ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

Τομέας Σημάτων, Ελέγχου και Ρομποτικής
Εργαστήριο Όρασης Υπολογιστών, Επικοινωνίας Λόγου και Επεξεργασίας Σημάτων

Τροπική Γεωμετρική Προσέγγιση Ζωνοτόπων με εφαρμογές στην Συμπίεση Νευρωνικών Δικτύων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

Παναγιώτη Μισιακού

Επιβλέπων: Πέτρος Μαραγκός
Καθηγητής ΕΜΠ

Αθήνα, Οκτώβριος 2021

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
Τομέας Σημάτων, Ελέγχου και Ρομποτικής
Εργαστήριο Όρασης Υπολογιστών, Επικοινωνίας Λόγου και
Επεξεργασίας Σημάτων

Τροπική Γεωμετρική Προσέγγιση Ζωνοτόπων με εφαρμογές στην Συμπίεση Νευρωνικών Δικτύων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

Παναγιώτη Μισιακού

Επιβλέπων: Πέτρος Μαραγκός
Καθηγητής ΕΜΠ

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 11η Οκτωβρίου 2021.

.....	.....	.....
Πέτρος Μαραγκός	Αλέξανδρος Ποταμιάνος	Δημήτριος Φωτάκης
Καθηγητής ΕΜΠ	Αναπληρωτής Καθηγητής ΕΜΠ	Αναπληρωτής Καθηγητής ΕΜΠ

Αθήνα, Οκτώβριος 2021.

.....
Παναγιώτης Μισιακός
Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Σύνοψη

Πρόσφατα, τα νευρωνικά δίκτυα είχαν σπουδαία ανάκαμψη στην Αναγνώριση Προτύπων και την Μηχανική Μάθηση με την προσέλευση των αρχιτεκτονικών βαθιάς μάθησης. Η πρωτοπορία αυτή έχει βελτιώσει τις state-of-the-art επιδόσεις σε ένα ευρύ φάσμα περιοχών εφαρμογής που περιλαμβάνει την Όραση Υπολογιστών και την Επεξεργασία Φυσικής Γλώσσας. Αυτό έχει δώσει κίνητρο στην περαιτέρω κατανόηση της θεωρίας των νευρωνικών δικτύων. Προς αυτή την κατεύθυνση, η παρούσα Διπλωματική εργασία συμβάλλει στην ενίσχυση του θεωρητικού πλαισίου μελέτης των νευρωνικών δικτύων μέσω τροπικής γεωμετρίας. Συγκεκριμένα, αποδεικνύουμε ένα άνω φράγμα στην προσέγγιση τροπικών πολυωνύμων που σχετίζεται με την απόσταση Hausdorff των εν λόγω επεκτεταμένων Newton πολυτόπων τους. Με αυτόν τον τρόπο γενικεύουμε την αμφιμονοσήμαντη αντιστοιχία των γραμμικών περιοχών ενός τροπικού πολυωνύμου με τις κορυφές του άνω φλοιού του επεκτεταμένου Newton πολυτόπου του. Το θεώρημα αυτό επιτυγχάνει την θεμελίωση ενός θεωρητικού πλαισίου κατασκευής αλγορίθμων ελαχιστοποίησης νευρωνικών δικτύων μέσω της γεωμετρικής ελαχιστοποίησης των ζωνοτόπων τους. Υπό αυτό το πρίσμα προτείνουμε 3 γεωμετρικούς αλγορίθμους συμπίεσης Zonotope K-means, Neural Path K-means και Convolutional Neural Path K-means που εφαρμόζουν τον αλγόριθμο συμπίεσης διανυσμάτων K-means. Ο Zonotope K-means περιορίζεται σε νευρωνικά μίας εξόδου, ο Neural Path K-means δεν έχει περιορισμό ως προς τον αριθμό εξόδων και προορίζεται για συμπίεση γραμμικών επιπέδων, ενώ ο Convolutional Neural Path K-means αφορά την συμπίεση συνελικτικών επιπέδων και αποτελεί τον πρώτο αλγόριθμο τροπικής γεωμετρίας που το επιτυγχάνει αυτό. Επίσης, στο πλαίσιο της συμπίεσης γραμμικών επιπέδων νευρωνικών δικτύων προτείνουμε και δύο ακόμη αλγορίθμους τους, AMM και semi-NMF, οι οποίοι εφαρμόζουν αριθμητικές μεθόδους. Ο AMM είναι πιθανοτικός και συμπίεζει το δίκτυο προσεγγίζοντας το γίνομενο των πινάκων δύο γραμμικών επιπέδων, ενώ ο semi-NMF πραγματοποιεί μη αρνητική παραγοντοποίηση πίνακα. Για την θεωρητική ανάλυση των αλγορίθμων μας χρησιμοποιούμε το θεώρημα προσέγγισης τροπικών πολυωνύμων. Επιπλέον, όλοι οι αλγόριθμοι εξετάζονται συγκριτικά με γνωστές τεχνικές συμπίεσης σε πειράματα που αφορούν σύγχρονες αρχιτεκτονικές συνελικτικών δικτύων. Για την εκτέλεση των πειραμάτων κάνουμε χρήση των συνόλων δεδομένων MNIST, Fashion-MNIST και CIFAR στα οποία εκπαιδεύουμε κάποια μικρά συνελικτικά δίκτυα, όπως το LeNet5, αλλά και μεγαλύτερα όπως τα CIFAR-VGG και AlexNet. Τα πειράματα αναδεικνύουν ότι οι μέθοδοι μας παρουσιάζουν βελτίωση έναντι άλλων τροπικών μεθόδων και βασικών μεθόδων συμπίεσης Random και L1 και επιπλέον παρουσιάζουν ανταγωνιστική επίδοση σε σχέση με την πιο σύγχρονης τεχνική συμπίεσης ThiNet.

Λέξεις Κλειδιά: Τροπική Άλγεβρα, Τροπική Γεωμετρία, Μηχανική Μάθηση, Νευρωνικά Δίκτυα, Ζωνότοπα, Προσέγγιση Hausdorff, Ελαχιστοποίηση Νευρωνικών Δικτύων

Abstract

Recently, neural networks have had an impressive comeback to Pattern Recognition and Machine Learning with the advent of deep learning architectures. This breakthrough has advanced the state-of-the-art in a broad spectrum of application areas including Computer Vision and Natural Language Processing. This has motivated an effort to better understand the theory of neural networks. In this direction, the present Diploma thesis contributes to the enhancement of the theoretical framework of neural networks through tropical geometry. In particular, we prove an upper bound on the approximation of tropical polynomials depending on the Hausdorff distance of their respective extended Newton polytopes. This way we generalize the one-to-one correspondence of the linear regions of a tropical polynomial with the vertices of the upper envelope of its extended Newton polytope. Based on this result, we construct tropical geometrical neural network compression algorithms through the geometric minimization of their zonotopes. We propose the geometrical algorithms Zonotope K-means, Neural Path K-means and Convolutional Neural Path K-means which employ the K-means vector compression algorithm. Zonotope K-means is limited to single output networks, Neural Path K-means generalizes to multiclass networks, but still applies for linear layer compression, while Convolutional Neural Path K-means is designed for compression of convolutional layers. In particular, Convolutional Neural Path K-means is the first tropical geometrical algorithm that achieves compression of convolutional layers. Furthermore, we propose two numerical algorithms for linear layer compression, namely AMM and semi-NMF. AMM is probabilistic and compresses the network by approximating the matrix product of the matrices corresponding to two consecutive linear layers, while semi-NMF performs non-negative matrix factorization. The geometrical algorithms and AMM are evaluated theoretically based on the tropical polynomial approximation theorem. Moreover, all of the algorithms are evaluated in comparison to related pruning techniques by conducting compression experiments in modern network architectures. The experiments are performed using MNIST, Fashion-MNIST and CIFAR datasets in which we train some small convolutional networks, such as LeNet5, as well as some larger ones such as CIFAR-VGG and AlexNet. We deduce that our methods show an improvement over other tropical methods and baseline pruning methods Random and L1, while at the same time they perform competitively in comparison with the modern ThiNet compression technique.

Keywords: Tropical Algebra, Tropical Geometry, Machine Learning, Neural Networks, Zonotopes, Hausdorff Approximation, Neural Network Compression

Ευχαριστίες

Στο σημείο αυτό θέλω να ευχαριστήσω όσους ανθρώπους με έχουν βοηθήσει στην μέχρι τώρα πορεία μου και να τους αφιερώσω την εκπόνηση αυτής της διπλωματικής.
Την οικογένεια μου: Τους γονείς μου και τους παππούδες μου που πάντοτε πιστεύουν σε μένα και μου δίνουν θάρρος και κίνητρο να συνεχίζω.
Την Γιώτα, που κάνει κάθε μέρα της ζωής μου χαρούμενη και με βοηθάει να γίνω ένας καλύτερος άνθρωπος.
Τους επιστημονικούς μου συνεργάτες: Τον επιβλέποντα καθηγητή μου κ. Πέτρο Μαραγκό ο οποίος μου έδωσε έμπνευση να καταφέρω να χρησιμοποιήσω τα μαθηματικά που τόσο αγαπώ στον τομέα του Μηχανικού. Τον Γιώργο Σμυρνή, που χωρίς αυτόν η διπλωματική δεν θα γινόταν πραγματικότητα. Ο ίδιος έδειχνε πάντοτε κατανόηση, μου έδινε ενδιαφέρουσες ιδέες και ήταν πάντα εκεί να με βοηθάει και να κάνει την διαδικασία της έρευνας ευχάριστη. Τον Γιώργο Ρετσινά που μου έδωσε χρήσιμες συμβουλές σε τομείς που εως τώρα μου ήταν άγνωστοι.
Τέλος, θα ήθελα να ευχαριστήσω όλους τους φίλους μου και τους ανθρώπους που γνώρισα στο Πολυτεχνείο και έκαναν τα χρόνια των σπουδών μου μία ευχάριστη εμπειρία που θα μείνει στην καρδιά μου.

Περιεχόμενα

Κατάλογος Σχημάτων

Κατάλογος Πινάκων

Κεφάλαιο 1

Εισαγωγή

1.1 Θεωρητικό Πλαίσιο Εργασίας

Η τροπική γεωμετρία [23] είναι ένα μαθηματικό πεδίο βασισμένο στην αλγεβρική γεωμετρία και στενά συνδεδεμένο με την Συνδυαστική Γεωμετρία και την Γεωμετρία Πολυτόπων. Η τροπική γεωμετρία θεμελιώνεται με τον *τροπικό ημιδακτύλιο* ο οποίος παραδοσιακά αναφέρεται στον min-plus ημιδακτύλιο $(\mathbb{R}_{\min}, \wedge, +)$, αλλά μπορεί να αναφέρεται και στον max-plus ημιδακτύλιο [7, 4]. Στην παρούσα εργασία, μάλιστα, ακολουθούμε την εκδοχή του max-plus ημιδακτυλίου $(\mathbb{R}_{\max}, \vee, +)$ ο οποίος ο οποίος αντικαθιστά τις συνήθεις πράξεις της πρόσθεσης και αφαίρεσης με τις πράξεις του μεγίστου και της πρόσθεσης αντίστοιχα. Η παραδοχή αυτή μετατρέπει τις καμπύλες των πολυωνύμων σε τμηματικά γραμμικές καμπύλες και τις γεωμετρικές πολλαπλότητες σε πολύεδρα. Η ιδιότητα αυτή της τροπικής γεωμετρίας της δίνει την δυνατότητα να εφαρμόζεται στην μελέτη των νευρωνικών δικτύων με τμηματικά γραμμικές ενεργοποιήσεις.

Ιστορικά, η τροπική άλγεβρα εισήχθηκε από τον Cuninghame-Green το 1979 με το ονομα minimax algebra, στο πλαίσιο της επιχειρησιακής έρευνας [7]. Ο ίδιος μελέτησε το τροπικό διοειδές και βρήκε λύση στο max-plus σύστημα εξισώσεων. Ο χαρακτηρισμός "τροπική" προέρχεται από τους Γάλλους Μαθηματικούς, μεταξύ των οποίων ήταν οι Dominique Perrin και Jean-Eric Pin. Εκείνοι απέδωσαν αυτό το όνομα προς τιμήν του Βραζιλιάνου Imre Simon, για την συμβολή του ως πρωτοπόρος στο πεδίο με εφαρμογές στα αυτόματα. Επιπλέον, οι συγγραφείς των [25, 26] προτείνουν μία εναλλακτική ερμηνεία βασισμένη στην ελληνική ετυμολογία της λέξης τροπικός, που προέρχεται από το ρήμα "τρέπω", το οποίο σημαίνει στρίβω. Αποδίδουν, λοιπόν, το όνομα αυτό στο γεγονός ότι οι τροπικές καμπύλες είναι τμηματικά ευθύγραμμες και σε ορισμένα σημεία στρίβουν.

Τα τροπικά μαθηματικά έχουν μεγάλο εύρος εφαρμογών το οποίο περιλαμβάνει Θεωρία παιγνίων [1], πλέγματα (lattices) [24, 25], τμηματικά γραμμική προσέγγιση επιφανειών [26], μηχανές πεπερασμένης κατάστασης [39, 40] και κυρτή βελτιστοποίηση και παλινδρόμηση [27, 41]. Πρόσφατα έχει υπάρξει σημαντική συνεισφορά της τροπικής γεωμετρίας στην μελέτη των νευρωνικών δικτύων και της μηχανικής μάθησης [25]. Στο [45] αποδεικνύεται η ισοδυναμία των ReLU ενεργοποιημένων νευρωνικών δικτύων με τροπικές ρητές απεικονίσεις. Επιπλεόν, παρατηρείται ότι τα ζωνότοπα αποτελούν την γεωμετρική δομή που αναπαριστά ένα επίπεδο του δικτύου. Με υπολογισμό των κορυφών του ζωνοτόπου γίνεται υπολογισμός άνω φράγματος στον αριθμό των γραμμικών περιοχών ενός νευρωνικού δικτύου. Ο υπολογισμός αυτός είχε προηγηθεί στο [29] χωρίς την χρήση τροπικής γεωμετρίας. Η τεχνική υπολογισμού γραμμικών περιοχών γενικεύεται στο [6] όπου γίνεται υπολογισμός άνω φράγματος στον αριθμό των γραμμικών περιοχών διαφορετικών ειδών επιπέδων νευρωνικών δικτύων, όπως είναι τα συνελικτικά και τα MaxOut επίπεδα. Μάλιστα, προτείνουν και έναν πιθανοτικό αλγόριθμο καταμέτρησης των γραμμικών περιοχών.

Η θεωρία των Πλεγμάτων είναι στενά συνδεδεμένη με μία πιο μοντέρνα κατηγορία Νευρωνικών Δικτύων, τα Μορφολογικά Νευρωνικά Δίκτυα [33, 34, 32, 44]. Τέτοια περίπτωση δικτύων μελετάται στο [5], όπου για έναν νευρώνα perceptron προτείνεται εκπαίδευση με αλγόριθμο convex-concave που προορίζεται εν γένει για συναρτήσεις που αποτελούν διαφορά κυρτών συναρτήσεων (Difference of Convex Programming). Στην εργασία [8] εξετάζονται τα μορφολογικά δίκτυα ως προς την εκπαίδευση και προτείνεται μία διαδικασία εκπαίδευσης για multiclass tasks. Επίσης, αναδεικνύεται η δυνατότητα συμπίεσης μορφολογικών δικτύων και εξετάζεται πως η αρχιτεκτονική τέτοιων δικτύων μπορεί να επιβάλλει ιδιότητες όπως η μονοτονία. Τέλος, στο [28] αναπτύσσεται η μορφολογική εκδοχή των Συνελικτικών Νευρωνικών Δικτύων με στόχο την οπτική αναγνώριση ψηφίων.

Στη δική μας εργασία θα εστιάσουμε σε Μηχανική Μάθηση και συγκεκριμένα στην

μελέτη των ιδιοτήτων των Νευρωνικών Δικτύων [13]. Τα μαθηματικά εργαλεία που θα χρησιμοποιήσουμε προέρχονται από την τροπική γεωμετρία [23], η οποία θα αποτελέσει ένα μέσο γεωμετρικής οπτικοποίησης των νευρωνικών δικτύων. Στόχος μας είναι να κατασκευάσουμε έναν γεωμετρικό τρόπο συμπίεσης των νευρωνικών δικτύων. Η συμπίεση αυτή θα αφορά την γεωμετρική αναπαράσταση των δικτύων που κατασκευάζεται μέσω των ζωνοτόπων. Ωστόσο, για να μπορέσει να αξιολογηθεί η ποιότητα της συμπίεσης, θα πρέπει να γνωρίζουμε τι σφάλμα επιφέρει στην πράξη η γεωμετρική συμπίεση των ζωνοτόπων. Η θεωρητική συμβολή αυτή είναι και η βασική συνεισφορά αυτής της εργασίας. Σημειώνουμε ότι η εργασία βασίζεται εκτενώς στα [38, 37, 2], και τις επεκτείνει επιτυγχάνοντας βελτιωμένη συμπίεση αλλά και συμπίεση σε συνελικτικά επίπεδα νευρωνικών δικτύων.

Η συμπίεση νευρωνικών δικτύων [3, 31] είναι ένας τομέας που έχει παρουσιάσει πρόσφατα έντονη ανάπτυξη λόγω της ανάγκης για συμπίεση των μεγάλων συνελικτικών δικτύων, αλλά και της εκπληκτικής ικανότητας ορισμένων τεχνικών να συμπιέζουν ένα νευρωνικό δίκτυο χωρίς αυτό να έχει επίπτωση στην επίδοση του. Δεδομένου ότι η τροπική γεωμετρία εκφράσει τις μαθηματικές ιδιότητες ενός νευρωνικού δικτύου, είναι εύλογο να αποτελεί εργαλείο και για την συμπίεση του. Πράγματι, στο [2] προτείνουν έναν αλγόριθμο συμπίεσης του νευρωνικού βασισμένο στους πίνακες που αναπαριστούν τα ζωνότοπα του δικτύου, ο οποίος επιχειρεί να διατηρήσει τις υπερεπιφάνειες διαχωρισμού του. Επιπλέον, στο [38] προτείνεται μία καινοτόμος μέθοδος τροπικής διαίρεσης τροπικών πολυωνύμων έχοντας εφαρμογές στην συμπίεση νευρωνικών δικτύων μίας εξόδου. Η μέθοδος αυτή επεκτείνεται καλύπτοντας και την περίπτωση των δικτών πολλών εξόδων [37].

Στην βιβλιογραφία της συμπίεσης νευρωνικών υπάρχουν πολλές μέθοδοι που εφαρμόζουν περίτεχνες ιδέες [3]. Μάλιστα υπάρχουν και υποθέσεις [12] για την ύπαρξη υπο-δικτύων με λιγότερες συνδέσεις που αποδίδουν εξίσου καλά με το αρχικό. Εμείς δεν θα επιχειρήσουμε να υλοποιήσουμε κάποια πρωτοπόρο μέθοδο συμπίεσης που να ξεπερνά όλες τις υπόλοιπες, αλλά θα κάνουμε συγκρίσεις με παρόμοιες μεθόδους για να αναδείξουμε ότι οι προτεινόμενες μέθοδοι πράγματι συμπιέζουν τα δίκτυα και έχουν την προοπτική να γίνουν ανταγωνιστικές με περαιτέρω μελέτη. Η εργασία μας θα εστιάσει στην εξέταση των τεχνικών συμπίεσης υπό το πρίσμα της τροπικής γεωμετρίας.

1.2 Θεωρητική Συμβολή Εργασίας

Στην παρούσα εργασία επιχειρούμε να ενισχύσουμε την θεωρητική υπόβαθρο της τροπικής γεωμετρίας για την μελέτη των νευρωνικών δικτύων. Η συμβολή μας για αυτόν τον σκοπό συνίσταται στα εξής.

- Αρχικά, αποδεικνύουμε ένα θεωρητικό άνω φράγμα στο σφάλμα της προσέγγισης ενός τροπικού πολυωνύμου από ένα άλλο, σχετίζοντας το με την Hausdorff απόσταση μεταξύ των επεκτεταμένων Newton πολυτόπων που τους αντιστοιχούν. Το Θεώρημα αυτό αποτελεί γενίκευση της εως τώρα γνωστής ένα προς ένα αντιστοιχίας των γραμμικών περιοχών ενός τροπικού πολυωνύμου με τις κορυφές του άνω φλοιού του επεκτεταμένου Newton πολυτόπου του.

- Προτείνουμε 3 νέους αλγορίθμους συμπίεσης νευρωνικών δικτύων Zonotope K-means, Neural Path K-means, Convolutional Neural Path K-means, και επιπλέον αναλύουμε τους αλγορίθμους AMM και semi-NMF. Οι πρώτοι 3 αλγόριθμοι είναι καθαρά γεωμετρικοί και βασίζονται στην προσέγγιση ενός ζωνοτόπου με χρήση λιγότερων γεννητόρων. Οι υπόλοιποι δύο αλγόριθμοι είναι αριθμητικής φύσεως. Ο AMM είναι πιθανοτικός

και προσεγγίζει το γινόμενο των πινάκων δύο διαδοχικών γραμμικών επιπέδων, ενώ ο semi-NMF εφαρμόζει μη-αρνητική παραγοντοποίηση πίνακα. Σημειώνουμε, ότι ο AMM παρουσιάζει θεωρητικό ενδιαφέρον στην μελέτη του μέσω τροπικής γεωμετρίας. Επιπλέον, τονίζουμε ότι ο Convolutional Neural Path K-means συμπιέζει συνελικτικά επίπεδα (convolutional layers), ενώ οι υπόλοιποι γραμμικά επίπεδα (fully connected layers) [1]. Μάλιστα, αποτελεί τον πρώτο αλγόριθμο τροπικής γεωμετρίας στην βιβλιογραφία που επιτυγχάνει αυτό το ζητούμενο.

