μας ρωτάνε αν υπάρχουν το πολύ k κορυφές που καλύπτουν όλες τις ακμές, δηλαδή για ακμή τουλάχιστον το ένα άκρο της να ανήκει σ'αυτές τις k το πολύ κορυφές. Γι'αυτό το πρόβλημα θα θεωρήσουμε ως πεδίο το $D = \{0, 1\}$, όπου τα στοιχεία θα αντιστοιχούν στην επιλογή ή μη μιας κορυφής. Ως μεταβλητές θα θεωρήσουμε πάλι τις κορυφές του γραφήματος, δηλαδή για το παραπάνω γράφημα θα έχουμε μεταβλητές το σύνολο $V = \{a, b, c, d\}$. Οι περιορισμοί θα είναι δύο ειδών. Πρώτον, θα υπάρχει ένας περιορισμός $C(x, y)$ για κάθε ακμή $\{x, y\}$ που θα αντιστοιχεί στη σχέση $C = \{(01), (10), (11)\}$ και θα εξασφαλίζει ότι για κάθε ακμή θα βάλουμε στο σύνολο μας τουλάχιστον ένα άκρο. Επίσης θα υπάρχει και ένας περιορισμός K που θα περιέχει όλες τις μεταβλητές και θα εξασφαλίζει ότι το πολύ k κορυφές έχουν επιλεγεί. Για το παραπάνω γράφημα ο τύπος που θα προκύψει είναι ο $C(a, b) \wedge C(a, c) \wedge C(b, c) \wedge C(a, d) \wedge K(a, b, c, d)$. Έστω ότι ζητάμε μια 2-κάλυψη κορυφών, οπότε μια λύση είναι η $a = b = 1, c = d = 0$.



Πέρα από το γενικό CSP υπάρχει και το παραμετρικό CSP. Η διαφορά του είναι ότι τώρα πλέον το πρόβλημα έχει και μια παράμετρο S που είναι ένα σύνολο σχέσεων στο πεδίο D. Στο πρόβλημα CSP(S) δίνεται πάλι ως είσοδος ένας τύπος ϕ αλλά αυτή τη φορά ο τύπος είναι πεπερασμένη σύζευξη περιορισμών που αντιστοιχούν αποκλειστικά σε σχέσεις από το S. Πάλι και σε αυτό το πρόβλημα το ερώτημα είναι αν ο ϕ είναι ικανοποιήσιμος.

Το γενικό CSP πρόβλημα είναι γνωστό ότι είναι NP-complete (αφού για παράδειγμα ο 3-χρωματισμός είναι ειδική περίπτωσή του). Το ενδιαφέρον με το παραμετρικό CSP είναι δοσμένου ενός πεδίου D υπάρχουν κάποιες παράμετροι S για τις οποίες το πρόβλημα CSP(S) ανήκει στο P. Ας θεωρήσουμε για παράδειγμα το δυαδικό πεδίο $D = \{0, 1\}$. Είδαμε ότι το CSP($\{C, K\}$) μοντελοποιεί το πρόβλημα k-κάλυψης κορυφών, άρα είναι NP-complete. Το πρόβλημα όμως CSP($\{\neq\}$) μοντελοποιεί το 2-χρωματισμό ενός γραφήματος άρα ανήκει στο P.

2.2 Δυαδικό (Boolean) CSP

Σ'αυτή την παράγραφο θα ασχοληθούμε με το δυαδικό πεδίο $D = \{0, 1\}$. Αυτή είναι η πιο απλή και ταυτόχρονα ενδιαφέρουσα περίπτωση πεδίου. Μοντελοποιεί αρκετά φυσικά προβλήματα και υπάρχουν και αρκετά ενδιαφέροντα αποτελέσματα για το δυαδικό πεδίο, όπως το θεώρημα του Schaefer για το οποίο θα μιλήσουμε παρακάτω.

Το πιο κλασικό παράδειγμα προβλήματος που μοντελοποιείται ως CSP στο δυαδικό πεδίο είναι το 3SAT.

Ορισμός 2.2.1. *Λέμε ότι ένας τύπος είναι σε κανονική συζευκτική μορφή (conjunctive normal form ή απλούστερα CNF) όταν είναι σύζευξη από όρους, όπου κάθε όρος είναι διάζευξη από literals, όπου κάθε literal είτε μια μεταβλητή x, είτε η άρνησή της $\bar{x}$.*

Στο πρόβλημα 3SAT δίνεται ως είσοδος ένας πεπερασμένος CNF τύπος ϕ στον οποίο κάθε όρος έχει το πολύ τρία literals και μας ρωτάνε αν είναι ικανοποιήσιμος. Είναι γνωστό ότι το 3SAT είναι NP-complete πρόβλημα (Karp 1972) [30].

Για να μοντελοποιήσουμε το 3SAT ως CSP χρειαζόμαστε τέσσερις σχέσεις, τις $R_0 = \{0, 1\}^3 \setminus \{000\}$, $R_1 = \{0, 1\}^3 \setminus \{001\}$, $R_2 = \{0, 1\}^3 \setminus \{011\}$ και $R_3 = \{0, 1\}^3 \setminus \{111\}$. Έτσι για παράδειγμα μπορούμε να φτιάξουμε τους περιορισμούς $(x \vee y \vee z) = R_0(x, y, z), (x \vee y \vee \bar{z}) = R_1(x, y, z), (x \vee \bar{y} \vee z) = R_1(x, z, y), (\bar{x} \vee y \vee \bar{z}) = R_2(y, x, z)$, κ.λ.π. Είναι φανερό λοιπόν ότι το 3SAT αντιστοιχεί στο CSP($\{R_0, R_1, R_2, R_3\}$).

Για να μπορέσουμε να διατυπώσουμε το θεώρημα του Schaefer πρέπει πρώτα να αναφέρουμε κάποιους ορισμούς για του περιορισμούς.

Ορισμός 2.2.2. *Ένας όρος θα ονομάζεται:*

- *Horn αν περιέχει το πολύ ένα θετικό literal,*
- *δυϊκός Horn αν περιέχει το πολύ ένα αρνητικό literal,*
- *διζευκτικός αν περιέχει το πολύ δύο literals*
- *αφινικός αν μπορεί να περιγραφεί από μια αφινική εξίσωση modulo 2.*

Επίσης ένας τύπος $\phi = c_1 \wedge \cdots \wedge c_p$ ονομάζεται Horn, δυϊκός Horn, διζευκτικός ή αφινικός αν κάθε όρος c_i είναι Horn, δυϊκός Horn, διζευκτικός ή αφινικός αντίστοιχα.

Για τον παραπάνω ορισμό θεωρούμε πάντα ότι οι τύποι είναι σε CNF μορφή. Ας δούμε τώρα μερικά παραδείγματα για τους παραπάνω ορισμούς:

- ο $c_1 := (\bar{x} \vee \bar{y} \vee z)$ είναι Horn,
- ο $c_2 := (\bar{x} \vee y \vee z)$ είναι δυϊκός Horn,

- ο $c_3 := (\bar{x} \vee y)$ είναι διζευκτικός, Horn και δυϊκός Horn,
- ο $c_4 := (x \vee y)$ είναι διζευκτικός και δυϊκός Horn,
- ο $c_5 := (x + y + z = 1) mod 2$ είναι αφινικός,
- ο $c_6 := (x = y)$ είναι αφινικός, διζευκτικός, Horn και δυϊκός Horn,
- ο $c_7 := (x \neq y)$ είναι αφινικός και διζευκτικός,

Ενώ το γενικό SAT πρόβλημα είναι NP-complete, υπάρχουν κάποιες ειδικές περιπτώσεις του που ανήκουν στο P. Μία από αυτές είναι το HornSat, στο οποίο μας δίνεται ένας πεπερασμένος Horn τύπος ϕ και μας ρωτάνε αν είναι ικανοποιήσιμος.

Όμοια μπορούμε να ορίσουμε και το πρόβλημα DualHornSat, στο οποίο ως είσοδος δίνεται ένας δυϊκός Horn τύπος. Ένας τύπος $\phi(x_1, \cdots, x_k)$ είναι Horn αν και μόνο αν ο τύπος $\phi(\bar{x_1}, \cdots, \bar{x_k})$ είναι δυϊκός Horn. Συμπεραίνουμε λοιπόν ότι και το DualHornSatλύνεται σε πολυωνυμικό χρόνο.

Αν τώρα περιορίσουμε τις εισόδους του προβλήματος να είναι μόνο διζευκτικοί τύποι, τότε προκύπτει το πρόβλημα 2SAT, το οποίο είναι γνωστό ότι ανήκει και αυτό στο P.

Τέλος, μπορούμε να ορίσουμε το πρόβλημα AffineSat, στο οποίο δίνεται ως είσοδος ένας πεπερασμένος αφινικός τύπος ϕ και μας ρωτάνε αν είναι ικανοποιήσιμος. Ένας άλλος τρόπος να δούμε το ίδιο πρόβλημα είναι ο εξής, μας δίνεται ως είσοδος ένα αφινικό σύστημα εξισώσεων $A\vec{x} = b$ στον δακτύλιο $\mathbb{Z}_2$ και μας ρωτάνε αν έχει λύση. Το AffineSat λύνεται κι αυτό σε πολυωνυμικό χρόνο αφού η ύπαρξει λύσης του συστήματος $A\vec{x} = b$ μπορεί να ελεγθεί με τη μέθοδο απαλοιφής Gauss.

Ας δούμε ένα παράδειγμα φυσικού προβλήματος που λύνεται ως AffineSat. Στο πρόβλημα του 2-χρωματισμού γραφήματος μας δίνεται ένα γράφημα $G = (V, E)$ με V το σύνολο κορυφών και το σύνολο των ακμών και μας ρωτάνε αν υπάρχει συνάρτηση $f : V \to \{0, 1\}$ τέτοια ώστε να ισχύει $f(x) \neq f(y)$ για κάθε ακμή (x, y) του γραφήματος. Θεωρούμε τη σχέση $2col = \{01, 10\}$, η οποία αντιστοιχεί στον αφινικό περιορισμό $2col(x, y) = (x + y = 1) mod 2$. Για κάθε ακμή λοιπόν $(x, y) \in E$ έχουμε έναν περιορισμό $(x + y = 1)$, και όλοι μαζί οι περιορισμοί φτιάχνουν ένα αφινικό σύστημα εξισώσεων $A\vec{x} = b$ στο $\mathbb{Z}_2$, που μπορούμε στη συνέχεια να λύσουμε σε πολυωνυμικό χρόνο με τη μέθοδο απαλοιφής Gauss και άρα να λύσουμε το πρόβλημα 2-χρωματισμού του G.

Ανακεφαλαιώνοντας, γνωρίζουμε ότι το πρόβλημα 3SAT είναι NP-complete, αλλά τα προβλήματα HornSat, DualHornSat, 2SAT και AffineSat ανήκουν όλα στο P.

Ένα άλλο πρόβλημα που θα μας απασχολήσει στην πορεία είναι το πρόβλημα Nae3Sat. Σ'αυτό το πρόβλημα ως είσοδο δεχόμαστε πάλι έναν πεπερασμένο τύπο

$\phi = c_1 \wedge \cdots \wedge c_p$ σε CNF μορφή με κάθε όρο να περιέχει ακριβώς 3 literals. Το ερώτημα όμως αυτή τη φορά είναι αν υπάρχει αποτίμηση I των μεταβλητών έτσι ώστε ο ϕ να είναι αληθής και σε κάθε όρο c_i να υάρχει τουλάχιστον ένα literal που αποτιμάται σε 1 και τουλάχιστον ένα literal που αποτιμάται σε 0. Ο Schaefer το 1978 [35] έδειξε ότι το NAE3SAT είναι NP-complete.

Ας δούμε τώρα πως μπορούμε να μοντελοποιήσουμε το NAE3SAT σαν CSP. Έστω ότι έχουμε μια παράμετρο S και ας υποθέσουμε ότι στο συνολο S περιέχεται η σχέση $nae = \{001, 010, 011, 100, 101, 110\}$. Εύκολα μέσω της nae μπορούμε να φτιάξουμε τον περιορισμό $neq(x, y) = nae(x, y, y)$, που δήλώνει ότι τα x, y είναι διαφορετικά μεταξύ τους. Με τη βοήθεια των neq και nae μπορούμε τώρα να ορίσουμε τους παρακάτω περιορισμούς

- $nae_1(x, y, z) = nae(x, y, \bar{z}) = \exists v(neq(z, v) \wedge nae(x, y, v))$
- $nae_2(x, y, z) = nae(x, \bar{y}, \bar{z}) = \exists v(neq(y, v) \wedge nae_1(x, v, z))$
- $nae_3(x, y, z) = nae(\bar{x}, \bar{y}, \bar{z}) = \exists v(neq(x, v) \wedge nae_2(v, y, z))$

Παρατηρούμε ότι παρόμοια με το 3SAT αυτοί οι περιορισμοί μας αρκούν για να μοντελοποιήσουμε το NAE3SAT. Το NAE3SAT λοιπόν είναι το CSP($\{nae\}$), άρα το CSP($\{nae\}$) είναι NP-complete.

Θα ορίσουμε τώρα δύο τετριμμένες περιπτώσεις σχέσεων.

Ορισμός 2.2.3. *Έστω ένα $d \in \{0, 1\}$. Μια σχέση R ονομάζεται d-έγκυρη αν περιέχει το διάνυσμα $(d, d, \cdots, d)$. Ένα σύνολο S από σχέσεις ονομάζεται d-έγκυρο αν κάθε σχέση $R \in S$ είναι d-έγκυρη.*

Είναι φανερό ότι αν το S είναι 0-έγκυρο ή 1-έγκυρο τότε το CSP(S) ανήκει στο P, αφού αν δώσουμε σε κάθε μεταβλητή την τιμή 0 ή 1 αντίστοιχα τότε προκύπτει μια προφανής λύση.

Τώρα πλέον μπορούμε να διατυπώσουμε το θεώρημα διχοτομίας του Schaefer [35].

Θεώρημα 2.2.4. *Έστω S ένα σύνολο από σχέσεις στο δυαδικό πεδίο $\{0, 1\}$. Αν το S είναι*

- *0-έγκυρο,*
- *ή 1-έγκυρο*
- *ή Horn*
- *ή δυϊκό Horn,*
- *ή διζευκτικό,*

- *ή αφινικό,*

τότε το CSP*(S) ανήκει στο* P*. Σε κάθε άλλη περίπτωση το* CSP*(S) είναι* NP-*complete.*

Η πορεία της απόδειξης έχει ως εξής, έχουμε ήδη δει όλες τις εύκολες περιπτώσεις, δηλαδή τις περιπτώσεις για τις οποίες το CSP(S) ανήκει στο P. θα δείξουμε ότι σε κάθε άλλη περίπτωση, δηλαδή αν το S δεν ικανοποιεί μια από τις έξι συνθήκες του θεωρήματος, τότε το S θα περιέχει τη σχέση nae που αναφέραμε παραπάνω. Σ'αυτή την περίπτωση και αφού το CSP($\{nae\}$) είδαμε ότι είναι NP-complete άρα και το CSP(S) θα είναι NP-complete.

