ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

Τεχνικές Σχεδίασης Προσεγγιστικών Αλγορίθμων βασισμένες στο Γραμμικό Προγραμματισμό

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Ηλίας Σ. Διακονικόλας

Επιβλέπων : Ευστάθιος Κ. Ζάχος
Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2004

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

Τεχνικές Σχεδίασης Προσεγγιστικών Αλγορίθμων βασισμένες στο Γραμμικό Προγραμματισμό

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Ηλίας Σ. Διακονικόλας

Επιβλέπων : Ευστάθιος Κ. Ζάχος
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την $22^{η}$ Ιουλίου 2004.

Αθήνα, Ιούλιος 2004

....	....	....
Ε. Ζάχος	Τ. Σελλής	Φ. Αφράτη
Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.	Καθηγήτρια Ε.Μ.Π.

....
Ηλίας Σ. Διακονικόλας
Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Περίληψη

Τα τελευταία χρόνια έχουν αναπτυχθεί πολλοί νέοι προσεγγιστικοί αλγόριθμοι για $NP-$hard προβλήματα βελτιστοποίησης. Ένα μεγάλο ποσοστό αυτών βασίζεται στη θεωρία του γραμμικού προγραμματισμού. Οι αλγόριθμοι αυτοί χρησιμοποιούν δύο βασικές τεχνικές, την τεχνική της στρογγυλοποίησης (rounding) και το σχήμα πρωτεύον – δυϊκό (primal-dual). Η διπλωματική αυτή εργασία μελετάει και τις δύο τεχνικές με ιδιαίτερη έμφαση στην πρώτη και στις διάφορες υποκατηγορίες της. Χάριν πληρότητας, δίνεται αρχικά μία εισαγωγή στη μαθηματική θεωρία του γραμμικού προγραμματισμού.

Οι προσεγγιστικοί αλγόριθμοι που βασίζονται στη μέθοδο στρογγυλοποίησης επιλύουν (βέλτιστα) ένα γραμμικό πρόγραμμα το οποίο αποτελεί χαλάρωση του προς επίλυση προβλήματος βελτιστοποίησης. Εν γένει, ωστόσο, η λύση που υπολογίζεται τοιουτοτρόπως έχει κλασματικές συνιστώσες. Οι περισσότεροι αλγόριθμοι λύνουν το εν λόγω γραμμικό πρόγραμμα μία φορά και κατασκευάζουν μία ακέραια (εφικτή) λύση μέσω κατάλληλης διαδικασίας στρογγυλοποίησης. Ωστόσο, αυτή η μέθοδος φαίνεται να μην εκμεταλλεύεται την πλήρη ισχύ του γραμμικού προγραμματισμού. Μία βελτιωμένη τεχνική είναι αυτή της επαναλαμβανόμενης στρογγυλοποίησης, στην οποία στρογγυλοποιούμε τη λύση σε φάσεις. Μετά από κάθε φάση, υπολογίζουμε εκ νέου την καλύτερη κλασματική λύση, διατηρώντας τη στρογγυλοποίηση που είχε πραγματοποιηθεί σε προηγούμενες φάσεις. Επιπλέον, η στρογγυλοποίηση μπορεί να γίνει τόσο με ντετερμινιστικό όσο και με randomized τρόπο. Ειδική έμφαση δίνεται στη randomized στρογγυλοποίηση.

Οι προσεγγιστικοί αλγόριθμοι που βασίζονται στη μέθοδο πρωτεύοντος-δυϊκού λειτουργούν με ένα ζεύγος πρωτεύοντος και δυϊκού γραμμικού προγράμματος. Στην παρούσα εργασία εφαρμόζουμε αυτό το σχήμα μόνο για περιπτώσεις που όλοι οι συντελεστές του πρωτεύοντος και του δυϊκού γραμμικού προγράμματος είναι μη αρνητικοί. Πρόσφατα, αποδείχθηκε ότι η εν λόγω μέθοδος μπορεί να εφαρμοστεί και σε περιπτώσεις που εμφανίζονται και αρνητικοί συντελεστές. Σημειώνεται ότι οι αλγόριθμοι αυτής της κατηγορίας είναι γενικά πιο γρήγοροι από αυτούς της πρώτης κατηγορίας.

Abstract

Over the last few years many new approximation algorithms have been developed for NP-hard optimization problems. A large fraction of these improved algorithms are based on linear programming theory. These algorithms use two basic techniques, rounding and primal – dual schema. This diploma thesis analyses both techniques with a special emphasis on the former and its subcategories. For completeness, an introduction to the mathematical theory of linear programming is also given.

Rounding based approximation algorithms solve an LP relaxation of the optimization problem. In general the solution obtained is fractional. Most algorithms solve the LP relaxation once and obtain an integral solution by a suitable rounding process. However, this method seems not to exploit the full power of linear programming. An enhanced technique is iterative rounding, in which the solution is rounded in phases. After each phase, we re - compute the best fractional solution, maintaining the rounding achieved in the previous phases. Furthermore, rounding can be accomplished in both a deterministic and a randomized way. Special emphasis is also given on randomized rounding.

Primal Dual approximation algorithms work with a primal and dual pair of linear programs. In this thesis we apply this schema only when there are no negative coefficients in either the primal or the dual programs. However, it has recently been shown that the method can also be applied in cases which have negative coefficients. Primal Dual schema based algorithms are generally faster that rounding based algorithms.

Κεφάλαιο 1

Εισαγωγή

Πολλά προβλήματα βελτιστοποίησης που εμφανίζονται στην πράξη είναι συνήθως $NP-\text{hard}$, π.χ. προβλήματα δρομολόγησης (routing), προβλήματα χρονοδρομολόγησης (scheduling), προβλήματα διαμέρισης (partitioning) κλπ. Σύμφωνα λοιπόν με τη γενικά αποδεκτή υπόθεση ότι $P \neq NP$, δεν υπάρχει ελπίδα εύρεσης αλγορίθμων πολυωνυμικού χρόνου για την επίλυση τέτοιων προβλημάτων. Ωστόσο, στην πράξη κάτι τέτοιο είναι επιβεβλημένο. Μάλιστα, στην πράξη, εμφανίζονται συχνά πολύ μεγάλα στιγμιότυπα τέτοιων προβλημάτων. Οι Βάσεις Δεδομένων, η Υπολογιστική Βιολογία, η σχεδίαση VLSI και το διαδίκτυο είναι μερικές από τις πηγές μεγάλων στιγμιοτύπων.

Η αποτυχία ακριβούς επίλυσης $NP-\text{hard}$ προβλημάτων βελτιστοποίησης σε πολυωνυμικό χρόνο, έχει οδηγήσει τους ερευνητές στην εύρεση διαφόρων ειδών λύσεων σε αυτά τα προβλήματα. Κάποιες από τις προσεγγίσεις που ακολουθούνται είναι:

1. Ευριστικές μέθοδοι (heuristics).
2. Ανάλυση της μέσης περίπτωσης (average case analysis).
3. «Σχεδόν» βέλτιστες λύσεις – Προσεγγιστικοί Αλγόριθμοι.

Η δυσκολία με την πρώτη προσέγγιση είναι ότι οι ευριστικές μέθοδοι, όπως η μέθοδος Branch and Bound και το simulated annealing [LA87], είναι αλγόριθμοι εκθετικού χρόνου και στηρίζονται στην υπόθεση ότι τα χειρότερα σενάρια δε θα εμφανιστούν στην πράξη. Ωστόσο, όταν έχουμε να κάνουμε με στιγμιότυπα αστρονομικού μεγέθους, αυτή η υπόθεση παύει να ισχύει και επομένως οι εν λόγω μέθοδοι γίνονται μη πρακτικές. Η δεύτερη προσέγγιση βρίσκει μια λύση αποδοτικά κατά μέσο όρο για όλα τα δυνατά στιγμιότυπα. Η δυσκολία με τη δεύτερη προσέγγιση είναι ότι απαιτεί να υποθέσει κανείς μία κατανομή πιθανότητας για τα δεδομένα εισόδου. Όμως, τα δεδομένα εισόδου ενδέχεται να μην ακολουθούν την υποτεθείσα κατανομή και ως εκ τούτου η ανάλυση να μην ισχύει.

Η τρίτη προσέγγιση χαλαρώνει τον ίδιο το χώρο των εφικτών λύσεων. Αντί λοιπόν να υπολογίζει κανείς μία βέλτιστη λύση, υπολογίζει *αποδοτικά* μία λύση τη οποίας η τιμή είναι *αποδεδειγμένα* κοντά στη βέλτιστη τιμή. Με άλλα λόγια, η εναλλακτική αυτή προσέγγιση, περιλαμβάνει την ανάπτυξη και την υλοποίηση προσεγγιστικών αλγορίθμων για το δοθέν πρόβλημα, οι οποίοι να συνοδεύονται από *αποδεδειγμένες* αποδόσεις.

Για τους σκοπούς μας, ένας αλγόριθμος θεωρείται αποδοτικός αν τρέχει σε πολυωνυμικό χρόνο (χρόνο που είναι φραγμένος από ένα *σταθερό* πολυώνυμο του μήκους της εισόδου). Για τους περισσότερους αλγορίθμους που παρουσιάζουμε (ως παραδείγματα εφαρμογής των διαφόρων τεχνικών) δεν υπολογίζουμε ξεχωριστά την πολυπλοκότητά τους. Στις περισσότερες περιπτώσεις είναι φανερό από τα συμφραζόμενα ότι είναι πολυωνυμικού χρόνου. Αξίζει να τονιστεί ότι δίνεται έμφαση στην ποιότητα της προσέγγισης που εγγυώνται οι αλγόριθμοί μας. Ακολούθως υπενθυμίζονται κάποιες βασικές έννοιες από τη θεωρία των προσεγγιστικών αλγορίθμων.

Δοθέντος ενός προβλήματος βελτιστοποίησης $\mathcal{P}$ και ενός στιγμιότυπου I του $\mathcal{P}$, ας συμβολίσουμε με $OPT(I)$ τη βέλτιστη τιμή της αντικειμενικής συνάρτησης για το I (υποθέτουμε ότι για κάθε εφικτή λύση του στιγμιότυπου I η τιμή της αντικειμενικής συνάρτησης είναι μη αρνητική).

Αν το $\mathcal{P}$ είναι πρόβλημα *μεγιστοποίησης*, ένας προσεγγιστικός αλγόριθμος $\mathcal{A}$ για το $\mathcal{P}$ είναι ένας αποδοτικός αλγόριθμος ο οποίος, για κάποιο $\lambda \geq 1$, παράγει μια εφικτή λύση με τιμή αντικειμενικής συνάρτησης τουλάχιστον $OPT(I)/\lambda$ *για όλα* τα στιγμιότυπα I. Ο $\mathcal{A}$ ονομάζεται λ - προσεγγιστικός αλγόριθμος για το $\mathcal{P}$. Το λ ονομάζεται (απόλυτος) λόγος προσέγγισης ή λόγος απόδοσης του αλγορίθμου. Για να διατηρήσουμε τη σύμβαση ότι $\lambda \geq 1$ και για προβλήματα ελαχιστοποίησης, λέμε ότι ένας αλγόριθμος $\mathcal{A}$ είναι λ - προσεγγιστικός αλγόριθμος για ένα πρόβλημα *ελαχιστοποίησης* $\mathcal{P}$, αν ο $\mathcal{A}$ παράγει μία εφικτή λύση με τιμή το πολύ $OPT(I)\cdot\lambda$ *για όλα* τα στιγμιότυπα I του $\mathcal{P}$. Συνεπώς, οι δύο φράσεις – κλειδιά στους προσεγγιστικούς αλγορίθμους είναι η *αποδοτικότητα* και η *αποδεδειγμένη απόδοση προσέγγισης*.

Είναι γνωστό ότι για μερικά προβλήματα ο *απόλυτος* λόγος απόδοσης δεν είναι το καλύτερο μέτρο για την αξιολόγηση της απόδοσης ενός αλγορίθμου. Για τέτοιες περιπτώσεις ορίζουμε τον *ασυμπτωτικό* λόγο απόδοσης. Ο τελευταίος ικανοποιεί τις ίδιες ιδιότητες με τον απόλυτο για όλα τα στιγμιότυπα I του προβλήματος που έχουν *αρκετά μεγάλη* τιμή για το $OPT(I)$. Τονίζεται ότι οι λόγοι που ορίστηκαν παραπάνω δεν είναι κατ' ανάγκην σταθερές, αλλά εν γένει εξαρτώνται από το μήκος της εισόδου του προβλήματος. Πολλές φορές μάλιστα θεωρούμε ως λόγο απόδοσης (απόλυτο ή ασυμπτωτικό) το *infimum* όλων αυτών που ικανοποιούν τις προαναφερθείσες ιδιότητες. Οι παραπάνω ορισμοί τροποποιούνται ελαφρά σε περίπτωση που ο υπό εξέταση προσεγγιστικός αλγόριθμος είναι *randomized* και όχι ντετερμινιστικός, όπως θα φανεί σε επόμενο κεφάλαιο.

Λέμε ότι ένα πρόβλημα βελτιστοποίησης έχει *προσεγγιστικό σχήμα πολυωνυμικού χρόνου* (polynomial time approximation scheme-*PTAS*) αν υπάρχει μια οικογένεια αλγορίθμων $\{A_\epsilon\}$ τέτοια ώστε ο αλγόριθμος A_ϵ να είναι $(1+\epsilon)$ - προσεγγιστικός για κάθε σταθερό $\epsilon > 0$. Ο χρόνος εκτέλεσης του A_ϵ είναι πολυωνυμικός ως προς το μήκος της εισόδου, όχι όμως κατ' ανάγκη και ως προς το $1/\epsilon$. Αν κάτι τέτοιο ισχύει, τότε μιλάμε για ένα προσεγγιστικό σχήμα πλήρως πολυωνυμικού χρόνου (fully polynomial time approximation scheme - *FPTAS*).

Υπάρχουν πολλές μέθοδοι για τη σχεδίαση αλγορίθμων σε αυτή την περιοχή, όπως η άπληστη (greedy) προσέγγιση, ο δυναμικός προγραμματισμός, η μέθοδος primal – dual και γενικότερα μέθοδοι που βασίζονται στο μαθηματικό προγραμματισμό (ειδικότερα στο γραμμικό και στο semidefinite), η μέθοδος τοπικής βελτίωσης κ. ά.

Ο όρος «προσεγγιστικός αλγόριθμος» φαίνεται να έχει εισαχθεί από τον David Johnson σε ένα πρωτοποριακό paper του 1974 [Jo74]. Σε αυτό το paper ο Johnson δίνει προσεγγιστικούς αλγορίθμους για αρκετά κλασικά προβλήματα, συμπεριλαμβανόμενου ενός PTAS για το πρόβλημα αθροίσματος υποσυνόλου, ενός 2-προσεγγιστικού αλγόριθμου για το πρόβλημα MAX SAT (εκδοχή βελτιστοποίησης του προβλήματος SAT), έναν $(\ln n+1)$ - προσεγγιστικό για το πρόβλημα set cover, όπως επίσης και ευριστικές μεθόδους για το πρόβλημα χρωματισμού γράφων και το πρόβλημα της μέγιστης κλίκας, παρόλο που δεν είχε μπορέσει να βρει λόγο προσέγγισης $O\left(n^{1-\epsilon}\right)$ για κανένα $\epsilon > 0$ για τα δύο αυτά προβλήματα, όπου n είναι ο αριθμός των κορυφών του γραφήματος.

Ωστόσο, οι προσεγγιστικοί αλγόριθμοι ήταν παρόντες στη βιβλιογραφία ακόμα και πριν την εισαγωγή της έννοιας της NP-πληρότητας. Ο Lovász αναφέρει ότι σε ένα paper του Erdös του 1967 [Er67] αποδεικνύεται ότι το μέγεθος μιας μέγιστης τομής (max cut) σε ένα γράφημα με μη αρνητικά βάρη στις ακμές είναι τουλάχιστον ίσο με το μισό του αθροίσματος των βαρών όλων των ακμών και ότι η απόδειξη είναι κατασκευαστική με την πολυωνυμική έννοια (δηλαδή μπορεί να μετατραπεί σε ένα 2-προσεγγιστικό αλγόριθμο για το πρόβλημα της μέγιστης τομής). Ο Graham [Gra66] έδωσε ένα σύνολο από 2-

προσεγγιστικούς αλγορίθμους για ένα πλήθος προβλημάτων χρονοδρομολόγησης το 1966. Τέλος, μόλις το 1964 ο Vizing [Vi64] έδωσε έναν αλγόριθμο για την εύρεση χρωματισμού ακμών σε ένα γράφημα ο οποίος χρησιμοποιεί το πολύ ένα χρώμα περισσότερο από το ελάχιστο δυνατόν.

Τα τελευταία χρόνια έλαβαν χώρα πολλές εντυπωσιακές ανακαλύψεις στην περιοχή των προσεγγιστικών αλγορίθμων. Δεν είναι εφικτό, στο χώρο που διατίθεται σε αυτή την εισαγωγή, να δοθεί μια ικανοποιητική επισκόπηση αυτών των αποτελεσμάτων. Από την άποψη της υπολογιστικής πολυπλοκότητας, μία ενδιαφέρουσα παρατήρηση που προέκυψε από τη μελέτη των προσεγγιστικών αλγορίθμων είναι ότι παρόλο που τα $NP-complete$ προβλήματα απόφασης είναι ισοδύναμα υπό την έννοια της ακριβούς επίλυσης, τα αντίστοιχα (φυσικά) προβλήματα βελτιστοποίησης ανήκουν σε ένα ιδιαίτερα ευρύ φάσμα υπό την έννοια της προσεγγισιμότητας.

Με άλλα λόγια, αποδείχθηκε ότι υπάρχουν προβλήματα για τα οποία η εύρεση ενός (πολυωνυμικού) προσεγγιστικού αλγορίθμου που επιτυγχάνει συγκεκριμένο λόγο απόδοσης είναι το ίδιο δύσκολο πρόβλημα με την εύρεση ακριβούς πολυωνυμικού αλγορίθμου για το ίδιο πρόβλημα. Τα αποτελέσματα αυτά έχουν τις ρίζες τους σε τμήματα της θεωρητικής πληροφορικής που δε σχετίζονται με τη βελτιστοποίηση. Η σύνδεση με τη βελτιστοποίηση έγινε από το κομβικής σημασίας paper των Feige, Goldwasser, Lovász, Safra και Szegedy [FGL+96] το οποίο ακολούθησαν δύο papers των Arora και Safra [AS92, AS98] και δύο των Arora, Lund, Motwani, Sudan και Szegedy [ALM+92, ALM+98] και τα οποία έδωσαν έναν εναλλακτικό ορισμό του NP.

Συνέπεια των προηγούμενων αποτελεσμάτων ήταν ότι μια συλλογή από προβλήματα βελτιστοποίησης δεν μπορούν να έχουν προσεγγιστικό σχήμα πολυωνυμικού χρόνου, εκτός κι αν $P = NP$. Σε αυτή τη συλλογή ανήκει το συμμετρικό πρόβλημα του πλανόδιου πωλητή (με τριγωνική ανισότητα φυσικά), το πρόβλημα της μέγιστης τομής, το πρόβλημα της ελάχιστης κάλυψης κορυφών (min vertex cover) και το πρόβλημα MAX SAT.

Τα εν λόγω αποτελέσματα ενισχύθηκαν και επεκτάθηκαν αργότερα με διάφορους τρόπους. Για παράδειγμα, με μία σειρά από papers [FGL+96, BGS98], που κατέληξαν σε ένα paper του Håstad [Ha99] αποδείχθηκε ότι για το πρόβλημα της μέγιστης κλίκας, δεν υπάρχει λόγος απόδοσης $O\left(n^{1-\epsilon}\right)$ για κανένα $\epsilon > 0$, εκτός κι αν $NP = RP$. Οι Feige και Killian επέκτειναν αυτό το αποτέλεσμα στο πρόβλημα του ελάχιστου χρωματικού αριθμού [FK98]. Μία άλλη σειρά από papers που ξεκίνησαν με τη δουλειά των Lund και Γιαννακάκη [LY94] και η οποία αργότερα βελτιώθηκε από άλλους ερευνητές, έδειξε ότι δεν υπάρχει $c \ln n$-προσεγγιστικός αλγόριθμος για το πρόβλημα set cover για $c < 1$ εκτός κι αν υπάρχει αλγόριθμος πολυπλοκότητας $O\left(n^{O(\log\log n)}\right)$ για κάθε NP-πλήρες πρόβλημα. Πρόσφατα κάποια άλλα papers έδειξαν ότι δεν υπάρχουν σταθεροί παράγοντες προσέγγισης καλύτεροι από συγκεκριμένα φράγματα για ορισμένα προβλήματα, εκτός κι αν $P = NP$. Για παράδειγμα $\frac{129}{128}$ για το συμμετρικό πρόβλημα του πλανόδιου πωλητή με τριγωνική ανισότητα [PV00], $\frac{16}{17}$ για το πρόβλημα της μέγιστης τομής [Ha97], $\frac{7}{6}$ για το πρόβλημα της ελαχίστου βάρους κάλυψης κόμβων [Ha97] και $\frac{7}{6}$ για το πρόβλημα MAX SAT [Ha97].

Κεφάλαιο 2

Εισαγωγή στο Γραμμικό Προγραμματισμό

Ο γραμμικός προγραμματισμός (Linear Programming - *LP*) περιλαμβάνει μια πολύ σημαντική κλάση προβλημάτων βελτιστοποίησης. Σε ένα πρόβλημα γραμμικού προγραμματισμού θέλουμε να βρούμε τιμές για ορισμένες μεταβλητές οι οποίες (*i*) να ικανοποιούν ένα σύνολο γραμμικών ισοτήτων και ανισοτήτων και (*ii*) ανάμεσα σε αυτές τις τιμές θέλουμε να επιλέξουμε εκείνη που μεγιστοποιεί ή ελαχιστοποιεί μια δοθείσα γραμμική αντικειμενική συνάρτηση.

Ο γραμμικός προγραμματισμός έχει πολλές εφαρμογές. Από την αλγοριθμική πλευρά, ο simplex προτάθηκε τη δεκαετία του ’40 και παρόλο που είχε πολύ καλά αποτελέσματα στην πράξη, ήταν από τότε γνωστό ότι έχει εκθετικό χρόνο εκτέλεσης στη χειρότερη περίπτωση. Απ’ την άλλη, από τις αρχές της δεκαετίας του ’70, οπότε και ορίστηκαν οι κλάσεις *P* και *NP*, παρατηρήθηκε ότι $\mathcal{LP} \in \Delta = NP \cap co-NP$ παρόλο που δεν ήταν γνωστός κανένας πολυωνυμικός αλγόριθμος για το πρόβλημα.

Ο πρώτος πολυωνυμικός αλγόριθμος, η ελλειψοειδής μέθοδος [Kha79], ανακαλύφθηκε στα τέλη της ίδιας δεκαετίας. Ο αλγόριθμος του Karmarkar στα μέσα της δεκαετίας του ’80 οδήγησε σε ενεργοποίηση της έρευνας στην περιοχή των μεθόδων εσωτερικού σημείου για το γραμμικό προγραμματισμό. Ο γραμμικός προγραμματισμός και ειδικά η έννοια της δυϊκότητας (duality) είναι μια πολύ σημαντική αποδεικτική τεχνική με σημαντικότατες εφαρμογές στο πεδίο τόσο των ακριβών όσο και των προσεγγιστικών αλγορίθμων. Για μια εις βάθος ανάλυση του γραμμικού προγραμματισμού, προτείνονται τα ακόλουθα συγγράμματα [Chv83, Schr86]. Η ανάλυση που ακολουθεί στηρίζεται στο [G94].

2.1 Εισαγωγή

Ένα γραμμικό πρόγραμμα είναι το πρόβλημα της βελτιστοποίησης μιας γραμμικής αντικειμενικής συνάρτησης ως προς ένα σύνολο μεταβλητών $x_1, x_2, ..., x_n$ (μεταβλητές απόφασης – decision variables) οι οποίες υπόκεινται σε περιορισμούς γραμμικών ισοτήτων και ανισοτήτων. Σε *standard* μορφή, ένα γραμμικό πρόγραμμα εκφράζεται ως εξής:

$$\min \sum_{j=1}^{n} c_j \cdot x_j \qquad \text{(αντικειμενική συνάρτηση)}$$

με τους περιορισμούς:

$$\sum_{j=1}^{n} a_{ij} \cdot x_j = b_i, \quad i \in [m] \qquad \text{(περιορισμοί)}$$

$$x_j \geq 0, \qquad j \in [n] \quad \text{(περιορισμοί μη αρνητικής τιμής)}$$

όπου τα $\{a_{ij}, b_i, c_j\}$ είναι δοσμένα. Από τούδε και στο εξής, το γράμμα *m* για να συμβολίζει το πλήθος των περιορισμών και το γράμμα *n* το πλήθος των μεταβλητών απόφασης.

Ένα γραμμικό πρόγραμμα εκφράζεται πιο βολικά με χρήση πινάκων:

$$\min c^T x \text{ με τους περιορισμούς : } \begin{cases} Ax = b \\ x \geq \mathbf{0}_{n\times 1} \end{cases}$$

όπου

$$x = \left(x_1, x_2, ..., x_n\right)^T \in \mathbb{R}^{n\times 1}$$

$$b = \left(b_1, b_2, ..., b_m\right)^T \in \mathbb{R}^{m\times 1} \text{ και } A = \begin{pmatrix} a_{11} & . & . & a_{1n} \\ . & . & . & . \\ . & . & . & . \\ a_{m1} & . & . & a_{mn} \end{pmatrix} \in \mathbb{R}^{m\times n}$$

$$c = \left(c_1, c_2, ..., c_n\right)^T \in \mathbb{R}^{n\times 1}$$

2.2 Βασική Ορολογία

Ορισμός 2.1 Αν το διάνυσμα x ικανοποιεί τους περιορισμούς $Ax = b,\ x \geq \mathbf{0}$, τότε το x ονομάζεται εφικτή λύση.

Ορισμός 2.2 Ένα γραμμικό πρόγραμμα (LP) ονομάζεται εφικτό αν έχει τουλάχιστον μία εφικτή λύση, ειδάλλως ονομάζεται μη εφικτό.

Ορισμός 2.3 Μια βέλτιστη λύση x^* είναι μια εφικτή λύση, τέτοια ώστε να ισχύει $c^T x^* = \min\left\{c^T x : Ax = b, x \geq \mathbf{0}_{n\times 1}\right\}$.

Ορισμός 2.4 Ένα γραμμικό πρόγραμμα είναι *μη* φραγμένο κάτω αν ισχύει $\forall \lambda \in \mathbb{R}\ \exists$ εφικτή λύση x^* τ.ω. $c^T x^* \leq \lambda$.

2.3 Ισοδύναμες μορφές

Ένα γραμμικό πρόγραμμα μπορεί να πάρει αρκετές μορφές. Ενδέχεται ο στόχος να είναι η μεγιστοποίηση της γραμμικής αντικειμενικής συνάρτησης. Επιπλέον, μπορεί να υπάρχει συνδυασμός από περιορισμούς ισότητας και ανισότητας. Κάποιες μεταβλητές μπορεί να έχουν περιορισμό μη θετικής τιμής ή ακόμα να μην έχουν περιορισμό πρόσημου. Δύο μορφές ονομάζονται *ισοδύναμες* αν είτε έχουν το ίδιο σύνολο βέλτιστων λύσεων, είτε είναι και οι δύο μη εφικτές, είτε και οι δύο μη φραγμένες. Είναι εύκολο να μετατρέψει κανείς τους περιορισμούς από τη μια μορφή στην άλλη:

1. Ένα πρόβλημα μεγιστοποίησης μπορεί να εκφραστεί ως πρόβλημα ελαχιστοποίησης ως εξής:
$$\max c^T x \Leftrightarrow \min\left(-c^T x\right)$$
2. Μια ισότητα μπορεί να αναπαρασταθεί ως ένα ζεύγος ανισοτήτων:
$$a_i^T x = b_i \Leftrightarrow \begin{Bmatrix} a_i^T x \leq b_i \\ a_i^T x \geq b_i \end{Bmatrix} \Leftrightarrow \begin{Bmatrix} a_i^T x \leq b_i \\ -a_i^T x \leq -b_i \end{Bmatrix}$$
3. Προσθέτοντας μια slack variable, μια ανισότητα μπορεί να αναπαρασταθεί ως συνδυασμός μιας ισότητας και περιορισμών μη αρνητικής τιμής:
$$a_i^T x \leq b_i \Leftrightarrow a_i^T x + s_i = b_i, s_i \geq 0$$

4. Περιορισμοί μη θετικής τιμής μπορούν να εκφρασθούν ως περιορισμοί μη αρνητικής τιμής. Πράγματι, για να εκφράσουμε τον περιορισμό $x_j \le 0$, αντικαθιστούμε κάθε εμφάνιση της μεταβλητής x_j με $-y_j$ και επιβάλλουμε τη συνθήκη $y_j \ge 0$.
5. Επίσης η μεταβλητή x μπορεί να μην έχει περιορισμό προσήμου. Εν τοιαύτη περιπτώσει αντικαθιστούμε παντού την εν λόγω μεταβλητή με την τιμή $x^+ - x^-$, όπου x^+, x^- είναι νέες μεταβλητές, προσθέτοντας τους περιορισμούς $x^+, x^- \ge 0$.

Εν γένει, λοιπόν, μια ανισότητα μπορεί να αναπαρασταθεί ισοδύναμα χρησιμοποιώντας ένα συνδυασμό περιορισμών ισότητας και μη αρνητικής τιμής, και αντιστρόφως.

Χρησιμοποιώντας αυτούς τους κανόνες, το $\min\left\{c^T x : Ax \ge b\right\}$ μπορεί να μετασχηματισθεί ισοδύναμα στην ακόλουθη μορφή:

$$\min\left\{c^T x^+ - c^T x^- : Ax^+ - Ax^- - I \cdot s = b,\ x^+,\ x^-,\ s \ge 0\right\}$$

Λέμε ότι ένα γραμμικό πρόγραμμα είναι σε *κανονική* μορφή (canonical form) όταν εκφράζεται με την πρώτη από τις παραπάνω δύο μορφές, ενώ όταν εκφράζεται με τη δεύτερη λέμε ότι είναι σε *standard* μορφή.

Αντίστροφα ένα γραμμικό πρόγραμμα σε standard μορφή μπορεί να γραφεί σε κανονική μορφή. Πράγματι το $\min\left\{c^T x : Ax = b\right\}$ είναι ισοδύναμο με το ακόλουθο:

$$\min\left\{c^T x : Ax \ge b,\ \text{-}Ax \ge -b,\ I \cdot x \ge \mathbf{0}\right\}$$

Η παραπάνω συνθήκη μπορεί ισοδύναμα να γραφεί ως $A'x \ge b'$, όπου $A' = \begin{pmatrix} A \\ -A \\ I \end{pmatrix}$ και $b' = \begin{pmatrix} b \\ -b \\ \mathbf{0} \end{pmatrix}$.

2.4 Η Γεωμετρία του Γραμμικού Προγραμματισμού

Έστω το σύνολο των εφικτών λύσεων $P = \left\{x\ :\ Ax = b,\ x \ge \mathbf{0}\right\} \subseteq \mathbb{R}^n$.

Ορισμός 2.5 Το διάνυσμα x είναι κορυφή του συνόλου P αν $\nexists y \ne \mathbf{0} : \left(\left(x + y \in P\right) \wedge \left(x - y \in P\right)\right)$.

Θεώρημα 2.1 Έστω ότι το $\min\left\{c^T x : x \in P\right\}$ είναι πεπερασμένο. Τότε $\forall x \in P$ $\exists$ κορυφή $x' : c^T x' \le c^T x$.

Απόδειξη:

Αν το x είναι κορυφή, τότε προφανώς το $x' = x$ ικανοποιεί τετριμμένα την παραπάνω σχέση.

Αν το x δεν είναι κορυφή, τότε, εξ ορισμού, $\exists y \ne \mathbf{0} : \left(\left(x + y \in P\right) \wedge \left(x - y \in P\right)\right)$. Άρα, έχουμε ότι $A\left(x + y\right) = A\left(x - y\right) = b$ και συνεπώς ότι $Ay = \mathbf{0}$.