- Τέλος, παραθέτουμε μέσω της τροπικής γεωμετρίας και του Θεωρήματος προσέγγισης τροπικών πολυωνύμων, θεωρητική ανάλυση για τον υπολογισμό άνω φράγματος στο σφάλμα της προσέγγισης των αλγορίθμων Zonotope K-means, Neural Path K-means, Convolutional Neural Path K-means και AMM. Για την εμπειρική εξέταση των αλγορίθμων εκτελούμε πειράματα συμπίεσης συνελικτικών δικτύων (CNN2D, LeNet5, CIFAR-VGG και AlexNet) τα οποία αναφέρονται είτε στα γραμμικά επίπεδα του δικτύου είτε στα συνελικτικά. Τα πειράματα μας αφορούν τα σύνολα δεδομένων MNIST, Fashion-MNIST και CIFAR. Συμπεραίνουμε ότι επιτυγχάνουμε βελτίωση έναντι άλλων τροπικών μεθόδων στην βιβλιογραφία αλλά και ανταγωνιστική επίδοση έναντι γενικότερων μεθόδων ελαχιστοποίησης νευρωνικών δικτύων (Random, L1 και ThiNet).

1.3 Συμβολισμός

Για την μαθηματική ανάλυση της παρούσας εργασίας θα κάνουμε χρήση των εξής συμβολισμών:

- Με bold-faced, πεζά γράμματα, π.χ. $\mathbf{v}$ αναπαριστούμε διανύσματα.
- Με κεφαλαία γράμματα, π.χ. A αναπαριστούμε πίνακες ή σύνολα ανάλογα με τα συμφραζόμενα.
- Με $A_{i,:}$ συμβολίζουμε την $i-$οστή γραμμή του πίνακα A, ενώ με $A_{:,j}$ την $j-$οστή στήλη του.
- Με A_{ij} συμβολίζουμε το (i, j) στοιχείο του A.
- Μεταξύ δύο γεωμετρικών συνόλων $P, Q \subseteq \mathbb{R}^d$ συμβολίζουμε με $\oplus$ το Minkowski άθροισμα των συνόλων που ορίζεται ως

$$P \oplus Q = \{\mathbf{u} + \mathbf{v} | \, \mathbf{u} \in P, \, \mathbf{v} \in Q\}$$

- Με $\text{conv}(A)$ συμβολίζουμε το κυρτό περίβλημα του $A \subseteq \mathbb{R}^d$.
- Ο συμβολισμός $\|\mathbf{v}\|$ αναπαριστά την ℓ_2 νόρμα του διανύσματος $\mathbf{v}$, ενώ ο συμβολισμός $\|\mathbf{v}\|_1$ την ℓ_1 νόρμα.
- Επίσης με $\|A\|$ συμβολίζουμε την νόρμα Frobenius του πίνακα A.
- Για φυσικό n συμβολίζουμε $[n] = \{1, 2, ..., n\}$.

[1] Ένα Fully Connected Layer δέχεται ως είσοδο το διάνυσμα $\mathbf{x}$ και δίνει στην έξοδο το διάνυσμα $A\mathbf{x} + \mathbf{b}$ που προκύπτει από τον γραμμικό μετασχηματισμό του πίνακα βαρών A και του bias $\mathbf{b}$ του επιπέδου.

1.4 Δομή Εργασίας

Η διπλωματική αυτή διαρθρώνεται στα Κεφάλαια με τον εξής τρόπο.

- Στο Κεφάλαιο 2 γίνεται η θεωρητική θεμελίωση της εργασίας με χρήση τροπικής γεωμετρίας. Αρχικά γίνεται μία ανασκόπηση των βασικών ιδιοτήτων των τροπικών πολυωνύμων και της σύνδεσής τους με τη γεωμετρία πολυτόπων. Έπειτα, παρουσιάζεται το Θεώρημα προσέγγισης τροπικών πολυωνύμων μέσω της απόστασης Hausdorff των αντιστοίχων πολυτόπων. Επιπλέον, παρουσιάζονται και κάποια θεωρητικά στοιχεία που έχουν τον ρόλο ανασκόπησης της πρόσφατης βιβλιογραφίας σχετικά με το πλήθος εδρών ή κορυφών πολυτόπων και διαίρεση τροπικών πολυωνύμων.

- Στο Κεφάλαιο 3 παρουσιάζεται η πρόσφατη συμβολή της τροπικής γεωμετρίας στην μελέτη των ιδιοτήτων των νευρωνικών δικτύων. Συγκεκριμένα, παρουσιάζουμε την ισοδυναμία των νευρωνικών δικτύων με τροπικές ρητές συναρτήσεις και την γεωμετρική μελέτη μέσω ζωνοτόπων. Στο σημείο αυτό προτείνουμε μία επέκταση στον υπολογισμό των κορυφών του ζωνοτόπου που αντιστοιχεί σε ένα δίκτυο, η οποία θα μας είναι χρήσιμη στο Κεφάλαιο 4. Τέλος, για τον ενδιαφερόμενο αναγνώστη παρατίθενται ορισμένα αποτελέσματα σε σχέση με την καταμέτρηση γραμμικών περιοχών νευρωνικών δικτύων, η οποία δεν είναι ο κύριος σκοπός του συγγράματος.

- Στο Κεφάλαιο 4 παρουσιάζεται η βασική συμβολή της εργασίας μας. Με βάση τα θεωρητικό αποτέλεσμα προσέγγισης τροπικών πολυωνύμων κατασκευάζουμε αλγορίθμους συμπίεσης (Zonotope K-means, Neural Path K-means και Convolutional Neural Path K-means) γραμμικών και συνελικτικών επιπέδων νευρωνικών δικτύων οι οποίοι προσεγγίζουν τα ζωνότοπα του δικτύου με χρήση K-means. Οι αλγόριθμοι συμπίεσης αναλύονται θεωρητικά και υπολογίζεται άνω φράγμα για το σφάλμα της προσέγγισης.

- Στο Κεφάλαιο 5 παρουσιάζουμε δύο εναλλακτικούς αλγορίθμους (AMM και semi-NMF) συμπίεσης για γραμμικά επίπεδα νευρωνικών οι οποίοι είναι αριθμητικοί και δεν βασίζονται στην γεωμετρία του δικτύου. Ο πρώτος αλγόριθμος προσεγγίζει το γινόμενο των πινάκων δύο διαδοχικών γραμμικών επιπέδων και η απόδοση του αναλύεται θεωρητικά με τροπική γεωμετρία. Ο δεύτερος αλγόριθμος βασίζεται σε μη-αρνητική παραγοντοποίηση πινάκων.

- Στο Κεφάλαιο 6 παραθέτουμε πειράματα για την εξέταση της απόδοσης των προτεινόμενων αλγορίθμων. Οι αλγόριθμοι Zonotope K-means, Neural Path K-means, Convolutional Neural Path K-means, AMM και semi-NMF αξιοποιούνται για την συμπίεση των νευρωνικών CNN2D, LeNet5, CIFAR-VGG και AlexNet σε σύνολα δεδομένων όπως τα MNIST, Fashion-MNIST και CIFAR. Οι αλγόριθμοι συγκρίνονται με αντίστοιχες τροπικές μεθόδους αλλά και γενικότερες μεθόδους pruning (ThiNet, Random και L1).

- Τέλος, στο Κεφάλαιο 7 παρουσιάζουμε εν συντομία τα συμπεράσματα της διπλωματικής, καθώς και πιθανές μελλοντικές επεκτάσεις αυτής.

Κεφάλαιο 2

Τροπική Γεωμετρία

Η τροπική γεωμετρία είναι μία εκδοχή της αλγεβρικής γεωμετρίας η οποία βασίζεται στον τροπικό ημιδακτύλιο. Στην εργασία μας θα μελετήσουμε την τροπική γεωμετρία υπό το πρίσμα του *max-plus ημιδακτυλίου* $(\mathbb{R}_{\max}, \vee, +)$ ο οποίος ορίζεται ως το σύνολο $\mathbb{R}_{\max} = \mathbb{R} \cup \{-\infty\}$ εφοδιασμένο με τις πράξεις $(\vee, +)$. Ο max-plus ημιδακτύλιος αντικαθιστά την πράξη της πρόσθεσης με μέγιστο $a \vee b = \max(a, b)$ και του πολλαπλασιασμού με κλασσική πρόσθεση $a + b$. Η δομή $(\mathbb{R}_{\max}, \vee, +)$ αποτελεί ημιδακτύλιο (semiring) και μάλιστα ημι-σώμα (semifield) αφού απουσιάζει ο αντίστροφος της πρόσθεσης.

Όπως θα δούμε η επιλογή του max-plus ημιδακτυλίου είναι σημαντική για την περίπτωση της μελέτης των νευρωνικών δικτύων. Αξίζει ωστόσο να σημειώσουμε ότι οι προτάσεις που θα παρουσιαστούν μπορούν χωρίς πολύ κόπο να μεταφερθούν και στην min-plus εκδοχή του ημιδακτυλίου. Κατά σύμβαση, θα αναφερόμαστε στον max-plus ημιδακτύλιο ως τροπικό ημιδακτύλιο.

2.1 Τροπικά Πολυώνυμα

Στον τροπικό ημιδακτύλιο τα τροπικά πολυώνυμα μπορούν να οριστούν με τρόπο όπως και στην κλασσική άλγεβρα. Τα τροπικά πολυώνυμα θα αξιοποιηθούν στην μελέτη των νευρωνικών δικτύων, οπότε είναι σημαντικό να ακολουθήσουμε την πολυμεταβλητή εκδοχή τους αφού στα νευρωνικά η είσοδος είναι εν γένει πολυδιάστατη.

Ορισμός. Ένα *τροπικό πολυώνυμο* f με d μεταβλητές $\mathbf{x} = (x_1, x_2, ..., x_d)^T$ ορίζεται ως η συνάρτηση

$$f(\mathbf{x}) = \max_{\mathbf{a} \in A} \{\mathbf{a}^T \mathbf{x} + c_{\mathbf{a}}\} \tag{2.1}$$

όπου A είναι ένα πεπερασμένο σύνολο διανυσμάτων στον $\mathbb{R}^d$ και $c_{\mathbf{a}}$ είναι ο αντίστοιχος συντελεστής μονωνύμων με τιμές στο $\mathbb{R}_{\max} = \mathbb{R} \cup \{-\infty\}$. Το σύνολο αυτών των πολυωνύμων αποτελεί τον ημιδακτύλιο $\mathbb{R}_{\max}[\mathbf{x}]$ των τροπικών πολυωνύμων.

Αξίζει να παρατηρήσουμε ότι κάθε όρος $\mathbf{a}^T \mathbf{x} + c_{\mathbf{a}}$ του πολυωνύμου αντιστοιχεί σε ένα υπερεπίπεδο του $\mathbb{R}^d$. Με αυτό το σκεπτικό, μπορούμε να αποκαλέσουμε τα διανύσματα $\mathbf{a} \in A$ τις *κλίσεις* του τροπικού πολυωνύμου και τον συντελεστή $c_{\mathbf{a}}$ τον αντίστοιχο *σταθερό όρο*.

Σημειώνουμε ότι, οι κλίσεις παίρνουν οποιαδήποτε τιμή στον $\mathbb{R}^d$ και όχι στο πλέγμα των μη-αρνητικών ακεραίων, όπως συμβαίνει με τους εκθέτες των κλασσικών πολυμεταβλητών πολυωνύμων. Στην πραγματικότητα τα πολυώνυμα αυτά συναντώνται στην βιβλιογραφία με τον όρο *signomials* [11], αλλά εμείς θα τα χρησιμοποιούμε αναφερόμενοι σε αυτά ως πολυώνυμα.

Με βάση τον ορισμό, τα πολυώνυμα στον $\mathbb{R}_{\max}[\mathbf{x}]$ διαθέτουν τις εξής ιδιότητες

$$f, g \in \mathbb{R}_{\max}[\mathbf{x}] \Rightarrow f \vee g = \max(f, g) \in \mathbb{R}_{\max}[\mathbf{x}]$$
$$f, g \in \mathbb{R}_{\max}[\mathbf{x}] \Rightarrow f + g \in \mathbb{R}_{\max}[\mathbf{x}]$$

Δηλαδή, ο $\mathbb{R}_{\max}[\mathbf{x}]$ είναι κλειστός ως προς την πράξη της πρόσθεσης και του πολλαπλασιασμού. Ωστόσο, δεν ισχύει το ίδιο για τις πράξεις της δύναμης και της σύνθεσης. Αυτό συμβαίνει επειδή δεχτήκαμε ότι οι κλίσεις παίρνουν τιμές στον $\mathbb{R}^d$. Συγκεκριμένα ισχύει

$$f^a = af \in \mathbb{R}_{\max}[\mathbf{x}] \Leftrightarrow a \in \mathbb{R}_{\geq 0}$$
$$f, g \in \mathbb{R}_{\max}[\mathbf{x}] \not\Rightarrow f \circ g \in \mathbb{R}_{\max}[\mathbf{x}]$$

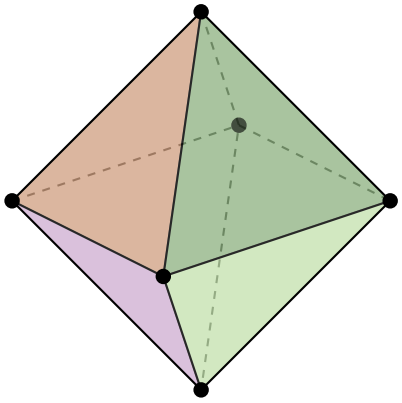
Δηλαδή, η τροπική δύναμη ενός πολυωνύμου δίνει τροπικό πολυώνυμο μόνο όταν είναι θετική. Αυτό πρακτικά σημαίνει ότι αν πολλαπλασιάσουμε ένα τροπικό πολυώνυμο με μία αρνητική σταθερά το αποτέλεσμα θα είναι ο αντίθετος ενός τροπικού πολυωνύμου, που δεν είναι τροπικό πολυώνυμο σύμφωνα με τον max-plus ορισμό. Πράγματι, ένα τροπικό max-plus πολυώνυμο πολλαπλασιασμένο με αρνητική σταθερά δεν μπορεί να γραφεί σαν μέγιστο γραμμικών όρων, αφού το $-$ δεν αντιμετατίθεται με τον max τελεστή. Ομοίως, η σύνθεση δύο τροπικών πολυωνυμικών απεικονίσεων $f \circ g$ δεν μας δίνει απαραίτητα τροπική πολυωνυμική απεικόνιση, διότι μπορεί η f να έχει αρνητικές κλίσεις, οπότε θα πάρουμε γραμμικούς συνδυασμούς των όρων της g που πιθανόν να έχουν αρνητικό πρόσημο και που δεν αντιμετατίθονται με τον max τελεστή. Όπως, θα δούμε στο Κεφάλαιο 3 η σύνθεση δύο τροπικών πολυωνυμικών απεικονίσεων μας δίνει μία τροπική ρητή απεικόνιση.

Τέλος, τονίζουμε ότι τα τροπικά πολυώνυμα είναι κυρτές και τμηματικά γραμμικές συναρτήσεις. Η παρατήρηση αυτή αποτελεί έναυσμα για την αποκάλυψη πολλών σημαντικών ιδιοτήτων των τροπικών πολυωνύμων που θα δούμε στην συνέχεια.

2.2 Πολύτοπα

Τα πολύτοπα θα είναι το βασικό εργαλείο της μελέτης μας και θα μας επιτρέψουν να οπτικοποιήσουμε γεωμετρικά τις ιδιότητες των τροπικών πολυωνύμων. Αυτά, ως γεωμετρικές δομές, έχουν μελετηθεί εκτενώς [46, 16] και παρέχουν έναν ενδιαφέρον και εύχρηστο τρόπο οπτικοποίησης με εφαρμογές σε πεδία όπως ο Γραμμικός Προγραμματισμός και η Βελτιστοποίηση. Θα ξεκινήσουμε την ανάλυσή μας παραθέτοντας τους ορισμούς και κάποιες σπουδαίες ιδιότητες που θα μας επιτρέψουν να τα χειριστούμε.

Ένα πολύτοπο, διαισθητικά, είναι ένα κυρτό γεωμετρικό στερεό σε οποιονδήποτε αριθμό διαστάσεων. Πρακτικά, τα πολύτοπα γενικεύουν τα γνωστά πολύγωνα (2 διαστάσεις) και πολύεδρα (3 διαστάσεις). Παρακάτω απεικονίζεται ένα πολύεδρο με 8 κορυφές, γνωστό και ως οκτάεδρο.



Σχήμα 2.1: Παράδειγμα Πολυτόπου (Οκτάεδρο) σε 3 διαστάσεις

Εν συνεχεία, ορίζουμε φορμαλιστικά το πολύτοπο. Ο ορισμός του μπορεί να γίνει με δύο τρόπους. Είτε ως το σύνολο των κυρτών συνδυασμών ενός πεπερασμένου συνόλου σημείων είτε ως την φραγμένη τομή ενός πεπερασμένου συνόλου ημιχώρων. Όπως θα δούμε οι δύο ορισμοί είναι ισοδύναμοι και περιγράφουν την ίδια γεωμετρική δομή που ονομάσαμε πολύτοπο.

Ορισμός. ($\mathcal{V}$-αναπαράσταση κυρτού Πολυτόπου) Ένα κυρτό πολύτοπο $P \subset \mathbb{R}^d$ ορίζεται ως το σύνολο των κυρτών συνδυασμών ενός πεπερασμένου συνόλου δοθέντων σημείων

$$P = \left\{ \sum_{i=1}^{m} \lambda_i \mathbf{v}_i : \sum_{i=1}^{m} \lambda_i = 1,\ \lambda_i \geq 0 \right\}$$

όπου $v_1, ..., v_m \in \mathbb{R}^d$.

Αξίζει να παρατηρήσουμε ότι οι κορυφές του πολυτόπου ανήκουν στο σύνολο των m διανυσμάτων $\mathbf{v}_1, ..., \mathbf{v}_m$. Ωστόσο, δεν είναι απαραίτητο ότι κάθε ένα από τα m διανύσματα ότι είναι κορυφή του πολυτόπου, αφού μπορεί να βρεθεί στο εσωτερικό του. Για πληρότητα παραθέτουμε έναν ορισμό για την ένοια της κορυφής του πολυτόπου.

Ορισμός. (Κορυφή Πολυτόπου) Ένα σημείο $\mathbf{v}$ που ανήκει σε ένα πολύτοπο P είναι κορυφή εάν για κάθε διάνυσμα $\mathbf{x} \in \mathbb{R}^d$ είτε $\mathbf{v} - \mathbf{x} \notin P$ ή $\mathbf{v} + \mathbf{x} \notin P$.

Επιπλέον, ένα πολύτοπο μπορεί να περιγραφεί από τον ακόλουθο ορισμό που αφορά ημιχώρους.

Ορισμός. ($\mathcal{H}$-αναπαράσταση κυρτού Πολυτόπου) Ένα κυρτό πολύτοπο P ορίζεται ως η φραγμένη τομή ενός πεπερασμένου συνόλου ημιχώρων

$$\{\mathbf{x} : A\mathbf{x} \leq \mathbf{b}\}$$

όπου $A \in \mathbb{R}^{m \times n}, \mathbf{b} \in \mathbb{R}^d$ και ο ημιχώρος είναι φραγμένος, δηλαδή υπάρχει υπερσφαίρα πεπερασμένης ακτίνας που τον καλύπτει.

Όπως προαναφέραμε, οι δύο ορισμοί είναι ισοδύναμοι, όπως περιγράφει το ακόλουθο θεώρημα.

Θεώρημα 2.1. *(Θεώρημα Αναπαράστασης Πολυτόπων [15, 46]) Ένα σύνολο P (πολύτοπο) μπορεί να γραφεί σε $\mathcal{V}-$αναπαράσταση αν και μόνον αν μπορεί να γραφεί σε $\mathcal{H}-$αναπαράσταση.*

Εν συνεχεία θα παρουσιάσουμε τις δομές που συνδέουν τα τροπικά πολυωνύμα με την γεωμετρία πολυτόπων. Αυτές είναι το πολυτόπου Newton και η επεκτεταμένη του μορφή. Συγκεκριμένα, υποθέτουμε ότι έχουμε το πολυώνυμο f της σχέσης (2.1). Τότε για το πολυώνυμο αυτό έχουμε τους εξής ορισμούς.

Ορισμός. (Πολύτοπο Newton) Το *πολύτοπο Newton* που αντιστοιχεί στο πολυώνυμο f, ορίζεται ως το κυρτό περίβλημα των κλίσεων του πολυωνύμου.

$$\mathrm{Newt}\,(f) := \mathrm{conv}\{\mathbf{a} : \mathbf{a} \in A\}$$

Ορισμός. Το *επεκτεταμένο πολύτοπο Newton* του πολυωνύμου f, ορίζεται ως το κυρτό περίβλημα των κλίσεων του πολυωνύμου επαυξημένους με τους αντίστοιχους σταθερούς όρους.

$$\mathrm{ENewt}\,(f) := \mathrm{conv}\{(\mathbf{a}^T, c_{\mathbf{a}}) : \mathbf{a} \in A\}$$

Τα πολύτοπα, όπως και κάθε άλλο σύνολο στον $\mathbb{R}^d$ επιδέχονται την πράξη του αθροίσματος Minkowski. Αυτή θα μας διευκολύνει να συνδέσουμε τις τροπικές πράξεις στα τροπικά πολυώνυμα με γεωμετρικές πράξεις πολυτόπων.

Ορισμός. Για δύο σύνολα A, B στον $\mathbb{R}^d$ ορίζουμε το *Minkowski άθροισμα* τους ως το σύνολο:

$$A \oplus B = \{\mathbf{a} + \mathbf{b} : \mathbf{a} \in A,\, \mathbf{b} \in B\}$$

Παρατήρηση. *Αξίζει να σημειώσουμε ότι το Minkowski άθροισμα δύο πολυτόπων είναι και αυτό πολύτοπο ([16], Κεφ. 3, Πρόταση 4).*

Όπως είδαμε προηγουμένως, η τροπική πρόσθεση και ο τροπικός πολλαπλασιασμός μεταξύ πολυωνύμων στον $\mathbb{R}_{\max}[\mathbf{x}]$ μας δίνει επίσης τροπικό πολυώνυμο. Αξίζει, λοιπόν, να μελετήσουμε το πως προκύπτει το πολύτοπο του τελικού πολυωνύμου από τα πολύτοπα των αρχικών. Η παρακάτω πρόταση θα μας φανεί χρήσιμη για αυτόν τον σκοπό.