Να σημειώσουμε εδώ ότι η απόδειξη που θα αναφέρουμε δεν είναι η πρωτότυπη απόδειξη του Schaefer, η οποία είναι πιο πολύπλοκη, αλλά βασίζεται στο [1] και πριέχει στοιχεία καθολικής άλγεβρας. Πριν προχωρήσουμε στο κύριο μέρος της απόδειξης θα πρέπει να αναφέρουμε μερικούς ακόμη ορισμούς.

Ορισμός 2.2.5. *Καρτεσιανό γινόμενο (cartesian product) δύο σχέσεων* $R_1 \subseteq D^k$ *και* $R_2 \subseteq D^m$ *ονομάζεται η σχέση* $R_1 \times R_2 = \{z \in D^{k+m} \mid \exists x \in R_1, y \in R_2 : z_i = x_i$ *για* $i = 1, 2, \cdots, k$ *και* $z_{k+j} = y_j$ *για* $j = 1, 2, \cdots, m\}$.

Για παράδειγμα αν έχουμε $R_1 = \{(01), (23)\}$ και $R_2 = \{(456), (789)\}$, τότε $R_1 \times R_2 = \{(01456), (01789), (23456), (23789)\}$.

Ορισμός 2.2.6. *Προβολή (projection) μιας σχέσης* $R \subseteq D^k$ *στην* t*-οστή συντεταγμένη είναι η σχέση* $\mathrm{pr}_t R = \{z \in D \mid \exists x \in R : x_t = z\}$.

Παρόμοια μπορεί να οριστεί και η προβολή μιας σχέσης σε περισσότερες της μιας συντεταγμένες. Για παράδειγμα αν $R = \{(123), (456), (789)\}$, τότε η προβολή της R στην πρώτη και τρίτη συντεταγμένη είναι η σχέση $\mathrm{pr}_{1,3} R = \{(13), (46), (79)\}$ ($= \mathrm{pr}_1 R \times \mathrm{pr}_3 R$).

Ορισμός 2.2.7. *Συνταύτιση (identification) μιας σχέσης* $R \subseteq D^k$ *στις συντεταγμένες* s *και* t *ονομάζεται η σχέση* $R' = \{z \in D^{k-1} \mid \exists x \in R : x_s = x_t$ *και* $z_i = x_i$ *για* $i = 1, 2, \cdots, t-1, t+1, \cdots, k\}$.

Για παράδειγμα αν $R = \{(122), (345), (777)\}$, τότε η συνταύτιση της R στις δύο τελευταίες συντεταγμένες είναι η σχέση $R' = \{(12), (77)\}$. Με άλλα λόγια η R' είναι το συνολο λύσεων του περιορισμού $R'(x, y)$ που ορίζεται ως $R'(x, y) = R(x, y, y)$.

Ορισμός 2.2.8. *Έστω ένα σύνολο* S *από σχέσεις. Θα λέμε ότι το* S *είναι κλειστό ως προς την πράξη op αν* $\forall R \in S, op(R) \in S$*. Με άλλα λόγια, δε μπορούμε εφαρμόζοντας την πράξη op μέσα στο* S *να προκύψει μια σχέση έξω από αυτό.*

Ορισμός 2.2.9. *Έστω ένα σύνολο σχέσεων S. Μπορούμε να κατασκευάσουμε ένα νέο σύνολο S' το οποίο θα περιέχει όλες τις σχέσεις του S, όλες τις προβολές των σχέσεων του S, όλες τις προβολές των προβολών των σχέσεων του S, κ.λ.π. Αυτό το νέο σύνολο S' το ονομάζουμε κλειστότητα προβολών (projection closure) του S.*

Για παράδειγμα αν έχουμε $S = \{R\}$ με $R = \{(123), (456), (789)\}$, τότε φτιάχνουμε τις σχέσεις

- $R_1 := \mathrm{pr}_{2,3}R = \{(23), (56), (89)\}$,
- $R_2 := \mathrm{pr}_{1,3}R = \{(13), (46), (79)\}$,
- $R_3 := \mathrm{pr}_{1,2}R = \{(12), (45), (78)\}$,
- $R_4 := \mathrm{pr}_2 R_1 = \{(3), (6), (9)\}$,
- $R_5 := \mathrm{pr}_1 R_1 = \{(2), (5), (8)\}$,
- $R_6 := \mathrm{pr}_1 R_2 = \{(1), (4), (7)\}$,

Η κλειστότητα προβολών του S είναι το σύνολο $S' = S \cup \bigcup_{1\leq i\leq 6} \{R_i\}$.

Έστω eq η σχέση ισότητας, δηλαδή $eq = \{(d, d) \mid d \in D\}$.

Ορισμός 2.2.10. *Σύγκλωνο (co-clone) ενός συνόλου σχέσεων S, που συμβολίζεται ως $\langle S\rangle$, ονομάζεται το μικρότερο (άπειρο) σύνολο σχέσεων τέτοιο ώστε:*

- *$S \subseteq \langle S\rangle$,*
- *$eq \in \langle S\rangle$,*
- *το $\langle S\rangle$ είναι κλειστό ως προς το καρτεσιανό γινόμενο,*
- *το $\langle S\rangle$ είναι κλειστό ως προς τις προβολές,*
- *το $\langle S\rangle$ είναι κλειστό ως προς την συνταύτιση.*

Ο λόγος που ενδιαφερόμαστε για τα σύγκλωνα είναι το παρακάτω θεώρημα.

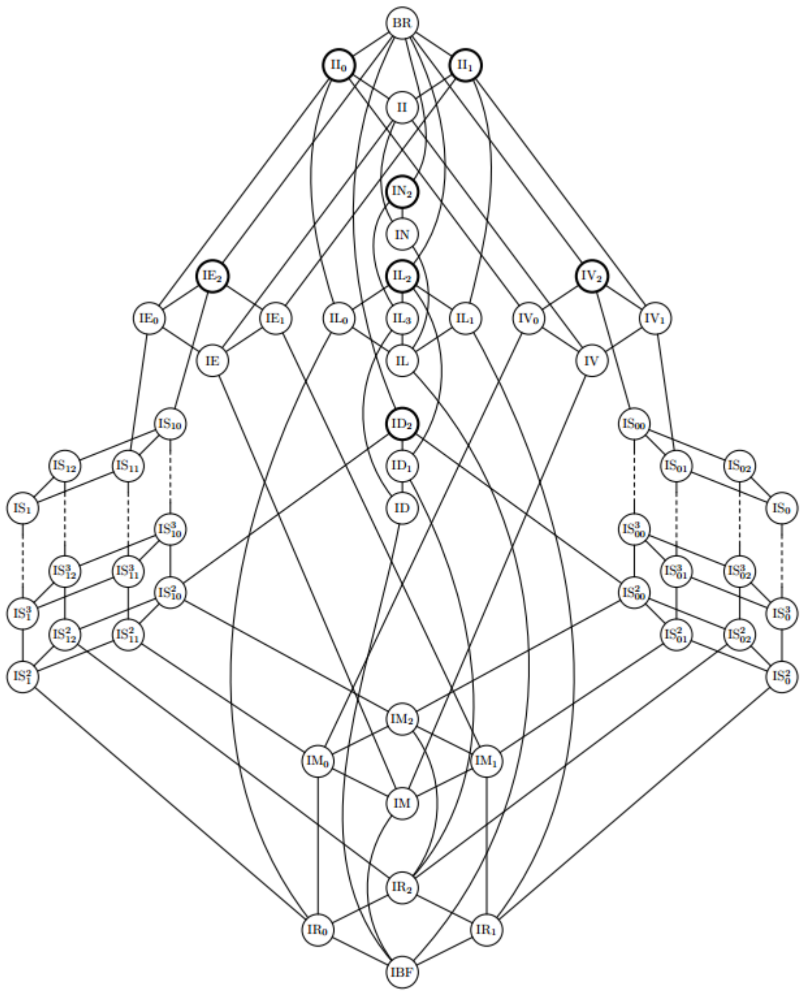
Θεώρημα 2.2.11. *Για κάθε σύνολο σχέσεων S ισχύει ότι $\mathrm{CSP}(S) \equiv_P \mathrm{CSP}(\langle S\rangle)$*

Με άλλα λόγια τα σύγκλωνα είναι κάτι σαν κανονικές μορφές των συνόλων σχέσεων. Μπορεί να έχουμε δύο διαφορετικά σύνολα S_1, S_2, αλλά να ισχύει $\langle S_1\rangle = \langle S_2\rangle$, οπότε λόγω του παραπάνω θεωρήματος δε χρειάζεται να τα αντιμετωπίσουμε ξεχωριστά όσον αφορά την πολυπλοκότητα του CSPτους.

Πόρισμα 2.2.12. *Έστω S_1 και S_2 δύο διαφορετικά μεταξύ τους σύνολα σχέσεων. Αν $\langle S_1 \rangle = \langle S_2 \rangle$ τότε ισχύει ότι CSP(S_1)$\equiv_P$ CSP(S_2).*

Ορισμός 2.2.13. *Δικτυωτό (lattice) ονομάζουμε ένα μερικώς διατεταγμένο σύνολο στο οποίο οποιαδήποτε δύο στοιχεία έχουν ένα μοναδικό ελάχιστο άνω φράγμα (supremum) και ένα μοναδικό μέγιστο κάτω φράγμα (infimum).*

Αν S είναι το σύνολο όλων των συγκλώνων όλων των διαφορετικών σχέσεων που ορίζονται στο πεδίο $D = \{0, 1\}$, τότε η δομή $(S, \subseteq)$, δηλαδή το S με τη σχέση διάταξης υποσυνόλου, αποτελεί ένα δικτυωτό. Το καλό με αυτό το δικτυωτό στο δυαδικό πεδίο είναι ότι γνωρίζουμε πλήρως τη δομή του.



Στο παραπάνω δικτυωτό κάθε κυκλάκι αντιστοιχεί σε ένα σύγκλωνο στο δυαδικό πεδίο. Αν δυο σύγκλωνα ενώνονται με γραμμή τότε αυτό σημαίνει ότι το σύγκλωνο που βρίσκεται πιο κάτω είναι υποσύνολο αυτού που βρίσκεται από πάνω.

Το επόμενο θεώρημα [29, 28, 27] είναι θεμέλιος λίθος για την απόδειξη του θεωρήματος του Schaefer.

Θεώρημα 2.2.14. *Έστω δύο σύνολα σχέσεων S_1, S_2 με το S_1 πεπερασμένο. Αν ισχύει ότι $S_1 \subseteq \langle S_2 \rangle$, τότε ισχύει και η αναγωγή* CSP*(S_1)*$\leq_P$ CSP*(S_2).*

Από το παραπάνω θεώρημα μπορούμε να συμπεράνουμε τα εξής:

1. Αν για ένα σύγκλωνο S ισχύει ότι CSP(S)$\in$ P, τότε για όλα τα σύγκλωνα S' που βρίσκονται κάτω από το S, ισχύει επίσης ότι CSP(S')$\in$ P.

2. Αν για ένα σύγκλωνο S ισχύει ότι CSP(S)$\in$ NP-complete, τότε για όλα τα σύγκλωνα S' που βρίσκονται πάνω από το S, ισχύει ότι CSP(S')$\in$ NP-complete.

Ας δούμε τώρα ξανά το θεώρημα του Schaefer σε συνδυασμό με το δικτυωτό των σύγκλωνων. Τα σύγκλωνα που αντιστοιχούν στις έξι εύκολες περιπτώσεις του θεωρήματος είναι τα εξής:

- 0-έγκυρο: iI_0,
- 1-έγκυρο: iI_1,
- Horn: iE_2
- δυϊκό Horn: iV_2,
- διζευκτικό: iD_2,
- αφινικό: iL_2.

Για αυτά λοιπόν τα σύγκλωνα έχουμε ήδη δει ότι τα αντίστοιχα CSP ανήκουν στο P. Λόγω της παραπάνω παρατήρησης και για όλα τα σύγκλωνα που βρίσκονται από κάτω τους θα ισχύει το ίδιο. Αυτό αφήνει ανεξερεύνητα δύο μόνο σύγκλωνα, το BR, που αντιστοιχεί σε όλες τις δυνατές σχέσεις στο δυαδικό πεδίο $D = \{0, 1\}$, και το iN_2, που είναι το σύγκλωνο της σχέσης nae. Έχουμε όμως δει ότι το CSP(nae) είναι NP-complete, άρα τα CSPπου αντιστοιχούν στα σύγκλωνα BR και iN_2 είναι NP-complete, και αυτό ολοκληρώνει την απόδειξη του θεωρήματος διχοτομίας του Schaefer.

Το θεώρημα του Schaefer μπορεί να εξειδικευτεί ακόμη πιο πολύ. Πιο συγκεκριμένα στο [1] μπορούμε να δούμε αποτελέσματα πληρότητας σε υποκλάσεις του P ($\oplus$L, NL, L) για τα CSP που ανήκουν στο P.

2.3 Το Γενικό Μοντέλο

Το 2002 ο Bulatov έδωσε ένα αποτέλεσμα διχοτομίας παρόμοιο με του Schaefer, αλλά για πεδίο με τρία στοιχεία, δηλαδή για $D = \{0, 1, 2\}$ [3]. Η γενική περίπτωση για οποιοδήποτε πεπερασμένο πεδίο παραμένει ανοιχτή μέχρι σήμερα [22].

Κεφάλαιο 3

Προβλήματα Μέτρησης

Σ' αυτό το κεφάλαιο θα παρουσιάσουμε τρεις σκελετούς προβλημάτων που μοντελοποιούν πολλά προβλήματα μέτρησης. Πιο συγκεκριμένα θα μιλήσουμε για ομομορφισμούς γραφημάτων (graph homomorphisms), προβλήματα μέτρησης ικανοποίησης περιορισμών ($\#CSP$) και προβλήματα Holant και θα δείξουμε ότι με τη σειρά που τα αναφέραμε κάθε ένα είναι γενίκευση του προηγούμενου. Γι' αυτά τα προβλήματα έχουν αποδειχθεί διάφορα θεωρήματα διχοτομίας τα οποία και θα δούμε. Τέλος θα κάνουμε μια σύντομη εισαγωγή στους ολογραφικούς αλγόριθμους (holographic algorithms) που είναι πολυωνυμικοί αλγόριθμοι συγκεκριμένου τύπου για προβλήματα μέτρησης.