Χωρίς βλάβη της γενικότητας μπορούμε να υποθέσουμε ότι $c^T y \leq 0$ (παίρνουμε είτε το y είτε το $-y$). Στην ειδική περίπτωση που ισχύει $c^T y = 0$, επιλέγουμε το y έτσι ώστε να έχει τουλάχιστον μια αρνητική συνιστώσα, δηλαδή $\exists j : y_j < 0$. Εφόσον ισχύει $y \neq \mathbf{0}$ και $c^T y = c^T(-y) = 0$, είτε το y είτε το $-y$ θα ικανοποιεί αυτή την ιδιότητα.

Θεωρούμε το διάνυσμα $x + \lambda \cdot y$, όπου $\lambda > 0$. Έχουμε ότι $c^T(x + \lambda \cdot y) = c^T x + \lambda c^T y \leq c^T x \quad (1)$, λόγω των παραπάνω υποθέσεων. Δηλαδή το εν λόγω διάνυσμα ικανοποιεί την επιθυμητή ιδιότητα, αλλά εν γένει δεν είναι κορυφή του *P*. Ακολούθως περιγράφουμε μια επαγωγική διαδικασία για να καταλήξουμε στην επιθυμητή κορυφή. Διακρίνουμε τις παρακάτω περιπτώσεις:

$1^η$ Περίπτωση $\exists j : y_j < 0$

Αυξανομένου του λ, η j - οστή συνιστώσα του διανύσματος $x + \lambda \cdot y$ μειώνεται και για $\lambda > \lambda_0$, όπου λ_0 μια οριακή τιμή για το λ, το εν λόγω διάνυσμα παύει να αποτελεί εφικτή λύση (διότι τουλάχιστον μία συνιστώσα του είναι αρνητική και ως εκ τούτου δεν ικανοποιεί τον περιορισμό μη αρνητικής τιμής). Επιλέγουμε ως τιμή του λ την προαναφερθείσα οριακή τιμή, δηλαδή $\lambda = \lambda_0 = \min_{\{j : y_j < 0\}} \left\{ {x_j}/{-y_j} \right\} = {x_k}/{-y_k}$. Η τιμή αυτή αντιστοιχεί προφανώς στη μέγιστη δυνατή τιμή του λ για την οποία εξακολουθεί να ισχύει $x + \lambda \cdot y \geq \mathbf{0}$. Εφόσον $Ay = \mathbf{0}$, έπεται ότι $A(x + \lambda \cdot y) = Ax + \lambda Ay = Ax \quad = b$. Συνεπώς, $x + \lambda_0 \cdot y \in P$ και επιπλέον το εν λόγω διάνυσμα έχει τουλάχιστον μία μηδενική συνιστώσα περισσότερη από το x, την $(x + \lambda_0 \cdot y)_k$.

Αντικαθιστούμε το x με το $x + \lambda_0 \cdot y$.

$2^η$ Περίπτωση $\forall j : y_j \geq 0$

Εξ υποθέσεως, έχουμε ότι $c^T y < 0$ και ότι το $x + \lambda \cdot y$ είναι εφικτή λύση *για κάθε* $\lambda \geq 0$, εφόσον αφενός $A(x + \lambda \cdot y) = Ax = b$ (όπως και στην $1^η$ περίπτωση) και αφετέρου $x + \lambda \cdot y \geq x \geq \mathbf{0}$. Ωστόσο, έχουμε ότι $c^T(x + \lambda \cdot y) \; = c^T x + \lambda \cdot c^T y \to -\infty$ καθώς το $\lambda \to +\infty$. Επομένως, καταλήγουμε στο ότι το υπό εξέταση γραμμικό πρόγραμμα είναι μη φραγμένο, γεγονός που αντιτίθεται στην υπόθεση του θεωρήματος (άτοπο).

Η $1^η$ περίπτωση μπορεί να συμβεί **το πολύ** *n* φορές, εφόσον το αρχικό σημείο $x \in \mathbb{R}^n$ έχει *n* συνιστώσες. Με επαγωγή στον αριθμό των μη μηδενικών συνιστωσών του *x*, τελικά καταλήγουμε σε μια κορυφή x'. Η ορθότητα της επαγωγής είναι φανερή λόγω της συνθήκης (1).

□

Παρατήρηση: Το παραπάνω θεώρημα διατυπώθηκε για το εξής σύνολο εφικτών λύσεων $P=\{x\ :\ Ax=b,\ x\geq \mathbf{0}\}$. Μιλώντας αυστηρά, το εν λόγω θεώρημα δεν ισχύει για το σύνολο $P=\{x\ :\ Ax\geq b\}$. Πράγματι, ένα τέτοιο σύνολο *P* ενδέχεται να μην έχει καμία κορυφή. Για παράδειγμα, θεωρείστε το σύνολο $P=\{(x_1,x_2):0\leq x_2\leq 1\}$. Το συγκεκριμένο πολύεδρο δεν έχει κορυφή, διότι για κάθε $x\in P$ και για $y=\begin{pmatrix}1 & 0\end{pmatrix}$ έχουμε $((x+y\in P)\wedge(x-y\in P))$. Γενικά μπορεί να αποδειχθεί ότι το εν λόγω σύνολο $P=\{x\ :\ Ax\geq b\}$ έχει κορυφή αν και μόνο αν ο βαθμός του πίνακα *A* είναι ίσος με *n*, $Rank(A)=n$, δηλαδή αν ο πίνακας A έχει *n* γραμμικά ανεξάρτητες στήλες (όλες οι στήλες του γραμμικά ανεξάρτητες). Μπορεί να παρατηρήσει κανείς ότι, αν μετασχηματίσουμε ένα γραμμικό πρόγραμμα από standard μορφή σε κανονική μορφή, οι περιορισμοί μη αρνητικής τιμής έπονται ότι ο πίνακας A' που προκύπτει είναι πλήρους βαθμού ως προς τις στήλες, εφόσον προφανώς $Rank\left(\begin{bmatrix}A & -A & I\end{bmatrix}^T\right)=n$.

Πόρισμα 2.2 Αν το ελάχιστο $\min\{c^Tx:Ax=b,\ x\geq\mathbf{0}\}$ είναι πεπερασμένο, υπάρχει βέλτιστη λύση x^*, η οποία είναι κορυφή.

Απόδειξη:

Άμεση συνέπεια του θεωρήματος 1.

□

Πόρισμα 2.3 Αν $P=\{x:Ax=b,x\geq\mathbf{0}\}\neq\varnothing$, τότε το *P* έχει τουλάχιστον μία κορυφή.

Το ακόλουθο θεώρημα παρέχει μια ικανή και αναγκαία συνθήκη για την αναγνώριση κορυφών.

Θεώρημα 2.4 Έστω $P=\{x:Ax=b,x\geq\mathbf{0}\}$. Για $x\in P$, έστω A_x ο υποπίνακας του *A* που αντιστοιχεί σε εκείνα τα *j* για τα οποία $x_j>0$. Τότε (*i*) το *x* είναι κορυφή αν και μόνο αν (*ii*) ο πίνακας A_x έχει γραμμικά ανεξάρτητες στήλες, δηλαδή ισοδύναμα αν είναι πλήρους βαθμού ως προς τις στήλες.

Απόδειξη:

Αρχικά αποδεικνύουμε ότι $ii\rightarrow i$. Ισοδύναμα αρκεί να αποδειχθεί ότι $\neg\, i\rightarrow\neg\, ii$. Πράγματι, έστω ότι το *x* δεν είναι κορυφή. Τότε εξ ορισμού έχουμε ότι $\exists y\neq\mathbf{0}:x+y,x-y\in P$. Έστω A_y ο υποπίνακας του *A* που αντιστοιχεί στις μη μηδενικές συνιστώσες του *y*.

Έχουμε ότι $A(x+y)=A(x-y)=b\Rightarrow Ay=\mathbf{0}$. Τότε από την προηγούμενη ισότητα και με δεδομένο ότι $y\neq\mathbf{0}$ έπεται εξ ορισμού ότι ο πίνακας A_y έχει γραμμικά εξαρτημένες στήλες.

Επιπλέον,

$$\left.\begin{aligned}x+y\geq\mathbf{0}\\ x-y\geq\mathbf{0}\end{aligned}\right\}\Rightarrow\left(\left(x_j=0\right)\rightarrow\left(y_j=0\right)\right)$$

Συνεπώς, οι στήλες του πίνακα A_y είναι και στήλες του πίνακα A_x. Άρα, ο A_x έχει γραμμικά εξαρτημένες στήλες.

Για να αποδείξουμε ότι $i \rightarrow ii$ αποδεικνύουμε ισοδύναμα ότι $\neg\, ii \rightarrow \neg\, i$. Έστω λοιπόν ότι ο πίνακας A_x έχει γραμμικά εξαρτημένες στήλες. Τότε εξ ορισμού έχουμε ότι $\exists y$ τ.ω. $A_x y = \mathbf{0},\ y \neq \mathbf{0}$. Μπορούμε να επεκτείνουμε το y στις n διαστάσεις προσθέτοντας μηδενικές συνιστώσες. Τότε έπεται ότι $\exists \tilde{y} \in \mathbb{R}^n$ τ.ω. $A\tilde{y} = \mathbf{0}$, $\tilde{y} \neq \mathbf{0}$ και μάλιστα $\tilde{y}_j = 0$ όπου $x_j = 0$.

Θεωρούμε το $y' = \lambda \cdot \tilde{y}$, $\lambda > 0$. Είναι εύκολο να δει κανείς ότι για κάθε τιμή του λ ισχύει $A\tilde{y} = \mathbf{0} \Rightarrow A\left(x \pm y'\right) = \mathbf{0}$. Επιπλέον, λόγω του ότι $x_j = 0 \Rightarrow \tilde{y}_j = 0$ έπεται ότι υπάρχει αρκετά μικρό $\lambda > 0$ τέτοιο ώστε $x \pm y' \geq \mathbf{0}$. Άρα, λοιπόν $x \pm y' \in P$, δηλαδή το x δεν είναι κορυφή.

□

2.5 Βάσεις

Έστω x μια κορυφή του συνόλου $P = \{x : Ax = b, x \geq \mathbf{0}\}$. Ας υποθέσουμε αρχικά ότι το πλήθος των θετικών συνιστωσών του x ισούται με m, δηλαδή ότι $\left\|\{j : x_j > 0\}\right\| = m$ (όπου ο πίνακας A είναι διαστάσεων $m \times n$). Εν τοιαύτη περιπτώσει συμβολίζουμε με $\mathcal{B}$ το εν λόγω σύνολο δεικτών, δηλαδή $\mathcal{B} = \{j : x_j > 0\}$. Επιπλέον, έστω $A_{\mathcal{B}} = A_x$. Χρησιμοποιούμε αυτό τον συμβολισμό όχι μόνο για τα A και $\mathcal{B}$, αλλά επίσης και για το x και για άλλα σύνολα δεικτών. Τότε ο πίνακας $A_{\mathcal{B}}$ είναι τετραγωνικός διαστάσεων $m \times m$ του οποίου οι στήλες είναι γραμμικά ανεξάρτητες, λόγω του θεωρήματος 4. Συνεπώς, ο πίνακας αυτός είναι αντιστρέψιμος. Από τα παραπάνω προκύπτει ότι μπορούμε να εκφράσουμε το x ως $x_j = 0$ αν $j \notin \mathcal{B}$, και (εφόσον ισχύει ότι $A_{\mathcal{B}} x_{\mathcal{B}} = b$) $x_{\mathcal{B}} = A_{\mathcal{B}}^{-1} b$. Οι μεταβλητές που αντιστοιχούν στο σύνολο δεικτών $\mathcal{B}$ ονομάζονται *βασικές*. Οι υπόλοιπες ονομάζονται *μη βασικές*. Το σύνολο δεικτών που αντιστοιχεί στις μη βασικές μεταβλητές συμβολίζεται $\mathcal{N} = [n] \setminus \mathcal{B}$. Συνεπώς, μπορούμε να γράψουμε εν συντομία τα παραπάνω ως $x_{\mathcal{B}} = A_{\mathcal{B}}^{-1} b$ και $x_{\mathcal{N}} = \mathbf{0}$.

Χωρίς βλάβη της γενικότητας θα θεωρήσουμε ότι ο πίνακας A είναι πλήρους βαθμού ως προς τις γραμμές, δηλαδή ότι $rank\left(A\right) = m$, πράγμα που σημαίνει ότι οι m το πλήθος περιορισμοί ισότητας είναι γραμμικά ανεξάρτητοι. Ειδάλλως, είτε υπάρχει κάποιος περιττός περιορισμός στο σύστημα $Ax = b$ (και μπορούμε να τον απαλείψουμε), είτε το σύστημα είναι αδύνατο.

Αν το πλήθος των θετικών συνιστωσών της κορυφής x είναι μικρότερο από m, $\left\|\{j : x_j > 0\}\right\| < m$, μπορούμε να επεκτείνουμε τον πίνακα A_x με επιπλέον γραμμικά ανεξάρτητες στήλες, μέχρι να προκύψει ένας $m \times m$ υποπίνακας του A πλήρους βαθμού, που τον συμβολίζουμε επίσης $A_{\mathcal{B}}$. Με άλλα λόγια, παρόλο που ενδέχεται να υπάρχουν λιγότερες από m θετικές συνιστώσες στο x, είναι βολικό να έχουμε πάντα μια βάση $\mathcal{B}$, τέτοια ώστε $\|\mathcal{B}\| = m$ και ο πίνακας $A_{\mathcal{B}}$ να είναι αντιστρέψιμος. Η ιδιότητα αυτή μας επιτρέπει να εκφράζουμε πάντα το x όπως και παραπάνω, δηλαδή $x_{\mathcal{B}} = A_{\mathcal{B}}^{-1} b$ και $x_{\mathcal{N}} = \mathbf{0}$.

Περίληψη Το x είναι μια κορυφή του P αν και μόνο αν υπάρχει ένα σύνολο δεικτών $\mathcal{B} \subseteq [n]$ τέτοιο ώστε $\|\mathcal{B}\| = m$ και να ισχύουν τα ακόλουθα:

1. $x_{\mathcal{N}} = \mathbf{0}$ για το σύνολο δεικτών $\mathcal{N} = [n] \setminus \mathcal{B}$.
2. Ο πίνακας $A_{\mathcal{B}}$ είναι αντιστρέψιμος.
3. $x_{\mathcal{B}} = A_{\mathcal{B}}^{-1} b \geq \mathbf{0}$.

Στην περίπτωση αυτή λέμε ότι το $x_{\mathcal{B}}$ είναι μια *βασική εφικτή λύση*. Παρατηρούμε ότι μια κορυφή μπορεί να έχει διάφορες αντίστοιχες βασικές εφικτές λύσεις (που προκύπτουν επεκτείνοντας το σύνολο $\{j : x_j > 0\}$ με διαφορετικούς τρόπους). Μια «τυχαία» βάση ενδεχομένως να μην οδηγεί σε καμία βασική εφικτή λύση εφόσον το διάνυσμα $A_{\mathcal{B}}^{-1} b$ δεν είναι κατ' ανάγκην μη αρνητικό.

Παράδειγμα:

$$\begin{aligned} &x_1 + x_2 + x_3 = 5 \\ &2x_1 - x_2 + 2x_3 = 1 \\ &x_1, x_2, x_3 \geq 0 \end{aligned}$$

Μπορούμε να επιλέξουμε ως βάση το σύνολο $\mathcal{B} = \{1,2\}$. Συνεπώς, $N = \{3\}$ και

$$A_{\mathcal{B}} = \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix}, \; A_{\mathcal{B}}^{-1} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} \\ \frac{2}{3} & -\frac{1}{3} \end{pmatrix}$$

$$A_{\mathcal{B}}^{-1} b = \begin{pmatrix} 2 \\ 3 \end{pmatrix} \text{ και } x = \begin{pmatrix} 2 & 3 & 0 \end{pmatrix}^T$$

Παρατήρηση: Ένα χονδρικό άνω φράγμα στον αριθμό των κορυφών του συνόλου των εφικτών λύσεων P είναι το $\binom{n}{m}$. Ο αριθμός αυτός είναι εκθετικός (είναι άνω φραγμένος από το n^m). Μπορούμε να δείξουμε ένα καλύτερο άνω φράγμα, το $\binom{n - m/2}{m/2}$, το οποίο είναι ωστόσο επίσης εκθετικό. Ο λόγος που ο πραγματικός αριθμός είναι πολύ μικρότερος είναι ότι οι περισσότερες βασικές λύσεις του συστήματος $Ax = b$ (τις οποίες προσμετρήσαμε) δεν είναι εφικτές, δηλαδή δεν ικανοποιούν τον περιορισμό $x \geq \mathbf{0}$.

2.6 Η μέθοδος Simplex

Ο αλγόριθμος Simplex επιλύει προβλήματα γραμμικού προγραμματισμού επικεντρώνοντας στις βασικές εφικτές λύσεις. Η βασική ιδέα είναι να ξεκινήσεις από κάποια κορυφή και να «κοιτάξεις» τις γειτονικές κορυφές. Αν είναι εφικτό να πετύχουμε βελτίωση στο κόστος μετακινούμενοι σε μια εκ των γειτονικών κορυφών, τότε το κάνουμε. Συνεπώς, θα ξεκινήσουμε με μια αναζήτηση σε πλάτος (*bfs*) που αντιστοιχεί σε μία βάση $\mathcal{B}$ και σε κάθε επανάληψη θα προσπαθούμε να βελτιώσουμε το κόστος της λύσης αφαιρώντας μια μεταβλητή από τη βάση και αντικαθιστώντας την με μια άλλη.

Ξεκινούμε τον αλγόριθμο Simplex ξαναγράφοντας πρώτα το γραμμικό πρόγραμμα στην ισοδύναμη μορφή:

$$\min \; c_{\mathcal{B}} x_{\mathcal{B}} + c_{\mathcal{N}} x_{\mathcal{N}}$$

με τους περιορισμούς:

$$A_{\mathcal{B}} x_{\mathcal{B}} + A_{\mathcal{N}} x_{\mathcal{N}} = b$$
$$x_{\mathcal{B}}, x_{\mathcal{N}} \geq \mathbf{0}$$

Εν προκειμένω, με $\mathcal{B}$ συμβολίζουμε τη βάση που αντιστοιχεί στην *bfs* από την οποία ξεκινάμε. Σημειώνουμε ότι για κάθε λύση *x*, ισχύει ότι $x_B = A_B^{-1} b - A_B^{-1} A_N x_N$ και ότι το συνολικό της κόστος, $c^T x$ μπορεί να καθοριστεί ως ακολούθως:

$$\begin{aligned} c^T x &= c_{\mathcal{B}} x_{\mathcal{B}} + c_{\mathcal{N}} x_{\mathcal{N}} \\ &= c_{\mathcal{B}} \left(A_{\mathcal{B}}^{-1} b - A_{\mathcal{B}}^{-1} A_{\mathcal{N}} x_{\mathcal{N}} \right) + c_{\mathcal{N}} x_{\mathcal{N}} \\ &= c_{\mathcal{B}} A_{\mathcal{B}}^{-1} b + \left(c_{\mathcal{N}} - c_{\mathcal{B}} A_{\mathcal{B}}^{-1} A_{\mathcal{N}} \right) x_{\mathcal{N}} \end{aligned}$$

Το συντελεστή του $x_{\mathcal{N}}$ στην παραπάνω εξίσωση τον ονομάζουμε *μειωμένο* κόστος των μη βασικών μεταβλητών και τον συμβολίζουμε με $\tilde{c}_{\mathcal{N}}$, δηλαδή $\tilde{c}_{\mathcal{N}} = c_{\mathcal{N}} - c_{\mathcal{B}} A_{\mathcal{B}}^{-1} A_{\mathcal{N}}$. Αν υπάρχει κάποιο $j \in \mathcal{N}$ τέτοιο ώστε $\tilde{c}_j < 0$, τότε αυξάνοντας την τιμή της μεταβλητής x_j, μειώνεται η τιμή της αντικειμενικής συνάρτησης (κόστος). Βέβαια, το $x_{\mathcal{B}}$ εξαρτάται από το $x_{\mathcal{N}}$, πράγμα που σημαίνει ότι μπορούμε να ελαττώσουμε την τιμή της μεταβλητής x_j μόνο υπό την προϋπόθεση ότι όλες οι συνιστώσες του μεταβληθέντος x_B παραμένουν θετικές.

Συνεπώς, σε ένα βήμα της μεθόδου βρίσκουμε κάποιο $j \in \mathcal{N}$ τέτοιο ώστε $\tilde{c}_j < 0$ και αυξάνουμε το x_j όσο το δυνατόν περισσότερο διατηρώντας τη συνθήκη $x_{\mathcal{B}} \geq \mathbf{0}$. Όταν μία τουλάχιστον συνιστώσα του διανύσματος $x_{\mathcal{B}}$ γίνει μηδέν, δεν είναι πλέον εφικτό να αυξήσουμε την τιμή της μεταβλητής x_j. Το αποτέλεσμα αυτής της διαδικασίας είναι ότι μια μη βασική μεταβλητή είναι τώρα θετική και την περιλαμβάνουμε στη βάση, ενώ μια μεταβλητή που ήταν βασική είναι τώρα ίση με μηδέν, οπότε την αφαιρούμε από τη βάση.

Αν από την άλλη δεν υπάρχει $j \in \mathcal{N}$ τέτοιο ώστε $\tilde{c}_j < 0$, τότε σταματάμε και η τρέχουσα βασική λύση είναι μια βέλτιστη λύση. Αυτό έπεται από τη νέα έκφραση για το $c^T x$ εφόσον το $x_{\mathcal{N}}$ είναι μη αρνητικό.

Παρατηρήσεις:

1. Παρατηρούμε ότι μερικές από τις βασικές μεταβλητές μπορεί να είναι μηδέν αρχικά και στην περίπτωση αυτή είναι πιθανό να μην μπορούμε να αυξήσουμε τα x_j καθόλου. Εν τοιαύτη περιπτώσει μπορούμε να αντικαταστήσουμε έστω το *j* από το *k* στη βάση, αλλά χωρίς να μετακινηθούμε από την κορυφή που αντιστοιχεί σε αυτή τη βάση. Στο επόμενο βήμα ενδέχεται να αντικαταστήσουμε το *k* από το *j* και έτσι να οδηγηθούμε σε ατέρμονα βρόχο. Συνεπώς, απαιτείται ο καθορισμός ενός κανόνα για pivoting, ο οποίος θα ορίζει ποιος δείκτης πρέπει να μπει στη βάση και ποιος πρέπει να αφαιρεθεί από αυτήν.
2. Ενώ πολλοί κανόνες για pivoting (συμπεριλαμβανόμενων και αυτών που χρησιμοποιούνται στην πράξη) μπορούν να οδηγήσουν σε ατέρμονες βρόχους, υπάρχει ένας κανόνας pivoting που εγγυημένα δεν το κάνει (γνωστός ως ο κανόνας του ελάχιστου δείκτη – επέλεξε τα ελάχιστα δυνατά *j* και *k*). Το γεγονός αυτό ανακαλύφθηκε από τον Bland το 1977. Υπάρχουν επίσης άλλες μέθοδοι (οι επονομαζόμενες «breaking ties») που εξαφανίζουν το εν λόγω πρόβλημα.

3. Δεν έχει βρεθεί μέχρι σήμερα κανόνας για pivoting για τον οποίο ο αριθμός των pivots στη χειρότερη περίπτωση είναι καλύτερος από εκθετικός.
4. Το ερώτημα της πολυπλοκότητας του αλγορίθμου Simplex και η τελευταία παρατήρηση οδηγούν στο ερώτημα ποιο είναι το μήκος του συντομότερου μονοπατιού μεταξύ δύο κορυφών ενός κυρτού πολυέδρου, όπου το μονοπάτι ορίζεται κατά μήκος ακμών και το μήκος του μετράται με βάση το πλήθος των κορυφών από τις οποίες διέρχεται.

Μια εποπτική απεικόνιση της γεωμετρίας του αλγορίθμου Simplex μπορεί να αποκτηθεί αν θεωρήσουμε τον αλγόριθμο στον τρισδιάστατο χώρο $\mathbb{R}^3$. Για ένα πρόβλημα στη μορφή $\min\left\{c^T x : Ax \le b\right\}$ το σύνολο εφικτών λύσεων είναι ένα πολύεδρο στο $\mathbb{R}^3$ και ο αλγόριθμος κινείται από κορυφή σε κορυφή σε κάθε βήμα (ή δεν κινείται καθόλου).

2.7 Πότε είναι ένα Γραμμικό Πρόγραμμα Εφικτό ;

Στο σημείο αυτό στρεφόμαστε σε ένα άλλο ερώτημα το οποίο θα μας οδηγήσει σε σημαντικές ιδιότητες του γραμμικού προγραμματισμού. Ας ξεκινήσουμε με μερικά παραδείγματα.

Θεωρούμε γραμμικά προγράμματα της μορφής $Ax = b, x \ge \mathbf{0}$. Με δεδομένο ότι η αντικειμενική συνάρτηση δεν έχει απολύτως καμία επίδραση στο εφικτό του προγράμματος, την αγνοούμε.

Αρχικά περιορίζουμε την προσοχή μας σε συστήματα εξισώσεων, δηλαδή αγνοούμε τους περιορισμούς προσήμου.

Παράδειγμα: Θεωρήστε το σύστημα εξισώσεων:

$$\begin{aligned} x_1 + x_2 + x_3 &= 6 \\ 2x_1 + 3x_2 + x_3 &= 8 \\ 2x_1 + x_2 + 3x_3 &= 0 \end{aligned}$$

και το γραμμικό συνδυασμό:

$$\begin{array}{rcrcrcrcl} -4 & \times & x_1 & + & x_2 & + & x_3 & = & 6 \\ 1 & \times & 2x_1 & + & 3x_2 & + & x_3 & = & 8 \\ 1 & \times & 2x_1 & + & x_2 & + & 3x_3 & = & 0 \end{array}$$

Ο συγκεκριμένος γραμμικός συνδυασμός οδηγεί στην εξίσωση:

$$0x_1 + 0x_2 + 0x_3 = -16$$

πράγμα που φυσικά σημαίνει ότι το σύστημα είναι αδύνατο.
Πράγματι, ένα στοιχειώδες θεώρημα της γραμμικής άλγεβρας λέει ότι αν ένα σύστημα δεν έχει λύση, υπάρχει πάντα ένα διάνυσμα y (εν προκειμένω $y = \begin{pmatrix} -4 & 1 & 1 \end{pmatrix}$) που αποδεικνύει ότι το σύστημα δεν έχει λύση.

Θεώρημα 2.5 *Ακριβώς ένα* από τα ακόλουθα είναι αληθές για το σύστημα $Ax = b$:
1. Υπάρχει x τέτοιο ώστε $Ax = b$.
2. Υπάρχει y τέτοιο ώστε $A^T y = \mathbf{0}$, αλλά $y^T b = 1$.

Ωστόσο, το παραπάνω θεώρημα δεν επαρκεί για τις ανάγκες μας, διότι ένα σύστημα μπορεί να είναι εφικτό αλλά να μην έχει μη αρνητικές λύσεις. Ευτυχώς, το ακόλουθο λήμμα παρέχει ισοδύναμα αποτελέσματα για το σύστημά μας $Ax = b, x \geq \mathbf{0}$.

Θεώρημα 2.6 (Λήμμα του Farka) Ακριβώς ένα από τα ακόλουθα είναι αληθές για το σύστημα $Ax = b, x \geq \mathbf{0}$:

1. Υπάρχει *x* τέτοιο ώστε $Ax = b, x \geq \mathbf{0}$.
2. Υπάρχει *y* τέτοιο ώστε $A^T y \geq \mathbf{0}$, αλλά $b^T y < 0$.

Απόδειξη:

Θα αποδείξουμε αρχικά ότι οι δύο συνθήκες δεν μπορούν να συμβούν ταυτόχρονα και ακολούθως ότι τουλάχιστον μία εξ αυτών πρέπει να είναι αληθής.
Ας υποθέσουμε λοιπόν ότι και οι δύο συνθήκες είναι αληθείς, δηλαδή υπάρχουν *x* και *y* με τις αντίστοιχες ιδιότητες. Έχουμε:

$$Ax = b \overset{y^T \cdot}{\Rightarrow} y^T A x = y^T b \overset{(\,)^T}{\Rightarrow} \left(y^T A x\right)^T = \left(y^T b\right)^T \Rightarrow x^T A^T y = b^T y$$

Ωστόσο, η παραπάνω σχέση αποτελεί αντίφαση, διότι $b^T y < 0$ και εφόσον $x \geq \mathbf{0}$ και $A^T y \geq \mathbf{0}$, θα είναι και $x^T A^T y \geq 0$.

Η άλλη κατεύθυνση είναι λιγότερο προφανής και συνήθως αποδεικνύεται με τη χρήση ιδιοτήτων του αλγορίθμου Simplex, κατά κύριο λόγο της δυικότητας. Θα χρησιμοποιήσουμε ένα άλλο εργαλείο και στα επόμενα θα χρησιμοποιήσουμε το θεώρημα του Farka για να αποδείξουμε ιδιότητες για τη δυικότητα στο γραμμικό προγραμματισμό. Το εργαλείο που θα χρησιμοποιήσουμε είναι το θεώρημα της προβολής, το οποίο διατυπώνουμε ακολούθως χωρίς απόδειξη:

Υπενθυμίζουμε ότι ένα σύνολο $S \subseteq \mathbb{R}^n$ ονομάζεται κυρτό αν $\forall x, y \in S$ το S περιέχει και όλα τα σημεία του ευθύγραμμού τμήματος που συνδέει τα x και y. Ειδάλλως ονομάζεται μη κυρτό.

Θεώρημα 2.7 (Θεώρημα Προβολής) Έστω *K* ένα κλειστό, κυρτό και μη κενό υποσύνολο του $\mathbb{R}^n$. Έστω επιπλέον ένα οποιοδήποτε σημείο $b \in \mathbb{R}^n$. Η προβολή του σημείου *b* επί του *K* είναι ένα σημείο $p \in K$ που ελαχιστοποιεί την ευκλείδεια απόσταση $\|b - p\|_2$. Τότε, το σημείο *p* έχει την ιδιότητα ότι $\forall z \in K\left(\left(z - p\right)^T \left(b - p\right) \leq 0\right)$.

Στο σημείο αυτό είμαστε έτοιμοι να αποδείξουμε την άλλη κατεύθυνση του λήμματος του Farka. Ας υποθέσουμε λοιπόν ότι δεν υπάρχει x τέτοιο ώστε $Ax = b, x \geq \mathbf{0}$. Θα αποδείξουμε ότι υπάρχει *y* τέτοιο ώστε $A^T y \geq \mathbf{0}$, αλλά $b^T y < 0$.

Έστω λοιπόν το σύνολο $K = \left\{Ax : x \geq \mathbf{0}\right\} \subseteq \mathbb{R}^m$, όπου *A* είναι ένας $m \times n$ πίνακας. Το σύνολο αυτό είναι ένας κώνος στο $\mathbb{R}^m$ και ως εκ τούτου είναι κυρτό, μη κενό και κλειστό. Σύμφωνα με την υπόθεση το σύστημα $Ax = b, x \geq \mathbf{0}$ δεν έχει λύση, οπότε το b δεν ανήκει στο K. Έστω p η προβολή του b στο K.

Αφού $p \in K$, υπάρχει $w \geq \mathbf{0}$ τέτοιο ώστε $Aw = p$. Σύμφωνα με το θεώρημα της προβολής, για κάθε $z \in K$ έχουμε $(z-p)^T (b-p) \leq 0$. Αυτό σημαίνει ότι για κάθε $x \geq \mathbf{0}$ ισχύει $(Ax-p)^T (b-p) \leq 0$.

Ορίζοντας $y = p - b$ η παραπάνω εξίσωση γράφεται $(Ax-p)^T y \geq 0$. Αντικαθιστώντας το p με τη βοήθεια του w έχουμε: $(Ax-Aw)^T y \geq 0 \Leftrightarrow (x-w)^T \left(A^T y\right) \geq 0$, για κάθε $x \geq \mathbf{0}$. Παρατηρούμε ότι το w είναι σταθερό για σταθερό b.

Θέτουμε $x = w + e_i$, όπου e_i είναι το μοναδιαίο διάνυσμα με μη μηδενική μόνο τη στήλη i. Είναι φανερό ότι ικανοποιείται ο περιορισμός $x \geq \mathbf{0}$, διότι $w \geq \mathbf{0}$.