Πρόταση 2.1. *([6, 45]) Έστω $f, g \in \mathbb{R}_{max}[\mathbf{x}]$ δύο τροπικά πολυώνυμα. Τότε για τα επεκτεταμένα Newton πολύτοπα ισχύει ότι*

$$\mathrm{ENewt}\,(f \vee g) = \mathrm{conv}\{\mathrm{ENewt}\,(f) \cup \mathrm{ENewt}\,(g)\}$$
$$\mathrm{ENewt}\,(f + g) = \mathrm{ENewt}\,(f) \oplus \mathrm{ENewt}\,(g)$$

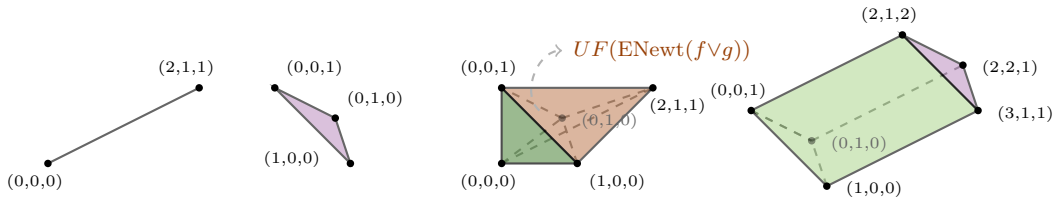
Απόδειξη. Για την πρώτη σχέση, παρατηρούμε ότι το $f \vee g$ είναι το max όλων των γραμμικών όρων των f, g. Επομένως, το επεκτεταμένο Newton πολύγωνο προκύπτει από την ένωση των κορυφών που αντιστοιχούν στα δύο πολυώνυμα. Συνεπώς θα είναι το κυρτό περίβλημα της ένωσης των κορυφών, ή ισοδύναμα το κυρτό περίβλημα της ένωσης των δύο επεκτεταμένων Newton πολυτόπων.

Για την δεύτερη σχέση γράφουμε αρχικά

$$f(\mathbf{x}) = \max_i\{\mathbf{a}_i^T\mathbf{x} + b_i\},\ \ g(\mathbf{x}) = \max_j\{\mathbf{c}_j^T\mathbf{x} + d_j\}$$

και παρατηρούμε ότι

$$f + g = \max_{i,j}\{\mathbf{a}_i^T\mathbf{x} + b_i + \mathbf{c}_j^T\mathbf{x} + d_j\} = \max_{i,j}\left\{(b_i + d_j) + (\mathbf{a}_i + \mathbf{c}_j)^T\mathbf{x}\right\}$$



ENewt (f) ENewt (g) ENewt $(f \vee g)$ ENewt $(f + g)$

Σχήμα 2.2: Αναπαράσταση των πολυτόπων που προκύπτουν μέσω τροπικών πράξεων. Η τροπική πρόσθεση $(\vee)$ ανστιστοιχεί στο κυρτό περίβλημα της ένωσης των δύο πολυτόπων και ο τροπικός πολλαπλασιασμός $(+)$ στο άθροισμα Minkowski των εν λόγω πολυτόπων.

Επομένως,

$$\text{ENewt}\,(f+g) = \text{conv}_{i,j}\{(\mathbf{a}_i^T + \mathbf{c}_j^T, b_i + d_j)\} = \text{ENewt}\,(f) \oplus \text{ENewt}\,(g)$$

Στην τελευταία σχέση χρησιμοποιήσαμε την ισότητα $\text{conv}(A \oplus B) = \text{conv}(A) \oplus \text{conv}(B)$ (Θεώρημα 1.1.2 [35]). □

Πόρισμα 2.1. *Η παραπάνω πρόταση μπορεί να γενικευθεί σε οποιονδήποτε πεπερασμένο αριθμό πολυωνύμων με επαγωγή. Έστω $\mathcal{F} \subset \mathbb{R}_{max}[\mathbf{x}]$ μία πεπερασμένη οικογένεια πολυωνύμων, τότε*

$$\text{ENewt}\left(\bigvee_{f\in\mathcal{F}} f\right) = \text{conv}\left\{\bigcup_{f\in\mathcal{F}} \text{ENewt}\,(f)\right\}$$

$$\text{ENewt}\left(\sum_{f\in\mathcal{F}} f\right) = \bigoplus_{f\in\mathcal{F}} \text{ENewt}\,(f)$$

Παράδειγμα 2.1. *Θεωρούμε τα ακόλουθα τροπικά πολυώνυμα 2 μεταβλητών*

$$f(x,y) = \max(2x + y + 1, 0), \quad g(x,y) = \max(x, y, 1)$$

Τότε οι τροπικές πράξεις μεταξύ αυτών δίνουν

$$f \vee g = \max(2x + y + 1, 0, x, y, 1)$$
$$f + g = \max(3x + y + 1, x, 2x + 2y + 1, y, 2x + y + 2, 1)$$

Στο Σχήμα 2.2 απεικονίζονται τα επεκτεταμένα Newton πολύτοπα των αρχικών πολυωνύμων, καθώς και αυτών που προκύπτουν από τις τροπικές πράξεις.

Τα Newton πολύτοπα προσφέρουν μία γεωμετρική περιγραφή για τα τροπικά πολυώνυμα. Στην συνέχεια θα παρουσιάσουμε τις προτάσεις που τεκμηριώνουν τον ακριβή τρόπο σύνδεσης των ιδιοτήτων των τροπικών πολυωνύμων με τα Newton πολύτοπα. Για τον σκοπό αυτό θα χρειαστούμε τον ορισμό του άνω φλοιού.

Ορισμός. Θεωρούμε ένα τροπικό πολυώνυμο f με d μεταβλητές. Ο *άνω φλοιός* του επεκτεταμένου Newton πολυτόπου ENewt (f) συμβολίζεται με $UF(\text{ENewt}\,(f))$ και αναπαριστά το σύνολο των ανώτατων σημείων του ENewt (f) ως προς την τελευταία διάσταση του $\mathbb{R}^{d+1}$.

$$UF(\text{ENewt}\,(f)) = \{(\mathbf{x}, x_{d+1}) \in \text{ENewt}\,(f) : (\mathbf{y}, y_{d+1}) \in \text{ENewt}\,(f) \Rightarrow y_{d+1} \leq x_{d+1}\}$$

Πρόταση 2.2. *([47], Θεώρημα 2.10) Οι τιμές που λαμβάνει το $f(x)$ καθορίζονται πλήρως από τις κορυφές του άνω φλοιού $UF(\mathrm{ENewt}(f))$ του επεκτεταμένου Newton Πολυγώνου.*

Απόδειξη. Έστω ότι $f(x) = \mathbf{a}^T\mathbf{x} + b$ για κάποιο $\mathbf{x} \in \mathbb{R}^d$ και $(\mathbf{a}^T, b)$ όχι κορυφή του $UF(\mathrm{ENewt}(f))$. Τότε, υπάρχει $b' \geq b$ ώστε $(\mathbf{a}^T, b') \in UF(\mathrm{ENewt}(f))$. Μάλιστα, το σημείο αυτό αφού βρίσκεται σε κάποια έδρα (face) του άνω φλοιού του πολυτόπου, μπορεί να γραφεί σαν γραμμικός συνδυασμός των κορυφών που την αποτελούν. Αν είναι, λοιπόν, $\mathbf{v}_1, ..., \mathbf{v}_m$ οι κορυφές της έδρας τότε

$$(\mathbf{a}^T, b') = \sum_{i=1}^{m} \lambda_i \mathbf{v}_i$$

όπου, $\sum_{i=1}^{m} \lambda_i = 1$ και $\lambda_i \geq 0$, $\forall i \in [m]$. Συμπεραίνουμε ότι

$$f(\mathbf{x}) = \mathbf{a}^T\mathbf{x} + b \leq \mathbf{a}^T\mathbf{x} + b' = \sum_{i=1}^{m} \lambda_i \mathbf{v}_i^T \begin{pmatrix} \mathbf{x} \\ 1 \end{pmatrix} \leq \sum_{i=1}^{m} \lambda_i (\mathbf{a}^T\mathbf{x} + b) = \mathbf{a}^T\mathbf{x} + b$$

όπου, η τελευταία ανισότητα προκύπτει διότι υποθέσαμε ότι το πολυώνυμο στο $\mathbf{x}$ παίρνει την τιμή $\mathbf{a}^T\mathbf{x} + b$, επομένως όλοι οι υπόλοιποι όροι του πολυωνύμου, έχουν μικρότερη ή ίση τιμή.

Για να ισχύει η ισότητα στην παραπάνω ανισότητα, θα πρέπει οι όροι που αντιστοιχούν στις κορυφές της έδρας να λαμβάνουν όλες την ίδια τιμή στο $\mathbf{x}$. Συμπεραίνουμε ότι, όποια τιμή και αν λαμβάνει το πολυώνυμο, τότε αυτή μπορεί να υπολογιστεί από κάποια κορυφή του πολυτόπου. Ο ισχυρισμός μας έπεται. □

Με βάση την προηγούμενη πρόταση συμπεραίνουμε ότι μπορούμε να γράψουμε ένα τροπικό πολυώνυμο ως το τροπικό άθροισμα των όρων που αντιστοιχούν σε κορυφές του άνω φλοιού του επεκτεταμένου Newton πολυγώνου. Αυτό σημαίνει πως μπορούμε να απαλείψουμε όρους, χωρίς να αλλάζει η αποτίμηση του πολυωνύμου. Με την παρακάτω πρόταση συμπεραίνουμε, μάλιστα, ότι δεν μπορούμε να απαλείψουμε περαιτέρω τους όρους του πολυωνύμου, αφού κάθε ένας αντιστοιχεί σε μία γραμμική περιοχή.

Θεώρημα 2.2. *([6]) Οι γραμμικές περιοχές ενός πολυωνύμου d μεταβλητών $f \in \mathbb{R}_{max}[\mathbf{x}]$ βρίσκονται σε αμφιμονοσήμαντη αντιστοιχία με τις κορυφές του $UF(\mathrm{ENewt}(f))$.*

Απόδειξη. Γράφουμε το δεδομένο πολυώνυμο στην μορφή

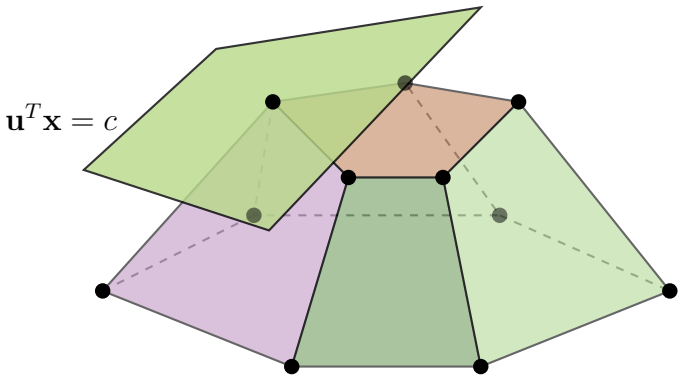
$$f(\mathbf{x}) = \max_{i \in [n]} \{\mathbf{a}_i^T\mathbf{x} + b\}$$

Θα δείξουμε ότι κάθε κορυφή $(\mathbf{a}_j^T, b) \in UF(\mathrm{ENewt}(f))$ μας δίνει μία γραμμική περιοχή d διαστάσεων για το f, δηλαδή

$$f(\mathbf{x}) = \mathbf{a}_j^T\mathbf{x} + b_j > \mathbf{a}_i^T\mathbf{x} + b_i, \ \forall \mathbf{x} \in \mathcal{D}$$

Θεωρούμε, λοιπόν, $(\mathbf{a}_j^T, b) \in UF(\mathrm{ENewt}(f))$ μία κορυφή του άνω φλοιού. Τότε, αυτή έχει την ιδιότητα ότι υπάρχει υπερεπίπεδο d διαστάσεων που διέρχεται από αυτή, δεν έχει κανένα

άλλο κοινό σημείο με το πολύτοπο και επιπλέον βρίσκεται πάνω από αυτό (ως προς την τελευταία διάσταση $d+1$). Η επιλογή αυτή απεικονίζεται γραφικά στην παρακάτω εικόνα.



Σχήμα 2.3: Αναπαράσταση του $UF(\text{ENewt}\,(f))$ και του επιπέδου που βρίσκεται πάνω απο αυτό και διέρχεται από μία κορυφή του.

Έστω ότι η εξίσωση του επιπέδου στις $d+1$ διαστάσεις είναι $\mathbf{u}^T\mathbf{x} = c$. Τότε αυτή γράφεται:

$$\mathbf{u}_d^T\mathbf{x}_d + u_{d+1}x_{d+1} = c \Leftrightarrow \frac{1}{u_{d+1}}\mathbf{u}_d^T\mathbf{x}_d + x_{d+1} = \frac{c}{u_{d+1}} \Leftrightarrow \mathbf{u}_d'^T\mathbf{x}_d + x_{d+1} = c'$$

Η παραπάνω σχέση είναι σωστή αφού $u_{d+1} \neq 0$, διότι διαφορετικά το επίπεδο θα ήταν κατακόρυφο ως προς την διάσταση $d+1$ με αποτέλεσμα να έχει κοινά σημεία τομής με το πολύτοπο. Επιπλέον, παρατηρούμε ότι αν αυξήσουμε το x_{d+1} τότε παίρνουμε τιμή μεγαλύτερη του c'. Συνεπώς, οι κορυφές του πολυτόπου πέραν της επιλεγμένης βρίσκονται στον ημίχωρο $\mathbf{u}_d'^T\mathbf{x}_d + x_{d+1} < c'$.

Το σημείο $(\mathbf{a}_j^T,\, b_j)$ θα ικανοποιεί την εξίσωση του επιπέδου, οπότε

$$\mathbf{a}_j^T\mathbf{u}_d' + b_j = c' > \mathbf{a}_i^T\mathbf{u}_d' + b_i$$

Δηλαδή, στο σημείο $\mathbf{x} = \mathbf{u}_d'$ από τους όρους της f την μεγαλύτερη τιμή λαμβάνει ο $\mathbf{a}_j^T\mathbf{x}+b_j = c'$. Για τους υπόλοιπους όρους $\mathbf{a}_i^T\mathbf{x}+b_i$ υποθέτουμε ότι η μεγαλύτερη τιμή που επιτυγχάνουν όταν υπολογίζονται στο $\mathbf{u}_d'$ είναι η $\lambda < c'$.

Αν προκαλέσουμε μία μικρή διαταραχή στο $\mathbf{u}_d'$

$$\mathbf{v} \leftarrow \mathbf{u}_d' + \epsilon = (u_1 + \epsilon_1, ..., u_d + \epsilon_d)$$

τότε

$$\mathbf{a}_j^T\mathbf{v} + b_j = \mathbf{a}_j^T\mathbf{u}_d' + b_j + \sum_{k=1}^{d} a_{jk}\epsilon_k = c' + \sum_{k=1}^{d} a_{jk}\epsilon_k$$

Για να εξακολουθεί να είναι αυτή η ποσότητα μεγαλύτερη από λ θα πρέπει

$$\sum_{k=1}^{d} a_{jk}\epsilon_k > \lambda - c' \Leftrightarrow -\sum_{k=1}^{d} a_{jk}\epsilon_k < |\lambda - c'| = c' - \lambda$$

Οπότε αρκεί να πάρουμε $|a_{jk}\epsilon_k| < \frac{|\lambda - c'|}{d} \Leftrightarrow |\epsilon_k| < \frac{|\lambda - c'|}{a_{jk}d}$ για $a_{jk} \neq 0$ και χωρίς περιορισμό για $a_{jk} = 0$. Τότε θα παίρναμε το επιθυμητό αποτέλεσμα από την τριγωνική ανισότητα

$$-\sum_{k=1}^{d} a_{jk} \leq \left|\sum_{k=1}^{d} a_{jk}\epsilon_k\right| \leq \sum_{k=1}^{d} |a_{jk}\epsilon_k| < |\lambda - c'|$$

Επομένως, για τα σημεία εντός της $d-$διάστατης υπερσφαίρας έχουμε ότι:

$$\|\mathbf{x}-\mathbf{u}'_d\|_\infty < \min_{a_{jk}\neq 0}\frac{|\lambda-c'|}{a_{jk}d} \Rightarrow \mathbf{a}_j^T\mathbf{x}+b_j > \mathbf{a}_i^T\mathbf{x}+b_i$$

Δηλαδή, πράγματι υπάρχει $d-$διάστατη γραμμική περιοχή $\mathcal{D}$ που αφορά την κορυφή $(\mathbf{a}_j^T,\, b_j) \in UF(\text{ENewt}\,(f))$. □

Παρατήρηση. *Με το Θεώρημα 2.2 συμπεραίνουμε πως κάθε κορυφή στο* $UF(\text{ENewt}\,(f))$ *μας δίνει μία γραμμική περιοχή του* f *η οποία είναι* d *διαστάσεων.*

Όπως καταλαβαίνουμε ένα τροπικό πολυώνυμο μπορεί να εκφραστεί πλήρως από τις κορυφές του άνω φλοιού του επεκτεταμένου Newton πολυτόπου του. Μάλιστα, αυτό είναι αρκετό ώστε να μπορέσει να συγκρίνει κανείς δύο πολυώνυμα, όπως φαίνεται από την πρόταση που ακολουθεί. Εκεί εξετάζουμε πότε δύο πολυώνυμα είναι ίσα ως συναρτήσεις, και όχι απαραίτητα όρο προς όρο.

Πρόταση 2.3. *Αν θεωρήσουμε την ισότητα ως συναρτησιακή ισοδυναμία δύο πολυωνύμων* $f, g \in \mathbb{R}_{max}[\mathbf{x}]$*, τότε ισχύει*

$$f = g \Leftrightarrow UF(\text{ENewt}\,(f)) = UF(\text{ENewt}\,(g))$$

Απόδειξη. Η μία κατεύθυνση είναι προφανής λόγω της Πρότασης 2.2. Πράγματι, αν θεωρήσουμε $UF(\text{ENewt}\,(f)) = UF(\text{ENewt}\,(g))$, τότε $f = g$ συναρτησιακά αφού οι τιμές των δύο συναρτήσεων καθορίζονται από τον ίδιο άνω φλοιό.

Για την αντίστροφη κατεύθυνση θεωρούμε ότι έχουμε μία γραμμική περιοχή $\mathcal{D} \subseteq \mathbb{R}^d$. Τότε, στο $\mathcal{D}$ έχουμε

$$f(\mathbf{x}) = g(\mathbf{x}),\ \forall \mathbf{x} \in \mathcal{D}$$

Αν γράψουμε $f(\mathbf{x}) = \mathbf{a}_f^T\mathbf{x}+\mathbf{b}_f$ και $g(\mathbf{x}) = \mathbf{a}_g^T\mathbf{x}+\mathbf{b}_g$, τότε για να είναι $f = g$ πρέπει

$$(\mathbf{a}_f-\mathbf{a}_g)^T\mathbf{x}+\mathbf{b}_f-\mathbf{b}_g = \mathbf{0}$$

που περιγράφει εξίσωση υπερεπιπέδου $d-1$ διαστάσεων για $\mathbf{a}_f \neq \mathbf{a}_g$. Επομένως, πρέπει $\mathbf{a}_f = \mathbf{a}_g$ και $\mathbf{b}_f = \mathbf{b}_g$.

Επαναλαμβάνοντας, το σκεπτικό μας για όλες τις γραμμικές περιοχές των πολυωνύμων και χρησιμοποιώντας το θεώρημα 2.2, βλέπουμε ότι τα $UF(\text{ENewt}\,(f)), UF(\text{ENewt}\,(g))$ αποτελούνται από τις ίδιες κορυφές και επομένως, συμπίπτουν. □

Παράδειγμα 2.2. *Χρησιμοποιώντας τα πολυώνυμα του Παραδείγματος 2.1 μπορούμε να υπολογίσουμε μία έκδοση του* $f \vee g$ *η οποία χρησιμοποιεί το ελάχιστο πλήθος μονωνύμων.*

$$f \vee g = \max(2x+y+1, 0, x, y, 1) = \max(2x+y+1, x, y, 1)$$

Πράγματι, αυτή η μορφή αντιστοιχεί επακριβώς στους όρους που εμφανίζονται ως κορυφές στον άνω φλοιό $UF(\text{ENewt}\,(f \vee g))$*, όπως υποδεικνύεται στο Σχήμα 2.2.*

Η πρόταση που ακολουθεί έχει εφαρμογές στην τροπική διαίρεση στην οποία θα αναφερθούμε σε επόμενη ενότητα. Συγκεκριμένα, μας επιτρέπει να μπορούμε να αποφανθούμε

πότε ένα τροπικό πολυώνυμο είναι μεγαλύτερο από ένα άλλο. Διαισθητικά, αφού μέχρι τώρα έχουμε συνδέσει στενά τα πολυώνυμα με τα επεκτεταμένα Newton πολύτοπα τους θα πρέπει και η ανισοτική σχέση μεταξύ αυτών να περιγράφεται με γεωμετρικό τρόπο. Αξίζει να σημειώσουμε ότι μία κατεύθυνση της παρακάτω πρότασης, συναντάται ως αναγκαία συνθήκη στο ([47], Θεώρημα 3.4).