3.1 Ορισμοί και Παραδείγματα

Το πρώτο και πιο απλό μοντέλο προβλήματος που θα δούμε είναι το πρόβλημα ομομορφισμών γραφήματος (graph homomorphisms problem). Έστω $\mathbf{A} = (A_{i,j}) \in \mathbb{C}^{q \times q}$ ένας πίνακας, ο οποίος εν γένει μπορεί να έχει μιγαδικά στοιχεία. Αυτός ο πίνακας είναι και η παράμετρος του προβλήματος. Το πρόβλημα ομομορφισμών γραφήματος, $\mathrm{EVAL}(\mathbf{A})$, είναι δοσμένου ενός γραφήματος $G = (V, E)$, να υπολογίσουμε τη συνάρτηση διαμέρισης

$$Z_{\mathbf{A}}(G) = \sum_{\sigma: V \to [q]} \prod_{(u,v) \in E} A_{\sigma(u),\sigma(v)}.$$

Αν ο $\mathbf{A}$ είναι συμμετρικός πίνακας και το G είναι μη κατευθυνόμενο γράφημα, τότε έχουμε το πρόβλημα ομομορφισμών μη κατευθυνόμενου γραφήματος.

Ανάλογα τώρα με το τι θα επιλέξουμε ως πίνακα $\mathbf{A}$, μπορεί να προκύψει και ένα διαφορετικό πρόβλημα και γενικά η συνάρτηση $Z_{\mathbf{A}}(G)$ μπορεί να εκφράσει πολλές ενδιαφέρουσες ιδιότητες. Ας δούμε ορισμένα παραδείγματα:

- Αν θέσουμε $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$, τότε η $Z_{\mathbf{A}}(G)$ μετράει το πλήθος των καλύψεων κορυφών του G.

- Αν θέσουμε $\mathbf{A} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$, τότε $Z_{\mathbf{A}}(G)$ μετράει το πλήθος των 3-χρωματισμών του G.

- Πιο γενικά, αν $\mathbf{A} = \begin{pmatrix} 0 & 1 & \cdots & 1 \\ 1 & 0 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 0 \end{pmatrix}$, τότε προκύπτει το πρόβλημα του k-χρωματισμού γραφήματος (ο πίνακας είναι $k \times k$).

- Επίσης ένα παράδειγμα στο οποίο ο πίνακας δε παίρνει μόνο τιμές 0 και 1 είναι το πρόβλημα του να μετρήσουμε το πλήθος των εναγόμενων υπογραφημάτων του G με άρτιο πλήθος ακμών. Αυτό το πρόβλημα είναι ισοδύναμο με το να θέσουμε $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$.

Το δεύτερο μοντέλο προβλημάτων που θα εξετάσουμε είναι το $\#CSP$. Έστω $D = \{1, 2, \cdots, d\}$ ένα πεδίο. Μια γλώσσα περιορισμών $\mathcal{L}$ είναι ένα πεπερασμένο σύνολο συναρτήσεων $\{f_1, \cdots, f_h\}$ για τις οποίες η $f_i : D^{r_i} \to \mathbb{C}$ είναι μια r_i-μελής συνάρτηση για κάποιο $r_i \geq 1$. Το πρόβλημα $\#\mathrm{CSP}(\mathcal{L})$ έχει ως είσοδο ένα σύνολο από n μεταβλητές $\mathbf{x} = (x_1, \cdots, x_n)$ και μια συλλογή I από m διανύσματα της μορφής $(f, i_1, \cdots, i_r)$ στα οποία η f είναι μια r-μελής συνάρτηση στο $\mathcal{L}$ και $i_1, \cdots, i_r \in [n]$. Το πρόβλημα ζητάει και πάλι να υπολογίσουμε τη συνάρτηση διαμέρισης η οποία τώρα είναι η

$$Z_{\mathcal{L}}(I) = \sum_{\mathbf{x} \in D^n} \prod_{(f, i_1, \cdots, i_r) \in I} f(x_{i_1}, \cdots, x_{i_r}).$$

Αν οι συναρτήσεις στην $\mathcal{L}$ έχουν σαν πεδίο τιμών το $\{0, 1\}$, τότε μπορούμε να τις δούμε σαν σχέσεις και έχουμε το πρόβλημα $\#\mathrm{CSP}(\mathcal{F})$ χωρίς βάρη.

Όπως βλέπουμε, το $\#CSP$ είναι το αντίστοιχο πρόβλημα μέτρησης του προβλήματος απόφασης που είδαμε στο προηγούμενο κεφάλαιο. Οπότε, όπως είναι αναμενόμενο, ένα παράδειγμα προβλήματος που μοντελοποιείται σαν $\#CSP$ είναι το $\#3SAT$. Πιο συγκεκριμένα μπορούμε να θέσουμε $D = \{0, 1\}$ και $\mathcal{L} =$

$\{f_0, f_1, f_2, f_3\}$, όπου

$$\begin{aligned} f_0(x,y,z) &= x \vee y \vee z, \\ f_1(x,y,z) &= \bar{x} \vee y \vee z, \\ f_2(x,y,z) &= \bar{x} \vee \bar{y} \vee z, \\ f_3(x,y,z) &= \bar{x} \vee \bar{y} \vee \bar{z}, \end{aligned}$$

οπότε βλέπουμε ότι το $\#\mathrm{CSP}(\mathcal{L})$ είναι ακριβώς το πρόβλημα $\#3SAT$.

Αναφέραμε στην εισαγωγή του κεφαλαίου ότι το $\#CSP$ είναι γενίκευση του προβλήματος ομομορφισμών γραφήματος. Αυτό συμβαίνει γιατί αν στην $\mathcal{L}$ θεωρήσουμε μία μόνο συνάρτηση η οποία είναι διμελής και εκφράζεται από τον πίνακα $\mathbf{A}$ και επίσης αντιστοιχήσουμε τις μεταβλητές $x_1, \cdots, x_n$ στις κορυφές του γραφήματος G και τους περιορισμούς του I στις ακμές του G, τότε έχουμε μοντελοποιήσει το πρόβλημα ομομορφισμών γραφήματος σαν $\#CSP$, άρα το πρώτο είναι ειδική περίπτωση του δεύτερου.

Το τρίτο και πιο γενικό μοντέλο προβλήματος που θα δούμε είναι πρόβλημα Holant. Έστω $[q]$ ένα πεδίο και F ένα πεπερασμένο σύνολο συναρτήσεων με μιγαδικές τιμές και ορίσματα από το $[q]$. Το πρόβλημα $\mathrm{Holant}(F)$ είναι το εξής: είσοδος είναι μια τριάδα $\Omega = (G, F, \pi)$ (signature grid), όπου $G = (V, E)$ είναι ένα γράφημα με ετικέτες στις κορυφές, και π είναι η συνάρτηση που σε κάθε κορυφή $v \in V$ αντιστοιχεί ως ετικέτα μια συνάρτηση $f_v \in F$ έτσι ώστε το πλήθος ορισμάτων που δέχεται η f_v να είναι ίσο με το βαθμό της v. Ζητούμενο του προβλήματος είναι ο υπολογισμός της συνάρτησης

$$Holant_\Omega = \sum_{\sigma : E \to [q]} \prod_{v \in V} f_v(\sigma \mid_{E(v)}).$$

Ένα παράδειγμα προβλήματος το οποίο μοντελοποιείται σαν Holant αλλά έχει αποδειχθεί ότι δεν μοντελοποιείται σαν πρόβλημα ομομορφισμών γραφήματος [23] είναι το πρόβλημα μέτρησης τέλειων ταιριασμάτων σε ένα γράφημα. Συγκεκριμένα για να το μοντελοποιήσουμε σαν πρόβλημα Holant θέτουμε $[q] = \{0, 1\}$ και F να περιέχει όλες τις ExactOne συναρτήσεις, δηλαδή τις συναρτήσεις που επιστρέφουν 1 αν ακριβώς ένα όρισμά τους είναι 1 και 0 διαφορετικά. Σε κάθε κορυφή του G αντιστοιχούμε μια ExactOne συνάρτηση με το κατάλληλο πλήθος ορισμάτων. Το γινόμενο $\prod_{v \in V} f_v(\sigma \mid_{E(v)})$ ισούται με 1 αν το $\sigma^{-1}(1) \subseteq E$ είναι ένα τέλειο ταίριασμα και ισούται με 0 διαφορετικά. Οπότε το Holant_Ω μετράει το πλήθος των τέλειων ταιριασμάτων του γραφήματος. Αν αντί για τις ExactOne συναρτήσεις χρησιμοποιήσουμε τις AtMostOne συναρτήσεις, τότε μετράμε το πλήθος όλων των (όχι απαραίτητα τέλειων) ταιριασμάτων.

Ορισμός 3.1.1. *Έστω A ένα σύνολο συναρτήσεων. Ορίζουμε το πρόβλημα* $\mathrm{Holant}^{\mathcal{A}}$ *ως υποπερίπτωση του Holant να είναι το πρόβλημα*

$$\mathrm{Holant}^{\mathcal{A}}(\mathcal{F}) = \mathrm{Holant}(\mathcal{F} \cup \mathcal{A}).$$

Σ'αυτή την υποπερίπτωση, καλούμε το $\mathcal{A}$, ελεύθερα διαθέσιμες συναρτήσεις (freely available functions).

Αν θεωρήσουμε όλες τις συναρτήσεις ισότητας ως ελεύθερα διαθέσιμες, τότε η υποπερίπτωση του Holant που δημιουργείται είναι ακριβώς το πρόβλημα #CSP. Με άλλα λόγια ισχύει ότι

$$\#\mathrm{CSP}(\mathcal{F}) = \mathrm{Holant}(\mathcal{F} \cup \mathit{Equalities}).$$

Για να δούμε ότι το Holant είναι γενίκευση του #CSP κάνουμε τα εξής: αναπαριστούμε ένα στιγμιότυπο του #CSP από ένα διμερές γράφημα, όπου το αριστερό μέρος (LHS) αντιστοιχεί στις μεταβλητές και το δεξί μέρος (RHS) αντιστοιχεί στους περιορισμούς (στις συναρτήσεις). Η τριάδα εισόδου Ω του Holant είναι η εξής: Σε κάθε κορυφή-μεταβλητή στο LHS αντιστοιχούμε μια συνάρτηση ισότητα (Equality function) και σε κάθε κορυφή-περιορισμό στο RHS δίνουμε ως ετικέτα τον περιορισμό-συνάρτηση που αντιστοιχεί σ' αυτή τη κορυφή. Η Equality συνάρτηση σε κάθε κορυφή-μεταβλητή αναγκάζει τα προσκείμενες ακμές αν πάρουν την ίδια τιμή. Αυτό κάνει τις αναθέσεις τιμών στις ακμές ουσιαστικά να είναι αναθέσεις τιμών στις κορυφές που με της σειρά τους είναι αναθέσεις τιμών στις μεταβλητές του LHS όπως στο #CSP. Βλέπουμε λοιπόν ότι το #CSP είναι ειδική περίπτωση του προβλήματος Holant.

Ας δούμε τώρα και κάποιες άλλες ειδικές περιπτώσεις του Holant.

Ορισμός 3.1.2. *Έστω U το σύνολο όλων των μονομελών συναρτήσεων. Τότε ορίζουμε*

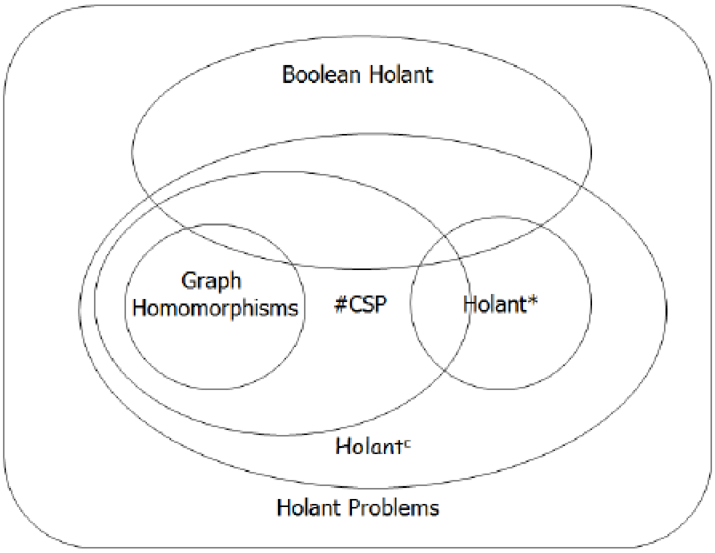
$$\mathrm{Holant}^{\star}(\mathcal{F}) = \mathrm{Holant}(\mathcal{F} \cup U).$$

Ορισμός 3.1.3. *Έστω Δ_i η μονομελής συνάρτηση που δίνει τιμή 1 σε είσοδο $i \in [q]$, και τιμή 0 σε κάθε άλλη είσοδο. Τότε ορίζουμε,*

$$\mathrm{Holant}^{c}(\mathcal{F}) = \mathrm{Holant}(\mathcal{F} \cup \{\Delta_1, \cdots, \Delta_q\}).$$

Είναι φανερό από τον ορισμό ότι το Holant^{c} πειέχει ως υποπερίπτωση το $\mathrm{Holant}^{\star}$. Κάτι άλλο όμως που επίσης ισχύει είναι ότι το Holant^{c} περιέχει ως υποπερίπτωση και το #CSP [18].

Ό λόγος που ορίζουμε υποπεριπτώσεις του Holant είναι γιατί ενώ ιδανικά θα θέλαμε να βρούμε θεωρήματα διχοτομίας για όλη την κλάση Holant, αν δε μπορούμε να το κάνουμε αυτό, τότε προσπαθούμε να βρούμε διχοτομίες σε όσο το δυνατόν μεγαλύτερες υποκλάσεις. Η εικόνα λοιπόν των κλασεων που έχουμε ορίσει, είναι η παρακάτω.



3.2 Ομομορφισμοί Γραφήματος

Θα δούμε τώρα συνοπτικά τα αποτελέσματα διχοτομίας που έχουν αποδειχθεί για τις κλάσεις προβλημάτων που ορίσαμε.

Πρώτα θα δούμε την περίπτωση των ομομορφισμών μη κατευθυνόμενου γραφήματος, δηλαδή την περίπτωση που ο πίνακας είναι συμμετρικός. Αν επιπλέον ο πίνακας περιέχει τιμές μόνο στο $\{0, 1\}$, τότε οι Dyer και Greenhill [20] έδειξαν ότι αν το γράφημα $H_{\mathbf{A}}$ που αντιστοιχεί στον πίνακα $\mathbf{A}$, δεν είναι πλήρες (complete) ή πλήρως διμερές (complete bipartite) τότε το πρόβλημα είναι στο P. Σε κάθε άλλη περίπτωση το πρόβλημα είναι #P-complete.

Οι Bulatov και Grohe [8] στη συνέχεια επέκτειναν το αποτέλεσμα για πίνακες με μη αρνητικά βάρη και στο επόμενο κεφάλαιο θα δούμε περισσότερα για το κριτήριό τους. Αργότερα οι Goldberg, Grohe, Jerrum και Thurley [24] απέδειξαν το αντίτοιχο θεώρημα διχοτομίας για όλους τους συμμετρικούς πίνακες με στοιχεία πραγματικούς αριθμούς.

Τέλος, οι Cai, Chen και Lu [12] έδειξαν το παρακάτω γενικότερο θεώρημα.