Έχουμε λοιπόν ότι $e_i^T \left(A^T y\right) \geq 0$. Όμως φανερά ισχύει ότι $e_i^T \left(A^T y\right) = \left(A^T y\right)_i$. Συνεπώς, για κάθε i ισχύει $\left(A^T y\right)_i \geq 0$ ή ισοδύναμα $A^T y \geq \mathbf{0}$.

Τώρα μένει να αποδειχθεί ότι $b^T y < 0$. Πράγματι, έχουμε ότι:

$$b^T y = (p-y)^T y = p^T y - y^T y$$

Εφόσον για κάθε $x \geq \mathbf{0}$ ισχύει $(Ax-p)^T y \geq 0$, αν πάρουμε για x το μηδενικό διάνυσμα προκύπτει ότι $p^T y \leq 0$. Εφόσον $b \notin K$, $y = p - b \neq \mathbf{0}$, άρα $yy^T > 0$. Τελικά, έπεται το ζητούμενο.

□

Χρησιμοποιώντας μια σχεδόν πανομοιότυπη απόδειξη μπορεί κανείς να δείξει το ίδιο αποτέλεσμα για την κανονική μορφή:

Θεώρημα 2.8 Ακριβώς ένα από τα ακόλουθα είναι αληθές για το σύστημα $Ax \leq b$:

1. Υπάρχει x τέτοιο ώστε $Ax \leq b$.
2. Υπάρχει $y \geq \mathbf{0}$ τέτοιο ώστε $A^T y = \mathbf{0}$, αλλά $y^T b < 0$.

Η διαίσθηση που κρύβεται πίσω από την ακριβή μορφή του δεύτερου σκέλους του προηγούμενου θεωρήματος βρίσκεται στο ότι δεν είναι δυνατόν να ισχύουν ταυτόχρονα και τα δύο. Η αντίφαση $0 = 0x = \left(y^T A\right)x = y^T (Ax) \leq y^T b < 0$ προκύπτει αν δεχθούμε ότι $y \geq \mathbf{0}$, $A^T y = \mathbf{0}$ και $y^T b < 0$.

2.8 Δυϊκότητα

Η δυϊκότητα είναι η πιο σημαντική έννοια στο γραμμικό προγραμματισμό. Η έννοια αυτή μας επιτρέπει να δίνουμε απόδειξη βελτιστότητας. Το γεγονός αυτό είναι σημαντικό όχι μόνο αλγοριθμικά, αλλά επίσης οδηγεί σε πολύ χρήσιμα συνδυαστικά συμπεράσματα. Για παράδειγμα, ας θεωρήσουμε την πρόταση:

> Σε ένα γράφο, ο μικρότερος αριθμός ακμών σε ένα μονοπάτι μεταξύ δύο δοσμένων κόμβων s και t ισούται με το μέγιστο αριθμό $s-t$ τομών (δηλαδή υποσυνόλων των ακμών των οποίων η αφαίρεση αποσυνδέει τις κορυφές s και t).

Αυτό το αποτέλεσμα είναι άμεση συνέπεια της δυϊκότητας για το γραμμικό προγραμματισμό.

Το κίνητρο για τη δυϊκότητα προέρχεται από το πρόβλημα της εύρεσης κάτω φραγμάτων για την τιμή της βέλτιστης λύσης σε ένα πρόβλημα γραμμικού προγραμματισμού (αν πρόκειται για πρόβλημα μεγιστοποίησης, τότε θέλουμε να βρούμε άνω φράγματα.). Για τους σκοπούς τόση αυτής της παραγράφου όσο και της επόμενης, θεωρούμε προβλήματα στην ακόλουθη μορφή:

$$\min\ c^T x$$

με τους περιορισμούς:

$$Ax \geq b$$
$$x \geq \mathbf{0}$$

Η επιλογή της παραπάνω μορφής στηρίζεται στο γεγονός ότι είναι αυτή που χρησιμοποιείται κατά κόρον στο πεδίο των προσεγγιστικών αλγορίθμων.

Ας υποθέσουμε ότι θέλουμε να υπολογίσουμε το καλύτερο δυνατό κάτω φράγμα για τη συνάρτησης κόστους. Πολλαπλασιάζοντας κάθε εξίσωση της μορφής $A_i x \geq b_i$, όπου $i = 1,...,m$ με κάποιο μη αρνητικό αριθμό y_i και αθροίζοντας τις εξισώσεις που προκύπτουν, παίρνουμε ότι $y^T \cdot Ax \geq b^T \cdot y$. Αν επιπλέον επιβάλλουμε τον περιορισμό ότι ο συντελεστής του x_j στην προκύπτουσα ανισότητα είναι άνω φραγμένος από τον c_j, τότε η τιμή $b^T y$ πρέπει να είναι ένα κάτω φράγμα για τη βέλτιστη τιμή (λαμβάνοντας υπόψη ότι τα x_j πρέπει να είναι μη αρνητικά). Για να βρούμε το καλύτερο δυνατό κάτω φράγμα, θέλουμε να λύσουμε το ακόλουθο πρόβλημα:

$$\max\ b^T y$$

με τους περιορισμούς:

$$A^T y \leq c$$
$$y \geq \mathbf{0}$$

Αυτό είναι άλλο ένα γραμμικό πρόγραμμα. Ονομάζουμε αυτό το πρόγραμμα δυϊκό (dual) του αρχικού, το οποίο ονομάζεται πρωτεύον (primal). Όπως είπαμε, η επίλυση του δυϊκού προγράμματος θα μας δώσει ένα κάτω φράγμα για τη βέλτιστη τιμή του αρχικού προβλήματος. Η ιδιότητα αυτή ονομάζεται ασθενής δυϊκότητα: αν συμβολίσουμε τη βέλτιστη λύση του primal με z, δηλαδή $z = \min c^T x$ και τη βέλτιστη λύση του dual με w, τότε $w \leq z$. Θα χρησιμοποιήσουμε το λήμμα του Farka για να αποδείξουμε την ισχυρή δυϊκότητα σύμφωνα με την οποία ισχύει $w = z$. Θα δούμε επίσης ότι το δυϊκό του δυϊκού είναι το αρχικό πρόβλημα.

Παράδειγμα:

$$\min\ 7x_1 + x_2 + 5x_3$$

με τους περιορισμούς:

$$\begin{aligned} x_1 - \ x_2 + 3x_3 &\geq 10 \\ 5x_1 + 2x_2 - \ x_3 &\geq 6 \\ x_1, x_2, x_3 &\geq 0 \end{aligned}$$

Η πρώτη ανισότητα δίνει ένα κάτω φράγμα ίσο με 10 για τη βέλτιστη τιμή z, εφόσον $7x_1 + x_2 + 5x_3 \geq x_1 - \ x_2 + 3x_3 \geq 10$, λόγω των περιορισμών για μη αρνητικά x_i. Μπορούμε να πάρουμε ένα ακόμα καλύτερο κάτω φράγμα παίρνοντας το άθροισμα των πρώτων δύο εξισώσεων. Αυτό με τον ίδιο τρόπο δίνει ότι $7x_1 + x_2 + 5x_3 \geq 6x_1 + x_2 + 2x_3 \geq 16$, το οποίο με τη σειρά του συνεπάγεται ένα κάτω φράγμα ίσο με 16 για τη βέλτιστη λύση z.

Ο μηχανισμός της παραγωγής κάτω φραγμάτων μπορεί να τυποποιηθεί από το δυϊκό γραμμικό πρόγραμμα:

$$\max\ 10y_1 + 6y_2$$

με τους περιορισμούς:

$$\begin{aligned} y_1 + 5\,y_2 &\leq 7 \\ -y_1 + 2y_2 &\leq 1 \\ 3\,y_1 - \ y_2 &\leq 5 \\ y_1, y_2 &\geq 0 \end{aligned}$$

Το y_1 αναπαριστά τον πολλαπλασιαστή του πρώτου περιορισμού και το y_2 τον πολλαπλασιαστή του δεύτερου.

Στο σημείο αυτό ορίζουμε τυπικά την έννοια της δυικότητας. Έστω P και D το ακόλουθο ζεύγος δυϊκών προγραμμάτων:

$$(P)\quad z = \min\left\{c^T x : Ax \geq b, x \geq \mathbf{0}\right\}$$

$$(D)\quad w = \max\left\{b^T y : A^T y \leq c, y \geq \mathbf{0}\right\}$$

Το (P) ονομάζεται πρωτεύον (primal) και το (D) δυϊκό (dual) γραμμικό πρόγραμμα.

Ακολούθως αποδεικνύουμε ότι το δυϊκό του δυϊκού ταυτίζεται με το πρωτεύον. Με άλλα λόγια, αν κάποιος μετατρέψει το (D) σε ένα ισοδύναμο γραμμικό πρόγραμμα σε standard μορφή (δηλαδή στην ίδια μορφή στην οποία βρίσκεται και το (P)), το δυϊκό του $D(D)$ είναι ισοδύναμο με το αρχικό primal (P). Σε κάθε πρόταση μπορούμε να εναλλάξουμε τους ρόλους πρωτεύοντος και δυϊκού.

Απόδειξη:

Το δυϊκό πρόγραμμα είναι ισοδύναμο με το $\min\left\{-b^T y : -A^T y \geq -c, y \geq \mathbf{0}\right\}$. Παίρνοντας το δυϊκό του δυϊκού έχουμε $\max\left\{-c^T x : -\left(A^T\right)^T x \leq -b, x \geq \mathbf{0}\right\}$. Αλλά αυτό είναι ισοδύναμο με το $\min\left\{c^T x : Ax \geq b, x \geq \mathbf{0}\right\}$, δηλαδή με το ζητούμενο.

□

Λήμμα 2.9 (Ασθενής Δυϊκότητα) $z \geq w$.
Απόδειξη:

Έστω x μια εφικτή λύση του primal και y μια εφικτή λύση του dual. Τότε, έχουμε: $c \geq A^T y \overset{(\cdot)^T}{\Rightarrow} c^T \geq \left(A^T y\right)^T = y^T A \Rightarrow c^T x \geq y^T Ax \geq y^T b$, διότι $Ax \geq b$ και $y \geq \mathbf{0}$.

□

Από το προηγούμενο λήμμα συμπεραίνουμε ότι οι ακόλουθες περιπτώσεις *δεν* είναι εφικτές (οι επόμενες προτάσεις είναι δυϊκές):

1. Το *P* είναι εφικτό και μη φραγμένο (κάτω) και το *D* είναι εφικτό.
2. Το *P* είναι εφικτό και το *D* είναι εφικτό και μη φραγμένο (άνω).

Ωστόσο, πρέπει να παρατηρήσει κανείς ότι ενδέχεται τόσο το primal όσο και το dual να είναι μη εφικτά.

Για να αποδείξουμε μια ισχυρότερη εκδοχή του λήμματος ασθενούς δυικότητας θα χρησιμοποιήσουμε το λήμμα του Farka το οποίο διατυπώνουμε εκ νέου ακολούθως χάριν πληρότητας:

Πόρισμα 2.10: Ακριβώς ένα από τα παρακάτω είναι αληθές:

1. $\exists x' : A'x' \leq b'$.
2. $\exists y' \geq \mathbf{0} : \left(A'\right)^T y' = \mathbf{0}$ και $\left(b'\right)^T y' < 0$

Θεώρημα 2.11 (Ισχυρή Δυϊκότητα) Αν ένα τουλάχιστον εκ των *P*, *D* είναι εφικτό, τότε $z = w$.
Απόδειξη:

Αρκεί να δείξουμε ότι $z \leq w$. Λόγω του ότι $D\left(D\right) \equiv P$, μπορούμε χωρίς βλάβη της γενικότητας να υποθέσουμε ότι το *P* είναι εφικτό. Αν το *P* είναι μη φραγμένο κάτω, τότε, λόγω της ασθενούς δυϊκότητας, έχουμε ότι $z = w = -\infty$.
Υποθέτουμε λοιπόν ότι το *P* είναι κάτω φραγμένο και έστω x^* μια βέλτιστη λύση, δηλαδή $Ax^* \geq b$, $x^* \geq \mathbf{0}$ και $c^T x^* = z$. Ισχυριζόμαστε ότι $\exists y \geq \mathbf{0} : A^T y \leq c$ και $b^T y \geq z$. Αν κάτι τέτοιο αποδειχθεί, τότε η απόδειξη θα είναι πλήρης.
Ας υποθέσουμε λοιπόν ότι τέτοιο y δεν υπάρχει. Τότε, εφαρμόζοντας το προηγούμενο λήμμα, με $A' = \begin{pmatrix} A^T \\ -b^T \end{pmatrix}$, $b' = \begin{pmatrix} c \\ -z \end{pmatrix}$, $x' = y$, $y' = \begin{pmatrix} x \\ \lambda \end{pmatrix}$ (και λαμβάνοντας υπόψη ότι η άρνηση της συνθήκης 1 είναι ισοδύναμη με την υπόθεση) έπεται ότι $\exists x \geq \mathbf{0}, \lambda \geq 0$ τέτοια ώστε

$$\left\{ \begin{matrix} Ax = \lambda b \\ \text{και} \\ c^T x < \lambda z \end{matrix} \right\}$$

Διακρίνουμε δύο περιπτώσεις:

- $1^{η}$ Περίπτωση: $\lambda \neq 0$. Εφόσον μπορούμε να κανονικοποιήσουμε ως προς λ, μπορούμε να υποθέσουμε ότι $\lambda = 1$. Αυτό σημαίνει ότι $\exists x \geq \mathbf{0}$ τέτοιο ώστε $Ax = b$ (δηλαδή x εφικτή λύση του P) και επιπλέον $c^T x < z$. Αλλά αυτό αντιτίθεται στην υπόθεση βελτιστότητας του x^*.
- $2^{η}$ Περίπτωση: $\lambda = 0$. Αυτό σημαίνει ότι $\exists x \geq \mathbf{0}$ τέτοιο ώστε $Ax = \mathbf{0}$ και $c^T x < 0$. Αν αυτό ισχύει, τότε $\forall \mu > 0$ το $x^* + \mu x$ είναι εφικτή λύση για το P και το κόστος της είναι $c^T\left(x^* + \mu x\right) = c^T x^* + \mu\left(c^T x\right) < z$ που επίσης αποτελεί αντίφαση.

□

2.9 Complementary Slackness

Έστω *P* και *D* τα ακόλουθα:

$$(P) \quad z = \min\left\{c^T x : Ax \geq b, x \geq \mathbf{0}\right\}$$
$$(D) \quad w = \max\left\{b^T y : A^T y \leq c, y \geq \mathbf{0}\right\}$$

και έστω *x* μια εφικτή λύση του *P* και *y* μια εφικτή λύση του *D*. Τότε, λόγω της ασθενούς δυϊκότητας, γνωρίζουμε ότι $c^T x \geq b^T y$. Η διαφορά $c^T x - b^T y$ ονομάζεται κενό δυϊκότητας (duality gap). Λόγω της ισχυρής δυικότητας έχουμε ότι το κενό δυικότητας είναι ίσο με μηδέν αν και μόνο αν το *x* είναι βέλτιστη λύση για το *P* και το *y* είναι βέλτιστη λύση για το *D*. Με άλλα λόγια, το κενό δυικότητας μπορεί να χρησιμοποιηθεί ως ένα καλό μέτρο του πόσο κοντά βρίσκονται οι εφικτές λύσεις *x* και *y* στις βέλτιστες λύσεις των *P* και *D* αντίστοιχα. Επιπλέον, το κενό δυικότητας θα χρησιμοποιηθεί στην περιγραφή της μεθόδου εσωτερικού σημείου για την παρακολούθηση της προόδου προς τη βελτιστότητα.

Είναι βολικό να γράψουμε τόσο το πρωτεύον όσο και δυϊκό γραμμικό πρόγραμμα ως εξής:

$$(P) \quad z = \min\left\{c^T x : Ax - s = b,\ x \geq \mathbf{0}_{n\times 1},\ s \geq \mathbf{0}_{m\times 1}\right\}$$
$$(D) \quad w = \max\left\{b^T y : A^T y + s' = c,\ y \geq \mathbf{0}_{m\times 1},\ s' \geq \mathbf{0}_{n\times 1}\right\}$$

Με βάση τον παραπάνω τρόπο γραφής, το κενό δυικότητας γράφεται ως εξής:

$$\begin{aligned} c^T x - b^T y &= c^T x - \left(Ax - s\right)^T y \\ &= c^T x - x^T A^T y + s^T y \\ &= x^T\left(c - A^T y\right) + s^T y \\ &= x^T s' + s^T y \end{aligned}$$

Το θεώρημα που ακολουθεί επιτρέπει να ελέγχει κανείς τη βελτιστότητα μιας λύσης είτε του primal είτε του dual προγράμματος.

Θεώρημα 2.12 (Complementary Slackness)
Έστω x^* και $\left(y^*, s^*\right)$ εφικτές λύσεις για τα (P) και (D) αντίστοιχα. Τα ακόλουθα είναι ισοδύναμα:

1. Το x^* είναι βέλτιστη λύση για το (P) και το $\left(y^*, s^*\right)$ είναι βέλτιστη λύση για το (D).
2. $\left(s'^*\right)^T x^* = 0$ και $\left(s^*\right)^T y^* = 0$.
3. $\forall j \in [n]\ \left(x_j^* s'^*_j = 0\right)$ και $\forall i \in [m]\ \left(y_i^* s_i^* = 0\right)$.
4. $\forall j \in [n]\ \left(\left(x_j^* = 0\right) \vee \left(s'^*_j = 0\right)\right)$ και $\forall i \in [m]\ \left(\left(y_j^* = 0\right) \vee \left(s_j^* = 0\right)\right)$.

Απόδειξη:

$(1) \rightarrow (2)$: Έστω ότι το (1) ισχύει, τότε λόγω της ισχυρής δυικότητας $c^T x^* = b^T y^*$. Δηλαδή το κενό δυικότητας είναι ίσο με το μηδέν, $\left(x^*\right)^T s'^* + \left(s^*\right)^T y^* = 0$ Εφόσον έχουμε ότι $s^*, x^* \geq \mathbf{0} \Rightarrow \left(s'^*\right)^T \cdot x^* \geq 0$ και ομοίως $\left(s^*\right)^T \cdot y^* \geq 0$. Συνεπώς, αναγκαστικά ισχύει ότι $\left(x^*\right)^T s'^* = \left(s^*\right)^T y^* = 0$.

$(2) \rightarrow (1)$: Αν ισχύει το (2), τότε έπεται ότι $c^T x^* = b^T y^*$ και συνεπώς λόγω της ασθενούς δυϊκότητας έπεται ότι x^* και y^* βέλτιστες λύσεις.

$(2) \leftrightarrow (3)$: Έστω ότι $\left(s'^*\right)^T x^* = 0$. Εφόσον ισχύει ότι $\forall j\left(x_j^* \geq 0 \wedge s'^*_j \geq 0\right)$ έπεται το ζητούμενο. Ομοίως για το δεύτερο σκέλος.

$(3) \leftrightarrow (4)$: Τετριμμένο.

Οι συνθήκες του complementary slackness παίζουν ζωτικό ρόλο στη σχεδίαση αποδοτικών αλγορίθμων, τόσο ακριβών όσο και προσεγγιστικών, όπως θα φανεί πολλές φορές σε επόμενα κεφάλαια.

2.10 Μέγεθος ενός Γραμμικού Προγράμματος

2.10.1 Μέγεθος της Εισόδου

Αν θέλουμε να επιλύσουμε ένα γραμμικό πρόγραμμα σε πολυωνυμικό χρόνο, απαιτείται να γνωρίζουμε τι εννοούμε, δηλαδή ποιο είναι το μέγεθος της εισόδου. Η αποσαφήνιση του παραπάνω είναι πολύ σημαντική. Πράγματι, θα μπορούσε κανείς *λανθασμένα* να υποθέσει ότι ένας αλγόριθμος που επιλύει γραμμικά προγράμματα ονομάζεται πολυωνυμικός αν έχει χρονική πολυπλοκότητα πολυωνυμική ως προς τα *m* και *n*, δηλαδή ως προς το πλήθος των περιορισμών και το πλήθος των μεταβλητών απόφασης.

Για το σκοπό αυτό εισάγουμε δύο έννοιες του μεγέθους της εισόδου ως προς το οποίο οι αλγόριθμοι που θα αναφέρουμε συνοπτικά στη συνέχεια τρέχουν σε πολυωνυμικό χρόνο. Το πρώτο μέτρο του μεγέθους της εισόδου θα είναι το *μέγεθος ενός γραμμικού προγράμματος LP*, αλλά θα εισάγουμε επίσης και ένα νέο μέτρο L το οποίο είναι πολυωνυμικά συσχετισμένο με το προηγούμενο και στην πράξη αποδεικνύεται πιο εύκολο στη χρήση. Επιπλέον, ισχύει $L \le \text{size}(LP)$, επομένως κάθε αλγόριθμος που τρέχει σε χρόνο πολυωνυμικό ως προς το L, θα τρέχει επίσης σε χρόνο πολυωνυμικό ως προς το $\text{size}(LP)$.

Ας θεωρήσουμε το γραμμικό πρόγραμμα της μορφής:

$$\min c^T x$$

με τους περιορισμούς :

$$Ax \ge b$$
$$x \ge \mathbf{0}$$

όπου μας δίνονται ως είσοδος οι *ρητοί* συντελεστές του A (πίνακας διαστάσεων $m \times n$), του b (ένα διάνυσμα $m \times 1$) και του c (ένα διάνυσμα $n \times 1$).

Μπορούμε επιπλέον να υποθέσουμε, χωρίς βλάβη της γενικότητας, ότι οι δοθέντες συντελεστές είναι όλοι ακέραιοι, εφόσον κάθε γραμμικό πρόγραμμα με ρητούς συντελεστές μπορεί εύκολα να μετασχηματιστεί σε ένα ισοδύναμο πρόγραμμα με ακέραιους συντελεστές (αν πολλαπλασιάσουμε κάθε ισότητα – ανισότητα με το ελάχιστο κοινό πολλαπλάσιο των παρονομαστών). Συνεπώς, στη συνέχεια υποθέτουμε ότι τα A, b και c έχουν ακέραιους συντελεστές.

Για κάθε ακέραιο n ορίζουμε το μέγεθός του ως ακολούθως:

$$\text{size}(n) \overset{\Delta}{=} 1 + \lceil \log_2(|n|+1) \rceil$$

όπου η μονάδα αντιπροσωπεύει την ανάγκη κωδικοποίησης του προσήμου του αριθμού n και το υπόλοιπό μέρος είναι το μήκος της δυαδικής αναπαράστασης της απόλυτης τιμής του. Με ανάλογο τρόπο ορίζουμε το μέγεθος ενός $p \times 1$ διανύσματος d και ενός $p \times l$ πίνακα M:

$$\text{size}(d) \overset{\Delta}{=} \sum_{i=1}^{p} \text{size}(d_i)$$
$$\text{size}(M) \overset{\Delta}{=} \sum_{i=1}^{p} \sum_{j=1}^{l} \text{size}(m_{ij})$$

Στο σημείο αυτό είμαστε έτοιμοι να ορίσουμε το μέγεθος ενός γραμμικού προγράμματος LP.

Ορισμός 2.6 (Μέγεθος ενός γραμμικού προγράμματος)

$$\text{size}(LP) \overset{\Delta}{=} \text{size}(A) + \text{size}(b) + \text{size}(c)$$

Ένα πιο βολικό μέτρο του μεγέθους ορίζεται ακολούθως

Ορισμός 2.7

$$L \overset{\Delta}{=} \operatorname{size}(det_{max}) + \operatorname{size}(b_{max}) + \operatorname{size}(c_{max}) + m + n$$

όπου

$$det_{max} \overset{\Delta}{=} \max_{A'}\left(\left|\det(A')\right|\right)$$

$$b_{max} \overset{\Delta}{=} \max_{i}\left(\left|b_i\right|\right)$$

$$c_{max} \overset{\Delta}{=} \max_{j}\left(\left|c_j\right|\right)$$

και A' είναι οποιοσδήποτε τετραγωνικός υποπίνακας του A.

Πρόταση 2.13 $\forall A,b,c\left(L(A,b,c) < \operatorname{size}(LP(A,b,c))\right)$.

Για την απόδειξη χρειαζόμαστε το ακόλουθο λήμμα:

Λήμμα 2.14

1. Αν $n \in \mathbb{Z}$, τότε $|n| \le 2^{\operatorname{size}(n)-1} - 1$.
2. Αν $v \in \mathbb{Z}^n$, τότε $\|v\|_2 \le \|v\|_1 \le 2^{\operatorname{size}(v)-n} - 1$.
3. Αν $A \in \mathbb{Z}^{n\times n}$, τότε $\left|\det(A)\right| \le 2^{\operatorname{size}(A)-n^2} - 1$.

Απόδειξη:

1. Εξ ορισμού.
2. Ως γνωστόν, στον ευκλείδειο χώρο η νόρμα - 2 είναι άνω φραγμένη από τη νόρμα – 1, δηλαδή $\|v\|_2 \le \|v\|_1$. Προσθέτοντας 1 και στα δύο μέλη της εν λόγω ανισότητας παίρνουμε:
$$1+\|v\|_2 \le 1+\|v\|_1 = 1+\sum_{i=1}^{n}|v_i| \le \prod_{i=1}^{n}\left(1+|v_i|\right) \le \prod_{i=1}^{n}2^{size(v_i)-1} = 2^{size(v)-n}.$$
3. Έστω $a_1, a_2, ..., a_n$ οι στήλες του A. Εφόσον η $\left|\det(A)\right|$ αναπαριστά τον όγκο του παραλληλεπιπέδου που ορίζεται από τα $a_1, a_2, ..., a_n$ έπεται ότι
$$\left|\det(A)\right| \le \prod_{i=1}^{n}\|a_i\|_2 \text{ (ανισότητα του Hadamard)}$$
Κατά συνέπεια, λόγω του 2
$$1+\left|\det(A)\right| \le 1+\prod_{i=1}^{n}\|a_i\|_2 \le \prod_{i=1}^{n}\left(1+\|a_i\|_2\right) \le \prod_{i=1}^{n}2^{\operatorname{size}(a_i)-n} = 2^{\operatorname{size}(A)-n^2}$$

□

Αποδεικνύουμε τώρα την πρόταση 2.13.
Απόδειξη:

Αν ο B είναι τετραγωνικός υποπίνακας του A τότε, εξ ορισμού, $\text{size}(B) \le \text{size}(A)$. Επιπλέον, από το λήμμα 14.1 $1+|\det(B)| \le 2^{\text{size}(B)-1}$. Άρα,

$$\left\lceil \log\left(1+|\det(B)|\right)\right\rceil \le \text{size}(B)-1 < \text{size}(B) \le \text{size}(A)$$

Έστω $v \in \mathbb{Z}^p$. Τότε $\text{size}(v) \ge \text{size}\left(\max_j |v_j|\right)+p-1 = \left\lceil \log_2\left(1+\max_j |v_j|\right)\right\rceil + p$.

Οπότε

$$\text{size}(b)+\text{size}(c) \ge \left\lceil \log_2\left(1+\max_j |b_j|\right)\right\rceil + \left\lceil \log_2\left(1+\max_i |c_i|\right)\right\rceil + m + n$$

Συνδυάζοντας τις προηγούμενες σχέσεις έπεται το επιθυμητό αποτέλεσμα.

□

Παρατήρηση: Ισχύει $det_{max} \times b_{max} \times c_{max} \cdot 2^{m+n} < 2^L$ εφόσον για κάθε ακέραιο n ισχύει ότι $2^{\text{size}(n)} > |n|$.

2.10.2 Μέγεθος της Εξόδου

Είναι γνωστό ότι υπάρχουν προβλήματα βελτιστοποίησης στα οποία το μήκος της αναπαράστασης μιας εφικτής λύσης δεν είναι πολυωνυμικά άνω φραγμένο ως προς το μήκος της εισόδου. Για τέτοια προβλήματα είναι φανερό ότι δεν υπάρχει πολυωνυμικός αλγόριθμος που τα επιλύει, διότι η απλή καταγραφή της λύσης δεν μπορεί να επιτευχθεί σε πολυωνυμικό χρόνο. Ως εκ τούτου, εν προκειμένω είναι απαραίτητο να εξασφαλίσει κανείς ότι η λύση μπορεί να αναπαρασταθεί σε μήκος πολυωνυμικό ως προς το L. Έχουμε ήδη αποδείξει ότι αν το *LP* είναι εφικτό, υπάρχει τουλάχιστον μία κορυφή η οποία αποτελεί μία βέλτιστη λύση. Συνεπώς, εφόσον αναζητούμε βέλτιστη λύση στο γραμμικό πρόγραμμα, είναι λογικό να περιορίσουμε την προσοχή μας μόνο σε κορυφές. Το ακόλουθο θεώρημα εξασφαλίζει ότι οι κορυφές είναι αναπαραστάσιμες με συμπαγή τρόπο.

Θεώρημα 2.15 Έστω x μια κορυφή του πολυέδρου που ορίζεται από τη σχέση $Ax = b$, $x \ge \mathbf{0}$. Τότε ισχύει:

$$x^T = \begin{pmatrix} \frac{p_1}{q} & \frac{p_2}{q} & . & . & \frac{p_n}{q} \end{pmatrix}$$

όπου $p_i, q \in \mathbb{N}$
και

$$0 \le p_i < 2^L$$
$$1 \le q < 2^L$$

Απόδειξη:

Εφόσον το x είναι μια βασική εφικτή λύση, έπεται ότι υπάρχει βάση B τέτοια ώστε $x_B = A_B^{-1}b$ και $x_N = \mathbf{0}$. Συνεπώς, μπορούμε να θέσουμε $p_j = 0$ για κάθε $j \in N$ και να επικεντρώσουμε την προσοχή μας στα x_j που είναι τέτοια ώστε $j \in B$.

Από τη γραμμική άλγεβρα είναι γνωστό ότι

$$x_B = A_B^{-1} b = \frac{1}{\det(A_B)} adj(A_B) b$$

όπου $adj(A_B)$ είναι ο adjoint πίνακας του A_B. Μπορούμε λοιπόν να θέσουμε $q = |\det(A_B)|$. Το q είναι φυσικός εφόσον τα στοιχεία του A_B είναι ακέραιοι. Επιπλέον, με δεδομένο ότι ο εν λόγω πίνακας είναι αντιστρέψιμος, έπεται ότι $q \geq 1$ και επιπλέον $q \leq det_{max} < 2^L$. Τελικά παρατηρούμε ότι $p_B = qx_B = |adj(A_B)b|$, οπότε $p_i \leq \sum_{j=1}^{m} \left|adj(A_B)_{i,j}\right| \cdot |b_j| \leq m \cdot det_{max} \cdot b_{max} < 2^L$

□

2.11 Πολυπλοκότητα του Γραμμικού Προγραμματισμού

Στην παράγραφο αυτή αποδεικνύουμε ότι το πρόβλημα του γραμμικού προγραμματισμού ανήκει στο $\Delta = NP \cap co-NP \supseteq P$. Ένα πρόβλημα απόφασης που ανήκει στο $NP \cap co-NP$ ονομάζεται καλά χαρακτηρισμένο (well characterized). Το αποτέλεσμα αυτό θα αποδειχθεί με βάση τη δυικότητα και τις εκτιμήσεις για το μέγεθος οποιασδήποτε κορυφής που δόθηκαν στην προηγούμενη παράγραφο.

Το πρόβλημα του γραμμικού προγραμματισμού έχει μέχρι στιγμής οριστεί ως πρόβλημα βελτιστοποίησης. Στη συνέχεια διατυπώνουμε το αντίστοιχο πρόβλημα απόφασης:

Ορισμός 2.8 ($\mathcal{LP}$)

Είσοδος: A, b, c με ακέραια στοιχεία και ένας ρητός αριθμός λ.

Ερώτηση: Ισχύει $\min\{c^T x : Ax = b, x \geq \mathbf{0}\} \leq \lambda$;

Θεώρημα 2.16 $\mathcal{LP} \in NP \cap co-NP$

Απόδειξη:

(i) $\mathcal{LP} \in NP$: Χρησιμοποιούμε έναν εναλλακτικό ορισμό σύμφωνα με τον οποίο το NP είναι το σύνολο των προβλημάτων απόφασης που έχουν πολυωνυμικού μήκους *Yes certificates* τα οποία μπορούν να επαληθεύσουν τα *yes instances* σε πολυωνυμικό χρόνο.