Πρόταση 2.4. *Για $f, g \in \mathbb{R}_{max}[\mathbf{x}]$ ισχύει ότι $f(\mathbf{x}) \geq g(\mathbf{x}), \forall \mathbf{x} \in \mathbb{R}^n$ αν και μόνον το $UF(\mathrm{ENewt}\,(f))$ βρίσκεται πάνω, ως προς την τελευταία διάσταση από το $UF(\mathrm{ENewt}\,(g))$, δηλαδή αν για κάθε $(\mathbf{x}, x_{d+1}) \in UF(\mathrm{ENewt}\,(f))$ είτε δεν υπάρχει y_{d+1}, ώστε $(\mathbf{x}, y_{d+1}) \in UF(\mathrm{ENewt}\,(g))$ είτε υπάρχει και $y_{d+1} \leq x_{d+1}$.*

Απόδειξη. Έχουμε ότι

$$f \geq g \Leftrightarrow f \vee g = f \overset{\text{Πρ. 2.3}}{\Longleftrightarrow} UF(\mathrm{ENewt}\,(f \vee g)) = UF(\mathrm{ENewt}\,(f)) \Leftrightarrow$$

$$UF(\mathrm{conv}\,\{\mathrm{ENewt}\,(f) \cup \mathrm{ENewt}\,(g)\}) = UF(\mathrm{ENewt}\,(f))$$

Η τελευταία σχέση μπορεί να ισχύει μόνο όταν το $UF(\mathrm{ENewt}\,(g))$ βρίσκεται κάτω από το $UF(\mathrm{ENewt}\,(f))$. Διαφορετικά, αν υπάρχει ένα σημείο που δεν βρίσκεται από κάτω τότε αυτό θα εμφανιστεί σίγουρα στο $UF(\mathrm{conv}\,\{\mathrm{ENewt}\,(f) \cup \mathrm{ENewt}\,(g)\})$, αλλά όχι στο $UF(\mathrm{ENewt}\,(f))$. □

Από την Πρόταση 2.4 προκύπτει εύκολα το παρακάτω πόρισμα, το οποίο συναντάμε στο ([47], Θεώρημα 3.3).

Πόρισμα 2.2. *Για $f, g \in \mathbb{R}_{max}[\mathbf{x}]$ αν ισχύει ότι $f(\mathbf{x}) \geq g(\mathbf{x}), \forall \mathbf{x} \in \mathbb{R}^n$ τότε*

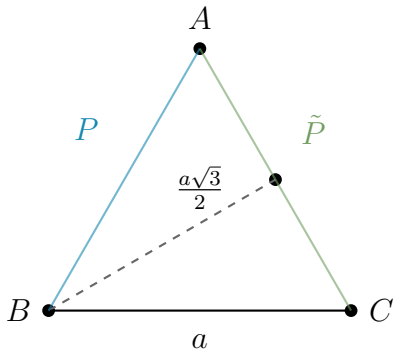
$$\mathrm{Newt}\,(g) \subseteq \mathrm{Newt}\,(f)$$

Απόδειξη. Αρκεί να παρατηρήσουμε ότι από την προηγούμενη πρόταση, ότι ο άνω φλοιός $UF(\mathrm{ENewt}\,(f))$ βρίσκεται πάνω από τον $UF(\mathrm{ENewt}\,(g))$, οπότε και η προβολή $\mathrm{Newt}\,(f)$ του $\mathrm{ENewt}\,(f)$ θα περιέχει το $\mathrm{Newt}\,(g)$. □

2.3 Προσέγγιση Πολυτόπων

Στην ενότητα αυτή θα παρουσιάσουμε το σημαντικότερο θεώρημα το οποίο αποτελεί τον θεμελιώδη λίθο της συμβολής μας και πάνω στο οποίο βασίζονται οι αλγόριθμοι συμπίεσης νευρωνικών δικτύων που προτείνουμε σε επόμενα Κεφάλαια. Ο στόχος μας είναι να μπορέσουμε να προσεγγίσουμε ένα τροπικό πολυώνυμο μέσω ενός άλλου, το οποίο πιθανότατα θα έχει λιγότερους όρους, ή απλώς διαφορετικούς. Το σφάλμα που θα προκύψει μεταξύ των δύο πολυωνύμων θα το φράξουμε από μία μετρική η οποία θα δείχνει πόσο κοντά βρίσκονται τα επεκτεταμένα Newton πολύτοπα των εν λόγω πολυωνύμων. Η μετρική που θα χρησιμοποιήσουμε για αυτόν τον σκοπό είναι η *απόσταση Hausdorff*.

Το κίνητρο που μας οδήγησε να φράξουμε γεωμετρικά, μέσω των πολυτόπων, την διαφορά δύο πολυωνύμων είναι η Πρόταση 2.3. Αυτή μας εξηγεί ότι δύο πολυώνυμα είναι ίσα αν και μόνο αν οι άνω φλοιοί των επεκτεταμένων Newton πολυτόπων τους είναι ίσοι. Είναι εύλογο, λοιπόν, να θεωρήσουμε ότι μπορούμε να γενικεύσουμε αυτήν την πρόταση σε μία προσεγγιστική εκδοχή της ισότητας των δύο πολυωνύμων. Συγκεκριμένα, θα δείξουμε ότι



Σχήμα 2.4: Ισόπλευρο τρίγωνο πλευράς a και πολύτοπα AB, AC με κοινή την κορυφή A.

εάν δύο πολυώνυμα έχουν "κοντινά" επεκτεταμένα Newton πολύτοπα, τότε η διαφορά τους είναι φραγμένη. Το ζήτημα αυτό εξετάζεται με το Θεώρημα 2.3.

Συμβολίζουμε την απόσταση ενός σημείου $\mathbf{u} \in \mathbb{R}^d$ από ένα πεπερασμένο σύνολο σημείων $\mathcal{V} \subset \mathbb{R}^d$ είτε ως $\operatorname{dist}(\mathbf{u}, \mathcal{V})$ είτε ως $\operatorname{dist}(\mathcal{V}, \mathbf{u})$. Αυτή υπολογίζεται ως $\min\{\|\mathbf{u} - \mathbf{v}\|, \mathbf{v} \in \mathcal{V}\}$ που πρακτικά είναι η Ευκλείδια απόσταση ενός σημείου $\mathbf{u}$ από το πλησιέστερό του σημείο $\mathbf{v} \in \mathcal{V}$.

Ορισμός. Η *Hausdorff απόσταση* $\mathcal{H}(\mathcal{V}, \mathcal{U})$ δύο πεπερασμένων συνόλων σημείων $\mathcal{V}, \mathcal{U} \subset \mathbb{R}^d$ ορίζεται με τον εξής τρόπο

$$\mathcal{H}(\mathcal{V}, \mathcal{U}) = \max\left\{\max_{\mathbf{v}\in\mathcal{V}} \operatorname{dist}(\mathbf{v}, \mathcal{U}), \max_{\mathbf{u}\in\mathcal{U}} \operatorname{dist}(\mathcal{V}, \mathbf{u})\right\}$$

Θεωρούμε δύο πολύτοπα $P, \tilde{P}$ με σύνολα κορυφών $\mathcal{V}_P, \mathcal{V}_{\tilde{P}}$ αντίστοιχα. Τότε, κατά σύμβαση ορίζουμε την Hausdorff απόσταση $\mathcal{H}\left(P, \tilde{P}\right)$ των δύο πολυτόπων να είναι ίση με την Hausdorff απόσταση των συνόλων των κορυφών τους $\mathcal{V}_P, \mathcal{V}_{\tilde{P}}$. Συγκεκριμένα, γράφουμε

$$\mathcal{H}\left(P, \tilde{P}\right) := \mathcal{H}\left(\mathcal{V}_P, \mathcal{V}_{\tilde{P}}\right) \tag{2.2}$$

Παράδειγμα 2.3. *Θεωρούμε το ισόπλευρο τρίγωνο ABC του σχήματος 2.4 με πλευρά μήκους a. Τα πολύτοπα $P, \tilde{P}$ ορίζονται ως τα ευθύγραμμα τμήματα AB, BC αντίστοιχα. Τότε η Hausdorff απόσταση των $P, \tilde{P}$ είναι ίση με a, αφού η κορυφή B απέχει a από τις κορυφές A, C του $\tilde{P}$.*

Παρατήρηση. *Αξίζει να σημειώσουμε ότι εαν στα πολύτοπα $P, \tilde{P}$ του σχήματος 2.4 χρησιμοποιήσουμε τον κλασσικό ορισμό της Hausdorff απόστασης για μη-πεπερασμένα γεωμετρικά σύνολα:*

$$\mathcal{H}\left(P, \tilde{P}\right) = \max\left\{\sup_{\mathbf{x}\in P} \operatorname{dist}\left(\mathbf{x}, \tilde{P}\right), \sup_{\mathbf{y}\in\tilde{P}} \operatorname{dist}(P, \mathbf{y})\right\}$$

τότε η τιμή που προκύπτει είναι $\frac{a\sqrt{3}}{2}$ που είναι διαφορετική από αυτήν που ορίσαμε εμείς συμβατικά, δηλαδή την $\mathcal{H}\left(\mathcal{V}_P, \mathcal{V}_{\tilde{P}}\right) = a$.

Είναι εμφανές ότι η Hausdorff απόσταση αποτελεί μία μετρική που δείχνει πόσο κοντά βρίσκονται μεταξύ τους τα δύο πολύτοπα, δηλαδή κατά πόσο είναι ίδια. Πράγματι, η απόσταση μεταξύ τους γίνεται 0 όταν τα δύο πολύτοπα ταυτίζονται. Με χρήση της μετρικής αυτής για την απόσταση μεταξύ των πολυτόπων λαμβάνουμε το ακόλουθο φράγμα για το σφάλμα μεταξύ δύο τροπικών πολυωνύμων.

Θεώρημα 2.3. *(Προσέγγιση Τροπικών Πολυωνύμων) Έστω* $p, \tilde{p} \in \mathbb{R}_{max}[\mathbf{x}]$ *δύο τροπικά πολυώνυμα* d *μεταβλητών και* $P = \mathrm{ENewt}(p), \tilde{P} = \mathrm{ENewt}(\tilde{p})$ *τα επεκτεταμένα Newton πολύτοπά τους. Τότε, ισχύει*

$$\max_{x \in \mathcal{B}} |p(\mathbf{x}) - \tilde{p}(\mathbf{x})| \leq \rho \cdot \mathcal{H}\left(P, \tilde{P}\right)$$

όπου $\mathcal{B} = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| \leq r\}$ *είναι η υπερσφαίρα ακτίνας* r*, και* $\rho = \sqrt{r^2 + 1}$.

Απόδειξη. Θεωρούμε ένα οποιοδήποτε σημείο $\mathbf{x} \in \mathcal{B}$ και υποθέτουμε ότι

$$p(\mathbf{x}) = \mathbf{a}^T\mathbf{x} + b,\ \tilde{p}(\mathbf{x}) = \mathbf{c}^T\mathbf{x} + d$$

Τότε,

$$p(\mathbf{x}) - \tilde{p}(\mathbf{x}) = p(\mathbf{x}) - \max_{(\tilde{\mathbf{a}}^T, \tilde{b}) \in \mathcal{V}_{\tilde{p}}} \{\tilde{\mathbf{a}}^T\mathbf{x} + \tilde{b}\} \leq \mathbf{a}^T\mathbf{x} + b - \left(\mathbf{u}^T,\ v\right)\begin{pmatrix}\mathbf{x}\\ 1\end{pmatrix}$$

όπου $\left(\mathbf{u}^T, v\right)$ μπορεί να είναι οποιαδήποτε κορυφή του $\tilde{P}$. Ομοίως, λαμβάνουμε

$$\left(\mathbf{r}^T,\ s\right)\begin{pmatrix}\mathbf{x}\\ 1\end{pmatrix} - \left(\mathbf{c}^T\mathbf{x} + d\right) \leq p(\mathbf{x}) - \tilde{p}(\mathbf{x})$$

για οποιαδήποτε κορυφή $\left(\mathbf{r}^T, s\right)$ του P. Επομένως, μπορούμε να επιλέξουμε οποιεσδήποτε κορυφές $\left(\mathbf{u}^T, v\right) \in \tilde{P}, \left(\mathbf{r}^T, s\right) \in P$ ώστε οι αντίστοιχες αποστάσεις τους από τις κορυφές $\left(\mathbf{a}^T, b\right)$ και $\left(\mathbf{c}^T, d\right)$, αντίστοιχα, να γίνονται ελάχιστες. Επιλέγοντας τις κορυφές κατά αυτόν τον τρόπο λαμβάνουμε

$$\begin{aligned} p(\mathbf{x}) - \tilde{p}(\mathbf{x}) &\leq \mathbf{a}^T\mathbf{x} + b - \left(\mathbf{u}^T,\ v\right)\begin{pmatrix}\mathbf{x}\\ 1\end{pmatrix} = \left(\left(\mathbf{a}^T,\ b\right) - \left(\mathbf{u}^T,\ v\right)\right)\begin{pmatrix}\mathbf{x}\\ 1\end{pmatrix} \leq \\ &\leq \left\|\left(\mathbf{a}^T,\ b\right) - \left(\mathbf{u}^T,\ v\right)\right\| \left\|\begin{pmatrix}\mathbf{x}\\ 1\end{pmatrix}\right\| \leq d\left(\left(\mathbf{a}^T,\ b\right), \tilde{P}\right)\sqrt{r^2 + 1} \end{aligned} \tag{2.3}$$

Με παρόμοιο τρόπο παίρνουμε

$$\begin{aligned} p(\mathbf{x}) - \tilde{p}(\mathbf{x}) &\geq \left(\mathbf{r}^T,\ s\right)\begin{pmatrix}\mathbf{x}\\ 1\end{pmatrix} - \mathbf{c}^T\mathbf{x} + d = \left(\left(\mathbf{r}^T,\ s\right) - \left(\mathbf{c}^T,\ d\right)\right)\begin{pmatrix}\mathbf{x}\\ 1\end{pmatrix} \geq \\ &\geq -\left\|\left(\mathbf{r}^T,\ s\right) - \left(\mathbf{c}^T,\ d\right)\right\| \left\|\begin{pmatrix}\mathbf{x}\\ 1\end{pmatrix}\right\| \geq -d\left(P, \left(\mathbf{c}^T,\ d\right)\right)\sqrt{r^2 + 1} \end{aligned} \tag{2.4}$$

Σημειώνουμε, ότι για τις σχέσεις (2.3) και (2.4) κάναμε χρήση της διανυσματικής Cauchy-Schwartz ανισότητας

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\|\|\mathbf{y}\| \Leftrightarrow -\|\mathbf{x}\|\|\mathbf{y}\| \leq \langle \mathbf{x}, \mathbf{y} \rangle \leq \|\mathbf{x}\|\|\mathbf{y}\|$$

Η ανισότητα (2.3) ισχύει σε κάθε σημείο $x \in \mathcal{B}$ για κάποια κορυφή $\left(a^T,\, b\right) \in P$, οπότε

$$p(\mathbf{x}) - \tilde{p}(\mathbf{x}) \leq \rho \cdot \max_{\left(\mathbf{a}^T, b\right) \in \mathcal{V}_P} d\left(\left(\mathbf{a}^T,\, b\right), \mathcal{V}_{\tilde{P}}\right) \tag{2.5}$$

για κάθε $x \in \mathcal{B}$. Ομοίως, προκύπτει

$$p(\mathbf{x}) - \tilde{p}(\mathbf{x}) \geq \min_{\left(\mathbf{c}^T, d\right) \in \mathcal{V}_{\tilde{P}}} -\rho \cdot d\left(\mathcal{V}_P, \left(\mathbf{c}^T,\, d\right)\right) = -\max_{\left(\mathbf{c}^T, d\right) \in \mathcal{V}_{\tilde{P}}} \rho \cdot d\left(\mathcal{V}_P, \left(\mathbf{c}^T,\, d\right)\right) \tag{2.6}$$

Συνδυάζοντας τις ανισότητες (2.5) και (2.6) καταλήγουμε στην εξής σχέση

$$-\max_{\left(\mathbf{c}^T, d\right) \in \mathcal{V}_{\tilde{P}}} \rho \cdot d\left(\mathcal{V}_P, \left(\mathbf{c}^T,\, d\right)\right) \leq p(\mathbf{x}) - \tilde{p}(\mathbf{x}) \leq \max_{\left(\mathbf{a}^T, b\right) \in \mathcal{V}_P} \rho \cdot d\left(\left(\mathbf{a}^T,\, b\right), \mathcal{V}_{\tilde{P}}\right) \Leftrightarrow$$

$$|p(\mathbf{x}) - \tilde{p}(\mathbf{x})| \leq \rho \cdot \max\left\{\max_{\left(\mathbf{a}^T, b\right) \in \mathcal{V}_P} \rho \cdot d\left(\left(\mathbf{a}^T,\, b\right), \mathcal{V}_{\tilde{P}}\right), \max_{\left(\mathbf{c}^T, d\right) \in \mathcal{V}_{\tilde{P}}} \rho \cdot d\left(\mathcal{V}_P, \left(\mathbf{c}^T,\, d\right)\right)\right\}$$

Ως εκ τούτου, από τον ορισμό της Hausdorff απόστασης δύο πολυτόπων λαμβάνουμε το επιθυμητό άνω φράγμα

$$|p(\mathbf{x}) - \tilde{p}(\mathbf{x})| \leq \rho \cdot \mathcal{H}\left(P, \tilde{P}\right), \forall \mathbf{x} \in \mathcal{B} \Rightarrow$$

$$\max_{\mathbf{x} \in \mathcal{B}} |p(\mathbf{x}) - \tilde{p}(\mathbf{x})| \leq \rho \cdot \mathcal{H}\left(P, \tilde{P}\right)$$

□

Παρατήρηση. *Επιλέγουμε να υπολογίσουμε το άνω φράγμα στο σφάλμα της προσέγγισης των δύο πολυωνύμων σε μία φραγμένη υπερσφαίρα $\mathcal{B}$. Η επιλογή αυτή γίνεται για δύο λόγους. Αρχικά, σε έναν μή φραγμένο χώρο το σφάλμα δύο πολυωνύμων που δεν ταυτίζονται εν γένει αποκλίνει και, δεύτερον, όταν θα δουλεύουμε με νευρωνικά δίκτυα ο υπόχωρος της εισόδου με τον οποίο εργαζόμαστε είναι φραγμένος. Πράγματι, αυτό συμβαίνει διότι οι τιμές στα pixel των εικόνων είναι περιορισμένες (π.χ. $0 - 255$ για RGB εικόνες).*

Αξίζει να σημειώσουμε ότι το Θεώρημα 2.3 μπορεί να διατυπωθεί με την χρήση των άνω φλοιών των πολυτόπων $P, \tilde{P}$, αφού όπως γνωρίζουμε από το Θεώρημα 2.2 τα πολυώνυμα αγνοούν τα μονώνυμα που δεν αντιστοιχούν σε κάποια κορυφή του άνω φλοιού. Το σχετικό φράγμα για τους άνω φλοιούς περιγράφεται με την ακόλουθη πρόταση, η απόδειξη της οποίας είναι εντελώς όμοια με του προηγούμενου θεωρήματος και παρέχεται για λόγους πληρότητας.

Πρόταση 2.5. *(Προσέγγιση Τροπικών Πολυωνύμων) Έστω $p, q \in \mathbb{R}_{max}[\mathbf{x}]$ δύο τροπικά πολυώνυμα d μεταβλητών με επεκτεταμένα Newton πολύτοπα $P = \mathrm{ENewt}\,(p)\,, \tilde{P} = \mathrm{ENewt}\,(\tilde{p})$. Τότε ισχύει*

$$\max_{x \in \mathcal{B}} |p(\mathbf{x}) - \tilde{p}(\mathbf{x})| \leq \rho \cdot \mathcal{H}\left(UF(P), UF(\tilde{P})\right)$$

όπου το maximum υπολογίζεται στην υπερσφαίρα $\mathcal{B} = \{x \in \mathbb{R}^d : \|x\| \leq r\}$ ακτίνας r, και $\rho = \sqrt{r^2 + 1}$.