Θεώρημα 3.2.1. *Έστω* $\mathbf{A}$ *ένας συμμετρικός πίνακας με στοιχεία μιγαδικούς αριθμούς. Τότε το να μετρήσουμε τους ομομορφισμούς στο* $H_{\mathbf{A}}$ *γίνεται είτε σε πολυωνυμικό χρόνο, είτε το πρόβλημα είναι #P-hard.*

Για την περίπτωση που έχουμε και μη κατευθυνόμενα γραφήματα, δηλαδή για την περίπτωση που ο πίνακας δεν είναι απαραίτητα συμμετρικός, οι Dyer, Goldberg και Paterson [19] έδειξαν ένα θεώρημα διχοτομίας για μια οικογένεια ακυκλικών κατευθυνόμενων γραφημάτων και για πίνακες με στοιχεία στο $\{0, 1\}$. Το 2010 οι Cai και Chen [10] επέκτειναν το αποτέλεσμά τους για όλα τα κατευθυνόμενα γραφήματα και για πίνακες με μη-αρνητικά στοιχεία.

3.3 $\#CSP$

Ας δούμε τώρα τα αποτελέσματα διχοτομίας για το δυαδικό $\#CSP$ όπου το πεδίο είναι το $D = \{0, 1\}$. Για την περίπτωση που δεν έχουμε βάρη, δηλαδή οι συναρτήσεις είναι απλά σχέσεις, οι Creignou και Hermann [17] έδειξαν ότι η μόνη εύκολη περίπτωση είναι όταν έχουμε αφινικές σχέσεις, ενώ σε κάθε άλλη περίπτωση το πρόβλημα είναι #P-hard.

Μπορούμε να γενικεύσουμε την έννοια των αφινικών σχέσεων και να ορίσουμε τις αγνά αφινικές συναρτήσεις, οι οποίες είναι σαν τις αφινικές, απλά αντί για τιμές 0,1 παίρνουν τιμές 0, c, για έναν παράγοντα c. Έστω επίσης $\mathcal{P}$ η κλάση όλων των συναρτήσεων που μπορούν να εκφραστούν ως γινόμενο μονομελών συναρτήσεων, διμελών συναρτήσεων ισότητας και διμελών συναρτήσεων ανισότητας. Οι Dyer, Goldberg και Jerrum [18] έδειξαν το παρακάτω θεώρημα.

Θεώρημα 3.3.1. *Έστω $\mathcal{F}$ ένα σύνολο από μη-αρνητικές συναρτήσεις στο δυαδικό πεδίο. Το* #CSP($\mathcal{F}$) *είναι* #P-*hard εκτός εάν όλες οι συναρτήσεις στο $\mathcal{F}$ είναι αγνά αφινικές ή ανήκουν στην κλάση $\mathcal{P}$, οπότε σε αυτή την περίπτωση το πρόβλημα ανήκει στο* P.

Το 2009 οι Bulatov, Dyer, Goldberg, Jalsenius και Richerby [7] απέδειξαν θεώρημα διχοτομίας για το δυαδικό $\#CSP$ όπου οι συναρτήσεις παίρνουν τιμές σε όλους του πραγματικούς αριθμούς, και ανεξάρτητα οι Cai, Lu και Xia [15] έκαναν το ίδιο για τους μιγαδικούς αριθμούς.

Για τη γενική περίπτωση στην οποία το πεδίο D μπορεί να είναι οποιοδήποτε πεπερασμένο σύνολο, ο Boulatov [4] απέδειξε διχοτομία για την περίπτωση του $\#CSP$ χωρίς βάρη. Το κριτήριο όμως που έδωσε για τη διχοτομία δεν ήταν γνωστό αν είναι αποφασίσιμο. Οι Dyer και Richerby [21] στη συνέχεια απέδειξαν το ίδιο θεώρημα δίνοντας ένα απλούστερο κριτήριο και αποδεικνύοντας ότι είναι αποφασίσιμο. Φυσικά το νέο κριτήριο ήταν ισοδύναμο με αυτό του Bulatov. Θα μιλήσουμε πιο αναλυτικά για αυτά τα δύο papers στο επόμενο κεφάλαιο. Λίγο αργότερα οι Bulatov, Dyer, Goldberg, Jalsenius, Jerrum και Richerby [6] επέκτειναν το αποτέλεσμα των Dyer και Richerby για θετικά ρητά βάρη. Στη συνέχεια οι Cai, Chen και Lu [13] έδειξαν διχοτομία για όλα τα μη-αρνητικά βάρη. Τέλος, οι Cai και Chen [11] απέδειξαν τη γενική περίπτωση για βάρη μιγαδικούς αριθμούς, δίνοντας τρία κριτήρια διχοτομίας που το αν είναι ή όχι αποφασίσιμα παραμένει ανοιχτό. Για το τελευταίο αποτέλεσμα θα δούμε πιο αναλυτικά την απόδειξη και τα κριτήρια στο επόμενο κεφάλαιο.

3.4 Προβλήματα Holant

Θα δούμε τώρα τι αποτελέσματα διχοτομίας υπάρχουν για τα γενικότερα μοντέλα των Holant* και Holant^c.

Οι Cai, Lu και Xia [15] απέδειξαν διχοτομία για το συμμετρικό, δυαδικό μοντέλο Holant* με μιγαδικά βάρη. Αργότερα [16] επέκτειναν το αποτέλεσμά τους και για τη μη συμμετρική περίπτωση.

Στο ίδιο paper [15] έδωσαν επίσης διχοτομία για το συμμετρικό, δυαδικό μοντέλο του Holant^c με βάρη πραγματικούς αριθμούς. Τέλος ,οι Cai, Huang και Lu [14] επέκτειναν αυτό το αποτέλεσμα για την περίπτωση με μιγαδικά βάρη.

Οι περισσότερες περιπτώσεις για τη γενικότερη κλάση Holant παραμένουν ανοιχτές.

3.5 Ολογραφικοί Αλγόριθμοι

Σ' αυτή την παράγραφο θα κάνουμε μια εισαγωγή στους ολογραφικούς αλγόριθμους (holographic algorithms). Τους αλγόριθμους αυτούς τους εισήγαγε ο Valiant [37] και είναι πολυωνυμικοί αλγόριθμοι για διάφορα "εξωτικά" προβλήματα για τα οποία πιο πριν δε γνωρίζαμε ότι ανήκαν στο P. Η γενική ιδέα είναι ότι ένα πρόβλημα μέτρησης το ανάγουμε στο πρόβλημα μέτρησης τέλειων ταιριασμάτων σε επίπεδα γραφήματα (ολογραφική αναγωγή), για το οποίο έχουμε πολυωνυμικό αλγόριθμο, και έτσι προκύπτει πολυωνυμικός αλγόριθμος και για το αρχικό μας πρόβλημα.

Ορισμός 3.5.1. *Το πρόβλημα #PerfMatch είναι το εξής:*
Είσοδος: Ένα μη κατευθυνόμενο γράφημα $G = (V, E, W)$ με βάρη.
Ζητούμενο:

$$\mathrm{PerfMatch}(G) = \sum_{E'} \prod_{(i,j)\in E'} w_{i,j},$$

όπου η άθροιση γίνεται για όλα τα τέλεια ταιριάσματα E' του G.

Αν τα βάρη στις ακμές του γραφήματος είναι όλα ίσα με 1, τότε το PerfMatch(G) μετάει το πλήθος των τέλειων ταιριασμάτων του γραφήματος G. Το αντίστοιχο πρόβλημα απόφασης γνωρίζουμε ότι ανήκει στο P, αλλά το παραπάνω πρόβλημα μέτρησης είναι #P-complete [36].

Για την περίπτωση όμως των επίπεδων γραφημάτων, το #PerfMatch ανήκει στο FP (Fisher, Kasteleyn, Temberley, 1961). Πιο συγκεκριμένα ισχύει το παρακάτω θεώρημα.

Θεώρημα 3.5.2. *Υπάρχει μια συνάρτηση υπολογίσιμη σε πολυωνυμικό χρόνο, η οποία δοσμένης μια εμβάπτισης ενός επίπεδου γραφήματος $G = (V, E, W)$ στο επίπεδο, ορίζει την $f : E \to \{-1, 1\}$ έτσι ώστε για τον αντισυμμετρικό πίνακα M που ορίζεται έτσι ώστε για κάθε $i < j$*

- *αν $(i, j) \notin E$ τότε $M_{i,j} = M_{j,i} = 0$, και*
- *αν $(i, j) \in E$ τότε $M_{i,j} = f(i, j) \cdot w_{i,j}$ και $M_{j,i} = -f(i, j) \cdot w_{i,j}$,*

να ισχύει ότι $\mathrm{PerfMatch}(G) = \sqrt{Det(M)}$.

Η στρατηγική μας για μια ολογραφική αναγωγή (holographic reduction) είναι η εξής:

- Κάθε στοιχείο ενός στιγμιοτύπου I νός προβλήματος μέτρησης (π.χ. κορυφές και ακμές) θα τα αντικαταστήσουμε από σκαριφήματα (gadgets) που ονομάζουμε matchgates (πύλες ταιριασμάτων).
- Με αυτό τον τρόπο μετατρέπουμε το στιγμιότυπο I σε ένα στιγμιότυπο Ω που το ονομάζουμε matchgrid (πλέγμα ταιριασμάτων).
- Το βεβαρημένο άθροισμα των τέλειων ταιριασμάτων στο Ω θα ισούται με το πλήθος λύσεων του I.

Συνεπώς έχοντας αλγόριθμο για το #PerfMatch προκύπτει και αλγόριθμος για το αρχικό πρόβλημα μέτρησης.

Πάμε τώρα να δούμε τους βασικούς ορισμούς για τα παραπάνω.

Ορισμός 3.5.3. *Ένα επίπεδο matchgate Γ είναι μια τριάδα (G, X, Y) όπου G είναι μια επίπεδη εμβάπτιση ενός επίπεδου γραφήματος (V, E, W), $X \subseteq V$ είναι ένα σύνολο από κορυφές εισόδου, $Y \subseteq V$ είναι ένα σύνολο από κορυφές εξόδου, και $X \cap Y = \emptyset$.*

- *Το arity του matchgate είναι το $|X| + |Y|$.*
- *Η κανονική υπογραφή (standard signature) του Γ ως προς το σύνολο $Z \subseteq X \cup Y$ είναι η* $\mathrm{PerfMatch}(G - Z)$.
- *Η κανονική υπογραφή (standard signature) του Γ είναι ο $2^{|X|} \times 2^{|Y|}$ πίνακας $u(\Gamma)$ του οποίου τα στοιχεία είναι οι κανονικές υπογραφές του Γ ως προς τα σύνολα Z για όλες τις πιθανές επιλογές του Z.*

Ορισμός 3.5.4. *Μια βάση (μεγέθους 1) είναι ένα σύνολο από δύο διαφορετικά μημηδενικά διανύσματα, που τα ονομάζουμε n και p. Η βάση $\boldsymbol{b0} = [n, p] = [(0, 1), (1, 0)]$ ονομάζεται συνήθης βάση. Γενικά, τα διανύσματα μια βάσης δεν είναι απαραίτητο να είναι ανεξάρτητα μεταξύ τους.*

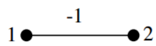
Παρακάτω θα χρησιμοποιήσουμε ως παράδειγμα τη βάση $\mathbf{b1} = [n, p] = [(-1, 1), (1, 0)]$.

Επίσης, εάν έχουμε δύο διανύσματα q, r με μήκη l, m αντίστοιχα, τότε ορίζουμε το τανυστικό γινόμενο τους $s = q \otimes r$ να είναι το διάνυσμα s με μήκος $l \cdot m$ για το οποίο $s_{i \cdot m + j} = q_i \cdot r_j$. Έτσι, για παράδειγμα, για τη βάση $\mathbf{b1}$, ισχύει $n \otimes p = (-1, 0, 1, 0)$.

Ορισμός 3.5.5. *Θα λέμε ότι ένα matchgate είναι γεννήτορας (generator) αν δεν έχει καθόλου κορυφές εισόδου, αλλά έχει κορυφές εξόδου.*

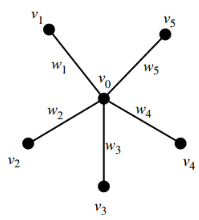
Παρόμοια, θα λέμε ότι ένα matchgate είναι αναγνωριστής (recognizer) αν δεν έχει καθόλου κορυφές εξόδου, αλλά έχει κορυφές εισόδου.

Ένα παράδειγμα ενός γεννήτορα Γ για τη βάση $\mathbf{b1}$ με κορυφές εξόδου $\{1, 2\}$ και μια ακμή με βάρος -1 είναι το ακόλουθο:



Η κανονική υπογραφή του παραπάνω γεννήτορα , $u(\Gamma)$, είναι το διάνυσμα $(-1, 0, 0, 1)$. Οπότε, θεωρούμε ότι ο γεννήτορας παράγει το $n \otimes n + n \otimes p + p \otimes n = (-1, 0, 0, 1)$. Η υπογραφή του γεννήτορα ως προς τη βάση $\mathbf{b1}$ θα λέμε ότι είναι το διάνυσμα (1,1,1,0). Για $x \in \{n, p\}^2$ ορίζουμε το $valG(\Gamma, x)$ να είναι το στοιχείο της υπογραφής που αντιστοιχεί στο x. Για το συγκεκριμένο παράδειγμα λοιπόν, θα είναι $valG(\Gamma, n \otimes p) = 1$ και $valG(\Gamma, p \otimes p) = 0$.

Ένα παράδειγμα ενός αναγνωριστή για τη βάση $\mathbf{b1}$ με κορυφές εισόδου τις $v_1, v_2, \cdots, v_5$ και βάρη στις ακμές τα $w_1, w_2, \cdots, w_5$ είναι το παρακάτω:



Ας φανταστούμε τώρα έναν αναγνωριστή σαν τον παραπάνω αλλά με k κορυφές εισόδου. Ο σκοπός ενός τέτοιου αναγνωριστή είναι να κάνουν το PerfMatch να παίρνει κατάλληλες τιμές όσο οι είσοδοι αλλάζουν ανάμεσα στα 2^k πιθανά διαφορετικά τανυστικά γινόμενα $x = x_1 \otimes x_2 \otimes \cdots \otimes x_k$, όπου $x_i \in \{n, p\}$. Αν u είναι η κανονική υπογραφή του Γ, τότε ορίζουμε $valR(\Gamma, x)$ να είναι το εσωτερικό γινόμενο των u και x.

Πρόταση 3.5.6. *Για κάθε $k > 0$ και για κάθε $w_1, \cdots, w_k$ υπάρχει ένας αναγνωριστής Γ με k εισοδους, τέτοιος ώστε σε είσοδο $x = x_1 \otimes x_2 \otimes \cdots \otimes x_k \in \{n, p\}^k$ με βάση **b1**, το $valR(\Gamma, x)$ ισούται με:*

- $-(w_1 + \cdots + w_k)$ *αν* $x_1 = \cdots = x_k = n$,
- w_i *αν* $x_i = p$, *και* $x_j = n$ *για κάθε* $j \neq i$,
- *0 σε κάθε άλλη περίπτωση.*

Συγκεκριμένα ο αναγνωριστής της πρότασης είναι του τύπου της παραπάνω εικόνας αλλά με k εισόδους.