Αν το γραμμικό πρόγραμμα είναι εφικτό και φραγμένο, το certificate προς επιβεβαίωση των yes instances (δηλαδή των στιγμιότυπων του προβλήματος για τα οποία ισχύει $\min\{c^T x : Ax = b, x \geq \mathbf{0}\} \leq \lambda$) είναι μια κορυφή x' του $\{Ax = b, x \geq \mathbf{0}\}$ τέτοια ώστε να ισχύει $c^T x' \leq \lambda$. Εφόσον υποθέσαμε ότι το ελάχιστον είναι πεπερασμένο, σύμφωνα με το Πόρισμα 3 υπάρχει πάντα κορυφή. Δοσμένης του x', είναι εύκολο να ελέγξουμε σε πολυωνυμικό χρόνο αν ισχύει $Ax' = b$ και $x' \geq \mathbf{0}$ και φυσικά αν $c^T x' \leq \lambda$. Χρειάζεται επίσης να δείξουμε ότι το μέγεθος ενός τέτοιου certificate είναι πολυωνυμικά φραγμένο ως προς το μήκος της εισόδου. Αυτό έχει ήδη αποδειχθεί στην προηγούμενη παράγραφο.

Αν το γραμμικό πρόγραμμα είναι εφικτό και μη φραγμένο, τότε, λόγω της ισχυρής δυικότητας, το δυϊκό είναι μη εφικτό. Χρησιμοποιώντας το λήμμα του Farka *για το δυϊκό*, έπεται ότι υπάρχει $\tilde{x}$: $A\tilde{x} = \mathbf{0}$, $\tilde{x} \geq \mathbf{0}$ και $c^T\tilde{x} = -1 < 0$. Το certificate σε αυτήν την περίπτωση περιέχει μια κορυφή του προγράμματος $\{Ax = b, x \geq \mathbf{0}\}$ (για να δειχθεί το εφικτό) και μια κορυφή του $\{Ax = b, x \geq \mathbf{0}, c^T x = -1\}$ (για να δειχθεί το μη φραγμένο σε περίπτωση που το πρόγραμμα είναι εφικτό). Επιλέγοντας μια κορυφή x' του δεύτερου προγράμματος, εξασφαλίζουμε το πολυωνυμικό μήκος.

(ii) $\mathcal{LP} \in co-NP \Leftrightarrow \overline{\mathcal{LP}} \in NP$: Το συμπλήρωμα $\overline{\mathcal{LP}}$ ορίζεται ως ακολούθως:

<u>*Είσοδος:*</u> A, b, c με ακέραια στοιχεία και ένας ρητός αριθμός λ.

<u>*Ερώτηση:*</u> Ισχύει $\min\{c^T x : Ax = b, x \geq \mathbf{0}\} > \lambda$;

Αν το σύνολο $\{Ax = b, x \geq \mathbf{0}\}$ είναι μη κενό, λόγω της ισχυρής δυϊκότητας, το $\overline{\mathcal{LP}}$ είναι ισοδύναμο με το ακόλουθο πρόβλημα απόφασης:

<u>*Είσοδος:*</u> A, b, c με ακέραια στοιχεία και ένας ρητός αριθμός λ.

<u>*Ερώτηση:*</u> Ισχύει $\max\{b^T y : A^T y \leq c\} > \lambda$;

Το παραπάνω πρόβλημα απόφασης προφανώς ανήκει στο *NP* (για τους ίδιους λόγους που ανήκει και το $\mathcal{LP}$).

Αν το primal είναι μη εφικτό, λόγω του λήμματος του Farka γνωρίζουμε ότι υπάρχει y: $A^T y \geq \mathbf{0}$ και $b^T y = -1 < 0$.

□

2.12 Επίλυση Γραμμικού Προγράμματος σε Πολυωνυμικό Χρόνο

Ο πρώτος πολυωνυμικός αλγόριθμος για το γραμμικό προγραμματισμό είναι ο επονομαζόμενος *ελλειψοειδής αλγόριθμος* που προτάθηκε από τον Khachian το 1979. Ο εν λόγω αλγόριθμος αναπτύχθηκε αρχικά για τον κυρτό προγραμματισμό (convex programming) του οποίου ο γραμμικός προγραμματισμός είναι ειδική περίπτωση. Παρόλο που έχει πολυωνυμική χρονική πολυπλοκότητα, ο αλγόριθμος αυτός δεν είναι πρακτικός για το γραμμικό προγραμματισμό. Ωστόσο, έχει εκτεταμένες θεωρητικές εφαρμογές στη συνδυαστική βελτιστοποίηση.

Το 1984 ο Karmarkar παρουσίασε έναν ακόμα πολυωνυμικό αλγόριθμο για το γραμμικό προγραμματισμό. Ο αλγόριθμός του αποφεύγει την εκθετική πολυπλοκότητα (που είναι εγγενής στη μέθοδο simplex) που συνδέεται με τις κορυφές, τις ακμές και τις επιφάνειες του πολυέδρου μένοντας στο εσωτερικό αυτού. Αξίζει να τονισθεί ότι ο αλγόριθμος αυτός οδήγησε σε πολλούς άλλους αλγορίθμους για το ίδιο πρόβλημα βασισμένους σε παρόμοιες ιδέες. Για ευνόητους λόγους, οι αλγόριθμοι αυτοί είναι γνωστοί ως *μέθοδοι εσωτερικού σημείου*.

Στη συνέχεια δίνουμε μια high –level περιγραφή μιας μεθόδου εσωτερικού σημείου:

1. Αν η τρέχουσα λύση x είναι κοντά στο σύνορο του πολυέδρου, απεικόνισε το πολύεδρο σε ένα άλλο τέτοιο ώστε η εικόνα του x να είναι σε «πιο εσωτερικό» σημείο του νέου πολυέδρου.
2. Κάνε ένα βήμα στο μετασχηματισμένο χώρο.
3. Επανάλαβε τα βήματα 1 και 2 μέχρι να είμαστε *αρκετά κοντά* σε μία βέλτιστη λύση.

Θα μπορούσε κανείς να αναρωτηθεί τι σημαίνει η φράση «αρκετά κοντά» της προηγούμενης πρότασης. Το επόμενο θεώρημα, που διατυπώνεται χωρίς απόδειξη (αν και η απόδειξή του είναι εύκολη και στηρίζεται σε πρώτες έννοιες) δίνει την απάντηση.

Θεώρημα 2.17 Έστω x_1, x_2 κορυφές του $Ax = b$, $x \geq \mathbf{0}$. Αν ισχύει ότι $c^T x_1 \neq c^T x_2$, τότε θα έχουμε ότι $\left| c^T x_1 - c^T x_2 \right| > 2^{-2L}$.

Πόρισμα 2.18 Έστω $z = \min \left\{ c^T x : \underbrace{Ax = b, x \geq \mathbf{0}}_{polyhedron\ P} \right\}$. Έστω επίσης ότι το x είναι εφικτή λύση του P και ότι $c^T x \leq z + 2^{-2L}$. Τότε, κάθε κορυφή x' τέτοια ώστε $c^T x' \leq c^T x$ είναι βέλτιστη λύση για το γραμμικό πρόγραμμα.

Αυτό που μας λέει το παραπάνω πόρισμα είναι ότι δε χρειάζεται να είμαστε πολύ ακριβείς όταν επιλέγουμε μια βέλτιστη κορυφή. Ακριβέστερα χρειάζεται απλά να υπολογίσουμε την αντικειμενική συνάρτηση με σφάλμα λιγότερο από 2^{-2L}. Αν βρούμε μια κορυφή που βρίσκεται εντός αυτού του περιθωρίου σφάλματος, τότε θα είναι και βέλτιστη.

Κεφάλαιο 3

3.1 Η έννοια της χαλάρωσης (relaxation) ενός προβλήματος βελτιστοποίησης

Μία πολύ σημαντική έννοια άρρηκτα συνδεδεμένη με τη σχεδίαση προσεγγιστικών αλγορίθμων είναι η έννοια της χαλάρωσης (relaxation) ενός $NP-hard$ προβλήματος βελτιστοποίησης $\mathcal{P}$, η οποία ορίζεται αμέσως παρακάτω. Χωρίς βλάβη της γενικότητας υποθέτουμε ότι το $\mathcal{P}$ είναι πρόβλημα μεγιστοποίησης. Η επέκταση για προβλήματα ελαχιστοποίησης είναι άμεση.

Έστω λοιπόν $S(I) \neq \varnothing$ το σύνολο των εφικτών λύσεων του προβλήματος $\mathcal{P}$ για το (δοθέν) στιγμιότυπο I και $f(\cdot)$ η αντίστοιχη αντικειμενική συνάρτηση. Στόχος είναι η μεγιστοποίηση της εν λόγω συνάρτησης, δηλαδή η εύρεση μιας λύσης $x^* \in S(I)$ τέτοιας ώστε

$$OPT(I) \doteq f(x^*) = \max_{x \in S(I)} f(x)$$

Με δεδομένο ότι δεν είναι γνωστός κανένας αποδοτικός αλγόριθμος για την εύρεση της βέλτιστης λύσης του αρχικού προβλήματος, θεωρούμε το ακόλουθο σχετικό πρόβλημα βελτιστοποίησης: Δοθέντος του στιγμιότυπου του προβλήματος, για ένα *κατάλληλα επιλεγμένο* σύνολο «χαλαρωμένων λύσεων» $R(I)$, αναζητούμε ένα $\tilde{x} \in R(I)$ τέτοιο ώστε να ισχύει

$$UB = g(\tilde{x}) = \max_{x \in R(I)} g(x)$$

Με $g(\cdot)$ συμβολίζουμε την αντικειμενική συνάρτηση για το νέο πρόβλημα.

Αν ισχύουν οι ακόλουθες συνθήκες:

$$R(I) \supseteq S(I)$$

και

$$\forall I \forall x \in S(I) \big(g(x) \geq f(x) \big)$$

το νέο πρόβλημα ονομάζεται χαλάρωση (relaxation) του αρχικού. Τονίζεται ότι στις περισσότερες περιπτώσεις οι συναρτήσεις f και g ταυτίζονται.

Είναι εύκολο να δει κανείς ότι για κάθε στιγμιότυπο του $\mathcal{P}$ η τιμή της βέλτιστης λύσης του χαλαρωμένου προβλήματος είναι ένα *άνω φράγμα* για την αντίστοιχη τιμή του $\mathcal{P}$. Πράγματι, οι παραπάνω συνθήκες έπονται ότι:

$$UB = \max_{x \in R(I)} g(x) \geq \max_{x \in S(I)} g(x) \geq \max_{x \in S(I)} f(x) = OPT(I)$$

Κατ' αντιστοιχία, στην περίπτωση προβλημάτων ελαχιστοποίησης, η δεύτερη από τις παραπάνω συνθήκες παίρνει τη μορφή $g \leq f$ και η βέλτιστη τιμή του χαλαρωμένου προβλήματος αποτελεί ένα *κάτω φράγμα* για τη βέλτιστη τιμή του αρχικού.

Τονίζεται ότι το μεγεθυσμένο σύνολο $R(I)$ επιλέγεται με τέτοιο τρόπο ώστε το χαλαρωμένο πρόβλημα να μπορεί να βελτιστοποιηθεί αποδοτικά, δηλαδή η βέλτιστη λύση $\tilde{x} \in R(I)$ να μπορεί να υπολογιστεί σε πολυωνυμικό χρόνο. Εν γένει, όπως είναι άλλωστε αναμενόμενο, η βέλτιστη αυτή χαλαρωμένη λύση *δεν* είναι εφικτή λύση για το αντίστοιχο στιγμιότυπο του προβλήματος $\mathcal{P}$, δηλ. ισχύει $\tilde{x} \in R(I) \setminus S(I)$. Ο προφανής στόχος είναι λοιπόν η χρησιμοποίηση του χαλαρωμένου προβλήματος (και ενδεχομένως της βέλτιστης λύσης του) με κατάλληλο τρόπο ώστε να πάρουμε μια καλή προσεγγιστική (εφικτή) λύση για το εκάστοτε στιγμιότυπο του $\mathcal{P}$.

Εξάλλου, το χαλαρωμένο πρόβλημα μπορεί να χρησιμοποιηθεί με διάφορους τρόπους για τη σχεδίαση προσεγγιστικών αλγορίθμων για το $\mathcal{P}$. Τα πιο κλασικά relaxations προκύπτουν με τη χρήση του γραμμικού προγραμματισμού. Ωστόσο, όπως θα φανεί πολλές φορές στα επόμενα, υπάρχουν περιορισμοί ως προς το πόσο καλή προσέγγιση μπορεί να παράγει ένα γραμμικό πρόγραμμα. Στη συνέχεια εξειδικεύουμε την έννοια του relaxation σε σχέση με το γραμμικό προγραμματισμό.

3.2 Η έννοια του LP-relaxation – Μια γενική Προσέγγιση σχεδίασης Προσεγγιστικών Αλγορίθμων

Πολλά προβλήματα συνδυαστικής βελτιστοποίησης μπορούν να εκφραστούν ως ακέραια γραμμικά προγράμματα (*Integer Linear Programs* - *ILPs*). Υπενθυμίζεται ότι το (γενικό) πρόβλημα βελτιστοποίησης του ακέραιου γραμμικού προγραμματισμού είναι $NP-hard$, ενώ το αντίστοιχο πρόβλημα χωρίς τους περιορισμούς για ακέραιες λύσεις επιλύεται σε πολυωνυμικό χρόνο. Αν λοιπόν χαλαρώσουμε τις απαιτήσεις για ακέραιες τιμές των μεταβλητών, παίρνουμε ένα γραμμικό πρόγραμμα το οποίο φανερά ικανοποιεί τις γενικές συνθήκες (βλέπε προηγούμενη σελίδα) και επομένως είναι η χαλάρωση του αρχικού ακέραιου προγράμματος (*LP* - relaxation). Για παράδειγμα, ένας περιορισμός της μορφής $x_i \in \{0,1\}$, στο αντίστοιχο *LP* - relaxation θα μετατραπεί σε $x_i \in [0,1]$.

Ας γίνουμε όμως λίγο πιο αυστηροί. Στη συνέχεια ορίζουμε δύο βασικές γενικές κατηγορίες ακέραιων γραμμικών προγραμμάτων που έχουν μεγάλη σημασία στο πεδίο των προσεγγιστικών αλγορίθμων και θα φανούν πολύ χρήσιμες στα επόμενα.

Ορισμός 3.1 Δεδομένου ενός πίνακα $A \in [0,1]^{m \times n}$ και των διανυσμάτων $b \in [1,\infty)^m$ και $c \in [0,1]^n$, έτσι ώστε $\max_j c_j = 1$, ένα *ακέραιο γραμμικό πρόγραμμα πακεταρίσματος* (αντίστοιχα *κάλυψης*) έχει στόχο τη μεγιστοποίηση (ελαχιστοποίηση) του $c^T \cdot x$ με τους περιορισμούς $x \in \mathbb{N}^n$ (ή $\forall i \in [n] (x_i \in \{0,1,...,d_i\})$, όπου $d_i \in \mathbb{N}^*$) και $Ax \leq b$ (αντίστοιχα $Ax \geq b$). Επιπλέον αν $A \in \{0,1\}^{m \times n}$ υποθέτουμε ότι κάθε συνιστώσα του διανύσματος b είναι ένας θετικός ακέραιος αριθμός. Επιπλέον ορίζουμε ως $B = \min_i b_i$ και υποθέτουμε χωρίς βλάβη της γενικότητας (*χ. β. τ. γ.*) ότι $B \geq 1$.

Θα συμβολίζουμε με τα αρχικά *PIP* ένα *ακέραιο γραμμικό πρόγραμμα πακεταρίσματος* (*packing integer program*) και αντίστοιχα *CIP* ένα *ακέραιο γραμμικό πρόγραμμα κάλυψης* (*covering integer program*). Οι δύο αυτές κατηγορίες ακέραιων προγραμμάτων μοντελοποιούν πολλά $NP-hard$ προβλήματα συνδυαστικής βελτιστοποίησης.

Παρόλο που συνήθως δεν υπάρχουν καθόλου περιορισμοί ως προς τα σύνολα τιμών των στοιχείων των A, b και c (εκτός φυσικά από τον περιορισμό να είναι μη αρνητικά), εύκολα μπορεί να δει κανείς ότι οι παραπάνω περιορισμοί έγιναν χωρίς βλάβη της γενικότητας. Κατ' αρχάς, μπορούμε να υποθέσουμε ότι $\forall i, j\ A_{ij} \leq b_i$. Αν αυτό δεν είναι αληθές για ένα *PIP*, μπορούμε προφανώς να θέσουμε $x_j := 0$. Αν πάλι δεν είναι αληθές για ένα *CIP*, μπορούμε να αλλάξουμε την τιμή του εν λόγω στοιχείου σε $A_{ij} := b_i$. Ακολούθως, κάνοντας scaling σε κάθε γραμμή του A, έτσι ώστε $\max_j A_{ij} = 1$ (για κάθε γραμμή i) και στο διάνυσμα c, έτσι ώστε $\max_j c_j = 1$, παίρνουμε την παραπάνω μορφή για τα A, b και c (συγκεκριμένα, παίρνουμε $B \geq 1$ εφόσον $\forall i, j\ A_{ij} \leq b_i$).

Η δε ανάγκη για το βασικό περιορισμό μη αρνητικών στοιχείων είναι φανερή διαισθητικά, με δεδομένο ότι στις περισσότερες περιπτώσεις στο πεδίο των προσεγγιστικών αλγορίθμων (και για όλες τις εφαρμογές που εξετάζει η παρούσα εργασία) χειριζόμαστε μη αρνητικές ποσότητες.

Ορισμός 3.2 Η (standard) χαλάρωση (*LP* - relaxation) των προγραμμάτων της μορφής *PIP* και *CIP* επιτρέπει στο διάνυσμα x να παίρνει πραγματικές (μη αρνητικές) τιμές, $x \in \Re_+^n$. Για περιορισμούς της μορφής $\forall i \in [n]\left(x_i \in \{0,1,...,d_i\}\right)$ ο αντίστοιχος χαλαρωμένος περιορισμός είναι ο εξής: $\forall i \in [n]\left(x_i \in [0, d_i]\right)$.

Βέβαια, υπό την προϋπόθεση ότι τα στοιχεία των A, b και c είναι ρητοί αριθμοί, *όπως συμβαίνει στην περίπτωσή μας*, η βέλτιστη λύση του χαλαρωμένου γραμμικού προγράμματος θα έχει ρητές συνιστώσες. Υπό αυτή την έννοια η βέλτιστη λύση του *LP*-relaxation μπορεί να θεωρηθεί κλασματική λύση (*fractional solution*) του αρχικού προβλήματος. Για το λόγο αυτό συνήθως συμβολίζεται με $OPT_f(\cdot)$. Κατ' αναλογία με όσα ελέχθησαν κατά την περιγραφή της γενικής μεθόδου της χαλάρωσης, στην περίπτωση $NP-hard$ προβλημάτων βελτιστοποίησης δεν μπορεί να περιμένει κανείς ότι η βέλτιστη λύση του *LP* – relaxation θα έχει (εν γένει) ακέραιες συνιστώσες. Κατά συνέπεια, για τέτοια δύσκολα προβλήματα, ο στόχος δεν είναι η αναζήτηση βέλτιστης λύσης για το *LP* – relaxation, αλλά μιας προσεγγιστικής ακέραιης λύσης (integral solution).

Αναφέρουμε ωστόσο ότι η έννοια του χαλαρωμένου γραμμικού προγράμματος είναι ιδιαίτερα χρήσιμη και για προβλήματα βελτιστοποίησης που επιλύονται σε πολυωνυμικό χρόνο. Στη συνέχεια δα δούμε δύο τέτοια προβλήματα για τα οποία η βέλτιστη λύση του *LP* – relaxation είναι πάντα ακέραια, δηλαδή ταυτίζεται με τη βέλτιστη λύση του ακέραιου προγράμματος (και συνεπώς του αρχικού προβλήματος).

Στο σημείο αυτό συνοψίζουμε σε τρία βήματα την παραπάνω περιγραφείσα γενική προσέγγιση για προσεγγιστικούς αλγορίθμους:

Γενική Προσέγγιση για Προσεγγιστικούς Αλγορίθμους

1. Διατύπωση του προβλήματος βελτιστοποίησης υπό μορφή ακέραιου γραμμικού προγράμματος (*ILP*).
2. Χαλάρωση του *ILP* σε ένα γραμμικό πρόγραμμα (*LP* - relaxation).
3. Χρήση του *LP* - relaxation *με κατάλληλο τρόπο*, ώστε να πάρουμε μία λύση του *ILP*.

Ουσιαστικά το κύριο τμήμα αυτής της εργασίας σκοπό έχει να αναλύσει τη φράση «με κατάλληλο τρόπο» στην πρόταση 3.

3.3 Integrality gap – Αξιολόγηση Επίδοσης Αλγορίθμων

Στη συνέχεια ορίζουμε ένα μέγεθος το οποίο εκφράζει το «μέτρο της καλοσύνης» ενός χαλαρωμένου γραμμικού προγράμματος για ένα πρόβλημα.

Ορισμός 3.3 Έστω ένα πρόβλημα βελτιστοποίησης $\mathcal{P}$, *ILP* ένα ακέραιο γραμμικό πρόγραμμα για αυτό και *LP* η χαλάρωση (*LP* - relaxation) του *ILP*. Ορίζουμε ως *integrality gap* του χαλαρωμένου γραμμικού προγράμματος για το πρόβλημα $\mathcal{P}$ τον ακόλουθο λόγο:

$$\text{integrality gap} \doteq \sup_I \left\{ \frac{OPT(I)}{OPT_f(I)}, \frac{OPT_f(I)}{OPT(I)} \right\}$$

όπου το *supremum* υπολογίζεται για όλα τα στιγμιότυπα του προβλήματος.

Από τον ορισμό είναι φανερό ότι για κάθε πρόβλημα το integrality gap είναι τουλάχιστον 1. Σε κάποιες περιπτώσεις, σε προβλήματα για τα οποία το relaxation δίνει τη βέλτιστη λύση (οπότε επιλύονται ακριβώς σε πολυωνυμικό χρόνο), η τιμή 1 επιτυγχάνεται. Χαρακτηριστικό παράδειγμα αποτελεί το πρόβλημα κομβικής κάλυψης (vertex cover) σε διμερείς γράφους (bipartite graphs), το οποίο θα αναλυθεί στη συνέχεια του κεφαλαίου.

Για την αξιολόγηση της απόδοσης των περισσοτέρων αλγορίθμων που σχεδιάζονται με βάση την προαναφερθείσα γενική προσέγγιση, τις περισσότερες φορές συγκρίνουμε το κόστος της λύσης που παρέχεται από τον αλγόριθμο με το κόστος της βέλτιστης κλασματικής λύσης του *LP*-relaxation.

Έστω λοιπόν ότι έχουμε σχεδιάσει έναν αλγόριθμο $\mathcal{A}$ για ένα *πρόβλημα ελαχιστοποίησης* και ότι (με κάποιο τρόπο) καταφέρνουμε να αποδείξουμε ότι για κάθε στιγμιότυπο I του προβλήματος ο εν λόγω αλγόριθμος δίνει μια λύση κόστους $SOL_{\mathcal{A}}(\cdot)$ τέτοια ώστε:

$$SOL_{\mathcal{A}}(I) \le a \cdot OPT_f(I)$$

Με δεδομένο το υπό εξέταση πρόβλημα είναι ελαχιστοποίησης είναι ήδη γνωστό ότι $OPT_f(I) \le OPT(I)$. Προκύπτει λοιπόν άμεσα ότι ο αλγόριθμος $\mathcal{A}$ έχει λόγο απόδοσης (το πολύ) a για το εν λόγω πρόβλημα.

Ωστόσο, έχουμε επίσης ότι $OPT(I) \le SOL_A(I)$. Συνδυάζοντας τις ανισότητες παίρνουμε ότι $OPT(I) \le a \cdot OPT_f(I)$, δηλαδή ότι ο λόγος απόδοσης είναι τέτοιος ώστε για κάθε στιγμιότυπο του προβλήματος να ισχύει

$$a \ge \frac{OPT(I)}{OPT_f(I)} \Rightarrow a \ge integrality\ gap$$

Δηλαδή με αυτή τον τρόπο φραξίματος του κόστους της λύσης ο καλύτερος λόγος προσέγγισης που μπορεί κανείς να αποδείξει είναι το integrality gap του χαλαρωμένου γραμμικού προγράμματος.

Παρόλο το παραπάνω είναι η συνήθης περίπτωση, κάτι τέτοιο δεν είναι απαραίτητο. Πράγματι, θεωρητικά μπορεί κανείς να αποδείξει, στηριζόμενος σε επιπλέον δομικές ιδιότητες του προβλήματος, ότι για εκείνα τα στιγμιότυπα που έχουν μεγάλο (δηλ. κακό) integrality gap ο αλγόριθμος δίνει μια λύση με κόστος πολύ μικρότερο από το $(integrality\ gap) \times OPT$. Με τον τρόπο αυτό ενδεχομένως να υπάρχει η δυνατότητα απόδειξης ενός καλύτερου λόγου απόδοσης για τον αλγόριθμο.

3.4 Εφαρμογή της τεχνικής για την επίλυση δύο πολυωνυμικών προβλημάτων

Στο σημείο αυτό κρίνεται σκόπιμο να εισαχθεί μια μικρή αλλά σημαντική παρένθεση. Για να εκτιμήσει κανείς το ρόλο της δυϊκότητας και του LP – relaxation στο πεδίο των προσεγγιστικών αλγορίθμων, είναι ιδιαίτερα χρήσιμο να κατανοήσει τη σημασία τους στο πεδίο των ακριβών αλγορίθμων.

Για το λόγο αυτό, στα επόμενα παρουσιάζονται δύο χαρακτηριστικά προβλήματα και ο τρόπος που σχετίζονται με τις παραπάνω έννοιες. Ακολουθούν κάποιοι χρήσιμοι ορισμοί και θεωρήματα.

Ορισμός 3.4.1 Ένας τετραγωνικός πίνακας B ονομάζεται *unimodular* αν $\det(B) \in \{+1, -1\}$. Ένας πίνακας A ονομάζεται *ολικά unimodular* αν για κάθε τετραγωνικός αντιστρέψιμος υποπίνακας B του A είναι unimodular.

Διατυπώνουμε τα ακόλουθα θεωρήματα:

Θεώρημα 3.4.2 Έστω το γραμμικό πρόγραμμα $\min\left\{c^T x : Ax \begin{Bmatrix} \ge \\ \le \end{Bmatrix} b, x \ge \mathbf{0}\right\}$. Όλες οι κορυφές του συνόλου των εφικτών λύσεων είναι ακέραιες *αν* ο πίνακας A είναι ολικά unimodular και το b έχει μόνο ακέραιες συνιστώσες.

Θεώρημα 3.4.3 Έστω A πίνακας με στοιχεία από το σύνολο $\{-1, 0, +1\}$, τέτοιος ώστε κάθε στήλη να έχει το πολύ δύο μη μηδενικές συνιστώσες. Ας υποθέσουμε επίσης ότι οι γραμμές του πίνακα μπορούν να διαμεριστούν σε δύο σύνολα, έστω I_1 και I_2 έτσι ώστε:

1. Αν μία στήλη περιέχει δύο μη μηδενικές συνιστώσες του ίδιου πρόσημου, τότε αυτές εμφανίζονται σε διαφορετικά σύνολα.
2. Αν μία στήλη περιέχει δύο μη μηδενικές συνιστώσες διαφορετικού πρόσημου, τότε αυτές εμφανίζονται στο ίδιο σύνολο.

Με αυτές τις προϋποθέσεις, ο πίνακας A είναι ολικά unimodular.

Απόδειξη: Με επαγωγή στο μέγεθος των υποπινάκων. Η βάση της επαγωγής είναι τετριμμένη.

Έστω C ένας $k \times k$ υποπίνακας. Διακρίνουμε τρεις περιπτώσεις:

- Αν ο C έχει μια μηδενική στήλη, τότε είναι μη αντιστρέψιμος.

- Αν ο C έχει μια στήλη με ακριβώς μία μη μηδενική συνιστώσα. Έστω ότι η συνιστώσα αυτή είναι στη θέση (i, j). Τότε,
$$\det(C) = (-1)^{i+j} \cdot M_{ij}$$
όπου M_{ij} είναι η ορίζουσα του πίνακα που προκύπτει αν από τον C διαγράψουμε την i - οστή γραμμή και την j - οστή στήλη. Από την επαγωγική υπόθεση έπεται ότι
$$\det(C) \in \{\pm 1, 0\}.$$

- Όλες οι στήλες του C έχουν δύο μη μηδενικές συνιστώσες. Από αυτό έπεται ότι το άθροισμα όλων των γραμμών του C στο I_1 είναι ίσο με το αντίστοιχο άθροισμα στο I_2. Επομένως, οι γραμμές του C είναι γραμμικά εξαρτημένες, οπότε $\det(C) = 0$.

□

Τα παραπάνω θεωρήματα χρησιμεύουν με τρόπο που θα φανεί παρακάτω στα δύο προβλήματα που μας ενδιαφέρουν.

3.4.1. Πρόβλημα Μέγιστου Ταιριάσματος σε Διμερές Γράφημα

Έστω $G = (V, E)$ ένα διμερές γράφημα με partitions V_1 και V_2 ($V = V_1 \bigcup V_2$). Μπορούμε εύκολα να εκφράζουμε το πρόβλημα εύρεσης μέγιστου ταιριάσματος στο G ως ένα ακέραιο γραμμικό πρόγραμμα. Ως σύνολο των μεταβλητών απόφασης παίρνουμε το χαρακτηριστικό διάνυσμα του συνόλου των ακμών E. Δηλαδή σε κάθε ακμή $e \in E$ αντιστοιχεί μια μεταβλητή $x_e \in \{0,1\}$ που δηλώνει τη παρουσία ή μη της ακμής στη λύση. Το αντίστοιχο *ILP* (πρόκειται μάλιστα για *PIP*) είναι το ακόλουθο:
$$\max \sum_{e \in E} x_e$$
με τους περιορισμούς:
$$\forall v \in V \left(\sum_{e \text{ incident at } v} x_e \le 1 \right)$$
$$\forall e \in E \left(x_e \in \{0,1\} \right)$$

Στο χαλαρωμένο γραμμικό πρόγραμμα αντικαθιστούμε κάθε περιορισμό της μορφής $x_e \in \{0,1\}$ με τον περιορισμό $x_e \ge 0$. Σημειώνεται ότι ο περιορισμός $x_e \le 1$ είναι εν προκειμένω περιττός, διότι επιβάλλεται από την πρώτη κατηγορία συνθηκών.

Είναι εύκολο να δει κανείς ότι ο πίνακας που αντιστοιχεί στην πρώτη κατηγορία περιορισμών (πίνακας A) είναι ο πίνακας πρόσπτωσης του γραφήματος (incidence matrix). Πράγματι, κάθε γραμμή του πίνακα αντιστοιχεί σε κόμβους του G και κάθε στήλη αντιστοιχεί σε ακμές. Ένα στοιχείο $A_{ve} \in \{0,1\}$ αναπαριστά αν η ακμή e προσπίπτει στον κόμβο v.

Με δεδομένο ότι το γράφημα G είναι διμερές, είναι επίσης εύκολο να δούμε ότι ο πίνακάς πρόσπτωσης του G (δηλ. ο A) ικανοποιεί τις προϋποθέσεις του θεωρήματος 3 με I_i το σύνολο των γραμμών που αντιστοιχούν στο V_i. Άρα, ο πίνακας A είναι ολικά unimodular. Επειδή επιπλέον το διάνυσμα b έχει μόνο ακέραιες συνιστώσες, από το θεώρημα 2 παίρνουμε τις ακόλουθες προτάσεις:

Πρόταση 1 Η βέλτιστη λύση του αντίστοιχου χαλαρωμένου γραμμικού προγράμματος είναι πάντα ακέραια (δηλ. έχει ακέραιες συνιστώσες).

Πρόταση 2 Ο πίνακας πρόσπτωσης ενός διμερούς γραφήματος είναι ολικά unimodular.