Απόδειξη. Θεωρούμε ένα οποιοδήποτε σημείο $\mathbf{x} \in \mathcal{B}$ και θεωρούμε

$$p(\mathbf{x}) = \mathbf{a}^T\mathbf{x} + b,\, \tilde{p}(\mathbf{x}) = \mathbf{c}^T\mathbf{x} + d$$

Λόγω του Θεωρήματος 2.2 τα πολυώνυμα $p, \tilde{p}$ παίρνουν τιμές που καθορίζονται από τις κορυφές των $UF(P)$, $UF(\tilde{P})$. Επομένως, ισχύει ότι $(\mathbf{a}^T, b) \in UF(P), (\mathbf{c}^T, d) \in UF(\tilde{P})$ και μπορούμε να επιλέξουμε $(\mathbf{u}^T, v) \in UF(\tilde{P}), (\mathbf{r}^T, s) \in UF(P)$, όπως και στην προηγούμενη απόδειξη ώστε

$$\begin{aligned} p(\mathbf{x}) - \tilde{p}(\mathbf{x}) &\leq \mathbf{a}^T\mathbf{x} + b - \left(\mathbf{u}^T, v\right)\begin{pmatrix}\mathbf{x}\\1\end{pmatrix} = \left(\left(\mathbf{a}^T, b\right) - \left(\mathbf{u}^T, v\right)\right)\begin{pmatrix}\mathbf{x}\\1\end{pmatrix} \leq \\ &\leq \left\|\left(\mathbf{a}^T, b\right) - \left(\mathbf{u}^T, v\right)\right\| \left\|\begin{pmatrix}\mathbf{x}\\1\end{pmatrix}\right\| \leq d\left(\left(\mathbf{a}^T, b\right), \tilde{P}\right)\sqrt{r^2+1} \end{aligned} \tag{2.7}$$

και ομοίως

$$p(\mathbf{x}) - \tilde{p}(\mathbf{x}) \geq \left(\mathbf{r}^T, s\right)\begin{pmatrix}\mathbf{x}\\1\end{pmatrix} - \mathbf{c}^T\mathbf{x} + d \geq -d\left(P, \left(\mathbf{c}^T, d\right)\right)\sqrt{r^2+1} \tag{2.8}$$

Επομένως, προκύπτει ότι

$$p(\mathbf{x})-\tilde{p}(\mathbf{x}) \leq \rho \cdot \max_{(\mathbf{a}^T,b)\in\mathcal{V}_P} d\left(\left(\mathbf{a}^T, b\right), \mathcal{V}_{UF(\tilde{P})}\right), \quad p(\mathbf{x})-\tilde{p}(\mathbf{x}) \geq -\max_{(\mathbf{c}^T,d)\in\mathcal{V}_{\tilde{P}}} \rho \cdot d\left(\mathcal{V}_{UF(P)}, \left(\mathbf{c}^T, d\right)\right) \tag{2.9}$$

για κάθε $x \in \mathcal{B}$. Συνδυάζοντας τις ανισότητες αυτές καταλήγουμε στην ζητούμενη

$$|p(\mathbf{x}) - \tilde{p}(\mathbf{x})| \leq \rho \cdot \max\left\{\max_{(\mathbf{a}^T,b)\in\mathcal{V}_{UF(P)}} \rho \cdot d\left(\left(\mathbf{a}^T, b\right), \mathcal{V}_{UF(\tilde{P})}\right), \max_{(\mathbf{c}^T,d)\in\mathcal{V}_{UF(\tilde{P})}} \rho \cdot d\left(\mathcal{V}_{UF(P)}, \left(\mathbf{c}^T, d\right)\right)\right\} \Leftrightarrow$$

$$|p(\mathbf{x}) - \tilde{p}(\mathbf{x})| \leq \rho \cdot \mathcal{H}\left(UF(P), UF(\tilde{P})\right), \forall \mathbf{x} \in \mathcal{B} \Rightarrow$$

$$\max_{\mathbf{x}\in\mathcal{B}} |p(\mathbf{x}) - \tilde{p}(\mathbf{x})| \leq \rho \cdot \mathcal{H}\left(UF(P), UF(\tilde{P})\right)$$

□

Παρατήρηση. *Η Πρόταση 2.5 αποτελεί γενίκευση για την μία κατεύθυνση της ισοδυναμίας στην σχέση 2.3. Συγκεκριμένα, εάν οι άνω φλοιοί δύο επεκτεταμένων Newton πολυτόπων ταυτίζονται, τότε το ίδιο ισχύει και για τα εν λόγω πολυώνυμα.*

Αξίζει να τονίσουμε ότι παρόλο που με την πρόταση 2.5 καταφέραμε να γενικεύσουμε την μία κατεύθυνσης της Πρότασης 2.3, δεν ισχύει το ίδιο και για την άλλη κατεύθυνση. Συγκεκριμένα, μπορεί να υπάρχουν τροπικά πολυώνυμα τα οποία είναι προσεγγιστικά ίσα, χωρίς να ισχύει το ίδιο και για τα πολύτοπά τους, υπό την έννοια της Hausdorff απόστασης που έχουμε ορίσει για τα πολύτοπα. Για παράδειγμα, αν θεωρήσουμε τα πολυώνυμα

$$p(x) = \max\{-\alpha x, \alpha x\}, \quad \tilde{p}(x) = \max\{-\alpha x, \epsilon, \alpha x\}$$

με $\epsilon \to 0$, τότε το μέγιστο σφάλμα των πολυωνύμων είναι

$$\max_x |p(x) - \tilde{p}(x)| = \epsilon$$

Ενώ τα επεκτεταμένα Newton πολύτοπά τους είναι τα $P = \text{conv}\{(-\alpha, 0), (\alpha, 0)\}$, $\tilde{P} = \text{conv}\{(-\alpha, 0), (0, \epsilon), (\alpha, 0)\}$ και για την Hausdorff απόσταση των πολυτόπων τους ισχύει

$$\mathcal{H}\left(P, \tilde{P}\right) = \sqrt{\alpha^2 + \epsilon^2}$$

που είναι μη φραγμένη, καθώς μπορεί $a \to \infty$.

2.4 Πλήθος Εδρών Πολυτόπου

Σε αυτήν την ενότητα θα επεκτείνουμε την μελέτη των πολυτόπων, περιγράφοντας ένα πρόβλημα συνδυαστικής γεωμετρίας. Το πρόβλημα αφορά την προσεγγιστική καταμέτρηση των εδρών ενός πολυτόπου αποδίδοντας θεωρητικά άνω και κάτω φράγματα. Αξίζει να σημειωθεί ότι έχει γίνει εργασία πάνω σε πιθανοτική καταμέτρηση κορυφών πολυτόπου [6]. Η ενότητα αυτή δεν έχει άμεση συσχέτιση με την συμπίεση νευρωνικών δικτύων, που είναι και το αντικείμενο αυτής της διπλωματικής, αλλά παρέχεται κυρίως ως βιβλιογραφική αναφορά, προσθέτοντας ορισμένα καινούρια στοιχεία.

Σε προηγούμενη ενότητα είδαμε πως υπάρχει αμφιμονοσήμαντη αντιστοιχία μεταξύ των γραμμικών περιοχών ενός τροπικού πολυωνύμου και των κορυφών του άνω φλοιού του επεκτεταμένου Newton πολυτόπου. Στην συνέχεια θα αναδείξουμε την στενή σύνδεση των τροπικών πολυωνύμων και των νευρωνικών δικτύων, και ως συνέπεια το πρόβλημα της καταμέτρησης θα βρει εφαρμογή στον προσδιορισμό του πλήθους των γραμμικών περιοχών των δικτύων. Με αυτόν τον τρόπο θα μπορέσουμε να προσεγγίσουμε το πλήθος των γραμμικών περιοχών γνωστών επιπέδων, όπως επιπέδων ReLU, συνελικτικά επίπεδα με ReLU ενεργοποιήσεις, Max-out επίπεδα αλλά και επίπεδα των state-of-the-art ResNet δικτύων. Οι υπολογισμοί αυτοί επεκτείνονται και σε βαθιά νευρωνικά δίκτυα αποτελούμενα από διαδοχικά τέτοια επίπεδα.

Η τεχνική που θα αναλύσουμε για την καταμέτρηση των εδρών ενός πολυτόπου προέρχεται από τα [45, 6]. Αρχικά, παραθέτουμε έναν ορισμό που περιγράφει μία πολύ χρήσιμη κλάση πολυτόπων, τα ζωνότοπα.

Ορισμός. Ένα πολύτοπο $P \subset \mathbb{R}^d$ το οποίο ορίζεται ως το Minkowski άθροισμα των ευθυγράμμων τμημάτων $P_1, ..., P_k$ ονομάζεται ζωνότοπο (zonotope).

Το ζωνότοπο είναι ένα πολύτοπο που αποτελεί την βάση για την κατασκευή ενός νευρωνικού. Ενδεικτικά αναφέρουμε ότι το επεκτεταμένο Newton πολύτοπο ενός ReLU επιπέδου είναι ζωνότοπο. Η ανάλυση των νευρωνικών με ζωνότοπα θα γίνει στο επόμενο Κεφάλαιο αναλυτικά.

Ξεκινάμε την ανάλυση μας με ένα Θεώρημα από τους Gritzmann και Sturmfels (1993) [15]. Το θεώρημα αυτό δεν περιορίζεται μόνο στην καταμέτρηση των κορυφών αλλά αφορά γενικότερα τις $i-$διάστατες έδρες ενός πολυτόπου. Μας παρέχει ένα σφιχτό (tight) άνω φράγμα για τον αριθμό των $i-$διάστατων εδρών του Minkowski αθροίσματος πολυτόπων.

Θεώρημα 2.4. *([15], Theorem 2.1.10) Έστω $P_1, ..., P_k$ πολύτοπα στον $\mathbb{R}^d$, που έχουν m μη-παράλληλες ανα δύο ακμές. Τότε, ο αριθμός των εδρών διάστασης i του $P = P_1 \oplus ... \oplus P_k$ είναι σε πλήθος το πολύ*

$$2\binom{m}{i}\sum_{j=0}^{d-1-i}\binom{m-1-i}{j}$$

Παρατήρηση. *Η ισότητα στο παραπάνω φράγμα ισχύει όταν το P είναι ένα ζωνότοπο που παράγεται από m μη-παράλληλα ανα δύο ευθύγραμμα τμήματα.*

Με βάση το προηγούμενο θεώρημα παίρνουμε το ακόλουθο πόρισμα που αφορά τις κορυφές του πολυτόπου.

Πόρισμα 2.3. *Το πλήθος των κορυφών ενός πολυτόπου με m μη-παράλληλες ανα δύο ακμές είναι*

$$2\sum_{j=0}^{n-1}\binom{m-1}{j}$$

Παρατήρηση. *Για την απόδειξη του παραπάνω πορίσματος αρκεί κανείς να παρατηρήσει ότι οι κορυφές ενός πολυτόπου είναι οι έδρες διάστασης 0 και να χρησιμοποιήσει το Θεώρημα 2.4.*

Για να μπορέσουμε να αποδώσουμε ένα άνω φράγμα στον αριθμό των γραμμικών περιοχών ενός επιπέδου νευρωνικού θα χρησιμοποιήσουμε την παρακάτω πρόταση η οποία φράσσει το πλήθος των κορυφών του άνω φλοιού ενός ζωνοτόπου.

Πρόταση 2.6. *([45, 6]) Έστω $P \in \mathbb{R}^{d+1}$ ένα ζωνότοπο που παράγεται από τα ευθύγραμμα τμήματα $P_1, ..., P_m$. Τότε το πληθος των κορυφών του άνω φλοιού του P είναι το πολύ*

$$\sum_{j=0}^{d}\binom{m}{j}$$

Απόδειξη. Θα μετρήσουμε τις κορυφές του P ως εξής. Αρχικά, θα θεωρήσουμε την προβολή π του P στον $\mathbb{R}^d$ που προκύπτει "ρίχνοντας" την τελευταία του διάσταση. Έστω, n_1 το πλήθος των κορυφών που βρίσκονται αποκλειστικά στον άνω φλοιό του P και n_2 αυτών που βρίσκονται ταυτόχρονα και στον άνω και στον κάτω φλοιό. Μάλιστα, οι κορυφές που είναι ταυτόχρονα στον άνω και κάτω φλοιό είναι οι κορυφές του πολυτόπου προβολής $\pi(P)$.

Λόγω, της κεντρικής συμμετρίας των ζωνοτόπων, το πλήθος των κορυφών που βρίσκονται αποκλειστικά στον κάτω φλοιό είναι επίσης n_1. Επομένως, ο αριθμός των κορυφών του άνω φλοιού που αναζητάμε είναι ίσος με $n_1 + n_2$.

Το πολύτοπο P γράφεται σαν το Minkowski άθροισμα m ευθυγράμμων τμημάτων. Την κατασκευή του ζωνοτόπου απο τα ευθύγραμμα τμήματα μπορούμε να την θεωρήσουμε με επαγωγικό τρόπο ως εξής. Το ζωνότοπο που προκύπτει από τα πρώτα i ευθύγραμμα τμήματα αποτελεί παράλληλη μετατόπιση του ζωνοτόπου που προκύπτει από τα πρώτα $i-1$ τμήματα και συνένωση με το αρχικό. Η παράλληλη μετατόπιση προκύπτει κατά την διεύθυνση που ορίζει το $i-$οστό τμήμα. Αυτό σημαίνει ότι το τελικό ζωνότοπο μπορεί να έχει το πολύ m μη-παράλληλες ανά δύο ακμές. Επιπλέον, το ίδιο ισχύει και για την προβολή του $\pi(P)$. Συνεπώς, για τα δύο αυτά πολύτοπα μπορούμε να εφαρμόσουμε το πόρισμα 2.3. Πράγματι, έχουμε

$$(2n_1 + n_2) + n_2 \leq 2\sum_{j=0}^{d}\binom{m-1}{j} + 2\sum_{j=0}^{d-1}\binom{m-1}{j} = 2\sum_{j=0}^{d}\left[\binom{m-1}{j} + \binom{m-1}{j-1}\right] \Leftrightarrow$$

$$n_1 + n_2 \leq \sum_{j=0}^{n}\binom{m}{j}$$

όπου η τελευταία σχέση ισχύει από την ταυτότητα Pascal για τους διωνυμικούς συντελεστές. Ο ισχυρισμός μας έπεται. □

Παρατήρηση. *Το παραπάνω φράγμα είναι σφιχτό (tight). Πράγματι, το πλήθος των κορυφών του άνω φλοιού του ζωνοτόπου επιτυγχάνει την μέγιστη τιμή όταν ικανοποιούνται οι συνθήκες:*

- *Τα $P_1, ..., P_m$ δεν είναι παράλληλα ανά δύο.*
- *Οι προβολές $\pi(P_1), ..., \pi(P_m)$ των τμημάτων στον $\mathbb{R}^d$ δεν είναι παράλληλες ανα 2.*

2.4.1 Γραμμικές περιοχές τροπικής απεικόνισης

Μέχρι τώρα έχουμε καταφέρει να υπολογίσουμε ένα άνω φράγμα στον αριθμό των γραμμικών περιοχών ενός τροπικού πολυωνύμου. Προκειμένου να μπορέσουμε να μελετήσουμε τις γραμμικές περιοχές των επιπέδων ενός νευρωνικού δικτύου, μένει να επεκτείνουμε την τεχνική μας σε μία πιο πολύπλοκη δομή, την τροπική πολυωνυμική απεικόνιση.

Ορισμός. Η απεικόνιση $f = (f_1, ..., f_m) : \mathbb{R}^d \to \mathbb{R}^m$, όπου κάθε $f_i \in \mathbb{R}_{\max}[\mathbf{x}]$ ονομάζεται *τροπική πολυωνυμική απεικόνιση.*

Ακολουθεί ένας ορισμός ο οποίος αποτελεί την τροπική έννοια η οποία αντιστοιχεί στο σύνολο των ριζών ενός πολυωνύμου της κλασσικής άλγεβρας.

Ορισμός. Η *τροπική υπερεπιφάνεια* $\mathcal{T}(f)$ ενός τροπικού πολυωνύμου που γράφεται $f(\mathbf{x}) = \max_i\{\mathbf{a}_i^T\mathbf{x} + b_i\}$ ορίζεται ως το σύνολο των σημείων του χώρου που το πολυώνυμο δεν είναι διαφορίσιμο, ή ισοδύναμα το σύνολο των σημείων όπου δύο όροι του πολυωνύμου ταυτίζονται

$$\mathcal{T}(f) = \{\mathbf{x} \in \mathbb{R}^d \mid f(\mathbf{x}) = \mathbf{a}_i^T\mathbf{x} + b_i = \mathbf{a}_j^T\mathbf{x} + b_j,\ (\mathbf{a}_i^T,\ b_i) \neq (\mathbf{a}_j^T,\ b_j)\}$$

Όμοια, για μία τροπική πολυωνυμική απεικόνιση $f = (f_1, ..., f_m)$ η τροπική της υπερεπιφάνεια ορίζεται το σύνολο των σημείων όπου η απεικόνιση δεν είναι διαφορίσιμη.

Η τροπική υπερεπιφάνεια ενός πολυωνύμου αποτελεί την δυική μορφή του Newton πολυτόπου, όπως περιγράφεται στο [23]. Ωστόσο, εδώ τα πολυώνυμα μας έχουν πραγματικούς συντελεστές και δεν μπορούμε να κάνουμε χρήση αυτής της πρότασης. Παρ' όλα αυτά δεν θα χρειαστεί για την αναλυσή μας. Παρουσιάζουμε το ακόλουθο λήμμα που συνδέει τις γραμμικές περιοχές μίας τροπικής πολυωνυμικής απεικόνισης με τις γραμμικές περιοχές τροπικών πολυωνύμων.

Λήμμα 2.1. *([6], Proposition 5) Θεωρούμε μία τροπική απεικόνιση $f = (f_1, ..., f_m)$, όπου κάθε f_i είναι ένα τροπικό πολυώνυμο d μεταβλητών στον $\mathbb{R}_{max}[\mathbf{x}]$. Τότε, το πλήθος των γραμμικών περιοχών της f δίνεται από τον άνω φλοιό του Minkowski αθροίσματος*

$$\bigoplus_{i=1}^{m} \mathrm{ENewt}(f_i)$$

Απόδειξη. Αρχικά, θα μελετήσουμε την υπερεπιφάνεια που επάγει η γραμμική απεικόνιση. Αν ένα σημείο $\mathbf{x}$ βρίσκεται στην $\mathcal{T}(f)$, τότε λόγω του ορισμού της $\mathcal{T}(f)$ θα υπάρχει i ώστε $\mathbf{x} \in \mathcal{T}(f_i)$. Αντίστροφα, αν $\mathbf{x} \in \mathcal{T}(f_i)$ τότε δύο όροι της f_i θα είναι ίσοι, οπότε και $\mathbf{x} \in \mathcal{T}(f)$. Έπεται λοιπόν ότι

$$\mathcal{T}(f) = \bigcup_{i=1}^{m} \mathcal{T}(f_i) = \mathcal{T}\left(\sum_{i=1}^{m} f_i\right)$$

Γνωρίζουμε, ότι το πλήθος των γραμμικών περιοχών είναι ίσο με το πλήθος των κορυφών του πολυτόπου που είναι δυικό με την τροπική υπερεπιφάνεια. Στην περίπτωση μας το πολύτοπο είναι το

$$\mathrm{ENewt}\left(\sum_{i=1}^{m} f_i\right) = \bigoplus_{i=1}^{m} \mathrm{ENewt}\left(f_i\right)$$

οπότε προκύπτει το ζητούμενο. □

Με το λήμμα 2.1 μπορούμε να επικεντρώσουμε το ενδιαφέρον μας αποκλειστικά στην καταμέτρηση των γραμμικών περιοχών τροπικών πολυωνύμων. Το λήμμα αυτό θα χρειαστεί εκτενώς στο επόμενο κεφάλαιο όπου θα καταμετρήσουμε γραμμικές περιοχές νευρωνικών δικτύων.

2.5 Τροπική Διαίρεση

Στην ενότητα αυτή θα κάνουμε μία μικρή εισαγωγή σε μία πρόσφατη θεωρητική συμβολή στην τροπική γεωμετρία. Αυτή είναι ο ορισμός και οι εφαρμογές της τροπικής διαίρεσης [36, 38]. Η τροπική διαίρεση αποτελεί το θεωρητικό πλαίσιο για την υλοποίηση και ερμηνεία αλγορίθμων συμπίεσης νευρωνικών δικτύων. Σε αυτήν την ενότητα θα κάνουμε μία απλή αναφορά στον αλγόριθμο της τροπικής διαίρεσης τροποποιώντας ελαφρώς τον ορισμό του, χωρίς να αναφερθούμε σε εφαρμογές. Αυτό έχει ως στόχο την αποτύπωση κάποιων πρώτων ιδεών για την τροπική διαίρεση που θα μπορούσαν να αποτελέσουν έναυσμα για κάποιον που επιθυμεί να επεκταθεί ερευνητικά στην θεωρητική της εξέλιξη.

Ορισμός. Για δύο τροπικά πολυώνυμα $f, d \in \mathbb{R}_{\max}[\mathbf{x}]$ n μεταβλητών, θα γράφουμε την *τροπική διαίρεση* του f με το d ως

$$f(\mathbf{x}) = \max\{q(\mathbf{x}) + d(\mathbf{x}),\, r(\mathbf{x})\}$$

Το q ονομάζεται πηλίκο της διαίρεσης και το r υπόλοιπο. Το r επιλέγεται έτσι ώστε το πλήθος των κοινών κορυφών του $UF(\mathrm{ENewt}\,(r))$ με το $UF(\mathrm{ENewt}\,(f))$ να γίνεται ελάχιστο.

Δοθέντων δύο τροπικών πολυωνύμων f, d, το πηλίκο q και το υπόλοιπο r της τροπικής τους διαίρεσης υπολογίζονται μέσω του αλγορίθμου 1. Σημειώνουμε ότι με $\mathcal{V}_f, \mathcal{V}_d$ συμβολίζουμε τα σύνολα των κορυφών των άνω φλοιών των $\mathrm{ENewt}\,(f)$, $\mathrm{ENewt}\,(d)$ αντίστοιχα.

Θεώρημα 2.5. *(Ύπαρξη πηλίκου και υπολοίπου) Ο αλγόριθμος 1 προσδιορίζει q, r που ικανοποιούν τον ορισμό της τροπικής διαίρεσης.*

Απόδειξη. Σύμφωνα με την Πρόταση 2.3 αρκεί να αποδείξουμε ότι

$$UF(\mathrm{ENewt}\,(f)) = UF(\mathrm{ENewt}\,(\max\{q+d,\, r\})) = UF(\mathrm{ENewt}\,(q+d) \cup \mathrm{ENewt}\,(r))$$

Όμως, λόγω κατασκευής, το $q+d$ περιέχει όρους που βρίσκονται είτε στο $UF(\mathrm{ENewt}\,(f))$ είτε κάτω από αυτό. Επίσης, το $UF(\mathrm{ENewt}\,(r))$ περιέχει όλες τις κορυφές του $UF(\mathrm{ENewt}\,(f))$ που δεν τις συναντάμε στο $UF(\mathrm{ENewt}\,(q+d))$. Οπότε, πράγματι $f = \max\{q+d,\, r\}$.

Για να ολοκληρωθεί η απόδειξη, μένει να δείξουμε ότι, το $UF(\mathrm{ENewt}\,(r))$ έχει τον ελάχιστο δυνατό αριθμό κοινών κορυφών με το $UF(\mathrm{ENewt}\,(f))$. Έστω ότι υπάρχουν $\tilde{q}, \tilde{r}$ με το $\tilde{r}$ να έχει λιγότερες κοινές κορυφές που ικανοποιούν τον ορισμό της διαίρεσης

$$f = \max\{\tilde{q} + d,\, \tilde{r}\}$$

Algorithm 1 Αλγόριθμος τροπικής διαίρεσης

1. Ορίζουμε
$$q(\mathbf{x}) = \max_{(\mathbf{a}^T, b)\in\mathcal{V}_f, (\mathbf{c}^T, d)\in\mathcal{V}_d} \left\{(\mathbf{a}-\mathbf{c})^T\mathbf{x} + (b-d)\right\}$$
όπου τα $(\mathbf{a}^T, b) \in \mathcal{V}_f$, $(\mathbf{c}^T, d) \in \mathcal{V}_d$ επιλέγονται έτσι ώστε ο άνω φλοιός του πολυτόπου $\mathrm{ENewt}\left((\mathbf{a}-\mathbf{c})^T\mathbf{x} + (b-d) + d(\mathbf{x})\right)$ να βρίσκεται κάτω από το $UF(\mathrm{ENewt}\,(f))$. Αν τέτοια $(\mathbf{a}^T, b), (\mathbf{c}^T, d)$ δεν υπάρχουν θέτουμε $q(\mathbf{x}) = -\infty$.