Ορίζουμε το matchgrid (πλέγμα ταιριασμάτων) με βάση $\mathbf{b}$ να είναι ένα μη κατευθυνόμενο επίπεδο γράφημα G με βάρη, που αποτελείται από:

- ένα σύνολο B από g γεννήτορες $B_1, \cdots, B_g$,
- ένα σύνολο A από r αναγνωριστές $A_1, \cdots, A_r$, και
- ένα σύνολο C από f ακμές $C_1, \cdots, C_f$ όπου η κάθε ακμή C_i έχει βάρος 1 και συνδέει μια κορυφή εξόδου ενός γεννήτορα με μια κορυφή εισόδου ενός αναγνωριστή.

Ας θεωρήσουμε ένα τέτοιο matchgrid $\Omega = (A, B, C)$ και ας συμβολίσουμε με $X = \mathbf{b}^f = (n, p)^f$ το σύνολο όλων των 2^f δυνατών συνδυασμών των στοιχείων της βάσης n, p που μπορούν να μεταδοθούν ταυτόχρονα μέσω των f ακμών του C.

Ορίζουμε ως τιμή του matchgrid την ποσότητα

$$Holant(\Omega) = \sum_{x \in b^f} \left[\prod_{1 \leq j \leq g} valG(B_j, x_j) \right] \cdot \left[\prod_{1 \leq i \leq r} valG(A_i, x) \right].$$

Το βασικό θεώρημα που απέδειξε ο Valiant [37] και είναι καθοριστικό για τους ολογραφικούς αλγορίθμους είναι το παρακάτω.

Θεώρημα 3.5.7. *Για κάθε matchgrid Ω με βάση $\boldsymbol{b}$, αν το Ω αντιστοιχεί στο γράφημα με βάρη G, τότε ισχύει*

$$\text{Holant}(\Omega) = \text{PerfMatch}(G).$$

Παραδείγματα προβλημάτων στα οποία μπορούμε να κάνουμε ολογραφικές αναγωγές και να τα λύσουμε σε πολυωνυμικό χρόνο είναι τα #X-Matchings(στην αναγωγή του οποίου χρησιμοποιούμε τα παραδείγματα του γεννήτορα και του αναγνωριστή που αναφέραμε παραπάνω), το #Pl-3-Nae-Sat και το $\#_7$Pl-Rtw-Mon-3-Cnf [38].

Κεφάλαιο 4

Θεωρήματα Διχοτομίας για το #CSP

Σ'αυτό το κεφάλαιο θα δούμε πιο αναλυτικά κάποια από τα δομικά συστατικά των αποδείξεων των αποτελεσμάτων που παρουσιάσαμε στο προηγούμενο κεφάλαιο. Συγκεκριμένα θα δούμε το αποτέλεσμα των Bulatov και Grohe [8] για το πρόβλημα μέτρησης ομομορφισμών μη-κατευθυνόμενων γραφημάτων για πίνακες με μη-αρνητικά στοιχεία, αυτό του Bulatov [4] για $\#CSP$ χωρίς βάρη και το βελτιωμένο αποτέλεσμα των Dyer και Richerby [21] για το ίδιο πρόβλημα. Τέλος θα επικεντρωθούμε περισσότερο στο ενοποιημένο αποτέλεσμα, που έχει όλα τα παραπάνω σαν ειδικές περιπτώσεις, των Cai και Chen [11] για το $\#CSP$ με βάρη μιγαδικούς αριθμούς.

4.1 Κάποιες Ειδικές Περιπτώσεις

Στα θεωρήματα διχοτομίας ένα βασικό συστατικό που αναζητούμε είναι τα κριτήρια για τη διχοτομία να είναι όσο πιο απλά γίνεται. Δυστυχώς αυτό δε συμβαίνει πάντα. Κάποιες φορές για παράδειγμα προκύπτει κριτήριο διχοτομίας για το οποίο δε γνωρίζουμε καν αν είναι αποφασίσιμο (decidable), δηλαδή αν υπάρχει αλγόριθμος που να του δίνεις μια γλώσσα $\mathcal{L}$ και να σου λέει σε πεπερασμένο χρόνο αν η $\mathcal{L}$ ικανοποιεί το κριτήριο ή όχι. Αν καταφέρουμε να εξασφαλίσουμε την αποφασισιμότητα τότε αρχίζει να μας ενδιαφέρει και η αποδοτικότητα αυτού του αλγορίθμου, δηλαδή αν ανήκει στο P, στο NP, κ.λ.π. Φυσικά για το μη-ομογενές $\#\mathrm{CSP}(\mathcal{L})$ μοντέλο, όπου η γλώσσα $\mathcal{L}$ και το πεδίο D θεωρούνται σταθερές δε μας ενδιαφέρει η αποδοτικότητα του αλγορίθμου παρά μόνο η αποφασισιμότητα. Παρ' όλα αυτά το ερώτημα της πολυπλοκότητας είναι ένα ενδιαφέρον ερώτημα από μόνο του.

Σ' αυτή την παράγραφο θα δούμε κάποια από τα κριτήρια διχοτομίας που θα μας χρειαστούν και στη συνέχεια και θα δούμε και πως γενικά εξελίσσονται τα κριτήρια διχοτομίας όταν προχωράμε από τις ειδικές περιπτώσεις στις πιο γενικές. Είναι χαρακτηριστικό ότι κάποιες φορές στη γενική περίπτωση, το κριτήριο διχοτομίας γίνεται πιο απλό.

Το 2005 οι Bulatov και Grohe βρήκαν ένα πολύ απλό κριτήριο για το πρόβλημα μέτρησης ομομορφισμών γραφημάτων με συμμετρικούς πίνακες με μη-αρνητικά στοιχεία. Το συγκεκριμένο κριτήριο θα δούμε ότι ξαναεμφανίστηκε και στην πορεία της έρευνας. Συγκεκριμένα δείξανε το παρακάτω θεώρημα.

Θεώρημα 4.1.1. *Έστω $\boldsymbol{A}$ ένας συμμετρικός και μη-αρνητικός πίνακας. Αν ο $\boldsymbol{A}$ είναι διαγώνιος κατά μπλοκς (block-diagonal) και κάθε μπλοκ έχει είτε βαθμίδα 1 είτε έχει τη μορφή*

$$\begin{pmatrix} \boldsymbol{0} & \boldsymbol{B} \\ \boldsymbol{B}^T & \boldsymbol{0} \end{pmatrix}, \text{ με τον } \boldsymbol{B} \text{ να έχει βαθμίδα 1,}$$

τότε το πρόβλημα EVAL($\boldsymbol{A}$) λύνεται σε πολυωνυμικό χρόνο. Σε κάθε άλλη περίπτωση το πρόβλημα είναι #P-hard.

Όταν λέμε ότι ένας πίνακας είναι διαγώνιος κατά μπλοκς εννοούμε ότι στην κύρια διαγώνιό του έχει πίνακες (μπλοκς) και παντού αλλού έχει μηδενικά.

Ας θυμηθούμε τώρα ξανά την περίπτωση του $\#CSP$ χωρίς βάρη. Σ' αυτό το μοντέλο, αντί για μια γλώσσα $\mathcal{L}$ έχουμε ένα σύνολο σχέσεων Γ και έχουμε το πρόβλημα #CSP(Γ).

Ο Bulatov το 2008 [4] έδειξε το παρακάτω θεώρημα για το #CSP(Γ) χωρίς βάρη. Όταν πρωτοδημοσιεύθηκε το αποτέλεσμα θεωρήθηκε πολύ σημαντικό και ήταν πραγματικά τομή στο χώρο των CSP's μέτρησης, γιατί ήταν το πρώτο θεώρημα διχοτομίας σε ένα αρκετά γενικό μοντέλο.

Θεώρημα 4.1.2 (Bulatov, 2008)**.** *Για κάθε σύνολο σχέσεων* Γ*, το* #CSP(Γ) *χωρίς βάρη ανήκει είτε στο* FP *είτε είναι* #P*-complete.*

Παρά το σημαντικό αυτό αποτέλεσμα, η απόδειξη έχει και κάποια αρνητικά στοιχεία. Κατ' αρχήν για να κατανοήσει κάποιος την απόδειξη ήθελε αρκετά βαθιά γνώση καθολικής άλγεβρας. Γενικότερη η απόδειξη ήταν αρκετά μεγάλη και δυσνόητη και λίγοι από τον χώρο της θεωρητικής πληροφορικής μπορούσαν να την κατανοήσουν σε βάθος. Ένα άλλο αρνητικό χαρακτηριστικό ήταν ότι ο FP αλγόριθμος του θεωρήματος απαιτούσε πρώτα να μετατρέψεις την είσοδο του προβλήματος σε μια αρκετά μεγαλύτερη και πολυπλοκότερη μορφή (subdirect product form) και αυτό έκανε πολύ δύσκολη την ανάλυση του αλγορίθμου, οπότε η ακριβής πολυπλοκότητα δεν ήταν γνωστή. Το σημαντικότερο όμως απ' όλα τα αρνητικά στοιχεία ήταν αυτό του κριτηρίου διχοτομίας. Το κριτήριο διχοτομίας του Bulatov (congruence singularity) ήταν δανεισμένο από την καθολική άλγεβρα και δεν ήταν γνωστό αν είναι αποφασίσιμο. Ένα από τα χαρακτηριστικά λοιπόν που θα θέλαμε σε ένα θεώρημα διχοτομίας ήταν πιθανό να μην ισχύει.

Το 2010, οι Dyer και Richerby [21] αποδείξαν ξανά το θεώρημα του Bulatov προσπαθώντας να βελτιώσουν τα αρνητικά χαρακτηριστικά που αναφέραμε παραπάνω. Πιο συγκεκριμένα η απόδειξή τους δε χρειαζόταν γνώσεις καθολικής άλγεβρας αφού και οι λίγες έννοιες από καθολική άλγεβρα που χρησιμοποιούνται μπορούν να εξηγηθούν σχετικά σύντομα. Δεύτερον, ο αλγόριθμός τους για την κλάση προβλημάτων στο FP είναι αρκετά φυσικός και με αποδεδειγμένη πολυπλοκότητα. Τρίτον και πιο σημαντικό, το κριτήριο διχοτομίας τους είναι φανερά πιο απλό και είναι και αποφασίσιμο, συγκεκριμένα ανήκει στο NP. Το κριτήριο διχοτομίας τους βασίζεται στις έννοιες ισχυρή ορθογωνιότητα (strong rectangularity) και ισχυρή ισορροπία (strong balance).

Ορισμός 4.1.3. *Έστω M ένας $m \times n$ μη-αρνητικός πίνακας. Θα λέμε ότι ο $\boldsymbol{M}$ είναι*

- *ορθογώνιος (rectangular) αν μπορούμε να αναδιατάξουμε τις γραμμές του και τις στήλες του ώστε ο $\boldsymbol{M}$ να γίνει διαγώνιος κατά μπλοκς,*
- *block-rank-1 αν είναι ορθογώνιος και κάθε μπλοκ έχει βαθμίδα 1.*

Ορισμός 4.1.4. *Θα λέμε ότι το σύνολο σχέσεων Γ είναι ισχυρά ορθογώνιο (strongly rectangular) αν για κάθε στιγμιότυπο R του $\#CSP(\Gamma)$, με μεταβλητές $x_1, \cdots, x_n$, και για κάθε 3-διαμέριση των μεταβλητών:*

$$\boldsymbol{u} = (u_1, \cdots, u_k), \boldsymbol{v} = (v_1, \cdots, v_l) \text{ και } \boldsymbol{w} = (w_1, \cdots, w_{n-k-l}),$$

ο ακόλουθος $|D|^k \times |D|^l$ πίνακας $\boldsymbol{M}$ είναι ορθογώνιος:

$$M(\boldsymbol{u}, \boldsymbol{v}) = \left|\{\boldsymbol{w} \in D^{n-k-l} \mid (\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) \in R\}\right|.$$

Παρακάτω επίσης θα δούμε την πολύ σημαντική έννοια του Mal'tsev πολυμορφισμού, η οποία τελικά προκύπτει ότι είναι ισοδύναμη με τη συνθήκη της ισχυρής ορθογωνιότητας. Οι Dyer και Richerby έδειξαν το παρακάτω θεώρημα για την ισχυρή ορθογωνιότητα.

Θεώρημα 4.1.5. *Αν το Γ δεν είναι ένα ισχυρά ορθογώνιο σύνολο σχέσεων, τότε το $\#CSP(\Gamma)$ είναι $\#$P-hard.*

Για να έχουμε ένα ολοκληρωμένο θεώρημα διχοτομίας θα πρέπει εκτός από το παραπάνω θεώρημα να έχουμε και το αντίστοιχο που λέει ότι αν ισχύει η συνθήκη για το Γ, τότε το $\#CSP(\Gamma)$ λύνεται σε πολυωνυμικό χρόνο. Δυστυχώς η ισχυρή ορθογωνιότητα δεν είναι αρκετή για να μας δώσει και αυτό το θεώρημα. Οι Dyer και Richerby όμως έδειξαν ότι αν έχουμε εξασφαλίσει τη συνθήκη της ισχυρής ορθογωνιότητας για ένα σύνολο Γ τότε μπορούμε να λύσουμε σε πολυωνυμικό χρόνο ένα άλλο δύσκολο πρόβλημα, το πρόβλημα του μάρτυρα (witness problem).

Το πρόβλημα του μάρτυρα είναι το εξής: Δοσμένων ενός στιγμιότυπου εισόδου R του $\#CSP(\Gamma)$ με n μεταβλητές $x_1, \cdots, x_n$, ενός συνόλου $S \subseteq [n]$ και $a_i \in D$ για κάθε $i \in S$, αποφάσισε αν υπάρχει ανάθεση $\mathbf{x} = (x_1, \cdots, x_n)$ στις μεταβλητές τέτοια ώστε $\mathbf{x} \in R$ και $x_i = a_i$ για κάθε $i \in S$. Αν υπάρχει, τότε δώσε ως έξοδο μια τέτοια ανάθεση.

Το συγκεκριμένο πρόβλημα είναι εν γένει δύσκολο, αλλά με τη συνθήκη της ισχυρής ορθογωνιότητας βλέπουμε ότι γίνεται εύκολο.

Αν τώρα θέσουμε ως σύνολο S το κενό σύνολο, δηλαδή δε δεσμεύσουμε καμιά μεταβλητή να πάρει συγκεκριμένη τιμή a_i τότε παρατηρούμε ότι το παραπάνω πρόβλημα είναι ισοδύναμο με το πρόβλημα απόφασης CSP(Γ). Αν λοιπόν $S = \emptyset$, τότε έχουμε μια εναλλακτική απόδειξη του αποτελέσματος των Bulatov και Dalmau [5] για το πρόβλημα απόφασης CSP(Γ) που λέει ότι υπάρχει πολυωνυμικός αλγόριθμος για κάθε ισχυρά ορθογώνια γλώσσα Γ.