Επομένως, η βέλτιστη λύση του *LP* – relaxation ταυτίζεται με τη βέλτιστη λύση του προβλήματος, δηλαδή δίνει σε κάθε περίπτωση ένα μέγιστο ταίριασμα για το διμερές γράφημα G. Το γεγονός αυτό δεν μας προκαλεί έκπληξη, διότι ως γνωστόν το εν λόγω πρόβλημα επιλύεται ακριβώς σε πολυωνυμικό χρόνο.

3.4.2 Πρόβλημα Ελάχιστης Κομβικής Κάλυψης σε Διμερές Γράφημα

Με αντίστοιχο τρόπο προκύπτει η κατασκευή του ακέραιου γραμμικού προγράμματος (τώρα πρόκειται για *CIP*) και του σχετικού *LP* - relaxation.

$$\min \sum_{v \in V} x_v$$

με τους περιορισμούς:

$$\forall e = (u, v) \in E\left(x_u + x_v \geq 1\right)$$
$$\forall v \in V\left(x_v \in \{0,1\}\right)$$

Ομοίως, το relaxation έχει ως δεύτερο περιορισμό τον $x_v \geq 0$.

Μπορεί κανείς να παρατηρήσει ότι ο πίνακας που αντιστοιχεί στην πρώτη κατηγορία περιορισμών είναι ο ανάστροφος του πίνακα πρόσπτωσης. Με τον ίδιο λοιπόν τρόπο όπως και προηγουμένως μπορεί κανείς να αποδείξει την Πρόταση 1 και για το πρόβλημα αυτό. Επομένως, και εν προκειμένω η βέλτιστη λύση του *LP* – relaxation δίνει τη βέλτιστη λύση του προβλήματος (είναι επίσης γνωστό ότι το συγκεκριμένο πρόβλημα λύνεται πολυωνυμικά).

Με όσα έχουν ειπωθεί στο κεφάλαιο 2, είναι εύκολο να δει κανείς ότι τα χαλαρωμένα γραμμικά προγράμματα για τα δύο υπό εξέταση προβλήματα είναι δυϊκά. Με βάση λοιπόν το θεώρημα της ισχυρής δυϊκότητας, προκύπτει ως παρεπόμενο το γνωστό θεώρημα των König και Egerváry.

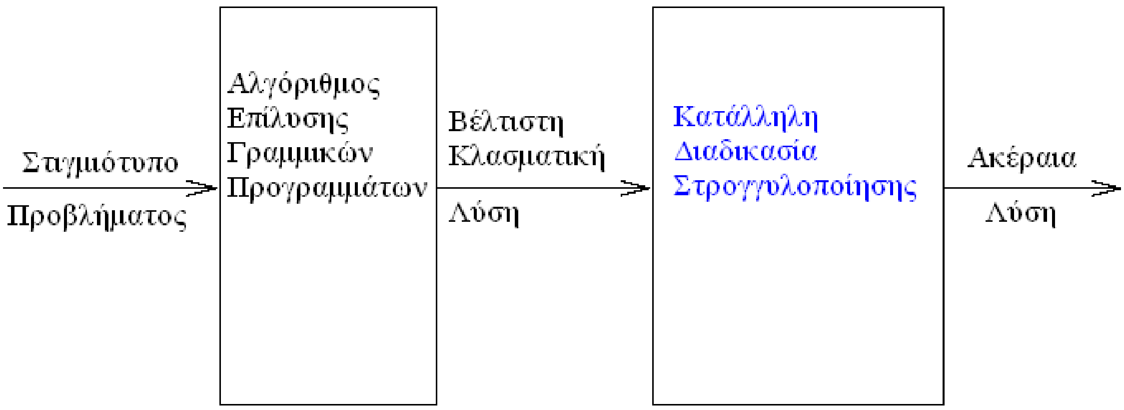
3.5 Η τεχνική της στρογγυλοποίησης

Τα παραπάνω (πολυωνυμικά) προβλήματα οδηγούν στον πρώτο (και πλέον προφανή) τρόπο χρήσης του *LP* – relaxation για τη σχεδίαση προσεγγιστικών αλγορίθμων, την τεχνική της στρογγυλοποίησης (*LP* – rounding ή απλά rounding). Σύμφωνα με την τεχνική αυτή, το βήμα 3 της γενικής μας προσέγγισης διασπάται στα ακόλουθα βήματα:

Στρογγυλοποίηση (*LP* - rounding)

3.1 Υπολογισμός μιας βέλτιστης λύσης του χαλαρωμένου γραμμικού προγράμματος (κλασματική λύση).
3.2 Μετατροπή της κλασματικής αυτής λύσης σε ακέραια προσπαθώντας να εξασφαλίσουμε ότι κατά τη μετατροπή δε θα μεταβληθεί «κατά πολύ» το κόστος.

Μια αναπαράσταση της τεχνικής αυτής δίνεται στο ακόλουθο σχήμα:



Είναι φανερό ότι κατά την χρησιμοποίηση της τεχνικής της στρογγυλοποίησης, ο λόγος απόδοσης των αλγορίθμων καθορίζεται συγκρίνοντας το κόστος της integral και της fractional λύσης. Επομένως ισχύουν τα όσα αναφέρθηκαν στα προηγούμενα περί integrality gap.

Ο τρόπος μετατροπής της κλασματικής βέλτιστης λύσης σε ακέραια (και επομένως εφικτή για το πρόβλημα) *εν γένει* λαμβάνει υπόψη δομικές ιδιότητες του υπό εξέταση προβλήματος και μπορεί να επιτευχθεί είτε με ντετερμινιστικό είτε με randomized τρόπο. Στην τελευταία περίπτωση η τεχνική ονομάζεται τυχαιοποιημένη στρογγυλοποίηση (randomized rounding) και αποτελεί το θέμα του επόμενου κεφαλαίου. Στο υπόλοιπό αυτού του κεφαλαίου περιοριζόμαστε σε ντετερμινιστικούς αλγορίθμους.

Επίσης, παρά το γεγονός ότι η μέθοδος του rounding είναι αρκετά εύκολη στην κατανόηση, περιέχει ένα σημαντικό μειονέκτημα: απαιτεί την επίλυση ενός γραμμικού προγράμματος. Βέβαια, αυτό μπορεί να επιτευχθεί σε πολυωνυμικό χρόνο στη χειρότερη περίπτωση, αλλά οι σχετικοί προσεγγιστικοί αλγόριθμοι γίνονται αργοί.

Ένα άλλο μειονέκτημα της μεθόδου είναι το γεγονός ότι δεν αφήνει πολλά περιθώρια να εκμεταλλευτεί κανείς τη δομή του εκάστοτε προβλήματος. Χαρακτηριστικό είναι ότι σύμφωνα με την περιγραφή της μεθόδου, βλέπουμε τον *LP* – solver σαν «μαύρο κουτί» και επεμβαίνουμε μόνο στη λύση του. Παρά τα προηγούμενα μειονεκτήματα, έχει αποδειχθεί στην πράξη ότι η εν λόγω τεχνική έχει οδηγήσει στη σχεδίαση πολύ σημαντικών αλγορίθμων για πληθώρα προβλημάτων και σε αρκετές περιπτώσεις με λόγους απόδοσης που (αποδεδειγμένα) δεν μπορούν να βελτιωθούν, με την υπόθεση ότι $P \neq NP$.

Υπάρχει ένα πρόβλημα με όλα τα παραπάνω που μέχρι στιγμής έχουμε αγνοήσει: ενδέχεται το ακέραιο γραμμικό πρόγραμμα που μοντελοποιεί το αρχικό πρόβλημα βελτιστοποίησης να είναι πολύ μεγάλο, δηλαδή να έχει εκθετικό πλήθος περιορισμών ή μεταβλητών (ως προς το μέγεθος της εισόδου). Εν τοιαύτη περιπτώσει, δεν είναι εν γένει αληθές ότι το *LP* – relaxation (που φανερά έχει το ίδιο πλήθος περιορισμών) μπορεί να επιλυθεί βέλτιστα σε πολυωνυμικό χρόνο.

Η ύπαρξη πολυωνυμικού αριθμού περιορισμών (και μεταβλητών) είναι φανερά ικανή συνθήκη για την επιλυσιμότητα του χαλαρωμένου γραμμικού προγράμματος σε πολυωνυμικό χρόνο (η σύνθεση πολυωνύμων είναι πολυώνυμο). Ωστόσο, ακόμα κι αν το relaxation έχει εκθετικό αριθμό περιορισμών, ενδέχεται να λύνεται σε πολυωνυμικό χρόνο, αν μπορούμε να βρούμε ένα *μαντείο διαχωρισμού* (separation oracle), δηλ. έναν πολυωνυμικό αλγόριθμο ο οποίος, δεδομένου κάποιου σημείου $x \in \Re^n$ (όπου n είναι ο αριθμός των μεταβλητών του relaxation) επιβεβαιώνει ότι το σημείο είναι εφικτή λύση (δηλ. ικανοποιεί όλους τους περιορισμούς) ή παράγει έναν (από τους εκθετικά πολλούς) περιορισμό που παραβιάζεται. Δεδομένου ενός τέτοιου μαντείου, μπορούμε να χρησιμοποιήσουμε την *ελλειψοειδή μέθοδο* για να λύσουμε το relaxation σε πολυωνυμικό χρόνο ανεξάρτητα από το πλήθος των περιορισμών. Εν τοιαύτη περιπτώσει, η πολυπλοκότητα της επίλυσης του LP - relaxation, αν και πολυωνυμική, είναι σε γενικές γραμμές απαγορευτική.

Μια ακόμα ειδική περίπτωση είναι να έχουμε (εν δυνάμει) εκθετικό αριθμό μεταβλητών. Ακόμα και στην περίπτωση αυτή ενδέχεται το *LP* - relaxation να λύνεται σε πολυωνυμικό χρόνο. Χαρακτηριστικό παράδειγμα αποτελεί ο αλγόριθμος των Karmarkar και Karp για το πρόβλημα bin packing που θα αναλυθεί στη συνέχεια του κεφαλαίου.

3.6 Εφαρμογή της τεχνικής της στρογγυλοποίησης για το πρόβλημα Set Cover

Θα χρησιμοποιήσουμε την παραπάνω γενική προσέγγιση και συγκεκριμένα την τεχνική της στρογγυλοποίησης με δύο διαφορετικούς τρόπους για το πρόβλημα Set Cover το οποίο ορίζουμε παρακάτω:

Set Cover

- **Είσοδος:**
  - Ένα σύνολο $T = \{t_1, t_2, ..., t_n\}$
  - Μια συλλογή από υποσύνολα του T, $\mathcal{S} = \{S_1, S_2, ..., S_m\}$
  - Μη αρνητικά βάρη $w_j \geq 0$ για κάθε υποσύνολο S_j
- **Έξοδος:** Ένα σύνολο δεικτών $I \subseteq \{1, 2, ..., m\}$ το οποίο ελαχιστοποιεί το άθροισμα $\sum_{j \in I} w_j$ με τον περιορισμό $\bigcup_{j \in I} S_j = T$. Δηλαδή αναζητούμε μία ελαχίστου κόστους συλλογή από υποσύνολα η οποία να καλύπτει το βασικό σύνολο T.

Εφαρμόζουμε τη γενική διαδικασία σχεδίασης για το εν λόγω πρόβλημα:

1. Διατυπώνουμε το πρόβλημα υπό μορφή ακέραιου γραμμικού προγράμματος. Εν προκειμένω, δημιουργούμε μία μεταβλητή x_j για κάθε υποσύνολο S_j. Αν $j \in I$, τότε $x_j = 1$, ειδάλλως $x_j = 0$.

$$\min \sum_{j=1}^{m} w_j x_j$$

   με τους περιορισμούς:

$$\forall t_i \in T \left(\sum_{j: t_i \in S_j} x_j \geq 1 \right)$$
$$\forall j \in \{1, 2, ..., m\} \left(x_j \in \{0, 1\} \right)$$

2. Χαλαρώνουμε την απαίτηση για ακέραια x_j μετατρέποντας τον τελευταίο περιορισμό σε $0 \leq x_j \leq 1$. Τονίζεται δε ότι με δεδομένο ότι τα βάρη είναι μη αρνητικά και ζητούμενο είναι η ελαχιστοποίηση της τιμής της αντικειμενικής συνάρτησης, ο περιορισμός $x_j \leq 1$ δεν είναι αναγκαίος. Έστω OPT η βέλτιστη τιμή της αντικειμενικής συνάρτησης για το ακέραιο πρόγραμμα. Έστω επίσης OPT_f η αντίστοιχη βέλτιστη τιμή για το χαλαρωμένο γραμμικό πρόγραμμα. Όπως έχει ήδη εξηγηθεί αναλυτικά, ισχύει ότι $OPT_f \leq OPT$.

Ας ορίσουμε την παράμετρο f ως εξής:

$$f = \max_i \left\| \{ j : t_i \in S_j \} \right\|$$

Δηλαδή f είναι ο μέγιστος αριθμός συνόλων που περιέχουν οποιοδήποτε δοθέν στοιχείο. Θεωρούμε την ακόλουθη διαδικασία στρογγυλοποίησης:

3.1 Επίλυσε το γραμμικό πρόγραμμα και έστω x^* η βέλτιστη κλασματική λύση. 3.2 $I \leftarrow \varnothing$. Για κάθε S_j Αν $x_j^* \geq 1/f$, τότε $I \leftarrow I \cup \{j\}$

Λήμμα 3.6.1 Ο παραπάνω αλγόριθμος παράγει ένα set cover.

Απόδειξη: Έστω ότι υπάρχει ένα στοιχείο t_i που δεν καλύπτεται, δηλαδή $t_i \notin \bigcup_{j \in I} S_j$. Τότε για κάθε σύνολο S_j που περιέχει το t_i θα έχουμε ότι $x_j^* < 1/f$. Επομένως,

$$\begin{aligned} \sum_{j : t_i \in S_j} x_j^* &< \frac{1}{f} \cdot \left\| \{ j : t_i \in S_j \} \right\| \\ &\leq 1 \end{aligned}$$

εφόσον $\left\| \{ j : t_i \in S_j \} \right\| \leq f$. Ωστόσο, η παραπάνω συνθήκη παραβιάζει τον περιορισμό του γραμμικού προγράμματος για το t_i (άτοπο).

□

Θεώρημα 3.6.2 [Hoch82] Ο παραπάνω αλγόριθμος είναι f - προσεγγιστικός για το set cover.

Απόδειξη: Κατ' αρχάς είναι φανερό ότι ο εν λόγω αλγόριθμος είναι πολυωνυμικού χρόνου. Επιπλέον,

$$\begin{aligned} \sum_{j \in I} w_j &\leq \sum_j w_j x_j^* f \\ &= f \cdot \underbrace{\sum_j w_j x_j^*}_{z_{LP}} \\ &\leq f \cdot OPT \end{aligned}$$

Η πρώτη ανισότητα έπεται από το γεγονός ότι $j \in I$ αν και μόνο αν $x_j^* f \geq 1$.

□

Μπορεί να παρατηρήσει κανείς ότι ο παραπάνω αλγόριθμος συνεπάγεται άμεσα έναν 2 – προσεγγιστικό αλγόριθμό για το πρόβλημα vertex cover το οποίο ορίζουμε ακολούθως.

Weighted Vertex Cover

- **Είσοδος:**
  - Ένας μη κατευθυνόμενος γράφος $G=(V,E)$
  - Μη αρνητικά βάρη $w_i \geq 0 \ \ \forall i \in V$
- **Έξοδος:** Ένα υποσύνολο των κόμβων $C \subseteq V$ που ελαχιστοποιεί το άθροισμα $\sum_{i \in C} w_i$ και τέτοιο ώστε κάθε ακμή να είναι γειτονική με κάποιον κόμβο του C.

Μπορεί εύκολα κανείς να μετασχηματίσει το παραπάνω πρόβλημα σε μία ειδική μορφή του προβλήματος set cover. Πράγματι, το σύνολο των ακμών αντιστοιχεί στο βασικό σύνολο T και το σύνολο των κόμβων αντιστοιχεί στη συλλογή από υποσύνολα του T (δηλαδή το υποσύνολο που αντιστοιχεί στον κόμβο i είναι όλες οι ακμές που προσπίπτουν στον κόμβο αυτό). Εφόσον κάθε ακμή βρίσκεται σε ακριβώς δύο σύνολα, εν τοιαύτη περιπτώσει ισχύει ότι $f=2$. Συνεπώς, οι f - προσεγγιστικοί αλγόριθμοι για το set cover μεταφράζονται αυτόματα σε 2 – προσεγγιστικούς αλγορίθμους για το vertex cover.

Ένας άλλος τρόπος να χρησιμοποιήσει κανείς τη μέθοδο της στρογγυλοποίησης είναι να την εφαρμόσει στη λύση του δυϊκού προγράμματος του *LP*-relaxation. Το δυϊκό του χαλαρωμένου γραμμικού προγράμματος για το πρόβλημα set cover είναι το ακόλουθο:

$$\max \sum_{i=1}^{n} y_i$$

με τους περιορισμούς:

$$\forall S_j \ \left(\sum_{i:t_i \in S_j} y_i \leq w_j \right)$$
$$\forall t_i \in T \left(y_i \geq 0 \right)$$

Λόγω της ασθενούς δυϊκότητας, αν y είναι μια εφικτή λύση για το παραπάνω πρόγραμμα, τότε

$$\sum_i y_i \leq OPT_f \leq OPT$$

Ακολούθως παραθέτουμε έναν αλγόριθμο για την εύρεση ενός set cover μικρού κόστους με χρήση του δυϊκού γραμμικού προγράμματος:

Δυϊκό Γραμμικό Πρόγραμμα

3.1 Επίλυσε το δυϊκό του χαλαρωμένου γραμμικού προγράμματος και υπολόγισε μια βέλτιστη λύση y^*.

3.2 $I \leftarrow \varnothing$.

Για κάθε S_j

Αν $\sum_{i:t_i \in S_j} y_i^* = w_j$, τότε $I \leftarrow I \bigcup \{j\}$

Λήμμα 3.6.3 Ο παραπάνω αλγόριθμος παράγει ένα set cover.

Απόδειξη: Έστω ότι $\exists t_i \notin \bigcup_{j\in I} S_j$. Τότε για κάθε σύνολο S_j που περιέχει το t_i θα ισχύει ότι

$$\sum\nolimits_{i:t_i \in S_j} y_i^* < w_j$$

οπότε μπορούμε να αυξήσουμε το y_i^* κατά μία θετική ποσότητα χωρίς να παραβιάσουμε τους περιορισμούς του γραμμικού προγράμματος, γεγονός που αντιτίθεται στη βελτιστότητα του y^*.

□

Θεώρημα 3.6.4 [Hoch82] Ο παραπάνω αλγόριθμος είναι f - προσεγγιστικός για το set cover.

Απόδειξη: Με δεδομένο ότι επιλέγουμε το σύνολο S_j μόνο αν ο αντίστοιχος $\le$ περιορισμός ισχύει με ισότητα, έχουμε:

$$\begin{aligned}
\sum_{j\in I} w_j &= \sum_{j\in I} \sum_{i:t_i\in S_j} y_i^* \\
&= \sum_i y_i^* \cdot \left\| \{ j \in I : t_i \in S_j \} \right\| \\
&\le f \cdot \sum_i y_i^* \\
&\le f \cdot OPT
\end{aligned}$$

□

Στο σημείο αυτό αναφέρουμε ότι, λόγω των συνθηκών του complementary slackness, όποτε ο αρχικός αλγόριθμός στρογγυλοποίησης συμπεριλαμβάνει ένα σύνολο S_j στην κάλυψη (επειδή ισχύει ότι $x_j^* \ge 1/f$), ο αντίστοιχος δυϊκός περιορισμός ($\sum_{i:t_i\in S_j} y_i^* \le w_j$) ισχύει με ισότητα. Άρα, ο παραπάνω αλγόριθμος επίσης περιλαμβάνει το S_j στη λύση του. Επομένως, ο αλγόριθμος που χρησιμοποιεί το δυϊκό γραμμικό πρόγραμμα δεν παρέχει ποτέ καλύτερη λύση από τον αρχικό αλγόριθμο στρογγυλοποίησης.

3.7 Εφαρμογή της τεχνικής στρογγυλοποίησης για το πρόβλημα minimum multi - cut

Με σκοπό να δείξουμε ότι η τεχνική της στρογγυλοποίησης μπορεί να οδηγήσει σε αρκετά περίπλοκούς αλγορίθμους, στρεφόμαστε στο πρόβλημα minimum multi – cut το οποίο ορίζουμε αμέσως:

Minimum Multi – cut

- **Είσοδος:**
  - Ένας μη κατευθυνόμενος γράφος $G=(V,E)$
  - k διακεκριμένα ζεύγη κόμβων (s_i,t_i), $i\in[k]$
  - Μια συνάρτηση κόστους $c:E\to\mathbb{N}^*$
- **Έξοδος:** Ένα σύνολο ακμών F *ελαχίστου κόστους* τέτοιο ώστε για κάθε $i\in[k]$, οι κόμβοι s_i και t_i να ανήκουν σε διαφορετικές συνεκτικές συνιστώσες του $G'=(V,E\setminus F)$

Το πρόβλημα είναι $NP-$hard ακόμα κι αν το γράφημα G είναι δέντρο. Ακολουθώντας τη γενική τεχνική σχεδίασης, σε πρώτη φάση μοντελοποιούμε το πρόβλημα ως ακέραιο γραμμικό πρόγραμμα. Για το σκοπό αυτό συμβολίζουμε με $\mathcal{P}_i$ το σύνολο όλων των μονοπατιών ανάμεσα στους κόμβους s_i και t_i. Το ακέραιο γραμμικό πρόγραμμα είναι το ακόλουθο:

$$\min\sum_{e\in E}c_e x_e$$

με τους περιορισμούς:

$$\forall i\in[k]\ \forall P\in\mathcal{P}_i\ \left(\sum_{e\in P}x_e\geq 1\right)$$

$$\forall e\in E\ \left(x_e\in\{0,1\}\right)$$

Στο επόμενο βήμα, κατά τα γνωστά, χαλαρώνουμε την απαίτηση για ακέραιες μεταβλητές $x_e\in\{0,1\}$ σε $x_e\in[0,1]$. Εύκολα μπορεί να δει κανείς ότι ούτε στην περίπτωση αυτή η απαίτηση $x_e\leq 1$ είναι απαραίτητη (ίδιο ακριβώς επιχείρημα με το αντίστοιχο για το πρόβλημα set cover). Άρα, λοιπόν στο χαλαρωμένο γραμμικό πρόγραμμα οι περιορισμοί της δεύτερης μορφής έχουν αντικατασταθεί από τον εξής περιορισμός $x_e\geq 0$, $e\in E$.

Μια διαισθητική ερμηνεία του χαλαρωμένου προγράμματος είναι η εξής: Προσπαθούμε να δώσουμε (μη αρνητικές) ετικέτες απόστασης σε κάθε ακμή, έτσι ώστε για κάθε μονοπάτι $P\in\bigcup_i\mathcal{P}_i$, οι ετικέτες των ακμών του να έχουν άθροισμα τουλάχιστον 1.

Είναι προφανές ότι το πλήθος των περιορισμών του παραπάνω ακέραιου προγράμματος (και επομένως και του αντίστοιχου χαλαρωμένου γραμμικού) είναι εν δυνάμει εκθετικό ως προς το $n=\|V\|$. Με άλλα λόγια, το μέγεθος του γραμμικού προγράμματος είναι εκθετικό ως προς το μέγεθος της εισόδου. Επομένως, τίθεται το ερώτημα αν το χαλαρωμένο πρόγραμμα μπορεί να λυθεί σε πολυωνυμικό χρόνο ως προς το n.

Η απάντηση σε αυτό το ερώτημα είναι θετική, διότι μπορούμε εύκολα να βρούμε ένα μαντείο διαχωρισμού για το πρόβλημα. Εν προκειμένω, το μαντείο διαχωρισμού είναι πολύ απλό: αρκεί ένας αλγόριθμος που επιλύει το πρόβλημα του συντομότερου μονοπατιού. Δοθέντος ενός διανύσματος $\hat{x}\in\Re_{+}^{\|E\|}$ (αν οποιαδήποτε συνιστώσα είναι αρνητική πρόκειται προφανώς για μη εφικτή λύση) υπολογίζουμε για κάθε ζεύγος (s_i,t_i) το μήκος του συντομότερου s_i-t_i μονοπατιού (ερμηνεύοντας την τιμή $x_{(i,j)}$, $(i,j)\in E$ σαν την απόσταση των κόμβων i και j). Αν για κάθε $i\in[k]$ το μήκος του συντομότερου μονοπατιού ανάμεσα στους κόμβους s_i, t_i (δηλ. το $\min_{P\in\mathcal{P}_i}\sum_{e\in P}\hat{x}_e$) είναι τουλάχιστον 1, έπεται ότι το $\hat{x}$ είναι εφικτή λύση. Ειδάλλως, αν για κάποιο $i\in[k]$ το μήκος του συντομότερου s_i-t_i μονοπατιού είναι <1 πρόκειται για μη εφικτή λύση και το συντομότερο μονοπάτι δίνει μια ανισότητα (περιορισμό) που δεν ικανοποιείται. Με δεδομένο ότι $k=O(n^2)$ (εφόσον υποθέσαμε διακεκριμένα ζεύγη) και $Shortest\ Path\in O(n^2)$, το μαντείο είναι πολυωνυμικού χρόνου.

Βέβαια, για το εν λόγω πρόβλημα, μπορεί κανείς να κατασκευάσει ένα ακέραιο πρόγραμμα με πολυωνυμικό πλήθος περιορισμών, με χρήση του δυϊκού προβλήματος. Έτσι, δεν είναι επιβεβλημένη η χρήση της ελλειψοειδούς μεθόδου. Ωστόσο, κάτι τέτοιο δε θα μας απασχολήσει περαιτέρω. Αυτό που μας ενδιαφέρει είναι ότι μπορούμε να επιλύσουμε το χαλαρωμένο γραμμικό πρόγραμμα σε πολυωνυμικό χρόνο ως προς το n.

Στη συνέχεια περιγράφουμε μια μέθοδο βασισμένη στη στρογγυλοποίηση που δίνει ένα $O(\log k)$ προσεγγιστικό αλγόριθμο για το πρόβλημα. Για το σκοπό αυτό δίνουμε μια φυσική ερμηνεία για το χαλαρωμένο γραμμικό πρόγραμμα. Βλέπουμε το γραμμικό πρόγραμμα ως ένα σύστημα σωληνώσεων, όπου $e\equiv(i,j)\in E$ σημαίνει ότι υπάρχει σωλήνας μεταξύ των i και j, x_e είναι το μήκος του σωλήνα και c_e είναι το εμβαδόν της διατομής του. Επομένως, $c_e x_e$ είναι ο όγκος του σωλήνα (εμβαδόν διατομής επί μήκος). Άρα, επίλυση του γραμμικού προγράμματος ισοδυναμεί με εύρεση του συστήματος σωληνώσεων ελαχίστου συνολικού όγκου, στο οποίο όλα τα ζεύγη (s_i,t_i) απέχουν τουλάχιστον μία μονάδα.

Δοθείσης μιας εφικτής λύσης x του χαλαρωμένου γραμμικού προγράμματος, έστω $dist_x(u,v)$ το μήκος του συντομότερου μονοπατιού μεταξύ των κόμβων u και v για την ανάθεση αποστάσεων x. Έστω επίσης $B_x(u,r)=\{v: dist_x(u,v)\le r\}$ η σφαίρα ακτίνας r γύρω από τον κόμβο u. Με βάση τα προηγούμενα προκύπτει ένας απλός προσεγγιστικός αλγόριθμος [GVY96] για το πρόβλημα:

GVY

3.1 Επίλυσε το LP – relaxation και πάρε τη βέλτιστη λύση x.
3.2 $F \leftarrow \varnothing$; $H \leftarrow G$
while $\exists$ συνδεδεμένο ζεύγος s_i, t_i στο τρέχον γράφημα H:
Έστω $S = B_x(s_i, r)$, για κάποια *κατάλληλη επιλογή* της ακτίνας r, τ.ω. $r < 1/2$.
Πρόσθεσε το $\delta_H(S)$ στο F.
Αφαίρεσε το S (και τις γειτονικές προς αυτό ακμές) από το τρέχον γράφημα H.
return F.

Με $\delta(S)$ συμβολίζουμε το σύνολο των ακμών στην τομή $(S, \overline{S})$. Δεν κρίνεται σκόπιμο να δώσουμε περισσότερες λεπτομέρειες, όπως για παράδειγμα πώς ακριβώς επιλέγουμε την περιοχή S σε κάθε επανάληψη κοκ.

Είναι εύκολο να δει κανείς ότι ο αλγόριθμος πάντα τερματίζει, διότι σε κάθε επανάληψη του βρόχου while αποσυνδέεται τουλάχιστον ένα ζεύγος (s_i, t_i). Χωρίς ιδιαίτερη προσπάθεια μπορεί επίσης να δει κανείς ότι ο αλγόριθμος επιστρέφει πράγματι μια πολλαπλή τομή (multi-cut).

Η μόνη δυσκολία προς αυτή την κατεύθυνση είναι η ακόλουθη: το ενδεχόμενο σε κάποια επανάληψη στην οποία διαχωρίζουμε το ζεύγος (s_i, t_i) να βρεθούν τα s_j και t_j (για κάποιο j) στο S. Αν κάτι τέτοιο ήταν εφικτό, τότε ο αλγόριθμος θα αφαιρούσε το S από το τρέχον γράφημα και ενδεχομένως τα s_j και t_j να μην ήταν διαχωρισμένα στη «λύση» του αλγορίθμου.

Έστω ότι κάτι τέτοιο συμβαίνει. Τότε από την κατασκευή του αλγορίθμου έπεται ότι $dist_x(s_i, s_j) \le r$ και $dist_x(s_i, t_j) \le r$. Αυτό σημαίνει ότι υπάρχει μονοπάτι $s_j \to \ldots \to s_i \ldots \to t_j$ μήκους λιγότερο από $2r$. Αλλά, εφόσον $r < \frac{1}{2}$, προκύπτει ότι υπάρχει μονοπάτι ανάμεσα στα s_j και t_j μήκους λιγότερο από 1, γεγονός που παραβιάζει το εφικτό του x (άτοπο). Άρα, κάτι τέτοιο δεν μπορεί να συμβεί και ο αλγόριθμος επιστρέφει μια εφικτή λύση στο πρόβλημα.

Τέλος, δίνουμε χωρίς απόδειξη το ακόλουθο θεώρημα. Η απόδειξη ξεφεύγει από τους σκοπούς της παρούσας εργασίας.

Θεώρημα 3.7.1 Ο αλγόριθμος GVY είναι $O(\log k)$-προσεγγιστικός αλγόριθμος για το πρόβλημα minimum multi-cut.

3.8 Εφαρμογή της τεχνικής στρογγυλοποίησης στο πρόβλημα bin packing

Θεωρούμε το πρόβλημα bin packing και εφαρμόζουμε μία βελτιωμένη τεχνική βασισμένη στη μέθοδο της στρογγυλοποίησης του γραμμικού προγραμματισμού για την κατασκευή ενός προσεγγιστικού αλγορίθμου για αυτό. Αξίζει να τονιστεί ότι ο εν λόγω προσεγγιστικός αλγόριθμος δίνει λύση που περιέχει μόνο ένα προσθετικό όρο σφάλματος και υπό αυτή την έννοια είναι ο μέχρι σήμερα καλύτερος προσεγγιστικός αλγόριθμος για το πρόβλημα.

Bin Packing

- **Είσοδος:** Ένα σύνολο από αντικείμενα $\{1,2,...,n\}$. Το αντικείμενο i έχει μέγεθος $s_i \in \mathbb{Q} \cap (0,1]$. Χ. β. τ. γ. υποθέτουμε ότι $1 \geq s_1 \geq s_2 \geq ... \geq s_n > 0$.
- **Έξοδος:** Ένας ελάχιστος αριθμός κάδων μοναδιαίας χωρητικότητας στους οποίους μπορούν να τοποθετηθούν όλα τα αντικείμενα. (δηλαδή μια διαμέριση του $\{1,2,...,n\}$ σε k σύνολα $B_1,...,B_k$ τέτοια ώστε $\forall j \in [k]\left(\sum_{i \in B_j} s_i \leq 1\right)$ και το k είναι το ελάχιστο δυνατό).