2. Έπειτα υπολογίζουμε
$$r(\mathbf{x}) = \max_{(\mathbf{a}^T, b)\in\mathcal{V}_f} \left(\mathbf{a}^T\mathbf{x} + b\right)$$
όπου τα $(\mathbf{a}^T, b) \in \mathcal{V}_f$ είναι οι κορυφές του $\mathcal{V}_f$ που δεν μπόρεσαν να καλυφθούν στο προηγούμενο βήμα. Δηλαδή, αυτές για τις οποίες δεν υπάρχει $(\mathbf{c}^T, d) \in \mathcal{V}_d$ ώστε ο άνω φλοιός του $\mathrm{ENewt}\left((\mathbf{a}-\mathbf{c})^T\mathbf{x} + (b-d) + d(\mathbf{x})\right)$ να βρίσκεται κάτω από το $UF(\mathrm{ENewt}\,(f))$. Αν τέτοια $(\mathbf{a}^T, b)$ δεν υπάρχουν θέτουμε $r(\mathbf{x}) = -\infty$.

3. Επιστρέφουμε ως έξοδο το πηλίκο q και το υπόλοιπο r της διαίρεσης του f με το d.

Τότε, λόγω της Πρότασης 2.3 έχουμε ισοδύναμα ότι

$$UF(\mathrm{ENewt}\,(f)) = UF(\mathrm{ENewt}\,(\max\{\tilde{q}+d, \tilde{r}\})) = UF(\mathrm{ENewt}\,(\tilde{q}+d) \cup \mathrm{ENewt}\,(\tilde{r}))$$

Για να είναι οι άνω φλοιοί των εν λόγω πολυτόπων ίδιοι, θα πρέπει να περιέχουν ακριβώς τις ίδιες κορυφές. Έστω ότι A είναι το σύνολο των κοινών κορυφών του $UF(\mathrm{ENewt}\,(q+d))$ με το $UF(\mathrm{ENewt}\,(f))$. Επίσης, ορίζουμε B το σύνολο των κοινών κορυφών του $UF(\mathrm{ENewt}\,(r))$ με το $UF(\mathrm{ENewt}\,(f))$. Ομοίως, ορίζουμε τα $\tilde{A}, \tilde{B}$ για τα $\tilde{q}, \tilde{r}$.

Η υπόθεση μας δίνει ότι $|B| > |\tilde{B}|$. Επιπλέον, όπως αναφέραμε, θα πρέπει $A \cup B = \tilde{A} \cup \tilde{B} = \mathcal{V}_f$. Συνεπώς, θα είναι $|A| < |\tilde{A}|$. Αυτό σημαίνει ότι υπάρχει μία κορυφή $(\mathbf{a}^T, b) \in \tilde{A} \subseteq \mathcal{V}_f$ η οποία δεν υπάρχει στο ανάπτυγμα του $q+d$, αλλά υπάρχει στο ανάπτυγμα του $\tilde{q}+d$.

Αυτό όμως θα σημαίνει ότι υπάρχει $(\mathbf{c}^T, d) \in \mathcal{V}_d$, ώστε το $(\mathbf{a}-\mathbf{c})^T x + (b-d) + d(\mathbf{x})$ να περιέχει την κορυφή $(\mathbf{a}^T, b)$ στο ανάπτυγμα του. Εφόσον, όμως αυτό δεν συμβαίνει στο $q+d$, και δεδομένου ότι η επιλογή του q λαμβάνει υπόψιν όλες τις δυνατές κορυφές, η κορυφή αυτή θα παραβιάζει τον περιορισμό ότι ο άνω φλοιός του $\mathrm{ENewt}\left((\mathbf{a}-\mathbf{c})^T x + (b-d) + d(\mathbf{x})\right)$ βρίσκεται κάτω από τον άνω φλοιό του $\mathrm{ENewt}\,(f)$. Τότε, λόγω της Πρότασης 2.4 θα πάρουμε ότι $\tilde{q}+d$ δεν είναι μικρότερο από το f, που είναι άτοπο. Το ζητούμενο έπεται. □

Πρόταση 2.7. *(Μοναδικότητα κορυφών διαίρεσης) Στην τροπική διαίρεση του f με το d, το πηλίκο και το υπόλοιπο δεν είναι μοναδικά ούτε πολυωνυμικά, ούτε ως προς συναρτησιακή ισοδυναμία. Ωστόσο, η διαμέριση των κορυφών του $\mathcal{V}_f$ που επάγουν είναι μοναδική. Δηλαδή αν έχουμε*

$$f(\mathbf{x}) = \max\{q(\mathbf{x}) + d(\mathbf{x}),\, r(\mathbf{x})\} = \max\{\tilde{q}(\mathbf{x}) + d(\mathbf{x}),\, \tilde{r}(\mathbf{x})\}$$

τότε $A = \tilde{A}$ και $B = \tilde{B}$.

Απόδειξη. Ορίζουμε $A, B, \tilde{A}, \tilde{B}$, όπως στην απόδειξη του θεωρήματος 2.5. Από τον ορισμό της τροπικής διαίρεσης το υπόλοιπο έχει τον ελάχιστο δυνατό αριθμό κοινών κορυφών με τον διαιρετέο, οπότε έχουμε ότι $|B| = |\tilde{B}|$. Αν $B \neq \tilde{B}$, τότε και $A \neq \tilde{A}$, οπότε υπάρχει $(\mathbf{a}^T, b) \in \tilde{A} \setminus A$. Ωστόσο, αυτό μας δίνει άτοπο για τον ίδιο λόγο που περιγράψαμε στην απόδειξη του Θεωρήματος 2.5. Συνεπώς, $B = \tilde{B}$ και $A = \tilde{A}$. □

2.5.1 Παραγοντοποίηση Τροπικών Πολυωνύμων μίας Μεταβλητής

Η παραγοντοποίηση τροπικών πολυωνύμων ως γινόμενο γραμμικών παραγόντων είναι NP-complete στην πολυμεταβλητή περίπτωση [21]. Ωστόσο, για την περίπτωση της μίας μεταβλητής μπορούμε να προσδιορίσουμε εύκολα την παραγοντοποίηση του πολυωνύμου [14]. Εδώ θα παρουσιάσουμε την απόδειξη για την παραγοντοποίηση των πολυωνύμων μίας μεταβλητής, στην γενική περίπτωση όπου οι συντελεστές είναι πραγματικοί αριθμοί. Αυτό θα αποτελέσει μία ευκαιρία να αναδείξουμε την χρησιμότητα των προτάσεων που έχουμε παρουσιάσει ως τώρα.

Θεωρούμε ένα πολυώνυμο μίας μεταβλητής

$$f(\mathbf{x}) = \max_i \{a_i x + b_i\}$$

όπου A είναι ένα πεπερασμένο σύνολο πραγματικών αριθμών. Ισχυριζόμαστε ότι μπορούμε σε αυτήν την περίπτωση να πάρουμε την εξής παραγοντοποίηση.

Θεώρημα 2.6. *Το τροπικό πολυώνυμο $f(x)$ μπορεί να παραγοντοποιηθεί στην μορφή*

$$f(x) = c + \lambda_1 \max\{x, \rho_1\} + ... + \lambda_m \max\{x, \rho_m\}$$

Απόδειξη. Θεωρούμε ότι ο άνω φλοιός $UF(\text{ENewt}\,(f))$, αποτελείται από τις κορυφές

$$\{(a_1, b_1), ..., (a_m, b_m)\}$$

με $a_1 \leq a_2 \leq ... \leq a_m$. Ο άνω φλοιός είναι κοίλος, συνεπώς θα πρέπει η ακολουθία

$$x_i = -\frac{b_{i+1} - b_i}{a_{i+1} - a_i}$$

να είναι αύξουσα. Το πολυώνυμο $f(x)$ λόγω της Πρότασης 2.2 μπορεί να γραφεί ως

$$f(x) = \max_{i=1,...,m} \{a_i x + b_i\}$$

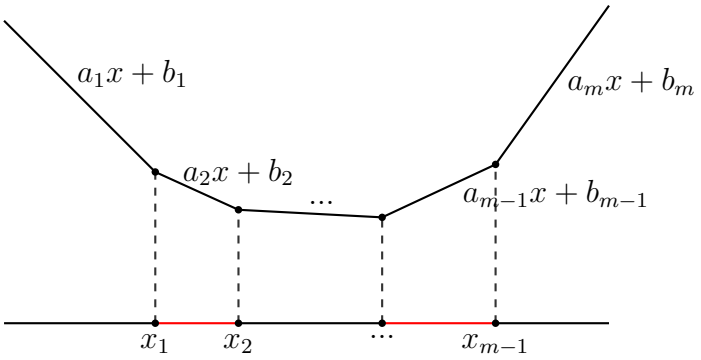
Αξίζει να παρατηρήσουμε ότι το πολυώνυμο για κάθε x παίρνει την τιμή μίας εκ των ευθειών $a_1 x + b_1, ..., a_m x + b_m$ και μάλιστα αυτής με την μεγαλύτερη τιμή. Επιπλέον, το σημείο x_i αποτελεί το σημείο τομής της ευθείας $a_{i+1} x + b_{i+1}$ με την $a_i x + b_i$.

Η σπουδαιότερη παρατήρηση είναι ότι η $i-$οστή ευθεία τέμνεται πρώτα με την $i+1$ από όλες τις ευθείες $i+1, i+2, ..., m$. Αυτό συμβαίνει διότι ο άνω φλοιός $UF(\text{ENewt}\,(f))$ είναι κοίλος, οπότε

$$-\frac{b_{i+1} - b_i}{a_{i+1} - a_i} \leq -\frac{b_j - b_i}{a_j - a_i}$$

Συνεπώς, η $f(x)$ θα καθορίζεται με τον εξής τρόπο. Αρχικά, στο διάστημα $(-\infty, x_1]$ θα είναι ίση με $a_1x + b_1$. Έπειτα, η πρώτη ευθεία που συναντάμε είναι η $a_2x + b_2$ η οποία για $x > x_1$ είναι μεγαλύτερη της $a_1x + b_1$. Συνεπώς, στο διάστημα $(x_1, x_2]$ η $f(x)$ παίρνει την τιμή της ευθείας 2. Στο διάστημα αυτό η ευθεία 2 είναι πάνω από την ευθεία 1 αλλά και από όλες τις υπόλοιπες ευθείες, αφού το σημείο τομής της με την 3η ευθεία είναι το x_2. Αξίζει να σημειώσουμε ότι εκμεταλλευόμαστε το γεγονός ότι οι κλίσεις των ευθειών αυξάνονται, οπότε η 3η ευθεία θα είναι μεγαλύτερη από την 2 έπειτα από το x_2.

Επεκτείνουμε τον συλλογισμό μας αυτό και βλέπουμε ότι η $f(x)$ παίρνει την τιμή της ευθείας i στο διάστημα $(x_{i-1}, x_i]$ με $x_0 = -\infty, x_m = \infty$. Τελικά, η γραφική παράσταση της $f(x)$ θα μοιάζει όπως παρακάτω.



Σχήμα 2.5: Γραφική παράσταση τροπικού πολυωνύμου μίας μεταβλητής.

Επομένως, θα έχουμε

$$\begin{aligned} f(x) &= (a_1x + b_m) + \max\{(a_2 - a_1)x, (b_1 - b_2)\} + ... + \max\{(a_m - a_{m-1})x, (b_{m-1} - b_m)\} \\ &= b_m + \max\{a_1x, -\infty\} + \max\{(a_2 - a_1)x, (b_1 - b_2)\} + ... + \max\{(a_m - a_{m-1})x, (b_{m-1} - b_m)\} \\ &= b_m + a_1 \max\{x, -\infty\} + (a_2 - a_1) \max\left\{x, \frac{b_1 - b_2}{a_2 - a_1}\right\} + ... + (a_m - a_{m-1}) \max\left\{x, \frac{b_{m-1} - b_m}{a_m - a_{m-1}}\right\} \end{aligned}$$

όπως ήταν επιθυμητό. □

Παρατήρηση. *Οι όροι* $\rho_i = \frac{b_i - b_{i+1}}{a_{i+1} - a_i}$, $i = 1, ..., m-1$ *αποτελούν τις ρίζες του πολυωνύμου, δηλαδή τα σημεία που ανήκουν στην τροπική υπερεπιφάνεια* $\mathcal{T}(f)$.

Κεφάλαιο 3

Τροπική Γεωμετρία Νευρωνικών Δικτύων

Στο Κεφάλαιο αυτό θα μελετήσουμε με τροπική γεωμετρία τις ιδιότητες των Νευρωνικών Δικτύων με ReLU ενεργοποιήσεις. Η ανάλυση αυτή θα αναδείξει την χρησιμότητα της τροπικής Γεωμετρίας στην θεωρητική κατανόηση των ιδιοτήτων των νευρωνικών δικτύων, αλλά και θα δώσει την βάση για την κατασκευή των αλγορίθμων συμπίεσης.

Η πορεία που θα ακολουθήσουμε είναι η εξής. Αρχικά, θα ορίσουμε το μοντέλο του δικτύου που μελετάμε και θα ορίσουμε τις εξισώσεις του δικτύου τροπικά. Οι εξισώσεις αυτές όπως θα δούμε ανάγουν το πρόβλημά μας στην μελέτη τροπικών ρητών απεικονίσεων. Όπως αποδείχθηκε στο [45], η οικογένεια των τροπικών ρητών απεικονίσεων είναι ισοδύναμη με την οικογένεια των νευρωνικών με ReLU ενεργοποιήσεις. Επομένως, αποτελεί ισχυρό εργαλείο για την μελέτη των νευρωνικών υπό το τροπικό πρίσμα.

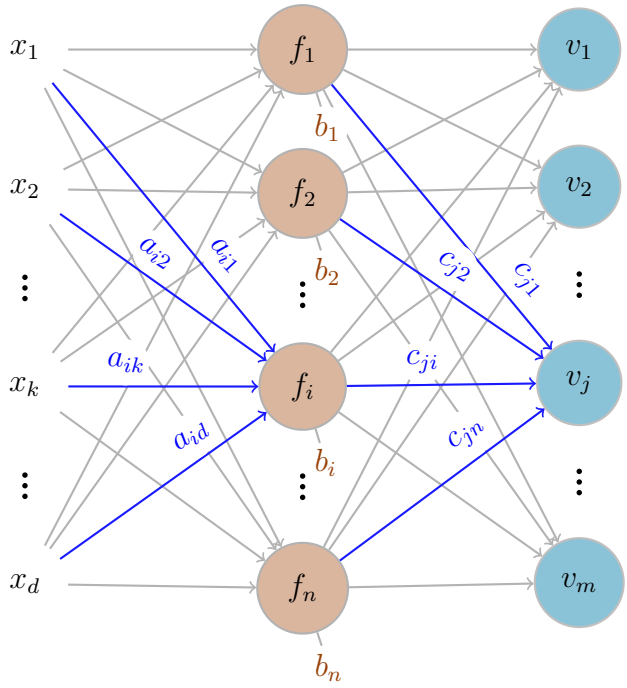
Έπειτα, θα αναπαραστήσουμε το επίπεδο ενός νευρωνικού δικτύου γεωμετρικά με χρήση ζωνοτόπων. Η αναπαράσταση αυτή έχει διπλό όφελος. Πρώτον θα μας εξυπηρετήσει στην υλοποίηση των γεωμετρικών αλγορίθμων συμπίεσης και δεύτερον, σε συνδυασμό με θεωρητικά αποτελέσματα καταμέτρησης γραμμικών περιοχών θα καταφέρουμε να προσεγγίσουμε το πλήθος των γραμμικών περιοχών βαθιών Νευρωνικών Δικτύων.

3.1 Τροπικές Εξισώσεις

Έχοντας αναλύσει τα βασικά εργαλεία της τροπικής γεωμετρίας που θα χρησιμοποιήσουμε, μπορούμε πλέον να δούμε πως αυτά εφαρμόζονται στην ανάλυση ενός Νευρωνικού Δικτύου με ReLU ενεργοποιήσεις. Η ανάλυση μας θα γίνει στο νευρωνικό δίκτυο με ένα κρυφό επίπεδο, του σχήματος 3.1.

Το δίκτυο αποτελείται από ένα επίπεδο εισόδου $\mathbf{x} = (x_1, ..., x_d)$ μεγέθους d, ένα κρυφό επίπεδο $f = (f_1, ..., f_n)$ μεγέθους n με ReLU ενεργοποιήσεις, καθώς και ένα επίπεδο εξόδου $v = (v_1, ..., v_m)$ μεγέθους m. Οι συνδέσεις μεταξύ των επιπέδων γίνονται μέσω δύο γραμμικών μετασχηματισμών (linear layers). Τα βάρη $\{a_{ik}\}_{i\in[n],k\in[d]}$ μαζί με τους σταθερούς όρους $\{b_i\}_{i\in[n]}$ καθορίζουν το πρώτο γραμμικό επίπεδο από την είσοδο στο κρυφό επίπεδο, ενώ τα $\{c_{ji}\}_{j\in[m],i\in[n]}$ αφορούν το δεύτερο γραμμικό επίπεδο. Σημειώνουμε ότι στο επίπεδο εξόδου θα αγνοήσουμε τους σταθερούς όρους στην ανάλυση μας.

Στο κείμενο, όπου αναφερόμαστε σε συμπίεση νευρωνικού δικτύου θα εννοούμε συμπίεση του δικτύου του σχήματος 3.1. Μάλιστα, θα χρησιμοποιούμε τον συμβολισμό A για τον πίνακα που αναπαριστά το πρώτο γραμμικό επίπεδο, με $A_{i,:} = (\mathbf{a}_i^T, b_i)$ και C για τον πίνακα του δεύτερου γραμμικού επιπέδου με $C_{ji} = c_{ji}$.



Σχήμα 3.1: Σχηματική αναπαράσταση Νευρωνικού Δικτύου με επίπεδο εισόδου διάστασης d, ένα hidden Layer διάστασης n και επίπεδο εξόδου διάστασης m.

Σύμφωνα με τις παραπάνω παραδοχές μπορούμε να υπολογίσουμε τις εξόδους των κόμβων του δικτύου. Αρχικά, για το output του $i-$οστού κόμβου του κρυμμένου επιπέδου έχουμε ότι

$$f_i(\mathbf{x}) = \max\left(\sum_{k=1}^{d} a_{ik}x_k + b_i, 0\right) = \max(\mathbf{a}_i^T\mathbf{x} + b_i, 0) \tag{3.1}$$

Αυτό δείχνει ότι το ReLU επίπεδο f ισοδυναμεί με μία τροπική πολυωνυμική απεικόνιση. Επιπλέον, συμπεραίνουμε ότι κάθε f_i είναι ένα τροπικό πολυώνυμο αποτελούμενο από 2 όρους. Αυτό γεωμετρικά σημαίνει ότι το επεκτεταμένο Newton πολύτοπό του είναι ένα ευθύγραμμο τμήμα στον $\mathbb{R}^{d+1}$. Για το επίπεδο εξόδου μπορούμε να υπολογίσουμε αντίστοιχα

$$v_j(\mathbf{x}) = \sum_{i=1}^{n} c_{ji}f_i(\mathbf{x}) = \sum_{c_{ji}>0} |c_{ji}|f_i(\mathbf{x}) - \sum_{c_{ji}<0} |c_{ji}|f_i(\mathbf{x}) = p_j(\mathbf{x}) - q_j(\mathbf{x}) \tag{3.2}$$

Οι συναρτήσεις p_j, q_j είναι γραμμικοί συνδυασμοί των τροπικών πολυωνύμων $\{f_i\}_{i\in[n]}$, οπότε αποτελούν και οι δύο τροπικά πολυώνυμα, λόγω της κλειστότητας των τροπικών πολυωνύμων ως προς τις πράξεις του τροπικού πολλαπλασιασμού και κλασσικού πολλαπλασιασμού με θετική σταθερά. Την έννοια αυτή την είχαμε αναφέρει στο κεφάλαιο με την εισαγωγή στα τροπικά πολυώνυμα. Συμπεραίνουμε, επομένως, ότι η έξοδος του $j-$οστού κόμβου v_j γράφεται ως την διαφορά δύο τροπικών πολυωνύμων. Μία τέτοια πράξη δίνει συνάρτηση η οποία δεν ανήκει στον $\mathbb{R}_{\max}[\mathbf{x}]$, αλλά αποτελεί μία γενικότερη κλάση συναρτήσεων, αυτή των τροπικών ρητών συναρτήσεων, όπως ορίζουμε παρακάτω.

Ορισμός. Μία συνάρτηση $v : \mathbb{R}^d \to \mathbb{R}$ θα ονομάζεται *τροπική ρητή συνάρτηση* εάν μπορεί να γραφεί στην μορφή $v(\mathbf{x}) = p(\mathbf{x}) - q(\mathbf{x})$, με $p, q \in \mathbb{R}_{\max}[\mathbf{x}]$ τροπικά πολυώνυμα.

Με φυσικό τρόπο προκύπτει και ο ακόλουθος ορισμός ο οποίος επεκτείνει τις τροπικές ρητές συναρτήσεις.

Ορισμός. Η απεικόνιση $v = (v_1, ..., v_m) : \mathbb{R}^d \to \mathbb{R}^m$, όπου κάθε v_i είναι μία τροπική ρητή συνάρτηση, ονομάζεται *τροπική ρητή απεικόνιση*.