Μιας και η συνθήκη της ισχυρής ορθογωνιότητας δε μας δίνει το επιθυμητό απότέλεσμα της διχοτομίας, πρέπει να επεκτείνουμε την έννοια ώστε να γίνει πιο ισχυρή. Έχουμε λοιπόν τον παρακάτω ορισμό της ισχυρής ισορροπίας.

Ορισμός 4.1.6. *Θα λέμε ότι το σύνολο σχέσεων* Γ *είναι ισχυρά ισορροπημένο (strongly balanced) αν για κάθε στιγμιότυπο* R *του* $\#CSP(\Gamma)$*, με μεταβλητές* $x_1, \cdots, x_n$*, και για κάθε 4-διαμέριση:*

$$\boldsymbol{u} = (u_1, \cdots, u_k), \boldsymbol{v} = (v_1, \cdots, v_l),$$

$$\boldsymbol{w} = (w_1, \cdots, w_t) \textit{ και } \boldsymbol{z} = (z_1, \cdots, z_{n-k-l-t}),$$

ο παρακάτω $|D|^k \times |D|^l$ *πίνακας* $\boldsymbol{M}$ *είναι block-rank-1:*

$$M(\boldsymbol{u}, \boldsymbol{v}) = \left|\{\boldsymbol{w} \mid \exists z \textit{ τέτοιο ώστε } (\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}, \boldsymbol{z}) \in R\}\right|.$$

Η νέα αυτή ιδιότητα είναι αρκετά ισχυρή ώστε οι Dyer και Richerby να τη χρησιμοποιήσουν ως κριτήριο διχοτομίας στο παρακάτω θεώρημα.

Θεώρημα 4.1.7 (Dyer, Richerby, 2010)**.** *Αν το σύνολο σχέσεων* Γ *είναι ισχυρά ισορροπημένο, τότε το* $\#CSP(\Gamma)$ *λύνεται σε πολυωνυμικό χρόνο. Σε κάθε άλλη περίπτωση, το πρόβλημα είναι* #P*-hard.*

Επίσης η ισχυρή ισορροπία είναι σαν έννοια ισοδύναμη με το congruence singularity (το κριτήριο διχοτομίας του Bulatov).

Το ότι η ισχυρή ισορροπία είναι ισοδύναμη με το congruence singularity είναι λίγο πολύ αναμενόμενο παρά το γεγονός ότι πρόκειται για διαφορετικούς ορισμούς, αφού αφορούν τη διχοτόμηση της ίδια κλάσης προβλημάτων.

4.2 Το $\#CSP$ με Βάρη Μιγαδικούς Αριθμούς

Οι Cai και Chen το 2012 έδειξαν ένα ενοποιημένο θεώρημα διχοτομίας [11]. Όλα τα αποτελέσματα διχοτομίας που αναφέραμε στην προηγούμενη παράγραφο είναι πλέον ειδικές περιπτώσεις του παρακάτω γενικού θεωρήματος.

Θεώρημα 4.2.1 (Cai, Chen, 2012). *Δοσμένου ενός οποιουδήποτε πεδίου D και μιας πεπερασμένης γλώσσας $\mathcal{L}$ απότελούμενης από συναρτήσεις που δίνουν εν γένει μιγαδικές τιμές, το $\#\mathrm{CSP}(\mathcal{L})$ είτε λύνεται σε πολυωνυμικό χρόνο, είτε είναι $\#\mathrm{P}$-hard.*

Παρακάτω θα δώσουμε μια σκιαγράφηση της απόδειξης αυτού του σπουδαίου αποτελέσματος.

Οι συνθήκες για τη διχοτομία είναι τρεις:

- η συνθήκη της Μπλοκ Ορθογωνιότητας (Block Orthogonality condition),
- η συνθήκη Mal'tsev και
- η συνθήκη Τυπικής Διαμέρισης (Type Partition condition)

Αν η $\mathcal{L}$ ικανοποιεί και τις τρεις αυτές συνθήκες, τότε υπάρχει πολυωνυμικός αλγόριθμος για το $\#\mathrm{CSP}(\mathcal{L})$. Σε κάθε άλλη περίπτωση το $\#\mathrm{CSP}(\mathcal{L})$ είναι $\#\mathrm{P}$-hard.

Έστω $F : D^n \to \mathbb{C}$ η συνάρτηση που ορίζεται από ένα στιγμιότυπο εισόδου I ως $F(\mathbf{x}) = \prod_{(f,i_1,\cdots,i_r)\in I} f(x_{i_1},\cdots,x_{i_r})$, είναι δηλαδή το γινόμενο όλων των περιορισμών.

Για κάθε $t \in [n]$, ορίζουμε τη συνάρτηση $F^{[t]} : D^t \to \mathbb{C}$ με τον τύπο

$$F^{[t]}(x_1,\cdots,x_t) = \sum_{x_{t+1},\cdots,x_n\in D} F(x_1,\cdots,x_t,x_{t+1},\cdots,x_n).$$

Θεωρούμε δηλαδή ως ορίσματα τις πρώτες t μεταβλητές και αθροίζουμε ως προς τις υπόλοιπες $n-t$.

Η συνάρτηση $F^{[t]}$ είναι μια πολύ χρήσιμη συνάρτηση για την απόδειξη και βοηθάει περισσότερο να τη βλέπουμε σαν έναν $d^{t-1} \times d$ πίνακα. Κάθε γραμμή του πίνακα θα έχει δείκτη ένα $\mathbf{x} = (x_1,\cdots,x_{t-1}) \in D^{t-1}$ και κάθε στήλη ένα $a \in D$. Το στοιχείο στη θέση $(\mathbf{x}, a)$ του πίνακα θα είναι το $F^{[t]}(\mathbf{x}, a)$.

Επίσης θα συμβολίζουμε με $F^{[t]}(\mathbf{x}, *)$ το d-διάστατο διάνυσμα γραμμή που έχει δείκτη το $\mathbf{x}$.

Ας φανταστούμε προς στιγμήν ότι έχουμε τη βοήθεια ενός μαντείου (oracle) που μας δίνει πληροφορίες για τα $F^{[2]},\cdots,F^{[n]}$ ως εξής: εμείς στέλνουμε στο μαντείο ένα $\mathbf{x} \in D^{t-1}$ και αυτό μας επιστρέφει ένα διάνυσμα $\mathbf{v}$ το οποίο είναι γραμμικά εξαρτημένο με το διάνυσμα γραμμή $F^{[t]}(\mathbf{x}, *)$. Αν $F^{[t]}(\mathbf{x}, *) = 0$, τότε θα επιστρέψει το μηδενικό διάνυσμα $\mathbf{v} = 0$. Επίσης θεωρούμε ότι το διάνυσμα $\mathbf{v}$ που μας επιστρέφει είναι κανονικοποιημένο έτσι ώστε το πρώτο μη μηδενικό στοιχείο του διανύσματος να είναι 1.

Η κατοχή ενός τέτοιου μαντείου θα μας έδινε πολύ μεγάλη δύναμη αφού θα μπορούσαμε να λύσουμε το πρόβλημά μας. Ας παρατηρήσουμε κατ' αρχήν ότι η ζητούμενη ποσότητα $Z(I)$ που θέλουμε να υπολογίσουμε δίνεται από τον τύπο

$$Z(I) = \sum_{a \in D} F^{[1]}(a).$$

Αρκεί λοιπόν να υπολογίσουμε το $F^{[1]}(a_1)$ για κάθε $a_1 \in D$. Για να το υπολογίσουμε αυτό κάνουμε το εξής: στέλνουμε το a_1 στο μαντείο και λαμβάνουμε ένα διάνυσμα $\mathbf{v}$ που είναι γραμμικά εξαρτημένο με το $F^{[2]}(a_1, *)$. Αν $\mathbf{v} = 0$, τότε έχουμε ότι $F^{[1]}(a_1) = 0$. Σε διαφορετική περίπτωση, έστω v_{a_2} το πρώτο μη μηδενικό στοιχείο του διανύσματος (δηλαδή $v_{a_2} = 1$), τότε ισχύει

$$F^{[1]}(a_1) = \sum_{b \in D} F^{[2]}(a_1, b) = F^{[2]}(a_1, a_2) \cdot \sum_{b \in D} v_b,$$

όπου η ισότητα προκύπτει από το γεγονός ότι το $F^{[2]}(a_1, *)$ και το $\mathbf{v}$ είναι γραμμικά εξαρτημένα.

Βλέπουμε λοιπόν ότι ο υπολογισμός του $F^{[1]}(a_1)$ ανάγεται στον υπολογισμό του $F^{[2]}(a_1, a_2)$.

Για να υπολογίσουμε τώρα το $F^{[2]}(a_1, a_2)$ εφαρμόζουμε παρόμοια διαδικασία: στέλνουμε στο μαντείο το (a_1, a_2) και λαμβάνουμε ένα διάνυσμα $\mathbf{w}$ το οποίο είναι γραμμικά εξαρτημένο με το διάνυσμα $F^{[3]}((a_1, a_2), *)$. Αν $\mathbf{w} = 0$, τότε και $F^{[2]}(a_1, a_2) = 0$. Αν από την άλλη, $\mathbf{w} \neq 0$, τότε έστω w_{a_3} η πρώτη μη αρνητική συντεταγμένη του διανύσματος $\mathbf{w}$ (οπότε $w_{a_3} = 1$), τότε ισχύει ότι

$$F^{[2]}(a_1, a_2) = \sum_{b \in D} F^{[3]}((a_1, a_2), b) = F^{[3]}(a_1, a_2, a_3) \cdot \sum_{b \in D} w_b.$$

Δηλαδή ο υπολογισμός του $F^{[2]}(a_1, a_2)$ ανάγεται στον υπολογισμό του $F^{[3]}(a_1, a_2, a_3)$.

Συνεχίζοντας με τον ίδιο τρόπο, ύστερα από $n - 1$ βήματα, ο υπολογισμός του $F^{[1]}(a_1)$ έχει αναχθεί στον υπολογισμό του $F^{[n]}(a_1, a_2, \cdots, a_n)$ για κάποια κατάλληλα $a_2, \cdots, a_n$ με τη βοήθεια του μαντείου. Παρατηρούμε τώρα ότι η συνάρτηση $F^{[n]}$ είναι η συνάρτηση F, δηλαδή ισχύει $F = F^{[n]}$, η οποία δοσμένης της εισόδου I μπορεί να υπολογιστεί αποτελεσματικά.

Αν είχαμε λοιπόν στα χέρια μας το παραπάνω μαντείο τότε θα είχαμε λύσει το πρόβλημα. Τί χρειάζεται όμως για να κατασκευάσουμε ένα τέτοιο μαντείο; Ας σταθεροποιήσουμε ένα $t \in [n]$. Κατ' αρχήν πρέπει να υπολογίσουμε ένα σύνολο από d-διάστατα διανύσματα

$$\mathbf{v}^{[t,1]}, \mathbf{v}^{[t,2]}, \cdots, \mathbf{v}^{[t,s_t]},$$

τα οποία είναι ανα δύο γραμμικά ανεξάρτητα και τέτοια ώστε κάθε μη μηδενικό διάνυσμα γραμμή $F^{[t]}(\mathbf{x},*)$ να είναι γραμμικά εξαρτημένο με κάποιο από αυτά. Αυτά τα διανύσματα θα είναι οι πιθανές έξοδοι του μαντείου.

Επίσης, πέρα από τα διανύσματα, χρειάζεται να γνωρίζουμε και τα σύνολα

$$S^{[t,1]}, S^{[t,2]}, \cdots, S^{[t,s_t]} \subseteq D^{t-1}$$

για τα οποία ισχύει ότι $\mathbf{x} \in S^{[t,j]}$ αν και μόνο αν το $F^{[t]}(\mathbf{x},*)$ είναι γραμμικά εξαρτημένο με το $v^{[t,j]}$.

Αν είχαμε τα παραπάνω τότε το μαντείο δοσμένου ενός $\mathbf{x}$ θα έβρισκε το σύνολο $S^{[t,j]}$ στο οποίο ανήκει και θα επέστρεφε το διάνυσμα $\mathbf{v}^{[t,j]}$.

Υπάρχουν όμως δύο δυσκολίες ως προς τον υπολογισμό αυτών των διανυσμάτων και των συνόλων. Κατ' αρχήν, ένας $m \times d$ πίνακας μπορεί να έχει εν γένει μέχρι και m γραμμικά ανεξάρτητες ανά δύο γραμμές. Για τον πίνακα που αντιστοιχεί στην $F^{[t]}$, που είναι ένας $d^{t-1} \times d$ πίνακας, τα διανύσματα που χρειάζεται να έχουμε αποθηκευμένα μπορεί να είναι εν γένει μέχρι και d^{t-1}, δηλαδή εκθετικά σε πλήθος στο t. Είναι φανερό όμως ότι για έναν πολυωνυμικό αλγόριθμο δεν μπορούμε να έχουμε δεσμευμένη τόσο μεγάλη μνήμη.

Παρόμοιο πρόβλημα συμβαίνει και με τα σύνολα $S^{[t,j]}$ που και αυτά εν γένει μπορεί να είναι σε πλήθος εκθετικά στο t, άρα και στο n.

Για να λύσουμε το πρώτο πρόβλημα θα χρησιμοποιήσουμε την έννοια της ορθογωνιότητας δύο διανυσμάτων. Συγκεκριμένα, αν οποιεσδήποτε δύο γραμμές του πίνακα $F^{[t]}$ είναι είτε γραμμικά εξαρτημένες είτε ορθογώνιες μεταξύ τους, τότε δε γίνεται να υπάρχουν περισσότερες από d γραμμικά ανεξάρτητες ανά δύο γραμμές. Αυτό συμβαίνει γιατί σε έναν d-διάστατο χώρο δε μπορούν να υπάρξουν περισσότερα από d ανά δύο ορθογώνια μεταξύ τους διανύσματα. Για παράδειγμα ο παρακάτω 4x4 πίνακας έχει 4 γραμμές ανά δύο ορθογώνιες.

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}$$

Δυστυχώς η συνθήκη της απλής ορθογωνιότητας δε φτάνει για το θεώρημά μας, οπότε θα ορίσουμε την έννοια της μπλοκ-ορθογωνιότητας (block-orthogonality).