Δοθέντος ενός στιγμιότυπου I του προβλήματος, ορίζουμε το $SIZE(I) \doteq \sum_{i=1}^{n} s_i$, το οποίο αποτελεί ένα προφανές κάτω φράγμα για το μέγεθος της βέλτιστης λύσης, δηλαδή ισχύει $SIZE(I) \leq OPT(I)$.

3.8.1 Ένα κάτω φράγμα

Στην παράγραφο αυτή δείχνουμε μια σύνδεση ανάμεσα στο πρόβλημα βελτιστοποίησης bin packing και στο $NP-complete$ πρόβλημα (απόφασης) Partition το οποίο ορίζουμε ακολούθως:

Partition

- **Είσοδος:** Ένα σύνολο από αντικείμενα $\{1,2,...,n\}$. Το αντικείμενο i έχει μέγεθος $s_i \in \mathbb{Q} \cap (0,1]$. Χ. β. τ. γ. υποθέτουμε ότι $1 \geq s_1 \geq s_2 \geq ... \geq s_n > 0$.
- **Ερώτηση:** Μπορούμε να διαμερίσουμε το σύνολο $\{1,2,...,n\}$ σε δύο σύνολα A και B τέτοια ώστε $\sum_{i \in A} s_i = \sum_{i \in B} s_i$;

Θεώρημα 3.8.1 Δεν υπάρχει a-προσεγγιστικός αλγόριθμος για το bin packing με $a < \frac{3}{2}$, με την υπόθεση ότι $P \neq NP$.

Απόδειξη: Δίνουμε μια gap – introducing αναγωγή από το $NP-complete$ πρόβλημα Partition στο bin packing με gap ίσο με $3/2$. Έστω ότι υπάρχει ένας a - προσεγγιστικός (πολυωνυμικός) αλγόριθμος για το bin packing με $a < \frac{3}{2}$.

Θεωρούμε ένα στιγμιότυπο I του Partition. Μπορούμε εύκολα με scaling στο μέγεθος των αντικειμένων να πάρουμε ένα νέο σύνολο n αντικειμένων με $\sum_{i=1}^{n} s_i^{'} = 2$ και να το θεωρήσουμε στιγμιότυπο I' του bin packing. Είναι φανερό ότι $OPT(I') \geq 2$. Είναι επίσης εύκολο να δει κανείς ότι υπάρχει Partition για το στιγμιότυπο I (δηλ. το I είναι *"yes instance"*) αν και μόνο αν $OPT(I') = 2$.

Έστω ότι $OPT(I') = 2$. Τότε ο προσεγγιστικός αλγόριθμος για το στιγμιότυπο I' θα δώσει μια λύση που χρησιμοποιεί το πολύ $2a < 3$ κάδους, δηλαδή ακριβώς 2 κάδους. Διαφορετικά θα πάρουμε μία λύση με τουλάχιστον 3 κάδους. Συνεπώς, ο προσεγγιστικός αλγόριθμος μπορεί να χρησιμοποιηθεί σαν μαντείο για να λύσουμε σε πολυωνυμικό χρόνο το πρόβλημα Partition (άτοπο, υποθέτοντας $P \neq NP$).

□

Ωστόσο, η κατάσταση δεν είναι και τόσο άσχημη. Έχει αποδειχθεί [CKST] ότι το εν λόγω πρόβλημα δεν ανήκει στα πιο δύσκολα προβλήματα της κλάσης APX και μάλιστα ότι δεν είναι $APX-complete$ (με βάση κάποια approximation preserving αναγωγή), εκτός κι αν καταρρέει η πολυωνυμική ιεραρχία.

Η φαινομενική αντίφαση μεταξύ των παραπάνω ερμηνεύεται ως εξής: Το εν λόγω πρόβλημα ανήκει σε εκείνη την κατηγορία προβλημάτων για τα οποία το μοντέλο του απόλυτου λόγου προσέγγισης δεν είναι ικανοποιητικό. Στην πραγματικότητα, υπάρχουν πολύ καλοί προσεγγιστικοί αλγόριθμοι για το πρόβλημα, αλλά περιλαμβάνουν και κάποιο προσθετικό όρο σφάλματος.

Ο διαισθητικά πιο προφανής αλγόριθμος είναι ο εξής: σε κάθε βήμα ελέγχουμε σειριακά όλους τους κάδους που έχουμε ανοίξει και τοποθετούμε το τρέχον αντικείμενο στον πρώτο κάδο που χωράει. Αυτός είναι ο αλγόριθμος του πρώτου ταιριάσματος (first fit - *FF*) και είναι εύκολο να δει κανείς ότι $\forall I \; FF(I) \leq 2SIZE(I) + 1 \leq 2OPT(I) + 1$ (αρκεί να παρατηρήσουμε ότι μπορεί να υπάρχει το πολύ ένας κάδος μισογεμάτος).

Αναφέρουμε επίσης ότι υπάρχει ασυμπτωτικό προσεγγιστικό σχήμα πολυωνυμικού χρόνου (asymptotic *PTAS*) για το πρόβλημα, όπως αναφέρει το επόμενο θεώρημα:

Θεώρημα 3.8.2 [FL81] Υπάρχει πολυωνυμικός αλγόριθμος FL για το πρόβλημα τέτοιος ώστε

$$FL(I) \leq (1+\epsilon)OPT(I) + 1$$

για κάθε στιγμιότυπο I και κάθε $\epsilon > 0$. Η χρονική πολυπλοκότητα του αλγορίθμου είναι $O\left(n^{4/\epsilon^2}\right)$.

Ο βασικός αλγόριθμος που θα εξετάσουμε αναλυτικά οφείλεται στους Karmarkar και Karp.

Θεώρημα 3.8.3 [KK82] Υπάρχει πολυωνυμικός αλγόριθμος KK τέτοιος ώστε $KK(I) \leq OPT(I) + O\left(\log^2\left(OPT(I)\right)\right)$, για κάθε στιγμιότυπο I.

Στην επόμενη ενότητα αναλύουμε αυτό τον αλγόριθμο.

3.8.2 Ο αλγόριθμος των Karmarkar και Karp

Σύμφωνα με την γενική προσέγγιση σχεδίασης προσεγγιστικών αλγορίθμων που εισάγαμε νωρίτερα στο κεφάλαιο έχουμε:

Βήμα 1

Μοντελοποιούμε το πρόβλημα ως ένα ακέραιο γραμμικό πρόγραμμα. Έστω ότι ένα στιγμιότυπο I αποτελείται από αντικείμενα m (το πλήθος) διακεκριμένων μεγεθών και ότι υπάρχουν $n_i \in \mathbb{N}^*$ αντικείμενα μεγέθους s_i ($i \in [m]$). Σε πρώτη φάση δίνουμε έναν ορισμό απαραίτητο για την κατασκευή του ακέραιου προγράμματος.

Ορισμός 3.8.1 Θεωρούμε τη διατεταγμένη $m-$άδα $(a_1,...,a_m)$, όπου $a_i \in \mathbb{N}$. Με $(a_1,...,a_m)$ συμβολίζουμε το σύνολο που περιέχει a_i το πλήθος αντικείμενα μεγέθους s_i. Το σύνολο $(a_1,...,a_m)$ ονομάζεται κατάσταση (*configuration*) αν «χωράει» σε ένα κάδο, δηλ. αν $\sum_i a_i s_i \le 1$.

Έστω N το πλήθος όλων των δυνατών configurations. Ένα χονδρικό άνω φράγμα για το N είναι το $O(n^m)$. Έστω επίσης $A_1, A_2, ..., A_N$ μια απαρίθμησή τους. Με a_{ij} συμβολίζουμε την $i-$οστή συνιστώσα του A_j, δηλαδή $A_j = (a_{1j},...,a_{mj})$.

Για κάθε configuration A_j, εισάγουμε μία μεταβλητή x_j η οποία συμβολίζει *το πλήθος των κάδων που έχουν ως κατάσταση την* A_j. Σύμφωνα με τα προηγούμενα, το αντίστοιχο ακέραιο πρόγραμμα (*CIP*) είναι το ακόλουθο:

$$\min \sum_{j=1}^{N} x_j$$

με τους περιορισμούς:

$$\forall i \in [m]\left(\sum_{j=1}^{N} a_{ij} x_j \ge n_i \right)$$

$$\forall j \in [N]\left(x_j \in \mathbb{N} \right)$$

Η πρώτη κατηγορία περιορισμών φανερά εκφράζει την απαίτηση κάθε αντικείμενο να έχει τοποθετηθεί σε κάποιο κάδο. Μάλιστα οι εν λόγω περιορισμοί ισχύουν με ισότητα.

Βήμα 2

Χαλαρώνοντας τις απαιτήσεις για ακέραια x_j σε $x_j \ge 0$ προκύπτει το αντίστοιχο χαλαρωμένο γραμμικό πρόγραμμα (*LP*-relaxation). Μπορεί να παρατηρήσει κανείς ότι το χαλαρωμένο γραμμικό πρόγραμμα έχει μεγάλο (εν δυνάμει εκθετικό) αριθμό μεταβλητών. Ωστόσο, μπορεί να λυθεί αποδοτικά, σύμφωνα με το επόμενο θεώρημα το οποίο διατυπώνεται χωρίς απόδειξη.

Θεώρημα 3.8.4 Το παραπάνω γραμμικό πρόγραμμα μπορεί να λυθεί με προσθετικό σφάλμα το πολύ 1 σε πολυωνυμικό χρόνο ως προς το m και το $\log\left(\frac{n}{s_m}\right)$, όπου s_m είναι το μικρότερο μέγεθος αντικειμένου στο στιγμιότυπο του προβλήματος.

Η ακολουθούμενη διαδικασία είναι η εξής:

1. Χωρίζουμε τα αντικείμενα σε «μεγάλα» (μεγέθους $\geq \epsilon_{\min} := \frac{1}{SIZE(I)}$) και «μικρά».
2. Χρησιμοποιούμε *κάποιο* αλγόριθμο για να πακετάρουμε τα μεγάλα αντικείμενα σε b το πολύ κάδους.
3. Χρησιμοποιούμε τον αλγόριθμο *FF* για να πακετάρουμε τα μικρά αντικείμενα (πιθανώς ανοίγοντάς νέους κάδους).

Στο βήμα 3 παραπάνω, αν ο αλγόριθμος του πρώτου ταιριάσματος δεν ανοίξει κανένα καινούριο κάδο, τότε (ο αλγόριθμος) χρησιμοποιεί συνολικά b το πολύ κάδους, και αν το b είναι κοντά στο βέλτιστο αριθμό κάδων για τα μεγάλα αντικείμενα *μόνο*, είναι κοντά στο βέλτιστο και για το πλήρες στιγμιότυπο (με όλα τα αντικείμενα). Αν πάλι ο *FF* ανοίξει νέους κάδους, είναι εύκολο να δει κανείς ότι (τελικά) όλοι οι κάδοι εκτός ενδεχομένως από έναν πρέπει να περιέχουν αντικείμενα συνολικού βάρους τουλάχιστον $1-\epsilon_{\min}$. Άρα ισχύει ότι

$$(1-\epsilon_{\min})\left(FF(I)-1\right) \leq SIZE(I)$$

από όπου παίρνουμε διαδοχικά

$$\begin{aligned} FF(I) &\leq \frac{1}{1-\epsilon_{\min}} SIZE(I)+1 \\ &\leq (1+2\epsilon_{\min}) SIZE(I)+1 \\ &\leq \left(1+\frac{2}{SIZE(I)}\right) SIZE(I)+1 \\ &\leq SIZE(I)+3 \\ &\leq OPT(I)+3 \end{aligned}$$

Η δεύτερη ανισότητα έπεται από το ότι για κάθε $x \in [0,\ 1/2]$ ισχύει $\frac{1}{1-x} \leq 1+2x$. Για να ισχύουν τα παραπάνω λοιπόν πρέπει $SIZE(I) \geq 2$.

Μπορούμε λοιπόν να εξασφαλίσουμε ότι $s_m \geq \frac{1}{SIZE(I)}$, προσθέτοντας στο τέλος σε κάδους όλα τα αντικείμενα μικρότερου μεγέθους με χρήση του αλγορίθμου του πρώτου ταιριάσματος. Αν ο *F.F.* ανοίξει καινούριο κάδο, θα πάρουμε μια λύση που χρησιμοποιεί το πολύ $OPT(I)+3$ κάδους.

Εφόσον λοιπόν $s_m \geq \frac{1}{SIZE(I)}$ από το προηγούμενο θεώρημα έπεται ότι το εν λόγω γραμμικό πρόγραμμα μπορεί να λυθεί σε πολυωνυμικό χρόνο (ως προς το μέγεθος της εισόδου του προβλήματος bin packing).

Βήμα 3

Παρόλο που ο αριθμός των μεταβλητών στο γραμμικό πρόγραμμα είναι πολύ μεγάλος, μπορεί εύκολα να δει κανείς ότι οποιαδήποτε κορυφή του συνόλου των εφικτών λύσεων έχει το πολύ m μη μηδενικές συνιστώσες. Μπορούμε επίσης να υποθέσουμε ότι έχουμε τη δυνατότητα να βρούμε μια τέτοια κορυφή. Συνεπώς, αν στρογγυλοποιήσουμε *προς τα πάνω* όλες τις μη μηδενικές μεταβλητές σε μία βέλτιστη λύση του *LP*-relaxation, ο αριθμός των κάδων που θα χρησιμοποιήσουμε είναι το πολύ $OPT(I)+m$. Ωστόσο, είναι δυνατόν να ισχύει $m=n$, οπότε στην περίπτωση αυτή το αποτέλεσμα αυτό είναι τετριμμένο.

Για να πετύχουμε κάτι καλύτερο από το παραπάνω, εισάγουμε την ακόλουθη τεχνική, η οποία στρογγυλοποιεί *προς τα κάτω* τη λύση του χαλαρωμένου γραμμικού προγράμματος. Τονίζεται ότι η τεχνική αυτή αποτελεί επέκταση της απλής στρογγυλοποίησης.

1. Δοθείσης της βέλτιστης λύσης x^* του χαλαρωμένου γραμμικού προγράμματος που αντιστοιχεί στο στιγμιότυπο I, πακετάρουμε $\lfloor x_j^* \rfloor$ το πλήθος κάδους σύμφωνα με την configuration A_j. Ας συμβολίσουμε με I_{int} (integral) το σύνολο των αντικειμένων που πακετάρονται με αυτόν τον τρόπο και με $I \setminus I_{\text{int}}$ τα αντικείμενα που απομένουν.
2. Επαναλαμβάνουμε τη διαδικασία του βήματος 1 αναδρομικά για το στιγμιότυπο $I \setminus I_{\text{int}}$.

Για να αποδείξουμε ότι η παραπάνω διαδικασία περιγράφει έναν καλό προσεγγιστικό αλγόριθμο, πρέπει να δείξουμε πώς προχωράμε σε κάθε επανάληψη. Αν συμβολίσουμε με $LP(I)$ τη βέλτιστη τιμή του γραμμικού προγράμματος με στιγμιότυπο I, τότε ισχύει το ακόλουθο λήμμα.

Λήμμα 3.8.5 $LP(I \setminus I_{\text{int}}) + LP(I_{\text{int}}) \leq LP(I)$.

Απόδειξη: Γνωρίζουμε ότι το διάνυσμα $\lfloor x^* \rfloor$ είναι εφικτή λύση για το γραμμικό πρόγραμμα με στιγμιότυπο I_{int} και ότι το $\left(x^* - \lfloor x^* \rfloor\right)$ είναι εφικτή λύση με στιγμιότυπο $I \setminus I_{\text{int}}$. Επομένως,

$$LP(I \setminus I_{\text{int}}) + LP(I_{\text{int}}) \leq \sum_j \left(x_j^* - \lfloor x_j^* \rfloor\right) + \sum_j \lfloor x_j^* \rfloor = \sum_j x_j^* = LP(I).$$

□

Λήμμα 3.8.6 $SIZE(I \setminus I_{\text{int}}) \le m$

Απόδειξη: Επειδή το x^* είναι κορυφή του συνόλου των εφικτών λύσεων, το πολύ m συνιστώσες του είναι μη μηδενικές. Έτσι, το πολύ m συνιστώσες του $x^* - \lfloor x^* \rfloor$ είναι μη μηδενικές και επιπλέον καθεμία από αυτές είναι μικρότερη της μονάδας. Εφόσον το $x^* - \lfloor x^* \rfloor$ είναι εφικτή λύση για το LP του bin packing στιγμιότυπου $I \setminus I_{\text{int}}$, με βάση τα παραπάνω (περί στρογγυλοποίησης προς τα πάνω) μπορούμε να πακετάρουμε το συγκεκριμένο στιγμιότυπο σε m κάδους. Συνεπώς, το συνολικό μέγεθος του $I \setminus I_{\text{int}}$ δεν μπορεί να είναι περισσότερο από m.

□

Το παραπάνω αποτέλεσμα δείχνει ότι αν ισχύει $SIZE(I) > m$ η αναδρομική διαδικασία προχωράει, δηλ. το μέγεθος της εισόδου φθίνει γνήσια σε κάθε βήμα. Για να υπάρχει πρόοδος γενικά, χρησιμοποιήσουμε ένα νέο τύπο «γεωμετρικής ομαδοποίησης» κάθε φορά πριν λύσουμε το γραμμικό πρόγραμμα. Θα δείξουμε ότι η ομαδοποίηση αυτή έχει ως αποτέλεσμα να ισχύει $m = SIZE(I)/2$, οπότε, λόγω του παραπάνω λήμματος, το μέγεθος του instance θα πέφτει κατά ένα παράγοντα 2 κάθε φορά που κάνουμε αναδρομή στο υπολειπόμενο («κλασματικό») τμήμα του instance.

Για να πραγματοποιήσουμε τη *γεωμετρική ομαδοποίηση*, ταξινομούμε τα αντικείμενα κατά φθίνουσα σειρά μεγέθους. Αρχίζοντας από το μεγαλύτερο μέγεθος προσθέτουμε αντικείμενα σε μία ομάδα μέχρι το μέγεθος της ομάδας να γίνει τουλάχιστον 2. Στη συνέχεια, δημιουργούμε την επόμενη ομάδα και ούτω καθεξής (μέχρις ότου όλα τα αντικείμενα να έχουν τοποθετηθεί σε μία ομάδα). Ας συμβολίσουμε τις ομάδες που προκύπτουν $G_1, G_2, ..., G_r$, όπου η ομάδα G_i αποτελείται από p_i αντικείμενα.

Πρόταση 3.8.7 Για κάθε $i \in \{2, ..., r-1\}$ ισχύει $p_i \ge p_{i-1}$.

Απόδειξη: Τα αντικείμενα που ανήκουν στην ομάδα G_{i-1} έχουν *τουλάχιστον* το ίδιο μέγεθος με οποιοδήποτε αντικείμενο της ομάδας G_i.

□

Κατασκευάζουμε ένα νέο στιγμιότυπο ομαδοποίησης I' από το I ως εξής:

Κατασκευή του I' από το I

Απορρίπτουμε τις ομάδες G_1 και G_r.

$\forall i \in \{2, ..., r-1\}$

- Απορρίπτουμε τα $p_i - p_{i-1}$ μικρότερα αντικείμενα της ομάδας G_i.
- Στρογγυλοποιούμε τα εναπομείναντα p_{i-1} το πλήθος αντικείμενα του G_i στο μέγεθος του μεγαλύτερου αντικειμένου της εν λόγω ομάδας.

Λήμμα 3.8.8 $LP(I') \le LP(I)$

Απόδειξη: Αρκεί να παρατηρήσουμε ότι αφενός το I' αποτελείται από λιγότερα αντικείμενα και αφετέρου ότι κάθε αντικείμενο στο I' μπορεί να απεικονιστεί σε ένα μοναδικό αντικείμενο του I τουλάχιστον του ίδιου μεγέθους.

□

Μπορούμε τώρα να δώσουμε τον πλήρη αλγόριθμο:

Κ. Κ.

Έστω ότι το I_0 περιέχει μεγάλα αντικείμενα, δηλαδή μεγέθους $\ge \frac{1}{SIZE(I)}$

$k \leftarrow 0$

while $SIZE(I_k) \ge 1$

Κάνε γεωμετρική ομαδοποίηση για να πάρεις το $I_k^{'}$.

Πακετάρισε τα αντικείμενα που απορρίφθηκαν από το grouping χρησιμοποιώντας *FF*.

Επίλυσε το *LP* που αντιστοιχεί στο I'_k : πακετάρισε το «ακέραιο» μέρος $I_{k,\text{int}}^{'}$ χρησιμοποιώντας τη λύση του *LP*.

$I_{k+1} \leftarrow I_k^{'} \setminus I_{k,\text{int}}^{'}$

$k \leftarrow k+1$

Πακετάρισε το υπόλοιπο I_k σε 1 κάδο.

Χρησιμοποίησε τον αλγόριθμο First Fit για να πακετάρεις τα μικρά αντικείμενα.

Παρατηρούμε ότι το πλήθος των ομάδων στο I' είναι το πολύ $\frac{SIZE(I)}{2}$ εφόσον κάθε ομάδα έχει μέγεθος τουλάχιστον 2. Στο I' όλα τα αντικείμενα που ανήκουν στο ίδιο group έχουν το ίδιο μέγεθος. Άρα,

$$\text{αριθμός διαφορετικών μεγεθών στο } I' \le \frac{SIZE(I)}{2}$$

και από το λήμμα 4 έπεται ότι

$$SIZE(I' \setminus I'_{\text{int}}) \le \frac{SIZE(I)}{2}.$$

Επομένως, το μέγεθος ενός instance μειώνεται κατά ένα παράγοντα 2 μετά την ομαδοποίηση, την επίλυση του γραμμικού προγράμματος, την στρογγυλοποίηση προς τα κάτω και την αναδρομή στο εναπομείναν τμήμα. Ο αλγόριθμος σταματάει όταν το $SIZE(I)$ γίνει μικρότερο από 1. Άρα, μπορούν να γίνουν το πολύ $O(\log(SIZE(I)))$ επαναλήψεις του αλγορίθμου.

Για να ξεκινήσουμε την ανάλυση του αλγορίθμου, πρέπει να φράξουμε το μέγεθος των απορριπτόμενων αντικειμένων κατά την ομαδοποίηση. Δείχνουμε παρακάτω ότι το συνολικό μέγεθος των απορριπτόμενων είναι το πολύ $O\left(\log\left(SIZE(I)\right)\right)$.

Λήμμα 3.8.9 Το συνολικό μέγεθος των αντικειμένων που απορρίπτονται κατά τη διαδικασία της γεωμετρικής ομαδοποίησης είναι $O\left(\log\left(SIZE(I)\right)\right)$.

Απόδειξη: Προφανώς ισχύει ότι $SIZE(G_1), SIZE(G_r) \leq 3$, εφόσον σταματάμε να προσθέτουμε στοιχεία σε μια ομάδα όταν το μέγεθός της φτάσει ή ξεπεράσει το 2 και το μέγεθος κάθε αντικειμένου είναι το πολύ 1.

Είναι επίσης εύκολο να δει κανείς ότι το συνολικό μέγεθος των $p_i - 1$ μεγαλύτερων αντικειμένων της ομάδας G_i είναι το πολύ 2, εφόσον πριν την προσθήκη του τελευταίου (και μεγαλύτερου) αντικειμένου, το συνολικό μέγεθος του G_i δεν ξεπερνούσε το 2. Χρησιμοποιώντας το επιχείρημα ότι ο αριθμητικός μέσος είναι το πολύ ίσος με τη μέγιστη τιμή παίρνουμε ότι

$$\text{μέγεθος των } p_i - p_{i\text{-}1} \text{ μικρότερων αντικειμένων στο } G_i \leq \frac{2}{p_i - 1}\left(p_i - p_{i\text{-}1}\right)$$

Συνεπώς,

$$\begin{aligned}
\text{συνολικό μέγεθος απορρίψεων} &\leq 6 + \sum_{i=2}^{r-1} \frac{2}{p_i - 1}\left(p_i - p_{i-1}\right) \\
&\leq 6 + 2\sum_{i=2}^{r-1}\left[\frac{1}{p_i - 1} + \frac{1}{p_i - 2} + \ldots + \frac{1}{p_i - \left(p_i - p_{i-1}\right)}\right] \\
&\qquad \left(\text{διότι για κάθε } k \geq 1 \ \ \frac{1}{p_i - 1} \leq \frac{1}{p_i - k}\right) \\
&= 6 + 2\sum_{j=p_1}^{p_{r-1}-1} \frac{1}{j} \\
&\leq 6 + 2H_{p_{r-1}}
\end{aligned}$$

Εφόσον κάθε αντικείμενο έχει μέγεθος $\geq \frac{1}{SIZE(I)}$ και κάθε ομάδα έχει μέγεθος το πολύ 3, ο αριθμός των αντικειμένων σε κάθε ομάδα είναι το πολύ $3SIZE(I)$. Συνεπώς, $p_{r-1} \leq 3SIZE(I)$, οπότε

$$\begin{aligned}
\text{συνολικό μέγεθος απορρίψεων} &\leq 6 + 2H_{p_{r-1}} \\
&= O\left(\log\left(p_{r-1}\right)\right) \\
&= O\left(\log\left(SIZE(I)\right)\right)
\end{aligned}$$

□

Στο σημείο αυτό ολοκληρώνουμε την ανάλυση του αλγορίθμου. Προφανώς απαιτούνται $O\left(\log^2\left(SIZE(I)\right)\right)$ κάδοι για το πακετάρισμα των απορριφθέντων αντικειμένων ($O\left(\log\left(SIZE(I)\right)\right)$ επαναλήψεις με $O\left(\log\left(SIZE(I)\right)\right)$ κάδους να χρησιμοποιούνται σε κάθε επανάληψη). Το υπόλοιπο της ανάλυσης δείχνει ότι με την στρογγυλοποίηση προς τα κάτω της λύσης του *LP* κάθε φορά, δεν χάνουμε τίποτα συνολικά: ο συνολικός αριθμός των κάδων που χρησιμοποιούνται για το πακετάρισμα του ακέραιου μέρους κάθε φορά δεν είναι μεγαλύτερος από την αρχική τιμή του *LP*, που αποτελεί ένα κάτω φράγμα για το $OPT(I)$.

Θεώρημα 3.8.10 Ισχύει $KK(I) \le OPT(I) + O\left(\log^2\left(SIZE(I)\right)\right)$ για κάθε στιγμιότυπο του bin packing.

Απόδειξη: Χρησιμοποιώντας το λήμμα 3.8.7 και το γεγονός ότι γίνονται το πολύ $O\left(\log\left(SIZE(I)\right)\right)$ επαναλήψεις, παίρνουμε:

$$
\begin{aligned}
KK(I) &= O\left(\log^2\left(SIZE(I)\right)\right) + \sum_i LP\left(I'_{i,\mathrm{int}}\right) \\
&\le O\left(\log^2\left(SIZE(I)\right)\right) + \sum_i \left(LP\left(I'_i\right) - LP\left(I'_i - I'_{\mathrm{int}}\right)\right) \\
&= O\left(\log^2\left(SIZE(I)\right)\right) + \sum_i \left(LP\left(I'_i\right) - LP\left(I_{i+1}\right)\right) \\
&\le O\left(\log^2\left(SIZE(I)\right)\right) + \sum_i \left(LP\left(I_i\right) - LP\left(I_{i+1}\right)\right) \\
&= O\left(\log^2\left(SIZE(I)\right)\right) + LP\left(I_0\right) \\
&\le O\left(\log^2\left(SIZE(I)\right)\right) + OPT(I)
\end{aligned}
$$

Η πρώτη ανισότητα προκύπτει από το λήμμα 3.8.5 και η δεύτερη από το λήμμα 3.8.7.

□

Κεφάλαιο 4

Randomized Στρογγυλοποίηση

Το κεφάλαιο αυτό επικεντρώνεται στην προσέγγιση της *randomized στρογγυλοποίησης* για τη σχεδίαση προσεγγιστικών αλγορίθμων. Υπενθυμίζεται η κλασική έννοια της χαλάρωσης (relaxation) ενός προβλήματος βελτιστοποίησης $\mathcal{P}$: δοθέντος ενός στιγμιότυπου I του $\mathcal{P}$, *επεκτείνουμε* το σύνολο των εφικτών λύσεων για το I με τέτοιο τρόπο ώστε η αντικειμενική συνάρτηση να μπορεί να βελτιστοποιηθεί αποδοτικά για το εκτεταμένο σύνολο.

Ας συμβολίσουμε με x^* μία (αποδοτικά υπολογίσιμη) βέλτιστη λύση για το πρόβλημα αυτό. Η τεχνική της randomized στρογγυλοποίησης συνίσταται στη χρήση του randomization για την απεικόνιση του x^* πίσω σε μία λύση που είναι εφικτή για το I. Αν με y^* συμβολίσουμε την (βέλτιστη) τιμή της αντικειμενικής συνάρτησης για το x^*, είναι φανερό ότι ισχύει $y^* \geq OPT(I)$ (αντίστοιχα $y^* \leq OPT(I)$) για προβλήματα μεγιστοποίησης (αντίστοιχα ελαχιστοποίησης).

Έστω λοιπόν ότι το πρόβλημά μας $\mathcal{P}$ είναι πρόβλημα μεγιστοποίησης και ότι μπορούμε να αναλύσουμε τη διαδικασία της randomized στρογγυλοποίησης (που εφαρμόζουμε για το πρόβλημα) για να δείξουμε ότι οδηγεί σε μία εφικτή λύση με τιμή της αντικειμενικής συνάρτησης τουλάχιστον y^*/λ (κατά μέση τιμή ή με μεγάλη πιθανότητα). Συνεπώς, εφόσον $y^* \geq OPT(I)$, το αποτέλεσμα είναι ένας λ - προσεγγιστικός αλγόριθμος (κατά μέση τιμή ή με μεγάλη πιθανότητα). Ανάλογα ισχύουν και για προβλήματα ελαχιστοποίησης.

4.1 Φράγματα μεγάλης απόκλισης

Στο σημείο αυτό παρουσιάζουμε μια βασική κατηγορία αποτελεσμάτων τα οποία είναι πολύ χρήσιμα στους randomized υπολογισμούς. Πρόκειται για (άνω) φράγματα στην πιθανότητα ορισμένες κατηγορίες τυχαίων μεταβλητών να αποκλίνουν σημαντικά από τη μέση τιμή τους, και ως εκ τούτου είναι γνωστά ως *άνω φράγματα μεγάλης απόκλισης* (*large deviation bounds* ή *tail probability bounds*). Θα παρουσιάσουμε μόνο τα αποτελέσματα εκείνα που σχετίζονται με τους randomized αλγόριθμους.

Για το υπόλοιπό αυτής της ενότητας, συμβολίζουμε με μ τη μέση τιμή $\mathbf{E}[X]$ μιας τ.μ. X.

Λήμμα 4.1.1 (ανισότητα Markov) Αν μια τ.μ. X παίρνει μόνο μη αρνητικές τιμές, τότε για κάθε $a > 0$ ισχύει $\Pr(X \geq a) \leq \mu/a$.

Η *ανισότητα του Markov* είναι καλύτερη δυνατή αν δε δίνεται καμία επιπλέον πληροφορία για την τ.μ. X. Ωστόσο, είναι συχνά πολύ αδύναμη, π.χ. όταν θέλουμε να φράξουμε άνω την πιθανότητα η τ.μ. X να πάρει τιμές πολύ μικρότερες από τη μέση τιμή της (*lower tail* probability) ακόμα κι αν γνωρίζουμε καλά άνω και κάτω φράγματα για το εύρος που το X είναι μη μηδενικό.

Περισσότερη πληροφορία για την εν λόγω τ.μ. X μπορεί να οδηγήσει σε σημαντικά καλύτερα φράγματα, όπως θα δούμε στη συνέχεια. Αξίζει να σημειωθεί ότι πίσω από την προσέγγιση που οδηγεί σε πολλά από αυτά τα (βελτιωμένα) φράγματα είναι κρυμμένη η ανισότητα του Markov.