Το νευρωνικό της εικόνας 3.1 ως συνάρτηση $v : \mathbb{R}^d \to \mathbb{R}^m$ προσδιορίζεται μοναδικά από την έξοδο του $v = (v_1, ..., v_m)$, η οποία σύμφωνα με τον ορισμό αποτελεί μία τροπική ρητή απεικόνιση. Συνεπώς, ένα νευρωνικό δίκτυο με ReLU ενεργοποιήσεις ισοδυναμεί με μία τροπική ρητή απεικόνιση. Αυτό μάλιστα όπως θα δούμε αποδεικνύεται και γενικότερα σε βαθύτερα νευρωνικά δίκτυα. Η παρατήρηση αυτή αποτελεί ισχυρό κίνητρο για να μελετήσουμε σε μεγαλύτερο βάθος τις ιδιότητες των τροπικών ρητών συναρτήσεων και απεικονίσεων.

3.2 Τροπικές Ρητές Απεικονίσεις

Τα τροπικά πολυώνυμα και οι τροπικές πολυωνυμικές απεικονίσεις αναλύθηκαν με την χρήση των επεκτεταμένων Newton πολυτόπων. Ωστόσο, όπως παρατηρήσαμε προηγουμένως, τα τροπικά πολυώνυμα δεν είναι αρκετά για να περιγράψουν την οικογένεια των Νευρωνικών Δικτύων και θα πρέπει να χρησιμοποιήσουμε ισχυρότερα εργαλεία. Στην ενότητα αυτή, επομένως, θα αναλύσουμε τις ιδιότητες των τροπικών ρητών συναρτήσεων και απεικονίσεων. Αυτές θα αξιοποιηθούν για την μελέτη των Νευρωνικών Δικτύων και την καταμέτρηση των γραμμικών περιοχών.

Μία σημαντική παρατήρηση που μας οδηγεί στην μελέτη των τροπικών ρητών απεικονίσεων είναι ότι η σύνθεση δύο τροπικών πολυωνύμων, όπως έχουμε προαναφέρει, δεν είναι απαραίτητα τροπικό πολυώνυμο. Αυτό συμβαίνει διότι έχουμε επιτρέψει στους κλίσεις (συντελεστές) να είναι οποιαδήποτε διανύσματα στον $\mathbb{R}^n$. Κατά αυτόν τον τρόπο, κατά την σύνθεση δύο τροπικών πολυωνύμων μπορεί να προκύψει πολλαπλασιασμός ενός αρνητικού συντελεστή με μία έκφραση max. Ο αρνητικός συντελεστής δεν επιμερίζεται με την έκφραση max με αποτέλεσμα το τελικό αποτέλεσμα να μην μπορεί να γραφεί σαν έκφραση μεγίστου γραμμικών όρων. Αυτό μπορεί θα γίνει καλύτερα κατανοητό με το ακόλουθο παράδειγμα.

Παράδειγμα 3.1. *Έστω* $f(x) = \max(x + 1, 2x)$ *και* $g(x) = max(-x + 1, 3)$ *δύο τροπικά πολυώνυμα μίας μεταβλητής. Τότε έχουμε*

$$g(f(x)) = \max(-f(x) + 1, 3) = \max(1, 3 + f(x)) - f(x) =$$

$$= \max(1, x + 4, 2x + 3) - \max(x + 1, 2x)$$

που δεν είναι τροπικό πολυώνυμο αφού το $-\max(x+1, 2x)$ *δεν μπορεί να γραφεί ως maximum γραμμικών όρων. Πράγματι, ισχύει γενικά ότι*

$$-\max\{a, b\} = \min\{-a, -b\} \neq \max\{-a, -b\}$$

Συνεπώς οι τροπικές ρητές συναρτήσεις αποτελούν μία επέκταση του ημιδακτυλίου των τροπικών συναρτήσεων. Θα δείξουμε ότι αυτές επιτυγχάνουν να διορθώσουν το πρόβλημα των τροπικών πολυωνυμικών απεικονίσεων που δεν είναι κλειστές ως προς την σύνθεση. Για να το αποδείξουμε την κλειστότητα των τροπικών ρητών απεικονίσεων ως προς την σύνθεση, θα πρέπει αρχικά να δείξουμε ορισμένα απλούστερα αποτελέσματα, ώστε σταδιακά να χτίσουμε το τελικό. Αρχικά, λοιπόν, θα δείξουμε ότι δύο τροπικές ρητές συναρτήσεις είναι κλειστές ως προς τις πράξεις του τροπικού ημιδακτυκλίου $\mathbb{R}_{\max}$, αλλά και την αφαίρεση.

Πρόταση 3.1. *Έστω, f, g δύο τροπικές ρητές συναρτήσεις. Τότε και οι*

$$f \vee g,\ f + g,\ f - g$$

είναι, επίσης, τροπικές ρητές συναρτήσεις.

Απόδειξη. Ας υποθέσουμε ότι οι ρητές συναρτήσεις γράφονται ως $f(\mathbf{x}) = u(\mathbf{x}) - v(\mathbf{x}), g(\mathbf{x}) = p(\mathbf{x}) - q(\mathbf{x})$, όπου $u, v, p, q \in \mathbb{R}_{\max}[\mathbf{x}]$ τροπικά πολυώνυμα. Τότε

$$f \vee g = \max\{f, g\} = \max\{u - v, p - q\} = \max\{u + q, p + v\} - (v + q) =$$

$$= ((u + q) \vee (p + v)) - (v + q)$$

$$f + g = f + g = u + p - v - q = (u + p) - (v + q)$$

$$f - g = f - g = u + q - p - v = (u + q) - (v + p)$$

Σημειώνουμε, όπως έχουμε αναφέρει στο προηγούμενο Κεφάλαιο, ότι εύκολα μπορεί κανείς να αποδείξει ότι οι τροπικές πράξεις μεταξύ τροπικών πολυωνύμων δίνουν τροπικό πολυώνυμο. Επομένως, οι συναρτήσεις που προκύπτουν είναι πράγματι τροπικές ρητές αφού γράφονται σαν διαφορές τροπικών πολυωνύμων και προκύπτει ο ισχυρισμός μας. □

Παρατήρηση. *Αξίζει, επιπλέον, να σημειωθεί ότι και ο πολλαπλασιασμός τροπικής ρητής με σταθερά λ δίνει τροπική ρητή συνάρτηση. Για παράδειγμα $\lambda f(\mathbf{x}) = \lambda u(\mathbf{x}) - \lambda v(\mathbf{x})$, όπου τα $\lambda u, \lambda v$ είναι τροπικά πολυώνυμα εάν $\lambda > 0$, αφού πολλαπλασιάζουμε κάθε μονώνυμο του πολυωνύμου με λ. Για $\lambda < 0$, εναλλακτικά γράφουμε $\lambda f(\mathbf{x}) = |\lambda| v(\mathbf{x}) - |\lambda| u(\mathbf{x})$, οπότε αποτελεί πάλι τροπική ρητή συνάρτηση.*

Με βάση την προηγούμενη πρόταση και παρατήρηση είμαστε σε θέση να αποδείξουμε το ακόλουθο ισχυρότερο αποτέλεσμα.

Πρόταση 3.2. *Έστω $f : \mathbb{R}^d \to \mathbb{R}^n$ μια τροπική ρητή απεικόνιση και $g : \mathbb{R}^n \to \mathbb{R}^1$ ένα τροπικό πολυώνυμο. Τότε η $g \circ f$ είναι τροπική ρητή συνάρτηση.*

Απόδειξη. Γράφουμε την f ως διάνυσμα τροπικών ρητών συναρτήσεων

$$f(\mathbf{x}) = \begin{pmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_n(\mathbf{x}) \end{pmatrix}$$

και θεωρούμε $g(\mathbf{x}) = \max_i\{\mathbf{a}_i^T \mathbf{x} + b_i\}$. Τότε προκύπτει

$$g(f(\mathbf{x})) = \max_i \left\{\mathbf{a}_i^T f(\mathbf{x}) + b_i\right\} = \max_i \left\{a_{i1} f_1(\mathbf{x}) + ... + a_{in} f_n(\mathbf{x}) + b_i\right\}$$

Η παράσταση $a_{i1} f_1(\mathbf{x}) + ... + a_{in} f_n(\mathbf{x}) + b_i$ αποτελεί για κάθε i τροπική ρητή συνάρτηση, αφού είναι γραμμένη σαν τροπικό γινόμενο τροπικών ρητών, οπότε και η $g \circ f$ είναι τροπική ρητή συνάρτηση, ως τροπικό άθροισμα τροπικών ρητών συναρτήσεων. □

Τέλος, αποδεικνύουμε το επιθυμητό αποτέλεσμα για την κλειστότητα των τροπικών ρητών απεικονίσεων.

Πρόταση 3.3. *(Κλειστότητα) Έστω $f : \mathbb{R}^d \to \mathbb{R}^n$ και $g : \mathbb{R}^n \to \mathbb{R}^m$, δύο τροπικές ρητές απεικονίσεις. Τότε και η $g \circ f$ είναι, επίσης, τροπική ρητή απεικόνιση.*

Απόδειξη. Γράφουμε την σύνθεση διανυσματικά

$$g(f(\mathbf{x})) = \begin{pmatrix} g_1(f(\mathbf{x})) \\ g_2(f(\mathbf{x})) \\ \vdots \\ g_m(f(\mathbf{x})) \end{pmatrix}$$

Αρκεί, λοιπόν, να δείξουμε ότι κάθε $g_i(f(\mathbf{x}))$ είναι τροπική ρητή συνάρτηση. Όμως, κάθε g_i γράφεται σαν διαφορά τροπικών πολυωνύμων, έστω $g_i = p_i - q_i = p_i - q_i$, οπότε

$$g_i(f(\mathbf{x})) = p_i(f(\mathbf{x})) - q_i(f(\mathbf{x}))$$

που είναι τροπική ρητή συνάρτηση, αφού λόγω της Πρότασης 3.2 τα $p_i(f(\mathbf{x})), q_i(f(\mathbf{x}))$ είναι τροπικές ρητές συναρτήσης και λόγω της 3.1 η διαφορά τους είναι, επίσης, τροπική ρητή συνάρτηση. Το ζητούμενο έπεται.

□

Στην περίπτωση του νευρωνικού δικτύου της εικόνας 3.1 με ένα κρυφό επίπεδο, είδαμε ότι η συνάρτηση εξόδου ισοδυναμεί με μία τροπική ρητή απεικόνιση. Με το ακόλουθο Θεώρημα γενικεύουμε αυτήν την παρατήρηση σε δίκτυα με περισσότερα κρυφά επίπεδα.

Θεώρημα 3.1. *([45]) Η απεικόνιση $F : \mathbb{R}^d \to \mathbb{R}^m$ ενός βαθιού νευρωνικού δικτύου, αποτελούμενο από L κρυφά επίπεδα με ReLU ενεργοποιήσεις, ισοδυναμεί με τροπική ρητή απεικόνιση.*

Απόδειξη. Αρκεί να παρατηρήσουμε ότι μπορούμε να γράψουμε την απεικόνιση του νευρωνικού δικτύου σαν σύνθεση των ReLU επιπέδων και ενός γραμμικού επιπέδου:

$$F = g \circ f_L \circ ... \circ f_1$$

όπου $f_l : \mathbb{R}^{n_{l-1}} \to \mathbb{R}^{n_l}$ με $n_0 = d$ η διάσταση εισόδου, $\{n_l\}_{l \in [L]}$ οι διαστάσεις των L κρυφών επιπέδων και $g : \mathbb{R}^{n_L} \to \mathbb{R}^m$ το τελικό γραμμικό επίπεδο που συνδέει το $L-$οστό κρυφό επίπεδο με το επίπεδο εξόδου.

Όπως διαπιστώσαμε προηγουμένως, κάθε f_l αναπαριστά μία τροπική πολυωνυμική απεικόνιση. Το ίδιο επίσης συμβαίνει και με την γραμμική απεικόνιση g, αφού πρακτικά αποτελείται από τροπικά πολυώνυμα με έναν μόνο γραμμικό όρο. Συνεπώς, από την Πρόταση 3.3, παίρνουμε ότι η σύνθεση τους $g \circ f_L \circ ... \circ f_1$ είναι μία τροπική ρητή απεικόνιση και λαμβάνουμε το ζητούμενο. □

3.3 Ζωνότοπα

Οι τροπικές ρητές συναρτήσεις, αν και θα μας φανούν χρήσιμες στον υπολογισμό των γραμμικών περιοχών νευρωνικών δικτύων, δεν έχουν κάποια προφανή γεωμετρική οπτικοποίηση. Στην πραγματικότητα, δεν είναι μέχρι τώρα γνωστό αν επιδέχονται μία πολυτοπική αναπαράσταση που να καθορίζει τις τιμές τους, όπως συμβαίνει στα τροπικά πολυώνυμα. Για να βρούμε μία γεωμετρική αναπαράσταση για τα νευρωνικά δίκτυα θα πρέπει να εστιάσουμε χωριστά στα πολυώνυμα που συνθέτουν την τροπική ρητή συνάρτηση του δικτύου.

Η γεωμετρική δομή που θα χρησιμοποιήσουμε είναι το ζωνότοπο, δηλαδή το Minkowski άθροισμα ευθυγράμμων τμημάτων. Συγκεκριμένα, θα δείξουμε ότι κάθε νευρώνας αντιστοιχεί σε ένα ευθύγραμμο τμήμα και κάθε κόμβος εξόδου αναπαρίσταται από δύο ζωνότοπα που καθορίζονται από το θετικό και αρνητικό μέλος της τροπικής ρητής συνάρτησης. Επομένως, το ζωνότοπο θα αποτελέσει τον θεμελιώδη γεωμετρικό λίθο για την αναπαράσταση του δικτύου.

Θεωρούμε πάλι το δίκτυο της εικόνας 3.1 και εστιάζουμε στην $j-$οστή έξοδο v_j. Αυτή όπως έχουμε αναφέρει γράφεται σαν τροπική ρητή συνάρτηση $v_j(\mathbf{x}) = p_j(\mathbf{x}) - q_j(\mathbf{x})$. Θα χρησιμοποιήσουμε τους συμβολισμούς P_j, Q_j για τα επεκτεταμένα Newton πολύτοπα των p_j, q_j. Τότε εύκολα μπορεί να δει κανείς ότι και τα δύο πολύτοπα έχουν την δομή ζωνοτόπων, αφού τα p_j, q_j γράφονται σαν γραμμικοί συνδυασμοί των πολυωνύμων $\{f_i\}_{i\in[n]}$ τα οποία αναπαρίστανται από ευθύγραμμα τμήματα. Θα αποκαλούμε το P_j το θετικό ζωνότοπο και Q_j το αρνητικό.

Πρόταση 3.4. *Τα πολύτοπα P_j, Q_j είναι ζωνότοπα στον $\mathbb{R}^{d+1}$.*

Απόδειξη. Όπως παρατηρήσαμε τα p_j, q_j γράφονται σαν γραμμικός συνδυασμός τροπικών πολυωνύμων που αποτελούνται από 2 όρους. Πράγματι, γράφουμε

$$p_j(\mathbf{x}) = \sum_{c_{ji}>0} c_{ji} \max(\mathbf{a}_i^T\mathbf{x} + b_i, 0) = \sum_{c_{ji}>0} \max(c_{ji}\mathbf{a}_i^T\mathbf{x} + c_{ji}b_i, 0) \xrightarrow{\text{Prop. 1}}$$
$$P_j = \bigoplus_{c_{ji}>0} \text{ENewt}\left(\max(c_{ji}\mathbf{a}_i^T\mathbf{x} + c_{ji}b_i, 0)\right)$$

Κάθε πολύτοπο $\text{ENewt}\left(\max(c_{ji}\mathbf{a}_i^T\mathbf{x} + c_{ji}b_i, 0)\right)$ είναι ένα ευθύγραμμο τμήμα με άκρα $\mathbf{0}$ και $\left(c_{ji}\mathbf{a}_i^T,\, c_{ji}b_i\right) = c_{ji}\left(\mathbf{a}_i^T,\, b_i\right)$. Επομένως, το P_j γράφεται σαν Minkowski άθροισμα ευθυγράμμων τμημάτων, που αποτελεί ζωνότοπο εξ' ορισμού. Όμοια αποδεικνύουμε ότι και το Q_j είναι ζωνότοπο. □

Παρατήρηση. *Το ζεύγος των ζωνοτόπων P_j, Q_j αποτελεί μία αναπαράσταση για την τροπική ρητή συνάρτηση v_j, η οποία ωστόσο δεν μπορεί να χαρακτηριστεί αμφιμονοσήμαντη. Για παράδειγμα η ρητή συνάρτηση $f(x) = \max\{x, 0\} - \max\{-x, 0\} = x$, μπορεί να αναπαρασταθεί με τα πολύτοπα $\text{conv}\{(1,0),(0,0)\}, \text{conv}\{(-1,0),(0,0)\}$ αλλά και με τα $\text{conv}\{(1,0)\}$, $\text{conv}\{\emptyset\}$. Δηλαδή, με αυτόν τον τρόπο δεν καταλήγουμε σε μοναδική αναπαράσταση.*

Με την Πρόταση 3.4 είδαμε ότι κάθε νευρώνας του κρυφού επιπέδου αντιστοιχεί γεωμετρικά σε ένα ευθύγραμμο τμήμα. Τα ευθύγραμμα τμήματα αυτά συνθέτουν το συνολικό ζωνότοπο και γι' αυτόν τον λόγο θα τα ονομάσουμε *γεννήτορες* του ζωνοτόπου. Έτσι, το διάνυσμα $c_{ji}\left(\mathbf{a}_i^T,\, b_i\right)$ θα αποτελεί τον $i-$οστό γεννήτορα του ζωνοτόπου της $j-$οστής εξόδου. Ο γεννήτορας αυτός συμβάλλει στην κατασκευή του θετικού ζωνοτόπου και ονομάζεται *θετικός*

γεννήτορας εάν $c_{ji} > 0$, διαφορετικά συμβάλλει στο αρνητικό ζωνότοπο και ονομάζεται *αρνητικός γεννήτορας*.

Είναι διαισθητικά σωστό να πούμε ότι ένα ζωνότοπο γίνεται πιο πολύπλοκο γεωμετρικά, όσο ο αριθμός των γεννητόρων του αυξάνει. Το διάνυσμα ενός γεννήτορα αναπαρίσταται από ένα ευθύγραμμο τμήμα με δύο άκρα, το $\mathbf{0}$ και $c_{ji}\left(\mathbf{a}_i^T, b_i\right)$. Πρακτικά, κάθε κορυφή του ζωνοτόπου αντιστοιχεί στην επιλογή ενός άκρου από κάθε γεννήτορα. Προφανώς, από κάθε γεννήτορα προκύπτουν δύο επιλογές, οπότε ο αριθμός των εν δυνάμει κορυφών είναι εκθετικός 2^k. Ωστόσο στην πραγματικότητα μπορεί μία επιλογή άκρων από τους γεννήτορες να καταλήγει σε κάποιο σημείο στο εσωτερικό του ζωνοτόπου. Με την ακόλουθη πρόταση περιγράφουμε φορμαλιστικά τα ανωτέρω.

Πρόταση 3.5. *Για κάθε κορυφή* $\mathbf{v}$ *του* P_j *υπάρχει ένα υποσύνολο δεικτών* I_+ *του* $\{1, 2, ..., n\}$ *με* $c_{ji} > 0, \forall i \in I_+$ *ώστε* $\mathbf{v} = \sum_{i \in I_+} c_{ji}\left(\mathbf{a}_i^T, b_i\right)$. *Ομοίως, μία κορυφή* $\mathbf{u}$ *του αρνητικού πολυτόπου* Q_j *μπορεί να γραφεί ως* $\mathbf{u} = \sum_{i \in I_-} c_{ji}\left(\mathbf{a}_i^T, b_i\right)$ *όπου* I_- *αντιστοιχεί σε* $c_{ji} < 0, \forall i \in I_-$.

Απόδειξη. Από τον ορισμό του αθροίσματος Minkowski, κάθε σημείο $\mathbf{v} \in P_j$ μπορεί να γραφεί ως $\sum_{c_{ji}>0} \mathbf{v}_i$, όπου κάθε $\mathbf{v}_i$ είναι ένα σημείο στο ευθύγραμμο τμήμα

$$\text{ENewt}\left(\max(c_{ji}\mathbf{a}_i^T\mathbf{x} + c_{ji}b_i, \mathbf{0})\right)$$

Μία κορυφή του P_j μπορεί να προκύψει μόνο εάν το $\mathbf{v}_i$ είναι ακραίο σημείο του τμήματος $\text{ENewt}\left(\max(c_{ji}\mathbf{a}_i^T\mathbf{x} + c_{ji}b_i, \mathbf{0})\right)$ για κάθε i που ισοδυναμεί με είτε $\mathbf{v}_i = \mathbf{0}$ ή $\mathbf{v}_i = c_{ji}\left(\mathbf{a}_i^T, b_i\right)$. Αυτό μας δείχνει ότι κάθε κορυφή του P_j αντιστοιχεί, πράγματι, σε ένα υποσύνολο $I_+ \subseteq [n]$ δεικτών i με $c_{ji} > 0$, για τους οποίους επιλέγουμε $\mathbf{v}_i = c_{ji}\left(\mathbf{a}_i^T, b_i\right)$ ενώ για τους υπόλοιπους $\mathbf{v}_i = \mathbf{0}$. Κατά αυτόν τον τρόπο προκύπτει,

$$\mathbf{v} = \sum_{i \in I_+} c_{ji}\left(\mathbf{a}_i^T, b_i\right)$$

που είναι και το επιθυμητό. Όμοια λαμβάνουμε την αντίστοιχη σχέση για την περίπτωση του αρνητικού ζωνοτόπου Q_j. □

Πόρισμα 3.1. *Το γεωμετρικό αποτέλεσμα που αφορά την δομή ζωνοτόπων που αποκτά η συνάρτηση εξόδου του δικτύου μπορεί να γενικευθεί και σε max-pooling επιπέδων. Για παράδειγμα σε ένα max-pooling επίπεδο με pooling μεγέθους* 2×2*, η συνάρτηση εξόδου αντιστοιχεί στο maximum* 4 *όρων. Γεωμετρικά, οι* 4 *όροι αναπαριστούν μία πυραμίδα και συνολικά οι κόμβοι εξόδου κατασκευάζουν ένα πολύτοπο το οποίο γράφεται ως Minkowski άθροισμα πυραμίδων. Μία τέτοια γεωμετρική δομή μπορεί να θεωρηθεί ως ένα γενικευμένο ζωνότοπο, που λόγου χάριν μπορεί να αποκαλεστεί πυραμιδο-ζωνότοπο. Πράγματι μία φυσική γενίκευση ενός ζωνοτόπου είναι ένα πολύτοπο το οποίο παράγεται από ομοειδή γεωμετρικά αντικείμενα, π.χ. να παράγεται μόνο από κύβους, ή μόνο τρίγωνα κ.λ.π.*

Η μελέτη των ζωνοτόπων θα συνεχιστεί στο επόμενο κεφάλαιο όπου θα παρουσιάσουμε προσεγγιστικές μεθόδους ελαχιστοποίησης της αναπαράστασης των ζωνοτόπων, με απώτερο στόχο την συμπίεση νευρωνικών δικτύων.