Ορισμός 4.2.2. *Έστω $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^d$ δύο μη μηδενικά d-διάστατα διανύσματα. Έστω τώρα $\boldsymbol{x'}$ και $\boldsymbol{y'}$ τα δύο διανύσματα τέτοια ώστε $x'_i = |x_i|$ και $y'_i = |y_i|$ για κάθε $i = 1, 2, \cdots, d$, και ας υποθέσουμε ότι τα $\boldsymbol{x'}$ και $\boldsymbol{y'}$ είναι γραμμικά εξαρτημένα. Αυτά τα τέσσερα διανύσματα θα έχουν κοινές μη μηδενικές συντεταγμένες, οπότε έστω $T \subseteq [d]$ το σύνολο των δεικτών που αντιστοιχούν σε αυτές τις μη μηδενικές συντεταγμένες. Έστω επίσης το σύνολο*

$$\{\mu_1, \cdots, \mu_l\} = \{x'_i : i \in T\},$$

για κάποιο $l \geq 1$, τέτοιο ώστε να ισχύει $\mu_1 > \cdots > \mu_l > 0$. Με βάση αυτό, το σύνολο T διαμερίζεται σε υποσύνολα $T_1, \cdots, T_l$ με

$$x'_i = \mu_k, \text{ για κάθε } i \in T_k \text{ και } k \in [l].$$

Παρόμοια μπορούμε να ορίσουμε διαμέριση για το $\boldsymbol{y}'$ που όμως τελικά θα είναι η ίδια, επειδή αυτό ειναι γραμμικά εξαρτημένο με το $\boldsymbol{x}'$.

Θα λέμε ότι τα $\boldsymbol{x}$ και $\boldsymbol{y}$ είναι μπλοκ-ορθογώνια αν για κάθε $k \in [l]$ ισχύει ότι

$$\sum_{i \in T_k} x_i \cdot \bar{y}_i = 0.$$

Είναι εύκολο να δειχθεί ότι αν τα $\mathbf{x}$ και $\mathbf{y}$ είναι μπλοκ-ορθογώνια τότε είναι και ορθογώνια. Πράγματι,

$$\sum_{i \in [d]} x_i \cdot \bar{y}_i = \sum_{i \in T} x_i \cdot \bar{y}_i = \sum_{k \in [l]} \sum_{i \in T_k} x_i \cdot \bar{y}_i = 0.$$

Αφού ορίσαμε πότε δύο διανύσματα είναι μπλοκ-ορθογώνια, πάμε να δούμε τη συνθήκη της Μπλοκ Ορθογωνιότητας.

Ορισμός 4.2.3. *Έστω $F : D^n \to \mathbb{C}$ η συνάρτηση που ορίζεται από ένα στιγμιότυπο εισόδου του $\#\mathrm{CSP}(\mathcal{L})$, και $t \in [n]$. Θα λέμε ότι η $\mathcal{L}$ ικανοποιεί τη συνθήκη της Μπλοκ Ορθογωνιότητας αν οποιεσδήποτε δύο γραμμές του πίνακα $F^{[t]}$ είναι είτε γραμμικά εξαρτημένες, είτε μπλοκ-ορθογώνιες.*

Την πρώτη δυσκολία την έχουμε ξεπεράσει αφού ισχύει το παρακάτω λήμμα.

Λήμμα 4.2.4. *Αν η $\mathcal{L}$ δεν ικανοποιεί τη συνθήκη της Μπλοκ Ορθογωνιότητας, τότε το πρόβλημα $\#\mathrm{CSP}(\mathcal{L})$ είναι $\#\mathrm{P}$-hard.*

Για να ξεπεράσουμε και τη δεύτερη δυσκολία θα χρειαστεί να ορίσουμε την έννοια του πολυμορφισμού.

Ορισμός 4.2.5. *Έστω $R \subseteq D^n$ μια n-μελής σχέση. Ένας πολυμορφισμός (polymorphism) της R είναι μια k-μελής συνάρτηση $\phi : D^k \to D$ τέτοια ώστε $\phi(R) \subseteq R$. Με άλλα λόγια, έστω $\boldsymbol{v}^1, \boldsymbol{v}^2, \cdots, \boldsymbol{v}^k \in R$ και $\boldsymbol{v} = \big(\phi(v_1^1, v_1^2, \cdots, v_1^k), \cdots, \phi(v_k^1, v_k^2, \cdots, v_k^k)\big)$,*

$$\begin{array}{cccccccc}
 & \boldsymbol{v}^1 & \boldsymbol{v}^2 & \cdots & \boldsymbol{v}^k & & & \boldsymbol{v} \\
 & \shortparallel & \shortparallel & & \shortparallel & & & \shortparallel \\
\phi(& v_1^1 & v_1^2 & & v_1^k &) & = & v_1 \\
\phi(& v_2^1 & v_2^2 & & v_2^k &) & = & v_2 \\
 & \vdots & \vdots & & \vdots & & & \vdots \\
\phi(& v_k^1 & v_k^2 & & v_k^k &) & = & v_k.
\end{array}$$

Αν ο ϕ είναι ένας πολυμορφισμός του R, τότε $\boldsymbol{v} \in R$.

Πολύ σημαντική έννοια επίσης, δανεισμένη από την απόδειξη του Bulatov, είναι ο πολυμορφισμός Mal'tsev.

Ορισμός 4.2.6. *Έστω $R \subseteq D^n$ μια n-μελής σχέση. Θα λέμε ότι η τριμελής συνάρτηση $\phi : D^3 \to D$ είναι ένας πολυμορφισμός Mal'tsev της R αν ο ϕ είναι ένας πολυμορφισμός της R και επιπλέον ισχύει ότι*

$$\phi(a, b, b) = \phi(b, b, a) = a, \quad \textit{για κάθε } a, b \in D.$$

Οι Dyer και Richerby απέδειξαν ότι αν η σχέση R έχει έναν πολυμορφισμό Mal'tsev ϕ, τότε έχει και μια συνοπτική αναπαράσταση, η οποία ονομάζεται συνάρτηση μάρτυρας (witness function). Αυτή η συνάρτηση μάρτυρας ω της R είναι γραμμικού μεγέθους στο n και επίσης δοσμένης της ω και ενός $\mathbf{x} \in D^n$ μπορούμε να αποφασίσουμε αν $\mathbf{x} \in R$ σε πολυωνυμικό χρόνο.

Αν λοιπόν είχαμε τέτοιες συναρτήσεις μάρτυρες για τα σύνολα $S^{[t,j]}$, τότε δε θα χρειαζόταν να τα έχουμε όλα αποθηκευμένα στη μνήμη για τον αλγόριθμό μας και έτσι θα αντιμετωπίζαμε και τη δεύτερη δυσκολία. Πριν δούμε πότε μπορούμε να βρούμε τέτοιες συναρτήσεις μάρτυρες για τα σύνολά μας, ας δούμε τι ακριβώς είναι αυτές οι συναρτήσεις.

Έστω μια -μελής σχέση $R \subseteq D^n$. Για κάθε $i \in [n]$ ορίζουμε την ακόλουθη σχέση ισοδυναμίας $\sim_i$ στην $pr_i R$ (η προβολή της R στην i-οστή συντεταγμένη):

$$a \sim_i b \iff \exists \mathbf{x} \in D^{i-1}, \mathbf{y}_a, \mathbf{y}_b \in D^{n-i} : \mathbf{x} \circ a \circ \mathbf{y}_a \in R \text{ και } \mathbf{x} \circ b \circ \mathbf{y}_b \in R.$$

Ορισμός 4.2.7. *Έστω $R \subseteq D^n$ μια σχέση η οποία έχει πολυμορφισμό Mal'tsev. Θα λέμε ότι η συνάρτηση $\omega : [n] \times D \to D^n \cup \{\perp\}$ είναι μια συνάρτηση μάρτυρας (witness function) της R αν ισχύουν τα παρακάτω:*

1. *Για κάθε $i \in [n]$ και $a \notin pr_i R, \omega(i, a) = \perp$.*
2. *Για κάθε $i \in [n]$ και $a \in pr_i R$,το $\omega(i, a) \in R$ και είναι ένας μάρτυρας για τα (i, a), δηλαδή η i-οστή συντεταγμένη του είναι ίση με a.*
3. *Για κάθε $i \in [n]$ και $a, b \in pr_i R$ τέτοια ώστε $a \sim_i b$, ισχύει ότι*

$$pr_{i-1}\omega(i, a) = pr_{i-1}\omega(i, b).$$

Τώρα λοιπόν έχουμε τα απαραίτητα στοιχεία για να ορίσουμε τη συνθήκη Mal'tsev, η οποία θα μας λύσει και τη δεύτερη δυσκολία.

Ορισμός 4.2.8. *Έστω $F : D^n \to \mathbb{C}$ η συνάρτηση που ορίζεται από ένα στιγμιότυπο εισόδου του $\#\mathrm{CSP}(\mathcal{L})$, και $t \in [n]$. Θα λέμε ότι η $\mathcal{L}$ ικανοποιεί τη συνθήκη Mal'tsev αν όλα τα σύνολα $S^{[t,j]} \subseteq D^{t-1}$ έχουν πολυμορφισμό Mal'tsev, και μάλιστα απαιτούμε να έχουν όλα έναν κοινό πολυμορφισμό Mal'tsev.*

Με την απαίτηση της συνθήκης Mal'tsev ξεπερνάμε και τη δεύτερη δυσκολία, αφού με αυτή τη συνθήκη εξασφαλίζουμε ότι όλα τα σύνολα έχουν μια σύντομη αναπαράσταση με τη βοήθεια της οποίας μπορούμε για κάποιο στοιχείο να ελέγξουμε αν ανήκει σε ένα σύνολο ή όχι, και έτσι δε μας ενοχλεί το πιθανό εκθετικό μέγεθος των συνόλων. Όπως ίσως ήταν αναμενόμενο μπορεί να αποδειχθεί και η δυσκολία του προβλήματος σε περίπτωση που δεν ισχύει η συνθήκη.

Λήμμα 4.2.9. *Αν η γλώσσα $\mathcal{L}$ δεν ικανοποιεί τη συνθήκη Mal'tsev, τότε το πρόβλημα* #CSP($\mathcal{L}$) *είναι* #P-*hard.*

Αφού εξασφαλίσουμε την ισχύ των πρώτων δύο συνθηκών, δηλαδή της συνθήκης της Μπλοκ Ορθογωνιότητας και τη συνθήκη Mal'tsev, τώρα προκύπτει ένα νέο πλάνο που έχουμε να εφαρμόσουμε.

Έστω $F : D^n \to \mathbb{C}$ η συνάρτηση που ορίζεται από την είσοδο του προβλήματος. Για κάθε $t \in [n]$ έχουμε να υπολογίσουμε:

1. τα $\mathbf{v}_1, \cdots, \mathbf{v}_h$, για κάποιο $h \leq d$, που είναι τέτοια ώστε κάθε διάνυσμα γραμμή $F^{[t]}(\mathbf{x}, *)$ είναι γραμμικά εξαρτημένο με ένα από τα $\mathbf{v}_i$,

2. μια συνάρτηση μάρτυρα ω_j για κάθε ένα από τα $S^{[t,j]}$.

Με τις δύο προηγούμενες συνθήκες εξασφαλίσαμε ότι αυτά τα αντικείμενα υπάρχουν. Το ερώτημα που προκύπτει τώρα είναι, πώς μπορούμε να τα υπολογίσουμε αποτελεσματικά;

Έστω $t = n$ και $F^{[n]} = F$. Εμείς χρειαζόμαστε τα $\mathbf{v}_1, \cdots, \mathbf{v}_h$ και τις ω_j για κάθε $S^{[t,j]}$. Οι Dyer και Richerby έδειξαν ότι αν ισχύει η συνθήκη Mal'tsev, μπορούμε να κατασκευάσουμε μια συνάρτηση μάρτυρα ω' για τη σχέση $R' \subseteq D^{n-1}$, που ορίζεται ως

$$\mathbf{x} \in R' \iff \exists b \in D, F(\mathbf{x}, b) \neq 0.$$

Με άλλα λόγια, από όλα τα διανύσματα που δεν μηδενίζουν την F, παίρνουμε τις προβολές στις πρώτες $n-1$ συντεταγμένες και φτιάχνουμε την R'. Για αυτή λοιπόν την R' έχουμε μια συνάρτηση μάρτυρα ω'. Εξ' ορισμού μπορούμε να δούμε ότι ισχύει $R' = S^{[n,1]} \cup S^{[n,2]} \cup \cdots \cup S^{[n,s_n]}$. Οπότε το ερώτημα τώρα είναι, μπορούμε από την ω' να υπολογίσουμε συναρτήσεις μάρτυρες ω_j για κάθε $S^{[n,j]}$; Θέλουμε δηλαδή μια διαδικασία διαχωρισμού.

Η κατάσταση είναι η εξής: έστω μια σχέση $R \subseteq D^n$ και σύνολα $S_1, \cdots, S_h$ τα οποία αποτελούν μια h-διαμέριση της R. Γνωρίζουμε ότι όλα αυτά τα σύνολα έχουν ένα κοινό πολυμορφισμό Mal'tsev. Από την άλλη δε γνωρίζουμε το h, έχουμε όμως εγγύηση ότι $h \leq d$. Επίσης έχουμε μια συνάρτηση μάρτυρα ω της R. Τέλος ας υποθέσουμε ότι έχουμε ένα μαύρο κουτί, στο οποίο αν δώσουμε ένα $\mathbf{x} \in R$, τότε αυτό επιστρέφει τον μοναδικό δείκτη $j \in [h]$ για τον οποίο ισχύει $\mathbf{x} \in S_j$. Με αυτά τα δεδομένα θέλουμε να υπολογίσουμε το h και επίσης μια συνάρτηση μάρτυρα ω_i για κάθε S_i.

Ορισμός 4.2.10. *Έστω η σχέση R και τα σύνολα $S_1, \cdots, S_h$ όπως παραπάνω. Για κάθε διάνυσμα $\boldsymbol{y} \in D^l, l \in [n]$, ορίζουμε το*

$$type(\boldsymbol{y}) = \{j \in [h] : \exists \boldsymbol{z} \in D^{n-l} \text{ τέτοιο ώστε } \boldsymbol{y} \circ \boldsymbol{z} \in S_j\} \subseteq [h].$$

Ορίζουμε επίσης τη συνθήκη διαμέρισης (partition condition) ως εξής: για κάθε $\boldsymbol{y}, \boldsymbol{y'} \in D^l$, τα $type(\boldsymbol{y})$ και $type(\boldsymbol{y'})$ είναι είτε ξένα μεταξύ τους, είτε ταυτίζονται.

Μπορεί να αποδειχθεί ότι αν ισχύει η συνθήκη διαμέρισης τότε έχουμε έναν αποδοτικό αλγόριθμο για τον διαχωρισμό. Συγκεκριμένα, υπάρχει ένας αναδρομικός αλγόριθμος ο οποίος δοσμένου ενός $\mathbf{x} \in D^l$, υπολογίζει το $type(\mathbf{x})$. Για αυτό τον αλγόριθμο είναι απαραίτητη η συνθήκη διαμέρισης. Επίσης υπάρχει ένας δεύτερος αναδρομικός αλγόριθμος ο οποίος δοσμένου ενός $\mathbf{x} \in D^l$ και ενός $j \in type(\mathbf{x})$, βρίσκει το κατάλληλο $\mathbf{y} \in D^{n-l}$ για το οποίο $\mathbf{x} \circ \mathbf{y} \in S_j$. Τέλος, με τη βοήθεια αυτών, μπορούμε να κατασκευάσουμε μια συνάρτηση μάρτυρα ω_i για κάθε S_i.