Μια πολύ γνωστή ισχυρότερη ανισότητα είναι η *ανισότητα του Chebysev*, η οποία μπορεί να εφαρμοστεί αν έχουμε ένα καλό άνω φράγμα για τη *μεταβλητότητα* (variance) $V[\cdot]$ της τ.μ. X. Υπενθυμίζεται ότι (εξ ορισμού) $V[X] \doteq \mathbf{E}\left[(X-\mu)^2\right]$ και ότι η θετική τετραγωνική ρίζα της μεταβλητότητας του X ονομάζεται *τυπική απόκλιση* (*standard deviation*) του X.

Λήμμα 4.1.2 (ανισότητα Chebysev) Για κάθε τ.μ. X και για κάθε $a > 0$, $\Pr\left(|X-\mu| \geq a\right) \leq \sigma^2/a^2$, όπου σ είναι η τυπική απόκλιση του X.

Αν το X είναι το άθροισμα ανεξάρτητων τυχαίων μεταβλητών, καθεμία εκ των οποίων παίρνει τιμές στο $[0,\ 1]$, δεν είναι δύσκολο να επιβεβαιώσει κανείς ότι $\mathbf{E}\left[(X-\mu)^2\right] \leq \mathbf{E}[X]$. Επομένως, παίρνουμε το εξής:

Πόρισμα 4.1.3 Έστω X το άθροισμα ανεξάρτητων τυχαίων μεταβλητών, καθεμία εκ των οποίων παίρνει τιμές στο $[0,\ 1]$. Τότε για κάθε $a > 0$, $\Pr\left(|X-\mu| \geq a\right) \leq \mu/a^2$.

Συνεπώς, τυχαίες μεταβλητές με μικρή τυπική απόκλιση έχουν καλή οριακή συμπεριφορά. Το αποτέλεσμα αυτό μπορεί να ενισχυθεί πολύ περισσότερο για μια κατηγορία τυχαίων μεταβλητών που εμφανίζονται πολύ συχνά στους randomized υπολογισμούς: αθροίσματα ανεξάρτητων τ.μ. X_i, όπου κάθε $X_i \in [0,\ 1]$. Η ιδέα είναι να παρατηρήσει κανείς ότι

$$\Pr(X \geq a) = \Pr\left(e^{tX} \geq e^{ta}\right) \leq E\left[e^{tX}\right]/e^{ta}$$

για κάθε $t > 0$, όπου το τελευταίο βήμα έπεται από την ανισότητα του Markov. Μπορούμε τότε να ελαχιστοποιήσουμε (έστω προσεγγιστικά) τον τελευταίο λόγος ως προς t, για να πάρουμε ισχυρότερα φράγματα. Σε γενικές γραμμές με την προαναφερθείσα προσέγγιση παίρνουμε το ακόλουθο αποτέλεσμα που προκύπτει ως συνδυασμός αποτελεσμάτων των Chernoff και Hoeffding (θα το συμβολίζουμε ως φράγμα *CH*).

Λήμμα 4.1.4 (Chernoff, Hoeffding) Έστω $X_1, X_2, ..., X_n$ ανεξάρτητες τ.μ. (όχι κατ' ανάγκη με την ίδια κατανομή), καθεμία εκ των οποίων παίρνει τιμές στο $[0,\ 1]$ με $X = \sum_{i=1}^{n} X_i$. Τότε για $\mu = \mathbf{E}[X]$ και για κάθε $\delta \geq 0$ έχουμε

$$\Pr\left(X \geq \mu(1+\delta)\right) \leq G(\mu,\delta) \doteq \left(\exp(\delta) \Big/ (1+\delta)^{1+\delta} \right)^{\mu}$$

και αν $0 \leq \delta < 1$

$$\Pr\left(X \le \mu(1-\delta)\right) \le H(\mu,\delta) \doteq \exp\left(-\mu\delta^2/2\right)$$

Εύκολα μπορεί κανείς να δει ότι ισχύουν τα ακόλουθα:

Πρόταση 4.1.5 Ισχύουν τα ακόλουθα:

(a) $G(\mu,\delta) \le \left(e/(1+\delta)\right)^{(1+\delta)\mu}$.

(b) $G(\mu,\delta) \le \exp\left(-\delta^2\mu/3\right)$, αν $\delta \le 1$.

(c) $G(\mu,\delta) \le \exp\left(-(1+\delta)\ln(1+\delta)\mu/4\right)$, αν $\delta \ge 1$.

(d) Αν $0 < \mu_1 \le \mu_2$,τότε $G(\mu_1,\delta) \ge G(\mu_2,\delta)$.

Στις περισσότερες περιπτώσεις που μας ενδιαφέρουν το X είναι άθροισμα τ.μ. X_i με τιμές στο $\{0,1\}$, δηλαδή οι μεταβλητές $X_1, X_2, ..., X_n$ είναι ανεξάρτητες δοκιμές Poisson με $\Pr(X_i = 1) = p_i$ και $\Pr(X_i = 0) = 1 - p_i$. Τα παραπάνω φράγματα απαντούν στο ερώτημα «ποια είναι η πιθανότητα το X να υπερβαίνει το $\mu(1+\delta)$;» και είναι χρήσιμα στην *ανάλυση* ενός (randomized) αλγορίθμου, συγκεκριμένα στο να δείξει κανείς ότι η πιθανότητα ο αλγόριθμος να αποτύχει στην επίτευξη συγκεκριμένης απόδοσης είναι μικρή.

Στη συνέχεια αναλύουμε μία εφαρμογή της υπό ανάλυση τεχνικής όπως αυτή ορίζεται στο [RagTh87] για την παραγωγή προσεγγιστικών αλγορίθμων για γενικά ακέραια προγράμματα. Ακολουθούμε την ανάλυση του [Rag88].

4.2 Εφαρμογή τη Randomized στρογγυλοποίησης για ακέραια προγράμματα τύπου PIP

Δοθέντος ενός θετικού ρητού αριθμού v, η ιδέα είναι να κοιτάμε το κλασματικό του μέρος ως πιθανότητα. Στρογγυλοποιούμε το v είτε στην τιμή $\lfloor v \rfloor + 1$ με πιθανότητα $v - \lfloor v \rfloor$, είτε στην τιμή $\lfloor v \rfloor$ με πιθανότητα $1 - (v - \lfloor v \rfloor)$. Η διαδικασία αυτή έχει την πολύ καλή ιδιότητα ότι η μέση τιμή του αποτελέσματος (που είναι φυσικά τυχαία μεταβλητή) ισούται με v.

Η ιδέα αυτή μπορεί να χρησιμοποιηθεί για ακέραια γραμμικά προγράμματα τύπου *PIP* (βλέπε ορισμό στο Κεφάλαιο 3) ως εξής: Λύνουμε το χαλαρωμένο γραμμικό πρόγραμμα και έστω x^* η υπολογιζόμενη βέλτιστη (κλασματική) λύση. Ακολούθως θέτουμε $x_i' := x_i^*/a$, για κάποια παράμετρο $a > 1$ που θα σταθεροποιηθεί αργότερα. Αυτή η κλιμάκωση προς τα κάτω κατά a γίνεται με σκοπό να αυξήσουμε την πιθανότητα όλοι οι περιορισμοί του *PIP* να ικανοποιούνται. Υπενθυμίζεται πρόκειται για περιορισμούς $\le$ και ότι όλα τα στοιχεία του πίνακα A είναι μη αρνητικά. Επομένως, ενδεχόμενη εφαρμογή της ακόλουθης διαδικασίας απευθείας στα x_i^* θα μπορούσε να παραβιάσει κάποιους από αυτούς τους περιορισμούς με πιθανότητα πολύ κοντά στο 1.

Τώρα ορίζουμε ένα τυχαίο διάνυσμα $z \in \mathbb{N}^n$, το οποίο αποτελεί το αποτέλεσμα της διαδικασίας του randomized rounding, ως ακολούθως. Οι τυχαίες μεταβλητές z_i είναι *ανεξάρτητες* μεταξύ τους και επιπλέον

$$z_i \in \left\{ \left\lfloor x_i' \right\rfloor, \left\lfloor x_i' \right\rfloor + 1 \right\} \text{ με}$$

$$\Pr\left(z_i = \left\lfloor x_i' \right\rfloor + 1\right) = x_i - \left\lfloor x_i' \right\rfloor \text{ και } \Pr\left(z_i = \left\lfloor x_i' \right\rfloor\right) = 1 - \left(x_i - \left\lfloor x_i' \right\rfloor\right)$$

Τώρα χρειάζεται να αποδείξουμε ότι όλοι οι περιορισμοί του *PIP* ικανοποιούνται *και* ότι η τιμή $c^T \cdot z$ της αντικειμενικής συνάρτησης δεν είναι «πολύ μικρότερη» από τη βέλτιστη ($y^* = c^T \cdot x^*$).

Όπως είδαμε παραπάνω μια σημαντική παρατήρηση είναι ότι $\forall i \in [n] \left(\mathbf{E}[z_i] = x_i' \right)$. Πράγματι,

$$\begin{aligned} \mathbf{E}[z_i] &= \left\lfloor x_i' \right\rfloor \cdot \left(1 - x_i' + \left\lfloor x_i' \right\rfloor\right) + \left(\left\lfloor x_i' \right\rfloor + 1\right) \cdot \left(x_i' - \left\lfloor x_i' \right\rfloor\right) \\ &= \cancel{\left\lfloor x_i' \right\rfloor} - \cancel{x_i' \cdot \left\lfloor x_i' \right\rfloor} + \cancel{\left\lfloor x_i' \right\rfloor^2} + \cancel{\left\lfloor x_i' \right\rfloor \cdot x_i'} - \cancel{\left\lfloor x_i' \right\rfloor^2} + x_i' - \cancel{\left\lfloor x_i' \right\rfloor} \\ &= x_i' \end{aligned}$$

Επομένως, λόγω της γραμμικότητας της μέσης τιμής παίρνουμε ότι

$$\forall i \in [n] \; \mathbf{E}\left[(Az)_i\right] = (Ax')_i \le b_i / a \text{ και } \mathbf{E}\left[c^T \cdot z\right] = y^* / a$$

Για κάποια παράμετρο $\beta > 1$ που επίσης θα σταθεροποιηθεί αργότερα, ορίζουμε $m+1$ το πλήθος «δυσάρεστα» ενδεχόμενα:

$$\forall i \in [m] \; E_i \equiv \left((Az)_i > b_i\right) \text{ και } E_{m+1} \equiv \left(c^T \cdot z < y^* / (a \cdot \beta)\right)$$

Τώρα, η λύση z είναι $(a\beta)$ - προσεγγιστική λύση για το *PIP* αν ισχύει

$$\bigwedge_{i=1}^{m+1} \overline{E_i} \ne \varnothing$$

Το ερώτημα που τίθεται είναι πόσο μικρή τιμή για το $(a\beta)$ μπορούμε να επιτύχουμε ώστε να ισχύει η παραπάνω σχέση.

Με βάση την (προφανή) ανισότητα

$$\Pr\left(\bigvee_{i=1}^{m+1} E_i\right) \le \sum_{i=1}^{m+1} \Pr(E_i)$$

μπορούμε να επιλέξουμε $a, \beta > 1$ τέτοια ώστε $\sum_{i=1}^{m+1} \Pr(E_i) < 1$, χρησιμοποιώντας τα Chernoff – Hoeffding bounds.

Αυτό μας δίνει μία $(a\beta)$ - προσεγγιστική λύση με μη μηδενική πιθανότητα, η οποία μπορεί να μετατραπεί σε ντετερμινιστικό αλγόριθμο με χρήση των απαισιόδοξων εκτιμητριών (pessimistic estimators) [Rag88]. Παρόμοιες ιδέες ισχύουν και για ακέραια προγράμματα τύπου *CIP* – η κλασματική λύση x^* κλιμακώνεται *προς τα πάνω*.

Στη συνέχεια ακολουθεί η κλασική ανάλυση της randomized στρογγυλοποίησης για PIPs. Υπενθυμίζεται ότι $B = \min_i b_i$. Επιλέγουμε $a = e(5m)^{1/B}$.

Αρχικά χειριζόμαστε μια εύκολη περίπτωση. Αν $y^* \le 3a$ επιλέγουμε απλά οποιοδήποτε j τέτοιο ώστε $c_j = 1$ (σημειώνεται ότι υπάρχει πάντα ένα τέτοιο j από τον ορισμό του PIP) και θέτουμε $x_j = 1$ και $\forall k \ne j\left(x_k = 0\right)$. Αυτός είναι φανερά ένας $O\left(m^{1/B}\right)$-προσεγγιστικός αλγόριθμος αν ισχύει $y^* \le 3a$.

Ας υποθέσουμε ότι $y^* > 3a$. Έχουμε τα δυσάρεστα γεγονότα $E_1, E_2, ..., E_{m+1}$ και επιλέγουμε $\beta = 3$. Αν αποφύγουμε όλα αυτά τα $(m+1)$ ενδεχόμενα, θα πάρουμε μια εφικτή λύση με τιμή της αντικειμενικής συνάρτησης τουλάχιστον $y^*/(3a)$. Ο στόχος τώρα είναι να αποδείξουμε ότι

$$\sum_{i\in[m+1]} \Pr\left(E_i\right) \le K_4$$

για κάποια **σταθερά** $K_4 < 1$.

Για το σκοπό της ανάλυσης, βλέπουμε κάθε z_j ως το άθροισμα $\left\lceil x_j' \right\rceil$ το πλήθος ανεξάρτητων τυχαίων μεταβλητών με τιμές στο $\{0,1\}$, εκ των οποίων οι πρώτες $\left\lceil x_j' \right\rceil - 1$ παίρνουν την τιμή 1 με πιθανότητα 1 και η τελευταία είναι 1 με πιθανότητα $1 - \left(\left\lceil x_j' \right\rceil - x_j'\right)$. Με δεδομένο ότι ισχύει $A_{i,j} \in [0,1]$, βλέπουμε, για κάθε $i \in [m]$, ότι το $(Az)_i$ είναι ένα άθροισμα ανεξάρτητων τυχαίων μεταβλητών, καθεμία εκ των οποίων παίρνει τιμές στο $[0,1]$. Επίσης, $\mathbf{E}\left[(Az)_i\right] = (Ax')_i / a \le b_i / a$.

Επομένως, για $i \in [m]$, η πιθανότητα $\Pr\left(E_i\right)$ μπορεί να φραχθεί άνω χρησιμοποιώντας τα *CH* φράγματα και το σκέλος (a) της Πρότασης 4.1.5. Έτσι,

$$\Pr\left(E_i\right) \doteq \Pr\left((Az)_i > b_i\right) \le (e/a)^{b_i} = (5n)^{-b_i/B} \le 1/(5n),$$

εφόσον $b_i \ge B$.

Για να φράξουμε άνω την πιθανότητα $\Pr\left(E_{m+1}\right)$, μας αρκεί η ανισότητα του Chebysev. Αρκεί να παρατηρήσουμε ότι το $c^T \cdot z$ είναι μια τυχαία μεταβλητή που προκύπτει ως άθροισμα τ.μ. με τιμές στο $[0,1]$ και με μέση τιμή ίση με $\mu \doteq \mathbf{E}\left[c^T \cdot z\right] = y^*/a \ge 3$ (υπενθυμίζεται ότι εξ υποθέσεως $y^* > 3a$). Το Πόρισμα 4.1.3 δίνει

$$\Pr\left(E_{m+1}\right) = \Pr\left(c^T \cdot z < \mu/3\right) \le \Pr\left(\left|c^T \cdot z - \mu\right| > 2\mu/3\right) \le 9/4\mu \le 3/4$$

Τα παραπάνω φράγματα δείχνουν ότι

$$\sum_{i\in[m+1]} \Pr\left(E_i\right) \le 1/5 + 3/4 = 0.95$$

Επαναλαμβάνοντας τον αλγόριθμο ένα σταθερό αριθμό φορών, η πιθανότητα αποτυχίας μπορεί να ελαττωθεί σε οποιαδήποτε επιθυμητή θετική σταθερά. Άρα, έχουμε ένα randomized προσεγγιστικό αλγόριθμο με λόγο απόδοσης $O\left(m^{1/B}\right)$ για γενικά PIPs.

Τώρα, ας υποθέσουμε ότι τα στοιχεία του πίνακα A ανήκουν στο $\{0,1\}$. Υπενθυμίζεται ότι εν προκειμένω κάθε b_i υποτίθεται ακέραιος. Συνεπώς, το ενδεχόμενο $E_i \equiv \left((Az)_i > b_i \right)$ είναι ισοδύναμο με το $\left((Az)_i \geq b_i + 1 \right)$. Επομένως, το B μπορεί να αντικατασταθεί από το $B+1$ στο πρώτο από τα παραπάνω άνω φράγματα. Μπορούμε λοιπόν να επιλέξουμε ως $a = \Theta\left(m^{1/(B+1)} \right)$ και να πάρουμε μία προσέγγιση $O(a)$ όπως παραπάνω.

4.3. Εφαρμογή στο Πρόβλημα MAX-SAT

Πρόκειται για τη «φυσική» εκδοχή βελτιστοποίησης του προβλήματος SAT. Δεδομένης μιας λογικής πρότασης F σε συζευκτική κανονική μορφή και ενός μη αρνητικού βάρους w_i συσχετισμένου με κάθε clause C_i, ο στόχος είναι η εύρεση μιας απονομής αλήθειας για τις μεταβλητές που μεγιστοποιεί το συνολικό βάρος των ικανοποιούμενων clauses. Το πρόβλημα είναι προφανώς $NP-hard$.

Ο πρώτος προσεγγιστικός αλγόριθμος που επιτυγχάνει λόγο προσέγγισης $4/3$ οφείλεται στο Γιαννακάκη [Yan94]. Θα παρουσιάσουμε δύο απλούστερους αλγορίθμους με λόγους απόδοσης $(1-1/e)^{-1}$ και 4/3. Οι αλγόριθμοι που θα παρουσιαστούν χρησιμοποιούν την τεχνική της randomized στρογγυλοποίησης και οφείλονται στους Goemans και Williamson [GW94].

Σε πρώτη φάση, διατυπώνουμε το πρόβλημα MAX – SAT:

MAX – SAT

- **Είσοδος:**
  - n λογικές μεταβλητές $x_1, x_2, ..., x_n$
  - m clauses $C_1, C_2, ..., C_m$
  - ένα μη αρνητικό βάρος w_i για κάθε clause C_i
- **Έξοδος:** Μία απονομή αληθείας για τις n λογικές μεταβλητές που μεγιστοποιεί το συνολικό βάρος των ικανοποιούμενων clauses.

Η πιο απλοϊκή χρήση του randomization για το εν λόγω πρόβλημα είναι η εξής: Θέσε κάθε μία από τις n μεταβλητές ανεξάρτητα στην τιμή *True* ή *False* με πιθανότητα $1/2$. Είναι εύκολο να δει κανείς ότι αυτός ο απλός αλγόριθμος είναι 2-προσεγγιστικός (κατά μέση τιμή) για το πρόβλημα. Πράγματι, αν ορίσουμε μια τ.μ. X_j με τιμή 1 αν το j-οστό clause ικανοποιείται και 0 ειδάλλως, η τ.μ. που εκφράζει το βάρος των ικανοποιούμενων clauses είναι η $W = \sum_{j=1}^{m} w_j X_j$. Αν συμβολίσουμε με l_j τον αριθμό των literals στο clause j, για τη μέση τιμή του W έχουμε:

$$\mathbf{E}[W] = \sum_j w_j \mathbf{E}\left[X_j\right] = \sum_j w_j \Pr\left(\underbrace{X_j = 1}_{C_j \text{ ικανοποιείται}} \right) = \sum_j w_j \left(1 - \left(\frac{1}{2} \right)^{l_j} \right) \geq \frac{1}{2} \sum_j w_j \geq \frac{1}{2} OPT$$

Παρατηρούμε ότι αυτός ο αλγόριθμος έχει καλή απόδοση όταν όλα τα clauses είναι μεγάλου μήκους.
Έχοντας εισάγει έναν απλό randomized προσεγγιστικό αλγόριθμο, μπορούμε να επιστρέψουμε στην υπό εξέταση τεχνική.

1. Μοντελοποιούμε το πρόβλημα MAX SAT με το ακόλουθο ακέραιο γραμμικό πρόγραμμα, στο οποίο εισάγουμε μια μεταβλητή z_j για κάθε φράση c_j και μία μεταβλητή y_i για κάθε λογική μεταβλητή x_i:

$$\max \sum_j w_j z_j$$

με τους περιορισμούς:

$$\begin{array}{ll} \forall C_j : \bigvee_{i \in I_j^+} x_i \vee \bigvee_{i \in I_j^-} \overline{x_i} & \left(\sum_{i \in I_j^+} y_i + \sum_{i \in I_j^-} (1 - y_i) \geq z_j \right) \\ \forall i \in [n] & \left(y_i \in \{0,1\} \right) \\ \forall j \in [m] & \left(z_j \in \{0,1\} \right) \end{array}$$

2. Για να πάρουμε ένα γραμμικό πρόγραμμα, χαλαρώνουμε τις απαιτήσεις $y_i \in \{0,1\}$ και $z_j \in \{0,1\}$ σε $y_i \in [0,1]$ και $z_j \in [0,1]$ αντίστοιχα. Κατά τα γνωστά, εφόσον πρόκειται για πρόβλημα μεγιστοποίησης, θα ισχύει $OPT_f \geq OPT$.
3. Με εφαρμογή της τεχνικής του randomized rounding παίρνουμε τον ακόλουθο αλγόριθμο:

3.1 Επίλυσε το χαλαρωμένο γραμμικό πρόγραμμα και βρες μια βέλτιστη λύση $\left(y^*, z^* \right)$.
3.2 $\forall i \in [n]$:

$x_i \leftarrow TRUE$ με πιθανότητα y_i^*

$x_i \leftarrow FALSE$ με πιθανότητα $1 - y_i^*$

Θεώρημα 4.3.1 [GW94] Ο παραπάνω randomized αλγόριθμος είναι $\left(1 - \frac{1}{e} \right)^{-1}$ προσεγγιστικός (κατά μέση τιμή).

Απόδειξη: Για την απόδειξη του θεωρήματος χρειαζόμαστε τα ακόλουθα:

Πρόταση 4.3.1.1 Ο γεωμετρικός μέσος κάθε ακολουθίας μη αρνητικών αριθμών είναι άνω φραγμένος από τον αριθμητικό τους μέσο, δηλαδή αν $\forall i \in [k] \left(a_i \geq 0 \right)$, τότε

$$\sqrt[k]{a_1 a_2 \ldots a_k} \leq \frac{1}{k} \left(a_1 + a_2 + \cdots + a_k \right)$$

□

Πρόταση 4.3.1.2 Αν η συνάρτηση $f(x)$ είναι κοίλη στο $[l,u]$ και επιπλέον ισχύουν $f(l) \geq al+b$ και $f(u) \geq au+b$, τότε

$$\forall x \in [l,u]\big(f(x) \geq ax+b\big)$$

□

Θεωρούμε αρχικά ένα clause C_j της μορφής $x_1 \vee x_2 \vee \ldots \vee x_k$. Παρατηρούμε ότι ο αντίστοιχος περιορισμός είναι $\sum_{i=1}^{k} y_i^* \geq z_j^*$. Έχουμε λοιπόν ότι

$$\begin{aligned}
&\Pr\big(\text{clause } C_j \text{ ικανοποιείται}\big) = 1 - \prod_{i=1}^{k}\big(1-y_i^*\big) \\
&(\text{πρόταση } 1) \qquad \geq 1 - \left(\frac{k-\sum_{i=1}^{k} y_i^*}{k}\right)^k \\
&(\text{περιορισμός}) \qquad \geq 1 - \left(1-\frac{z_j^*}{k}\right)^k \\
&(\text{πρόταση } 2) \qquad \geq \left[1-\left(1-\frac{1}{k}\right)^k\right] z_j^*
\end{aligned}$$

Η τελευταία ανισότητα προκύπτει με εφαρμογή της πρότασης 2 για την κοίλη συνάρτηση $f(x) = 1-\left(1-\frac{x}{k}\right)^k$ στο $[0,1]$. Είναι $f(0)=0$ και $f(1) = 1-\left(1-\frac{1}{k}\right)^k$. Επομένως η αντίστοιχη ευθεία προκύπτει για $a = f(1)$ και $b=0$.

Είναι εύκολο να δει κανείς ότι η υπόθεση πως το clause C_j δεν περιέχει αρνήσεις μεταβλητών έγινε χωρίς βλάβη της γενικότητας. Το παραπάνω αποτέλεσμα λοιπόν ισχύει γενικά. Έτσι, για το μέσο κόστος των ικανοποιούμενων clauses παίρνουμε:

$$\begin{aligned}
\mathbf{E}[W] &= \sum_j w_j \Pr\big(\text{clause } C_j \text{ ικανοποιείται}\big) \\
&\geq \inf_k \left[1-\left(1-\frac{1}{k}\right)^k\right] \cdot \sum_j w_j z_j^* \\
&\geq \inf_k \left[1-\left(1-\frac{1}{k}\right)^k\right] \cdot OPT \\
&\geq \left(1-\frac{1}{e}\right) \cdot OPT
\end{aligned}$$

□

Παρατηρούμε ότι ο εν λόγω αλγόριθμος έχει καλή συμπεριφορά όταν όλα τα clauses έχουν μικρό μήκος. Ο δεύτερος αλγόριθμος στηρίζεται σε μία παραλλαγή της προηγούμενης τεχνικής που ονομάζεται μη γραμμική randomized στρογγυλοποίηση.

Non-linear randomized rounding

Η μη γραμμική randomized στρογγυλοποίηση στηρίζεται στην παρατήρηση ότι δεν μας υποχρεώνει κανείς να χρησιμοποιήσουμε την βέλτιστη λύση του χαλαρωμένου γραμμικού προγράμματος ως έχει για την κατασκευή του αποτελέσματος. Μπορούμε να χρησιμοποιήσουμε μια συνάρτηση της λύσης αυτής για την παραγωγή των σχετικών πιθανοτήτων στρογγυλοποίησης. Θεωρούμε τον ακόλουθο αλγόριθμο:

Μη γραμμική randomized στρογγυλοποίηση
3.1 Επίλυσε το χαλαρωμένο γραμμικό πρόγραμμα και βρες μια βέλτιστη λύση $\left(y^*, z^*\right)$.
3.2 Επέλεξε οποιαδήποτε συνάρτηση $g(\cdot)$ τέτοια ώστε $1-4^{-y} \le g(y) \le 4^{y-1}$ για $y \in [0,1]$.
$\forall i \in [n]$:
$x_i \leftarrow TRUE$ με πιθανότητα $g\left(y_i^*\right)$
$x_i \leftarrow FALSE$ με πιθανότητα $1-g\left(y_i^*\right)$

Θεώρημα 4.3.2 [GW94] Ο παραπάνω randomized αλγόριθμος είναι $4/3$-προσεγγιστικός για το πρόβλημα MAX-SAT.

Απόδειξη: Στην απόδειξη χρησιμοποιούμε την Πρόταση 4.3.1.2 της προηγούμενης απόδειξης (που αφορά σε κοίλες συναρτήσεις).

Υποθέτουμε αρχικά ότι έχουμε ένα clause C_j χωρίς αρνήσεις, δηλαδή ότι ισχύει $C_j \equiv x_1 \vee x_2 \vee \ldots \vee x_k$. Ο αντίστοιχος περιορισμός του χαλαρωμένου γραμμικού προγράμματος είναι πάλι $\sum_{i=1}^{k} y_i^* \ge z_j^*$. Επομένως,

$$\begin{aligned}
\Pr\left(\text{clause } C_j \text{ ικανοποιείται}\right) &= 1-\prod_{i=1}^{k}\left(1-g\left(y_i^*\right)\right) \\
&\ge 1-\left(\prod_{i=1}^{k} 4^{-y_i^*}\right) \\
&= 1-4^{-\sum_{i=1}^{k} y_i^*} \\
(\text{περιορισμός}) \qquad &\ge 1-4^{-z_j^*} \\
(\text{πρόταση } 4.3.1.2) \qquad &\ge \frac{3}{4}\cdot z_j^*
\end{aligned}$$

Η τελευταία ανισότητα προκύπτει με εφαρμογή της πρότασης 2 για την κοίλη συνάρτηση $g(x)=1-4^{-x}$ στο $[0,1]$. Με δεδομένο ότι $g(0)=0$ και $g(1)=3/4$ έπεται ότι $\forall x \in [0,1]$ ισχύει $g(x) \ge \frac{3}{4} x$.

Μπορεί εύκολα να δει κανείς ότι το αποτέλεσμα αυτό ισχύει γενικότερα για κάθε clause. Έστω για παράδειγμα ότι έχουμε το $C_j \equiv x_1 \vee x_2 \vee \ldots \vee \overline{x}_k$. Τότε

$$\begin{aligned}
\Pr\left(\text{clause } C_j \text{ ικανοποιείται}\right) &= 1 - \prod_{i=1}^{k-1}\left(1 - g\left(y_i^*\right)\right) \times g\left(y_k^*\right) \\
&\geq 1 - \prod_{i=1}^{k-1} 4^{-y_i^*} \times 4^{-\left(1-y_k^*\right)} \\
&= 1 - 4^{-\left(\sum_{i=1}^{k-1} y_i^* + \left(1-y_k^*\right)\right)} \\
(\text{περιορισμός}) \quad &\geq 1 - 4^{-z_j^*} \\
(\text{πρόταση } 2) \quad &\geq \frac{3}{4} \cdot z_j^*
\end{aligned}$$

Υπενθυμίζεται ότι εν προκειμένω ο αντίστοιχος περιορισμός του χαλαρωμένου γραμμικού προγράμματος θα είναι ο $\sum_{i=1}^{k-1} y_i^* + \left(1 - y_k^*\right) \geq z_j^*$. Με παρόμοιο τρόπο προκύπτει το ίδιο αποτέλεσμα για όλες τις μορφές φράσεων.

Επομένως,

$$\mathbf{E}\left[W\right] = \sum_j w_j \Pr\left(\text{clause } C_j \text{ ικανοποιείται}\right) \geq \frac{3}{4} \cdot \sum_j w_j z_j^* \geq \frac{3}{4} \cdot OPT$$

□

Είναι εύκολο να δει κανείς ότι το integrality gap του χαλαρωμένου γραμμικού προγράμματος για το MAX SAT είναι ίσο με $4/3$. Επομένως, δεν μπορούμε να ελπίζουμε σε καλύτερο αλγόριθμο με βάση την υπό συζήτηση τεχνική.

4.4 Το Πρόβλημα MAX CUT σε πυκνά γραφήματα.

Με σκοπό να δείξουμε ότι η χρήση του randomization μπορεί να γίνει αρκετά περίπλοκη, στρεφόμαστε στο πρόβλημα MAX CUT.

MAX CUT

- **Είσοδος:** Ένας μη κατευθυνόμενος γράφος $G = \left(V, E\right)$ με μη αρνητικά βάρη $w_{ij} \geq 0$ για κάθε $\left(i, j\right) \in E$ (μπορούμε να υποθέσουμε ότι $\forall\left(i, j \notin E\right)$ ισχύει $w_{ij} = 0$).
- **Έξοδος:** Ένα υποσύνολο S του συνόλου των κόμβων ($S \subseteq V$) που μεγιστοποιεί το $w\left(S\right) = \sum_{i \in S, j \notin S} w_{ij}$.

Κατ' αναλογία με τα όσα αναφέρθηκαν για το πρόβλημα MAX SAT, μπορούμε εύκολα να δείξουμε ότι ο απλός αλγόριθμος που επιλέγει κάθε κόμβο ανεξάρτητα με ίση πιθανότητα είναι 2-προσεγγιστικός για το πρόβλημα. Σε αντίθεση με την περίπτωση του MAX SAT, δεν μπορούμε να πετύχουμε κάτι καλύτερο με τα εργαλεία του γραμμικού προγραμματισμού.

Ωστόσο, μπορούμε με τα εργαλεία αυτά να πετύχουμε κάτι καλύτερο για μια ειδική περίπτωση του προβλήματος. Θεωρούμε λοιπόν την περίπτωση στην οποία τα βάρη όλων των ακμών είναι ίσα με 1 και ο γράφος είναι πυκνός, δηλ. $\|E\| \geq an^2$ για κάποιο $a > 0$, όπου $n = \|V\|$. Αναλύουμε ένα αποτέλεσμα των Arora, Karger και Karpinski [AKK98] το οποίο δίνει ένα προσεγγιστικό σχήμα πολυωνυμικού χρόνου για την ειδική αυτή περίπτωση του προβλήματος.