Ας ορίσουμε τώρα το τρίτο κριτήριο διχοτομίας, το οποίο είναι η συνθήκη Τυπικής Διαμέρισης (Type Partition Condition).

Ορισμός 4.2.11. *Στην ουσία αυτή η συνθήκη απαιτεί κάθε φορά που χρειάζεται να εφαρμόσουμε τη διαδικασία διαχωρισμού για τον αλγόριθμο του μαντείου, τα αντίστοιχα σύνολα R και $S_1, \cdots, S_h$ να ικανοποιούν τη συνθήκη διαμέρισης.*

Επίσης ισχύει και εδώ το αντίστοιχο λήμμα με τα άλλα δυο κριτήρια διχοτομίας.

Λήμμα 4.2.12. *Αν η γλώσσα $\mathcal{L}$ δεν ικανοποιεί τη συνθήκη της Τυπικής Διαμέρισης, τότε το πρόβλημα $\#\mathrm{CSP}(\mathcal{L})$ είναι $\#\mathrm{P}$-hard.*

Έχοντας εξασφαλίσει όλες τις προηγούμενες συνθήκες, τώρα μπορούμε να δούμε ολοκληρωμένο τον αλγόριθμο που λύνει το πρόβλημα. Έστω $F : D^n \to \mathbb{C}$ η συνάρτηση που ορίζεται από την είσοδο.

Επαγωγικά, για t από n έως 2:

Χρησιμοποίησε τα μαντεία για τα $F^{[t+1]}, F^{[t+2]}, \cdots, F^{[n]}$ ώστε

1. να υπολογίσεις τα $\mathbf{v}^{[t,1]}, \cdots, \mathbf{v}^{[t,s_t]}$, όπου $s_t \leq d$, τέτοια ώστε κάθε $F^{[t]}(\mathbf{x}, *)$ να είναι γραμμικά εξαρτημένο με ένα από τα $\mathbf{v}^{[t,j]}$,

2. να υπολογίσεις μια συνάρτηση μάρτυρα ω_j για κάθε $S^{[t,j]}$.

Και τα δύο παραπάνω θα γίνουν με χρήση του αλγορίθμου του διαχωρισμού.

Τέλος, υπολόγισε το $\sum_{\mathbf{x} \in D^n} F(\mathbf{x})$ χρησιμοποιώντας τα μαντεία όπως περιγράψαμε στην αρχή της παραγράφου.

Δυστυχώς το paper των Cai και Chen αφήνει και ένα αναπάντητο ερώτημα σχετικά με την αποφασισιμότητα των τριών κριτηρίων διχοτομίας. Συγκεκριμένα, δοσμένου ενός πεπερασμένου συνόλου $\mathcal{L}$ από συναρτήσεις με μιγαδικές τιμές, υπάρχει αλγόριθμος που να αποφασίζει σε πεπερασμένο χρόνο αν η $\mathcal{L}$ ικανοποιεί τις τρεις συνθήκες διχοτομίας; Εκ πρώτης όψεως φαίνεται να μη γίνεται κάτι τέτοιο, γιατί οι τρεις συνθήκες αφορούν άπειρα αντικείμενα. Θα μπορούσε όμως να γίνει κάτι ανάλογο με την περίπτωση της απόδειξης του Bulatov, δηλαδή να βρεθούν μέσω κάποιας άλλης απόδειξης κάποια αποφασίσιμα κριτήρια και να αποδειχθεί η ισοδυναμία τους με τα συγκεκριμένα τρία. Ουσιαστικά αυτό είναι και το μόνο ανοιχτό πρόβλημα που αν απαντηθεί θα κλείσει το κεφάλαιο των $\#CSP$'s.

Γλωσσάρι

Αγγλο-Ελληνικό

affine	αφινικός
bijunctive	διζευκτικός
binary	διμελής
block-diagonal	διαγώνιος κατά μπλοκς
block-orthogonality	μπλοκ-ορθογωνιότητα
boolean	δυαδικό
cardinality	πληθικός αριθμός
cartesian product	καρτεσιανό γινόμενο
clause	όρος
co-clone	σύγκλωνο
complete	πλήρες
complete bipartite	πλήρως διμερές
conjunctive normal form	κανονική συζευκτική μορφή
constraint satisfaction problem	πρόβλημα ικανοποίησης περιορισμών
counting	μετρητικός
decidable	αποφασίσιμος
dichotomy	διχοτομία
domain	πεδίο
dual	δυϊκός
embedding	εμβάπτιση
forest	δάσος
formula	τύπος
freely available	ελεύθερα διαθέσιμος
gadget	σκαρίφημα
Gaussian elimination	μέθοδος απαλοιφής Gauss
generator	γεννήτορας
graph homomorphisms	ομομορφισμοί γραφήματος
holographic algorithms	ολογραφικοί αλγόριθμοι

identification	συνταύτιση
infimum	μέγιστο κάτω φράγμα
instance	στιγμιότυπο
lattice	δικτυωτό
matchgate	πύλη ταιριασμάτων
matchgrid	πλέγμα ταιριασμάτων
oracle	μαντείο
parsimonious	φειδωλός
partition	διαμέριση
polymorphism	πολυμορφισμός
projection	προβολή
projection closure	κλειστότητα προβολών
pure affine	αγνά αφινική
rank	βαθμίδα
recognizer	αναγνωριστής
rectangular	ορθογώνιος
relation	σχέση
spanning tree	δενδροπαράγοντας
standard basis	συνήθης βάση
standard signature	κανονική υπογραφή
strong balance	ισχυρή ισορροπία
strong rectangularity	ισχυρή ορθογωνιότητα
strongly balanced	ισχυρά ισορροπημένο
supremum	ελάχιστο άνω φράγμα
unary	μονομελής
valid	έγκυρος
valuation	αποτίμηση
witness	μάρτυρας

Ελληνο-Αγγλικό

αγνά αφινική	pure affine
αναγνωριστής	recognizer
αποτίμηση	valuation
αποφασίσιμος	decidable
αφινικός	affine
βαθμίδα	rank
γεννήτορας	generator
δάσος	forest
δενδροπαράγοντας	spanning tree

διαγώνιος κατά μπλοκς	block-diagonal
διαμέριση	partition
διζευκτικός	bijunctive
δικτυωτό	lattice
διμελής	binary
διχοτομία	dichotomy
δυαδικό	boolean
δυϊκός	dual
έγκυρος	valid
ελάχιστο άνω φράγμα	supremum
ελεύθερα διαθέσιμος	freely available
εμβάπτιση	embedding
ισχυρά ισορροπημένο	strongly balanced
ισχυρή ισορροπία	strong balance
ισχυρή ορθογωνιότητα	strong rectangularity
κανονική συζευκτική μορφή	conjunctive normal form
κανονική υπογραφή	standard signature
καρτεσιανό γινόμενο	cartesian product
κλειστότητα προβολών	projection closure
μαντείο	oracle
μάρτυρας	witness
μέγιστο κάτω φράγμα	infimum
μέθοδος απαλοιφής Gauss	Gaussian elimination
μετρητικός	counting
μονομελής	unary
μπλοκ-ορθογωνιότητα	block-orthogonality
ολογραφικοί αλγόριθμοι	holographic algorithms
ομομορφισμοί γραφήματος	graph homomorphisms
ορθογώνιος	rectangular
όρος	clause
πεδίο	domain
πλέγμα ταιριασμάτων	matchgrid
πληθικός αριθμός	cardinality
πλήρες	complete
πλήρως διμερές	complete bipartite
πολυμορφισμός	polymorphism
πρόβλημα ικανοποίησης περιορισμών	constraint satisfaction problem
προβολή	projection
πύλη ταιριασμάτων	matchgate
σκαρίφημα	gadget

στιγμιότυπο	instance
σύγκλωνο	co-clone
συνήθης βάση	standard basis
συνταύτιση	identification
σχέση	relation
τύπος	formula
φειδωλός	parsimonious

Βιβλιογραφία

[1] Eric Allender, Michael Bauland, Neil Immerman, Henning Schnoor, and Heribert Vollmer. The complexity of satisfiability problems: Refining Schaefer's theorem. *Electronic Colloquium on Computational Complexity (ECCC)*, (100), 2004. 19, 22

[2] Sanjeev Arora and Boaz Barak. *Computational Complexity - A Modern Approach*. Cambridge University Press, 2009. 3, 5

[3] Andrei A. Bulatov. A dichotomy theorem for constraints on a three-element set. In *FOCS*, pages 649–658, 2002. 23

[4] Andrei A. Bulatov. The complexity of the counting constraint satisfaction problem. In *ICALP (1)*, pages 646–661, 2008. 30, 35, 36

[5] Andrei A. Bulatov and Víctor Dalmau. A simple algorithm for Mal'tsev constraints. *SIAM J. Comput.*, 36(1):16–27, 2006. 38

[6] Andrei A. Bulatov, Martin E. Dyer, Leslie Ann Goldberg, Markus Jalsenius, Mark Jerrum, and David Richerby. The complexity of weighted and unweighted #CSP. *CoRR*, abs/1005.2678, 2010. 30

[7] Andrei A. Bulatov, Martin E. Dyer, Leslie Ann Goldberg, Markus Jalsenius, and David Richerby. The complexity of weighted boolean #CSP with mixed signs. *Theor. Comput. Sci.*, 410(38-40):3949–3961, 2009. 30

[8] Andrei A. Bulatov and Martin Grohe. The complexity of partition functions. *Theor. Comput. Sci.*, 348(2-3):148–186, 2005. 29, 35

[9] Jin-Yi Cai. Holographic algorithms: guest column. *SIGACT News*, 39(2):51–81, 2008.

[10] Jin-Yi Cai and Xi Chen. A decidable dichotomy theorem on directed graph homomorphisms with non-negative weights. In *FOCS*, pages 437–446, 2010. 30

[11] Jin-Yi Cai and Xi Chen. Complexity of counting CSP with complex weights. In *STOC*, pages 909–920, 2012. 30, 35, 39

[12] Jin-Yi Cai, Xi Chen, and Pinyan Lu. Graph homomorphisms with complex values: A dichotomy theorem. In *ICALP (1)*, pages 275–286, 2010. 29

[13] Jin-Yi Cai, Xi Chen, and Pinyan Lu. Non-negatively weighted #CSP: An effective complexity dichotomy. In *IEEE Conference on Computational Complexity*, pages 45–54, 2011. 30

[14] Jin-Yi Cai, Sangxia Huang, and Pinyan Lu. From Holant to #CSP and back: Dichotomy for Holant^c problems. In *ISAAC (1)*, pages 253–265, 2010. 31

[15] Jin-Yi Cai, Pinyan Lu, and Mingji Xia. Holant problems and counting CSP. In *STOC*, pages 715–724, 2009. 30, 31

[16] Jin-Yi Cai, Pinyan Lu, and Mingji Xia. Dichotomy for Holant* problems of boolean domain. In *SODA*, pages 1714–1728, 2011. 31

[17] Nadia Creignou and Miki Hermann. Complexity of generalized satisfiability counting problems. *Inf. Comput.*, 125(1):1–12, 1996. 30

[18] Martin E. Dyer, Leslie Ann Goldberg, and Mark Jerrum. The complexity of weighted boolean #CSP. *CoRR*, abs/0704.3683, 2007. 28, 30

[19] Martin E. Dyer, Leslie Ann Goldberg, and Mike Paterson. On counting homomorphisms to directed acyclic graphs. *J. ACM*, 54(6), 2007. 30

[20] Martin E. Dyer and Catherine S. Greenhill. The complexity of counting graph homomorphisms. *Random Struct. Algorithms*, 17(3-4):260–289, 2000. 29

[21] Martin E. Dyer and David Richerby. On the complexity of #CSP. In *STOC*, pages 725–734, 2010. 30, 35, 37

[22] Tomás Feder and Moshe Y. Vardi. The computational structure of monotone monadic SNP and constraint satisfaction: A study through datalog and group theory. *SIAM J. Comput.*, 28(1):57–104, 1998. 23

[23] Michael Freedman, László Lovász, Alexander Schrijver, and Alexander Schrijver. Reflection positivity, rank connectivity, and homomorphism of graphs. 2007. 27

[24] Leslie Ann Goldberg, Martin Grohe, Mark Jerrum, and Marc Thurley. A complexity dichotomy for partition functions with mixed signs. *CoRR*, abs/0804.1932, 2008. 29

[25] Pavol Hell and Jaroslav Nesetril. On the complexity of H-coloring. *J. Comb. Theory, Ser. B*, 48(1):92–110, 1990. 10

[26] Francois Jaeger, Dirk L. Vertigan, and Dominic J. A. Welsh. On the computational complexity of the Jones and Tutte polynomials. *Mathematical Proceedings of the Cambridge Philosophical Society*, 108(1):35–53, 1990. 8

[27] Peter Jeavons, David A. Cohen, and Martin C. Cooper. Constraints, consistency and closure. *Artif. Intell.*, 101(1-2):251–265, 1998. 22

[28] Peter Jeavons, David A. Cohen, and Marc Gyssens. Closure properties of constraints. *J. ACM*, 44(4):527–548, 1997. 22

[29] Peter Jeavons and Martin C. Cooper. Tractable constraints on ordered domains. *Artif. Intell.*, 79(2):327–339, 1995. 22

[30] Richard M. Karp. Reducibility among combinatorial problems. In *Complexity of Computer Computations*, pages 85–103, 1972. 16

[31] Richard E. Ladner. On the structure of polynomial time reducibility. *J. ACM*, 22(1):155–171, 1975. 5

[32] Pinyan Lu. Complexity dichotomies of counting problems. *Electronic Colloquium on Computational Complexity (ECCC)*, 18:93, 2011.

[33] Ugo Montanari. Networks of constraints: Fundamental properties and applications to picture processing. *Inf. Sci.*, 7:95–132, 1974. 13

[34] Christos H. Papadimitriou. *Computational complexity*. Addison-Wesley, 1994. 3

[35] Thomas J. Schaefer. The complexity of satisfiability problems. In *STOC*, pages 216–226, 1978. 18

[36] Leslie G. Valiant. The complexity of computing the permanent. *Theor. Comput. Sci.*, 8:189–201, 1979. 4, 31

[37] Leslie G. Valiant. Holographic algorithms (extended abstract). In *FOCS*, pages 306–315, 2004. 31, 34

[38] Leslie G. Valiant. Holographic algorithms. *SIAM J. Comput.*, 37(5):1565–1594, 2008. 34

[39] Dirk Vertigan and Dominic J. A. Welsh. The compunational complexity of the Tutte plane: the bipartite case. *Combinatorics, Probability & Computing*, 1:181–187, 1992. 9

[40] Dominic J. A. Welsh. *Complexity: Knots, Colourings and Counting.* London Mathematical Society Lecture Note Series. Cambridge University Press, 1993. 8