Προκύπτει άμεσα ότι για την ειδική αυτή περίπτωση ισχύει $OPT \geq \frac{a}{2}n^2$. Πράγματι, ο προφανής randomized αλγόριθμος για το (γενικό) πρόβλημα δίνει μια τομή με μέσο βάρος τουλάχιστον $\frac{1}{2}\sum_{(i,j)\in E} w_{ij}$. Επομένως, το βάρος της μέγιστης τομής πρέπει να είναι τουλάχιστον τόσο.

Ας θεωρήσουμε ένα συγκεκριμένο μοντέλο του προβλήματος MAX CUT. Το πρόβλημα μπορεί να διατυπωθεί ισοδύναμα ως το πρόβλημα εύρεσης μιας συνάρτησης $x : V \to \{0,1\}$ τέτοιας ώστε $x_i \equiv x(i) = 1$ αν και μόνο αν $i \in S$. Η αντικειμενική συνάρτηση γίνεται

$$\max_{x_i \in \{0,1\}} \sum_{i\in V} x_i \cdot \sum_{(i,j)\in E} \left(1 - x_j\right)$$

Για να γίνει φανερή η ορθότητα της παραπάνω σχέσης, μπορεί κανείς εύκολα να παρατηρήσει ότι το άθροισμα $\sum_{(i,j)\in E}\left(1-x_j\right)$ μετράει το πλήθος των ακμών που γειτονεύουν με τον κόμβο i και το άλλο τους άκρο παίρνει την τιμή 0 (δηλ. ακμές τη ς μορφής (i,j) με $x_j = 0$). Εφόσον πολλαπλασιάζουμε κάθε τέτοιο όρο με x_i, μετράμε μόνο αυτές τις ακμές που έχουν $x_i = 1$. Συνεπώς, για κάθε κόμβο, μετράμε όλες τις γειτονικές προς αυτόν ακμές που έχουν το άλλο άκρο τους στο άλλο σύνολο διαχωρισμού.

Εφόσον, θα χρησιμοποιούμε τον όρο $\sum_{(i,j)\in E}\left(1-x_j\right)$ αρκετά συχνά στα επόμενα, ορίζουμε κάποιο συμβολισμό για αυτόν. Έστω λοιπόν

$$ZN(x,i) \doteq \sum_{(i,j)\in E}\left(1-x_j\right)$$

Τα αρχικά συμβολίζουν τη φράση “zero neighbors”.

Έστω x^* μία βέλτιστη λύση. Ας υποθέσουμε ότι υπάρχει ένα μαντείο που μας δίνει τιμές Z_i τέτοιες ώστε $\left|Z_i - ZN\left(x^*,i\right)\right| \leq \epsilon \cdot n$, $n = \|V\|$. Μπορούμε να χρησιμοποιήσουμε αυτή την πληροφορία για να υπολογίσουμε μια προσεγγιστική λύση.

Η τεχνική που χρησιμοποιούμε είναι αυτή του randomized rounding. Θεωρούμε το ακόλουθο γραμμικό πρόγραμμα:

$$\max \sum_{i\in V} Z_i y_i$$

με τους περιορισμούς:

$$\forall i \in V \left(Z_i - \epsilon \cdot n \leq \sum_{(i,j)\in E} \left(1 - y_j\right) \leq Z_i + \epsilon \cdot n \right)$$

$$\forall i \in V \left(0 \leq y_i \leq 1\right)$$

Μπορεί να παρατηρήσει κανείς ότι από τον ορισμό των Z_i, η βέλτιστη λύση x^* είναι εφικτή για το παραπάνω γραμμικό πρόγραμμα. Επίσης, η τιμή της αντικειμενικής συνάρτησης (του γραμμικού προγράμματος) για το x^* είναι κοντά στο OPT, όπως αποδεικνύουμε αμέσως παρακάτω:

$$\begin{aligned}\sum_{i\in V} Z_i \cdot x_i^* &\geq \sum_{i\in V}\left(ZN\left(x^*,i\right)-\epsilon\cdot n\right)\cdot x_i^* \\ &= OPT-\epsilon\cdot n\cdot\sum_{i\in V} x_i^* \\ &\geq OPT-\epsilon\cdot n^2 \\ &\geq \left(1-\frac{2\epsilon}{a}\right)OPT\end{aligned}$$

Επομένως, η βέλτιστη λύση του γραμμικού προγράμματος έχει κέρδος τουλάχιστον $\left(1-\frac{2\epsilon}{a}\right)OPT$. Θεωρούμε τώρα τον ακόλουθο αλγόριθμο που βασίζεται στην τεχνική της randomized στρογγυλοποίησης.

ΑΚΚ '98

3.1 Πάρε τα Z_i από το μαντείο. Λύσε το γραμμικό πρόγραμμα και υπολόγισε μια βέλτιστη λύση y^*.

3.2 $\forall i \in V$:

$x_i' \leftarrow 1$ με πιθανότητα y_i^*

$x_i' \leftarrow 0$ με πιθανότητα $1-y_i^*$

Αν συμβολίσουμε με x' το τυχαίο διάνυσμα που προκύπτει από τη διαδικασία της στρογγυλοποίησης, τότε η τιμή της τομής που προκύπτει από την προηγούμενη διαδικασία (και φυσικά είναι τυχαία μεταβλητή) είναι προφανώς $\sum_{i\in V} x_i' \cdot ZN\left(x',i\right)$.

Αποδεικνύεται με χρήση των CH bounds ότι με μεγάλη πιθανότητα η μέση τιμή της παραπάνω τ.μ. είναι κοντά στην βέλτιστη τιμή.

Ας υπολογίσουμε αρχικά τη μέση τιμή του ZN για τη λύση x'.

$$\begin{aligned}\mathbf{E}\left[ZN\left(x',i\right)\right] &= \mathbf{E}\left[\sum_{(i,j)\in E}\left(1-x_j'\right)\right] \\ &= \sum_{(i,j)\in E}\left(1-\mathbf{E}\left[x_j'\right]\right) \\ &= \sum_{(i,j)\in E}\left(1-y_j^*\right) \\ &= ZN\left(y^*,i\right)\end{aligned}$$

Τώρα αποδεικνύουμε ότι με μεγάλη πιθανότητα αυτή η μέση τιμή είναι κοντά στην τιμή της λύσης του γραμμικού προγράμματος εφαρμόζοντας τα CH bounds. Θέτουμε $\delta_i = \sqrt{\dfrac{2c\ln n}{\max\left\{ZN\left(y^*,i\right),2c\ln n\right\}}}$. Τότε

$$\Pr\left(ZN\left(x^{'},i\right) < \left(1-\delta_i\right)\cdot \underbrace{ZN\left(y^*,i\right)}_{\mu} \right) \le \exp\left(-\frac{\mu\delta_i^2}{2}\right)$$

$$= \exp\left(-ZN\left(y^*,i\right)\cdot \frac{2c\ln n}{\max\left\{ZN\left(y^*,i\right),2c\ln n\right\}} \right)$$

$$\le e^{-c\ln n} = \frac{1}{n^c}$$

Άρα, λόγω του φράγματος της ένωσης, με πιθανότητα τουλάχιστον $1-1/n^{c-1}$, έχουμε:

$$\sum_i x_i^{'} ZN\left(x^{'},i\right) \ge \sum_i x_i^{'}\left(1-\delta_i\right)\cdot ZN\left(y^*,i\right)$$

$$\ge \sum_i x_i^{'}\left(ZN\left(y^*,i\right) - \sqrt{2c\ln n\cdot ZN\left(y^*,i\right)} \right)$$

$$\ge \sum_i x_i^{'}\left(Z_i - \epsilon\cdot n - \sqrt{2c\ln n\cdot ZN\left(y^*,i\right)} \right)$$

$$\ge \sum_i x_i^{'} Z_i - \left(\epsilon\cdot n + \sqrt{2c\ln n}\right)\cdot \sum_i x_i^{'}$$

Στο σημείο αυτό κρίνεται αναγκαίο να φράξουμε το $\sum_i x_i^{'} Z_i$. Εφόσον

$$\mathbf{E}\left[\sum_i x_i^{'} Z_i\right] = \sum_i Z_i\cdot \mathbf{E}\left[x_i^{'}\right] = \sum_i Z_i\cdot y_i^*$$

Χρησιμοποιώντας το CH bound με $Z = \max_i Z_i$ και $\delta = \sqrt{\dfrac{2c\ln n}{\max\left\{\sum_i y_i^* \dfrac{Z_i}{Z}, 2c\ln n\right\}}}$ έπεται

ότι

$$\Pr\left(\sum_i x_i^{'} \frac{Z_i}{Z} < \left(1-\delta\right)\sum_i \frac{Z_i}{Z} y_i^* \right) \le \frac{1}{n^c}$$

Επομένως, με μεγάλη πιθανότητα ισχύει ότι

$$\sum_i x_i' Z_i \geq (1-\delta)\sum_i Z_i \cdot y_i^*$$

$$= \left(1-\sqrt{\frac{2c\ln n}{\max\left\{\sum_i y_i^* \frac{Z_i}{Z}, 2c\ln n\right\}}}\right)\sum_i Z_i \cdot y_i^*$$

$$\geq \sum_i Z_i \cdot y_i^* - \sqrt{2\cdot Z\cdot c\cdot \ln n \cdot \sum_i Z_i \cdot y_i^*}$$

$$\geq \sum_i Z_i \cdot y_i^* - n\sqrt{2c\ln n}$$

Χρησιμοποιώντας αυτό το αποτέλεσμα, από την προηγούμενη σχέση παίρνουμε:

$$\sum_i x_i' ZN\left(x', i\right) \geq \sum_i x_i' Z_i - n\sqrt{2c\ln n} - \left(\epsilon\cdot n + \sqrt{2c\ln n}\right)\cdot \sum_i x_i'$$

$$\geq \left(1-\frac{2\cdot\epsilon}{a}\right)OPT - n\sqrt{2c\ln n} - \epsilon\cdot n^2 - n\sqrt{2c\ln n}$$

$$\geq \left(1-\frac{2\cdot\epsilon}{a}\right)OPT - \frac{2\cdot\epsilon}{a}OPT - o(1)OPT$$

$$\geq \left(1-\frac{5\cdot\epsilon}{a}\right)OPT$$

Η τελευταία ανισότητα ισχύει για αρκετά μεγάλο n, έτσι ώστε ο όρος $o(1)$ να έχει γίνει μικρότερος από $\frac{\epsilon}{a}OPT$. Αν λοιπόν θέσουμε $\epsilon' = \frac{5\epsilon}{a}$ παίρνουμε μια λύση με αντίστοιχη τιμή αντικειμενικής συνάρτησης $\geq (1-\epsilon')OPT$ με μεγάλη πιθανότητα για αρκετά μεγάλο n.

Μέχρι το σημείο αυτό αγνοήσαμε τον τρόπο λειτουργίας του μαντείου που παράγει τα Z_i. Κι αυτό διότι σκοπός μας ήταν να δώσουμε ένα παράδειγμα εφαρμογής της τεχνικής του randomized rounding. Ο τρόπος λειτουργίας του μαντείου βασίζεται στο ακόλουθο θεώρημα:

Θεώρημα 4.4.1 Δοθείσης μιας δυαδικής ακολουθίας $a_1, a_2, \ldots, a_n \in \{0,1\}$, ορίζουμε το $Z = \sum_{i=1}^n a_i$. Αν επιλέξουμε ένα τυχαίο σύνολο (δεικτών) $S \subseteq [n]$, με $\|S\| = c\log n/\epsilon^2$, τότε με μεγάλη πιθανότητα ισχύει

$$\left| Z - \frac{n}{\|S\|}\cdot\sum_{i\in S} a_i \right| \leq \epsilon\cdot n$$

Το μαντείο λοιπόν λειτουργεί ως εξής: Επιλέγουμε ένα τυχαίο υποσύνολο S του συνόλου των κόμβων του γραφήματος με $\|S\| = c\log n/\epsilon^2$. Θέτουμε

$$Z_i = \frac{n}{\|S\|}\sum_{(i,j)\in E, j\in S}\left(1-x_j^*\right)$$

και από το προηγούμενο θεώρημα παίρνουμε ότι $\left|Z_i - ZN\left(x^*, i\right)\right| \le \epsilon \cdot n$.

Το πρόβλημα είναι ότι δεν γνωρίζουμε τη βέλτιστη λύση x^*. Για να υπερβούμε αυτό το πρόβλημα, τρέχουμε τον αλγόριθμο για όλες τις $2^{\|S\|} = n^{O\left(1/\epsilon^2\right)}$ αναθέσεις του x_j^*, όπου $j \in S$. Δεν γνωρίσουμε ποια από αυτές τις αναθέσεις δίνει τη βέλτιστη λύση, αλλά αυτό δεν έχει σημασία. Απλά επιστρέφουμε τη μεγαλύτερη τομή που βρίσκουμε, η οποία εγγυημένα είναι τουλάχιστον $\left(1-\epsilon'\right) \cdot OPT$, εφόσον τουλάχιστον μία από τις τομές θα είναι αυτού του μεγέθους.

Κεφάλαιο 5

Η τεχνική Primal –Dual για Προσεγγιστικούς Αλγορίθμους

5.1 Εισαγωγή

Κάθε πρόβλημα βελτιστοποίησης έχει τη δική του δομή η οποία μπορεί να χρησιμοποιηθεί για δύο σκοπούς. Ο πρώτος είναι για να επιτύχουμε καλό χρόνο εκτέλεσης και ο δεύτερος για να επιτύχουμε καλό λόγο προσέγγισης. Οι προσεγγιστικοί αλγόριθμοι που βασίζονται στη μέθοδο της στρογγυλοποίησης χρησιμοποιούν αυτή τη δομή μόνο για να επιτύχουν το δεύτερο. Αυτό οφείλεται στο γεγονός ότι όλοι οι αλγόριθμοι στρογγυλοποίησης απαιτούν τη βέλτιστη επίλυση ενός γραμμικού προγράμματος, εφόσον αυτό μπορεί να επιτευχθεί σε πολυωνυμικό χρόνο. Οι προσεγγιστικοί αλγόριθμοι που βασίζονται στην τεχνική primal-dual δε βελτιστοποιούν κανένα γραμμικό πρόγραμμα. Στην πραγματικότητα, οι αλγόριθμοι αυτοί δε χρησιμοποιούν τη βέλτιστη τιμή ενός γραμμικού προγράμματος ως κάτω φράγμα.

5.2 Θεμελίωση της μεθόδου

Στόχος αυτού του κεφαλαίου είναι η παρουσίαση των βασικών ιδεών που διέπουν τη μέθοδο primal-dual για προσεγγιστικούς αλγορίθμους. Δε θα ασχοληθούμε εκτενώς με τη μέθοδο primal-dual που αφορά στη συνδυαστική βελτιστοποίηση. Μία πολύ καλή επισκόπηση για αυτό το θέμα μπορεί να βρεθεί στο βιβλίο των Παπαδημητρίου και Steiglitz [PS82]. Η βασική ιδέα οφείλεται στον Kuhn, ο οποίος τη χρησιμοποίησε για τη σχεδίαση ενός αλγορίθμου για το πρόβλημα εύρεσης τέλειου ταιριάσματος ελαχίστου κόστους σε διμερή γραφήματα. Ακολούθως η ιδέα επεκτάθηκε από τους Dantzig, Ford και Fulkerson σε ένα γενικό αλγόριθμο για το γραμμικό προγραμματισμό. Η βασική ιδέα είναι ότι δοθείσης μιας εφικτής λύσης y του dual, προσπαθούμε να βρούμε μια εφικτή λύση x του primal η οποία να υπακούει τις συνθήκες complementary slackness σε σχέση με το y. Αν μπορούμε να βρούμε ένα τέτοιο x, τότε έχουμε βέλτιστες λύσεις. Αν δεν υπάρχει τέτοιο x, τότε μπορούμε να μεταβάλλουμε τη δυϊκή λύση, έτσι ώστε η νέα δυϊκή λύση να είναι πιο κοντά στο βέλτιστο (δηλ. αν υποθέσουμε standard μορφή, να δίνει μεγαλύτερη τιμή δυϊκής αντικειμενικής συνάρτησης).

Η μέθοδος primal-dual για προσεγγιστικούς αλγορίθμους θεωρεί μία μοντελοποίηση του υπό εξέταση ($NP-$hard) προβλήματος υπό μορφή ακέραιου προγράμματος και επίσης το δυϊκό του αντίστοιχού *LP*-relaxation. Η παραπάνω μέθοδος τροποποιείται χαλαρώνοντας τις συνθήκες complementary slackness. Όπως θα δούμε παρακάτω, η χαλάρωση αυτών των περιορισμών με κατάλληλο τρόπο μπορεί να οδηγήσει σε αποδεδειγμένα καλούς αλγόριθμους για $NP-$hard προβλήματα συνδυαστικής βελτιστοποίησης.

Η μέθοδος παράγει μία λύση για το (πρωτεύον) ακέραιο πρόγραμμα με κόστος το πολύ a φορές επί την τιμή της εφικτής *δυϊκής* λύσης που κατασκευάζεται, πράγμα που συνεπάγεται, λόγω της ασθενούς δυϊκότητας ότι η πρωτεύουσα λύση που υπολογίζεται είναι το πολύ a φορές επί το κόστος της βέλτιστης λύσης του πρωτεύοντος ακέραιου προγράμματος.

Εφόσον ο λόγος απόδοσης αποδεικνύεται συγκρίνοντας την τιμή μιας λύσης του πρωτεύοντος ακέραιου προγράμματος με την τιμή μιας εφικτής δυϊκής λύσης, κατ' αναλογία με όσα έχουν ειπωθεί στο κεφάλαιο 3, ο καλύτερος λόγος που μπορούμε να αποδείξουμε δεν μπορεί να είναι καλύτερος από το integrality gap του *LP*-relaxation για το πρόβλημα. Αντίστροφα, αν αποδείξουμε ότι ένας αλγόριθμος που κατασκευάζεται με αυτή την τεχνική, έχει λόγο απόδοσης a, έπεται ότι το integrality gap του *LP*-relaxation είναι το πολύ a.

5.3 Ένα διαισθητικό παράδειγμα

Ως εισαγωγικό παράδειγμα εφαρμογής της μεθόδου, θεωρούμε το πρόβλημα κομβικής κάλυψης ελαχίστου βάρους (min weight vertex cover) σε γενικά γραφήματα. Υπενθυμίζεται ότι για την έκδοση του προβλήματος χωρίς βάρη στους κόμβους, έχουμε ήδη (Κεφάλαιο 3) κατασκευάσει το ακέραιο και το αντίστοιχο χαλαρωμένο γραμμικό πρόγραμμα. Κατ' αναλογία λοιπόν παίρνουμε:

$$\min \sum_{i \in V} w_i x_i$$

με τους περιορισμούς:

$$x_i + x_j \geq 1 \qquad \forall (i,j) \in E$$

$$x_i \in \{0,1\} \qquad \forall i \in V$$

Χαλαρώνοντας τους περιορισμούς για ακέραιες μεταβλητές σε $x_i \geq 0$, κατά τα γνωστά, οδηγούμαστε στο *LP*-relaxation. Αν πάρουμε το δυϊκό του γραμμικού αυτού προγράμματος, παίρνουμε:

$$\max \sum_{(i,j) \in E} y_{(i,j)}$$

με τους περιορισμούς:

$$\sum_{k:(i,k) \in E} y_{(i,k)} \leq w_i \qquad \forall i \in V$$

$$y_{(i,j)} \geq 0 \qquad \forall (i,j) \in E$$

Ο primal-dual αλγόριθμος ξεκινάει με τη δυϊκή εφικτή λύση $y = \mathbf{0}$ και την πρωτεύουσα μη εφικτή λύση $x = \mathbf{0}$. Για όσο η πρωτεύουσα «λύση» x είναι μη εφικτή, θα πρέπει να υπάρχει μια τουλάχιστον ακμή $(i,j) \in E$ που δεν καλύπτεται (για την οποία $x_i + x_j = 0$). Αυξάνουμε λοιπόν την αντίστοιχη (προς αυτή την ακμή) δυϊκή μεταβλητή $y_{(i,j)}$ *όσο το δυνατόν περισσότερο*, διατηρώντας εφικτή τη δυϊκή λύση που προκύπτει. Αυτό σημαίνει ότι *τουλάχιστον ένας* από τους περιορισμούς του δυϊκού προγράμματος που αντιστοιχούν στις κορυφές i και j πρέπει να ισχύει με ισότητα. Αν λοιπόν ισχύει $\sum_{k:(i,k) \in E} y_{(i,k)} = w_i$, τότε θέτουμε $x_i = 1$, δηλ. επιλέγουμε στην κάλυψη την κορυφή i, ενώ αν ισχύει $\sum_{k:(j,k) \in E} y_{(j,k)} = w_j$, τότε θέτουμε $x_j = 1$. Τελικά, επιτυγχάνουμε μια εφικτή λύση x του primal τέτοια ώστε:

$$\begin{aligned}
\sum_{i \in V} w_i x_i &= \sum_{i \in V} \left(\sum_{k:(i,k) \in E} y_{(i,k)} \right) x_i \\
&= \sum_{(i,j) \in E} \left(x_i + x_j \right) y_{(i,j)} \\
&\leq 2 \cdot \sum_{(i,j) \in E} y_{(i,j)}
\end{aligned}$$

Σημειώνεται ότι η πρώτη ισότητα έπεται από το γεγονός ότι $\sum_{k:(i,k)\in E} y_{(i,k)} = w_i$ για κάθε κόμβο i που έχει επιλεγεί (δηλ. ισχύει $x_i = 1$), η δεύτερη ισότητα προκύπτει αν εναλλάξουμε τη σειρά του διπλού αθροίσματος και η τρίτη προκύπτει από το γεγονός ότι $x_i + x_j \le 2$.

Λόγω της *ασθενούς δυϊκότητας*, η δυϊκή αντικειμενική συνάρτηση $\sum_{(i,j)\in E} y_{(i,j)}$ είναι ένα κάτω φράγμα για την τιμή της βέλτιστης ακέραιης λύσης. Άρα, η παραπάνω ανισότητα δείχνει ότι η λύση μας είναι 2-προσεγγιστική.

5.4 Η μέθοδος Primal-Dual για το Πρόβλημα Set Cover

Στη συνέχεια αναπτύσσουμε τις βασικές ιδέες που δόθηκαν παραπάνω για το πρόβλημα set cover που ως γνωστόν αποτελεί γενίκευση του vertex cover.

Set Cover: Primal – Dual

$I \leftarrow \emptyset$

$\forall i \left(\widetilde{y_i} \leftarrow 0\right)$

while $\exists t_k : t_k \notin \bigcup_{j\in I} S_j$

$\quad l = \arg\min_{j:t_k\in S_j} \left\{ w_j - \sum_{i:t_i\in S_j} \widetilde{y_i} \right\}$

$\quad \epsilon_l \leftarrow w_l - \sum_{i:t_i\in S_l} \widetilde{y_i}$

$\quad \widetilde{y_k} \leftarrow \widetilde{y_k} + \epsilon_l$

$\quad I \leftarrow I \bigcup \{l\}$

Η συνάρτηση $\arg\min(\cdot)$ επιστρέφει το όρισμα (δείκτη εν προκειμένω) που ελαχιστοποιεί την έκφραση.

Λήμμα 5.4.1 Ο αλγόριθμος primal – dual επιστρέφει ένα set cover.

Απόδειξη: Τετριμμένη, εφόσον αποτελεί τη συνθήκη τερματισμού του βρόχου.

□

Λήμμα 5.4.2 Ο εν λόγω αλγόριθμος κατασκευάζει μια εφικτή λύση του δυϊκού γραμμικού προγράμματος.

Απόδειξη: Η απόδειξη με επαγωγή στο πλήθος των επαναλήψεων του βρόχου. Η βάση της επαγωγής ισχύει τετριμμένα εφόσον αρχικά

$$\forall j \left(\sum_{i:t_i\in S_j} \widetilde{y_i} = 0 \le w_j \right)$$

Για το επαγωγικό βήμα, υποθέτουμε ότι κατά την είσοδό μας στο βρόχο ισχύει ότι

$$\forall j \left(\sum_{i:t_i \in S_j} \widetilde{y_i} \le w_j \right)$$

Η μοναδική μεταβλητή του δυϊκού προγράμματος που μεταβάλλεται από το βρόχο είναι η $\widetilde{y_k}$, οπότε οι ανισότητες που αφορούν σε σύνολα S_j που *δεν* περιέχουν το t_k μένουν ανεπηρέαστες. Αν $t_k \in S_j$, τότε λόγω της επιλογής του l παίρνουμε:

$$\begin{aligned}
\sum_{i:t_i \in S_j} \widetilde{y_i} + \epsilon_l &= \sum_{i:t_i \in S_j} \widetilde{y_i} + \left(w_l - \sum_{i:t_i \in S_l} \widetilde{y_i} \right) \\
&\le \sum_{i:t_i \in S_j} \widetilde{y_i} + \left(w_l - \sum_{i:t_i \in S_j} \widetilde{y_i} \right) \\
&\le w_j
\end{aligned}$$

□

Λήμμα 5.4.3 Αν $j \in I$, τότε $\sum_{i:t_i \in S_j} \widetilde{y_i} = w_j$

Απόδειξη: Από τον αλγόριθμο είναι φανερό ότι στο βήμα κατά το οποίο το j προσετέθη στο σύνολο I, η τιμή του $\widetilde{y_k}$ αυξήθηκε ακριβώς τόσο ώστε ο περιορισμός που αντιστοιχεί στο S_j να ισχύει με ισότητα.

□

Θεώρημα 5.4.4 [BYE81] Ο αλγόριθμος primal–dual είναι f-προσεγγιστικός αλγόριθμος για το set cover.

Απόδειξη: Έχουμε διαδοχικά ότι

$$\begin{aligned}
\sum_{j \in I} w_j &= \sum_{j \in I} \sum_{i:t_i \in S_j} \widetilde{y_i} \\
&\le \sum_{1 \le i \le n} \widetilde{y_i} \cdot \left\| \{ j : t_i \in S_j \} \right\| \\
&\le f \cdot \sum_i \widetilde{y_i} \\
&\le f \cdot OPT
\end{aligned}$$

Η πρώτη ισότητα έπεται από το λήμμα 3. Η επόμενη ανισότητα προκύπτει από το γεγονός ότι κάθε $\widetilde{y_i}$ μπορεί να εμφανίζεται στο διπλό άθροισμα το πολύ $\left\| \{ j : t_i \in S_j \} \right\|$ φορές. Η προτελευταία ανισότητα έπεται από τον ορισμό του f και η τελευταία λόγω ασθενούς δυϊκότητας.

□

Βιβλιογραφία

[AKK98] S. Arora, D. Karger, and M. Karpinski. *Polynomial time approximation schemes for Dense Instances of NP-hard problems*. Preliminary version STOC 1995, p. 284-293. Final version in Journal of Computer and System Sciences, 58(1): 193-210,1999.

[ALM+92] S. Arora, C. Lund, R. Motwani, M. Sudan and M. Szegedy. *Proof Verification and the hardness of approximation problems*, FOCS 1992, p. 14-23.

[ALM+98] S. Arora, C. Lund, R. Motwani, M. Sudan and M. Szegedy. *Proof Verification and the hardness of approximation problems*, Journal of the ACM, 45: 501-555, 1998.

[AS92] S. Arora and S. Safra. *Probabilistic checking of proofs; a new characterization of NP*. FOCS 1992, p. 2-13.

[AS98] S. Arora and S. Safra. *Probabilistic checking of proofs; a new characterization of NP*. Journal of the ACM, 45: 70-122, 1998.

[BGS98] M. Bellare, S. Goldwasser, and M. Sudan. *Free bits, PCPs, and non-approximability – towards tight results*, SIAM Journal on Computing, 27:804-915, 1998.

[BYE81] R. Bar-Yehuda and S. Even. *A linear – time aproximation algorithm for the weighted vertex cover problem*. Journal of Algorithms, 2:198-203, 1981.

[Chv83] V. Chvatal. *Linear Programming*. W.H. Freeman and Company, 1983.

[Er67] P. Erdös. *On bipartite subgraphs of graphs* (in Hungarian). Mat. Lapok, 18:283-288, 1967.

[FGL+96] U. Feige, S. Goldwasser, L. Lovász, S. Safra and M. Szegedy. *Interactive proofs and the hardness of approximating cliques*. Journal of the ACM, 43: 268-292, 1996.

[FK98] U. Feige and J. Killian. *Zero Knowledge and the chromatic number*. Journal of Computer and System Sciences, 57: 187-199, 1998.

[G94] M.X. Goemans. *An Introduction to Linear Programming*. MIT, 1994.

[GJ79] M.R. Garey and D.S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W.H. Freeman and Co., New York, NY, 1979.

[Gra66] R. Graham. *Bounds for certain multiprocessor anomalies*. Bell System Technical Journal, 45:1563-1581, 1966.

[GVY96] N. Garg, V.V. Vazirani, and M. Yannakakis. *Approximate max-flow min-(multi)cut theorems and their applications*. SIAM Journal on Computing, 25:235-251, 1996.

[GW94] M.X. Goemans and D.P. Williamson. *New 3/4 – approximation algorithms for the maximum satisfiability problem*. SIAM Journal on Discrete Mathematics, 7:656-666, 1994.

[Ha97] J. Håstad. Some optimal inapproximability results. STOC 1997, p.1-10.

[Ha99] J. Håstad. *Clique is hard to approximate within* $n^{1-\varepsilon}$, Acta Math., 182:105-142, 1999.

[Hoc82] D.S. Hochbaum. *Approximation algorithms for the set covering and vertex cover problems*. SIAM Journal on Computing, 11:555-556, 1982.

[Kar84] N. Karmarkar. *A New Polynomial-Time Algorithm for Linear Programming*. STOC 1984, p. 302-311.

[KK82] N. Karmarkar, R.M. Karp. *An Efficient Approximation Scheme for the One-Dimensional Bin-Packing Problem*. FOCS 1982, p. 312-320.

[Kha79] L. Khachian. *A polynomial algorithm for linear programming*. Doklady Acad. Nauk USSR, 244(5): 1093-1096, 1979.

[Jo74] D.S. Johnson. *Approximation algorithms for combinatorial problems*. Journal of Computer and System Sciences, 9:256-278, 1974.

[LA87] P.J.M. van Laarhoven and E.H.L. Aarts. *Simulated Annealing: Theory and Applications*. Mathematics and its Applications. Reidel Publishing Company, 1987.

[LY94] C. Lund and M. Yannakakis. *On the hardness of approximating minimization problems*. Journal of the ACM, 41:960-981, 1994.

[MR95] R. Motwani, and P. Raghavan. *Randomized Algorithms*. Cambridge University Press, London, 1995.

[P94] C.H. Papadimitriou. *Computational Complexity*. Addison – Wesley, MA, 1994.

[PS82] C.H. Papadimitriou and K. Steiglitz. *Combinatorial optimization: algorithms and complexity.* Prentice-Hall, Englewood Cliffs, NJ, 1982.

[PV00] C. H. Papadimitriou and S. Vempala. *On the approximability of the traveling salesman problem*. STOC 2000.

[Rag88] P. Raghavan. *Probabilistic construction of deterministic algorithms: Approximating packing integer programs*. Journal of Computer and System Sciences, 37:130-143, 1988.

[Rag94] P. Raghavan. *Randomized approximation algorithms in Combinatorial Optimization*. Lecture Notes in Computer Science, 880:300-317, 1994.

[RagTh87] P. Raghavan and C. D. Thomson. Randomized rounding: A technique for provably good algorithms and algorithmic proofs. Combinatorica, 7:365-374, 1987.

[Schr86] A. Schrijver. *Theory of Linear and Integer Programming*. John Wiley & Sons, 1986.

[Sri99] A. Srinivasan. *Improved approximation guarantees for packing and covering integer programs*. SIAM Journal on Computing 29:648-670,1999.

[Vi64] V. G. Vizing. *On an estimate of the chromatic class of a p-graph* (in Russian). Discret. Analiz., 3:23-30, 1964.

[Yan94] M. Yannakakis. On the approximation of maximum satisfiability. Journal of Algorithms, 3:475-502, 1994.