3.4 Γραμμικές Περιοχές Νευρωνικών Δικτύων

Στην ενότητα αυτή θα αξιοποιήσουμε τις τροπικές ρητές και πολυωνυμικές απεικονίσεις σε συνδυασμό με τα αποτελέσματα του πλήθους εδρών πολυτόπων από το προηγούμενο κεφάλαιο, για να υπολογίσουμε το πλήθος των γραμμικών περιοχών ορισμένων γνωστών νευρωνικών δικτύων.

Ορισμός. Ένας υποσύνολο $\mathcal{D} \subset \mathbb{R}^d$ ονομάζεται γραμμική περιοχή ενός νευρωνικού δικτύου με συνάρτηση εξόδου $v : \mathbb{R}^d \to \mathbb{R}^m$ εάν υπάρχουν $A, \mathbf{b}$ ώστε

$$v(\mathbf{x}) = A\mathbf{x} + \mathbf{b},\ \forall \mathbf{x} \in \mathcal{D}$$

Γενικά, η μελέτη των γραμμικών περιοχών ενός νευρωνικού δικτύου μας δίνει ένα μέτρο της εκφραστικότητας ενός δικτύου και της ικανότητας του να μπορεί να προσαρμόζεται σε ζητούμενα δεδομένα, τα οποία συνήθως είναι μη-γραμμικά. Ο υπολο-γισμός των γραμμικών περιοχών ενός Feed-Forward νευρωνικού δικτύου έχει προηγηθεί στο [29] και έχει επαναϋπολογιστεί στο [45] με χρήση τροπικής γεωμετρίας και πολυτόπων. Επιπλέον, στο [43] έχει γίνει προσέγγιση των γραμμικών περιοχών ενός συνελικτικού νευρωνικού δικτύου αγνοώντας ωστόσο τα max-pooling επίπεδα. Εδώ θα επαναλάβουμε την μέτρηση των περιοχών σε feed-forward νευρωνικά, θα κάνουμε χρήση τροπικής γεωμετρίας για υπολογισμό γραμμικών περιοχών σε συνελικτικά δίκτυα, αλλά και θα επεκταθούμε σε πιο σύγχρονες αρχιτεκτονικές, τα Residual Nets (ResNets) που θα παρουσιάσουμε στην συνέχεια.

Αρχικά, πριν επεκταθούμε σε προτάσεις που αφορούν βαθιά νευρωνικά, θα εστιάσουμε σε επίπεδα νευρωνικών δικτύων. Με βάση το λήμμα 2.1 αποδεικνύουμε τα παρακάτω άνω φράγματα, όπως παρουσιάζονται στο [6, 25]. Αυτά θα αποτελέσουν την βάση για να υπολογίσουμε μετέπειτα τις γραμμικές περιοχές βαθιών νευρωνικών δικτύων.

Πρόταση 3.6. *Ένα ReLU επίπεδο με n εισόδους και m εξόδους έχει το πολύ*

$$\sum_{j=0}^{n} \binom{m}{j}$$

γραμμικές περιοχές.

Απόδειξη. Το ReLU επίπεδο αποτελείται από μία συστοιχία m τροπικών πολυωνύμων $f = (f_1, f_2, ..., f_m)$, όπου $f_i(\mathbf{x}) = \max\{\mathbf{w}_i^T\mathbf{x} + b_i, 0\}$. Επομένως, το πλήθος των γραμμικών περιοχών του επιπέδου είναι, σύμφωνα με το λήμμα 2.1, ίσο με το πλήθος των άνω κορυφών του

$$P = \bigoplus_{i=1}^{m} \mathrm{ENewt}\,(f_i)$$

Όμως, τα $\mathrm{ENewt}\,(f_i)$ είναι ευθύγραμμα τμήματα, οπότε το P είναι ζωνότοπο και εξ' ορισμού έχει το πολύ m μή-παράλληλες ακμές. Ο ισχυρισμός μας έπεται από την Πρόταση 2.6. □

Πρόταση 3.7. *Το πλήθος των γραμμικών περιοχών ενός συνελικτικού φίλτρου διάστασης $k \times k$ με padding p και ReLU ενεργοποιήσεις που εφαρμόζεται σε τετραγωνικές εικόνες διαστάσεων $d \times d$ είναι σε πλήθος το πολύ*

$$\sum_{j=0}^{d^2} \binom{(d-k+2p+1)^2}{j}$$

Απόδειξη. Αρκεί να παρατηρήσουμε ότι το συνελικτικό επίπεδο με τις ReLU ενεργοποιήσεις ισοδυναμεί με μία συστοιχία ReLU $f_i = \max\{\mathbf{w}_i^T \mathbf{x}, 0\}$, όπου το $\mathbf{x}$ είναι διαστάσεων $d \times d$ (όσο και η εικόνα) και το $\mathbf{w}_i$ έχει τις $k \times k$ τιμές του φίλτρου και στις υπόλοιπες θέσεις μηδενικά. Η διάσταση εξόδου θα είναι ίση με το πλήθος των εφαρμογών του φίλτρου στην εικόνα, δηλαδή $(d-k+2p+1)^2$. Επομένως, όπως και προηγουμένως, παίρνουμε το επιθυμητό αποτέλεσμα. □

Πρόταση 3.8. *Ένα Max-Out επίπεδο με n εισόδους και m εξόδους έχει το πολύ*

$$2\sum_{j=0}^{n} \binom{m\binom{k}{2}}{j}$$

γραμμικές περιοχές.

Απόδειξη. Το MaxOut επίπεδο ισοδυναμεί με την τροπική απεικόνιση $f = (f_1, ..., f_m)$ όπου

$$f_i(\mathbf{x}) = \max_{j \in [k]} \{\mathbf{w}_j^T \mathbf{x} + b_j\}$$

Το πλήθος των γραμμικών περιοχών της απεικόνισης είναι, λόγω του λήμματος 2.1, ίσο με το πλήθος των κορυφών του άνω φλοιού του

$$P = \bigoplus_{i=1}^{m} \mathrm{ENewt}\,(f_i)$$

Κάθε maxout unit σχηματίζει $\binom{k}{2}$ ευθύγραμμα τμήματα, οπότε συνολικά μπορούμε να έχουμε το πολύ $m\binom{k}{2}$ μη-παράλληλες ακμές στο πολύτοπο P. Επομένως, το πόρισμα 2.3 μας δίνει το ζητούμενο. □

Έχοντας υπολογίσει τα άνω φράγματα για επίπεδα νευρωνικών, μένει να βρούμε ένα τρόπο υπολογισμό των γραμμικών περιοχών στην περίπτωση όπου έχουμε βαθιά νευρωνικά αποτελούμενα από διαδοχικά επίπεδα. Η πράξη που θα μας επιτρέψει να πραγματοποιήσουμε τους επιθυμητούς υπολογισμούς είναι οι σύνθεση απεικονίσεων. Για αυτόν τον σκοπό εισάγουμε τους εξής ορισμούς και συμβολισμούς.

Ορισμός. Με $\mathcal{L}\,(f)$ θα συμβολίζουμε το πλήθος των γραμμικών περιοχών μίας τμηματικά γραμμικής συνάρτησης f. Στην δική μας περίπτωση η f θα είναι μία τροπική ρητή ή πολυωνυμική απεικόνιση.

Ορισμός. Για μία τροπική ρητή απεικόνιση $f : \mathbb{R}^n \to \mathbb{R}^m$ ορίζουμε

$$\mathcal{L}(f|_d) = \max\{\mathcal{L}(f|_\Omega) : \Omega \subseteq \mathbb{R}^n, \dim(\Omega) = d\}$$

το πλήθος των γραμμικών περιοχών της όταν αυτή περιορίζεται στις d διαστάσεις.

Είναι διαισθητικά εύλογο να υποθέσουμε ότι η σύνθεση δύο τροπικών ρητών απεικονίσεων έχει πλήθος γραμμικών περιοχών ίσο με το γινόμενο των αριθμών των γραμμικών περιοχών των δύο επιμέρους απεικονίσεων. Το ακόλουθο λήμμα δίνει μία φορμαλιστική εκδοχή αυτής της ιδέας. Επαγωγικά, το λήμμα αυτό μπορεί να χρησιμοποιηθεί για τον υπολογισμό των γραμμικών περιοχών μίας τροπικής ρητής απεικόνισης που προκύπτει από την σύνθεση πολλών τροπικών ρητών απεικονίσεων. Αυτή θα μας χρησιμεύσει για τα νευρωνικά δίκτυα αφού τα ίδια αποτελούν σύνθεση διαχοχικών τροπικών πολυωνυμικών απεικονίσεων.

Λήμμα 3.1. *(Κανόνας Αλυσίδας για γραμμικές περιοχές) Έστω $f : \mathbb{R}^n \to \mathbb{R}^m$, $g : \mathbb{R}^d \to \mathbb{R}^n$ δύο τροπικές ρητές απεικονίσεις. Τότε, ισχύει*

$$\mathcal{L}(f \circ g) \leq \mathcal{L}(f|_d)\,\mathcal{L}(g)$$

Απόδειξη. Έστω $G_1, G_2..., G_{\mathcal{L}(g)}$ οι γραμμικές περιοχές της g. Ας υποθέσουμε ότι $\mathbf{x} \in G_i \subseteq \mathbb{R}^d$, τότε αφού η g είναι γραμμική θα υπάρχει A ώστε $g(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$. Το $g(\mathbf{x})$ ανήκει στον γραμμικό χώρο των στηλών του πίνακα A ο οποίος έχει διαστάσεις $n \times d$. Συνεπώς, τα $g(x)$ ανήκουν σε υπόχωρο διάστασης το πολύ d. Από τον ορισμό του περιορισμού σε γραμμικό υπόχωρο:

$$\mathcal{L}(f \circ (g|_{G_i})) \leq \mathcal{L}(f|_d)$$

Σαν επόμενο βήμα, θα αποδείξουμε ότι

$$\mathcal{L}(f \circ g) \leq \sum_{i=1}^{\mathcal{L}(g)} \mathcal{L}(f \circ (g|_{G_i}))$$

Για να γίνει αυτό θα κατασκευάσουμε μία "επί" αντιστοίχιση των γραμμικών περιοχών των $\{f \circ (g|_{G_i})\}_i$ στις γραμμικές περιοχές της $f \circ g$. Πράγματι, αν $\mathcal{D} \subseteq G_i$ μία γραμμική περιοχή της $f \circ (g|_{G_i})$ για κάποιο i, τότε $f \circ g$ γραμική στο $\mathcal{D}$, οπότε και $\mathcal{D} \subseteq \mathcal{R}$ για κάποια γραμμική περιοχή $\mathcal{R}$ της $f \circ g$. Για κάθε τέτοια περιοχή $\mathcal{D}$ θεωρούμε την αντιστοίχιση $\mathcal{D} \mapsto \mathcal{R}$.

Η αντιστοίχιση αυτή είναι "επί" αφού αν θεωρήσουμε οποιαδήποτε γραμμική περιοχή $\mathcal{R}$ της $f \circ g$, τότε, για κάποιο i θα ισχύει $\mathcal{D} = \mathcal{R} \cap G_i \neq \emptyset$. Η $\mathcal{D}$ είναι γραμμική περιοχή της $f \circ g|_{G_i}$. Πράγματι, αν δεν ήταν γραμμική περιοχή, αυτό σημαίνει ότι θα μπορούσε να επεκταθεί σε $\mathcal{D} \subset \mathcal{D}' \subseteq G_i$, όπου η $\mathcal{D}'$ θα είναι γραμμική περιοχή. Αυτό όμως μας δίνει άτοπο αφού τότε η γραμμική περιοχή $\mathcal{R}$ της $f \circ g$ θα μπορούσε να επεκταθεί στην $\mathcal{R}' = \mathcal{R} \cup (\mathcal{D}' \setminus \mathcal{D})$ η οποία είναι πράγματι γραμμική δεδομένου ότι η g παίρνει την ίδια τιμή στα $\mathcal{D}', \mathcal{D}$. Άρα, το $\mathcal{D}$ είναι γραμμική περιοχή και $\mathcal{D} \mapsto \mathcal{R}$, δηλαδή η αντιστοιχία που δημιουργήσαμε είναι πράγματι "επί".

Συμπεραίνουμε ότι, για κάθε γραμμική περιοχή της $f \circ g|_{G_i}$ για κάποιο i παίρνουμε μία γραμμική περιοχή της $f \circ g$. Ως εκ τούτου

$$\mathcal{L}(f \circ (g|_{G_i})) \leq \sum_{i=1}^{\mathcal{L}(g)} \mathcal{L}(f \circ (g|_{G_i})) \leq \sum_{i=1}^{\mathcal{L}(g)} \mathcal{L}(f|_d) = \mathcal{L}(f|_d)\,\mathcal{L}(g)$$

□

Εφοδιασμένοι με τις παραπάνω προτάσεις, είμαστε πλέον σε θέση να τις εφαρμόσουμε για να πραγματοποιήσουμε υπολογισμούς σε διάφορες περιπτώσεις νευρωνικών δικτύων. Σημειώνουμε ότι οι Προτάσεις 3.11, 3.12, και 3.13 αποτελούν συμβολή της διπλωματικής αυτής.

3.4.1 Feed-Forward Νευρωνικό Δίκτυο

Αρχικά, θα εφαρμόσουμε τις προτάσεις που παρουσιάσαμε για τον υπολογισμό των γραμμικών περιοχών ενός κλασσικού feed-forward με ReLU ενεργοποιήσεις νευρωνικού δικτύου. Ένα τέτοιο νευρωνικό δίκτυο μπορεί να ειδωθεί ως σύνθεση τροπικών πολυωνυμικών απεικονίσεων που προκύπτουν από τα διαδοχικά ReLU επίπεδα. Επομένως, για να γίνει ο υπολογισμός του άνω φράγματος θα κάνουμε χρήση του λήμματος 3.1.

Για αυτόν τον σκοπό, θα χρειαστεί να υπολογίσουμε ένα άνω φράγμα στις γραμμικές περιοχές ενός ReLU επίπεδο σε έναν περιορισμένο υπόχωρο. Παρουσιάζουμε, λοιπόν, την παρακάτω πρόταση η οποία αποτελεί εναλλακτική μορφή της Πρότασης 3.6. Η διαφορά εδώ είναι ότι περιορίζουμε το ReLU επίπεδο να παίρνει είσοδο διανύσματα τα οποία προέρχονται από έναν $d-$διάστατο υπόχωρο. Ο λόγος που επιλέγουμε να εκτελέσουμε τον παρακάτω υπολογισμό θα διαφανεί καλύτερα μέσω του τελικού υπολογισμό των γραμμικών περιοχών του δικτύου.

Πρόταση 3.9. *Έστω $f : \mathbb{R}^n \to \mathbb{R}^m$ ένα ReLU επίπεδο με n εισόδους και m εξόδους. Τότε*

$$\mathcal{L}\left(f|_d\right) \leq \sum_{j=0}^{\min(d,n)} \binom{m}{j}$$

Απόδειξη. Για $d \geq n$ η πρόταση γίνεται ισοδύναμη με την Πρόταση 3.6. Διαφορετικά υποθέτουμε ότι $d < n$ και θεωρούμε γραμμικό υπόχωρο Ω διάστασης d ώστε $\mathcal{L}\left(f|_\Omega\right) = \mathcal{L}\left(f|_d\right)$. Θεωρούμε $\mathbf{w}_1, ..., \mathbf{w}_d$ τα διανύσματα βάσης του Ω, οπότε το input της f μπορεί να γραφεί στην μορφή:

$$\mathbf{x} = x_1\mathbf{w}_1 + ... + x_d\mathbf{w}_d = \mathbf{W}\tilde{\mathbf{x}}$$

όπου $\mathbf{W} \in \mathbb{R}^{n\times d}, \tilde{\mathbf{x}} \in \mathbb{R}^d$.

Αν γράψουμε $f = (f_1, ..., f_m)$ και $f_i(\mathbf{x}) = \max\{\mathbf{a}_i^T\mathbf{x} + b, 0\}$ προκύπτει ότι

$$f_i(\mathbf{x}) = \max\{\mathbf{a}_i^T\mathbf{x} + b, 0\} = \max\{\mathbf{a}_i^T\mathbf{W}\tilde{\mathbf{x}} + b, 0\} = \max\{\tilde{\mathbf{a}}_i^T\tilde{\mathbf{x}} + b, 0\} = \tilde{f}_i$$

Δηλαδή, το $f|_\Omega$ ισοδυναμεί με ένα ReLU επίπεδο $\tilde{f} = (\tilde{f}_1, ..., \tilde{f}_m)$ με d εισόδους και m εξόδους. Οπότε παίρνουμε πάλι το ζητούμενο αποτέλεσμα από την Πρόταση 3.6. □

Η πρόταση που μόλις αποδείξαμε θα αξιοποιηθεί διαδοχικά στα επίπεδα του βαθιού νευρωνικού. Όπως, είδαμε σε προηγούμενη ενότητα, ένα ReLU επίπεδο αντιστοιχεί σε μία τροπική πολυωνυμική απεικόνιση, και κατ' επέκταση το νευρωνικό μπορεί να γραφεί ως σύνθεση αυτών των απεικονίσεων. Το λήμμα 3.1, με απλά λόγια, μας δίνει το συμπέρασμα ότι το άνω φράγμα στο πλήθος των γραμμικών περιοχών είναι το γινόμενο των γραμμικών περιοχών κάθε επιπέδου του νευρωνικού. Παρακάτω, παραθέτουμε την πρόταση που αφορά τον υπολογισμό αυτό.

Πρόταση 3.10. *Έστω $F : \mathbb{R}^d \to \mathbb{R}^n$ ένα Deep Feed-Forward νευρωνικό με L επίπεδα, όπου στο $i-$οστό επίπεδο έχουμε n_i νευρώνες και στο τελευταίο $n_L = n$. Τότε:*

$$\mathcal{L}(F) \leq \prod_{l=1}^{L} \sum_{j=0}^{\min(d,n_{l-1})} \binom{n_l}{j}$$

με την σύμβαση πως $n_0 = d$.

Απόδειξη. Γράφουμε $F = f_L \circ f_{L-1} \circ ... \circ f_1$, όπου f_i τα ReLU επίπεδα. Τότε με χρήση του λήμματος 3.1 παίρνουμε

$$\mathcal{L}(F) = \mathcal{L}(f_L \circ ... \circ f_1) \leq \mathcal{L}(f_L|_d)\,\mathcal{L}(f_{L-1} \circ ... \circ f_1) \leq ... \leq \left(\prod_{l=2}^{L} \mathcal{L}(f_l|_d)\right) \mathcal{L}(f_1) \Leftrightarrow$$

$$\mathcal{L}(F) \leq \prod_{l=1}^{L} \mathcal{L}(f_l|_d) \leq \prod_{l=1}^{L} \sum_{j=0}^{\min(d,n_{l-1})} \binom{n_l}{j}$$

όπου στην τελευταία ανισότητα χρησιμοποιήσαμε την Πρόταση 3.9 καθώς και $\mathcal{L}(f_1) = \mathcal{L}(f_1|_d)$ αφού η f_1 έχει διάσταση εισόδου $n_0 = d$. □

3.4.2 Συνελικτικό Νευρωνικό Δίκτυο

Με παρόμοια τεχνική με το κλασσικό feed-forward νευρωνικό δίκτυο θα υπολογίσουμε ένα άνω φράγμα στο πλήθος των γραμμικών περιοχών ενός συνελικτικού δικτύου. Επομένως, αρχικά θα υπολογίσουμε το πλήθος των γραμμικών περιοχών ενός συνελικτικού επιπέδου περιορισμένο σε υπόχωρο και εν συνεχεία με πολλαπλασιασμό για όλα τα επίπεδα θα προκύψει το τελικό άνω φράγμα.

Στο συνελικτικό επίπεδο αξίζει να παρατηρήσουμε ότι έχουμε 2 στελέχη. Το πρώτο είναι το συνελικτικό φίλτρο το οποίο όπως έχουμε δείξει στην Πρόταση 3.7 ισοδυναμεί με ένα ReLU επίπεδο καταλλήλου μεγέθους. Το δεύτερο στέλεχος είναι το max-pooling επίπεδο το οποίο μειώνει τις διαστάσεις εξόδου παίρνοντας το μέγιστο από τα pixel μίας γειτονιάς. Όπως θα δούμε το max-pooling επίπεδο είναι ειδική περίπτωση max-out επιπέδου, και συνεπώς για αυτό μπορούμε να κάνουμε χρήση της Πρότασης 3.8.

Υπενθυμίζουμε ότι στην ανάλυση μας περιορίζουμε τα στελέχη του επιπέδου σε είδοδο που προέρχεται από $d-$διάστατο υπόχωρο. Αν απορρίψουμε αυτήν την παραδοχή, απλώς μεγαλώνει το άνω φράγμα.

Πρόταση 3.11. *Θεωρούμε ένα επίπεδο ενός δικτύου CNN το οποίο γράφεται ως* $f \circ g : \mathbb{R}^{n\times n} \to \mathbb{R}^{m\times m}$. *Η* $g : \mathbb{R}^{n\times n} \to \mathbb{R}^{q\times q}$ *είναι ένα συνελικτικό φίλτρο διάστασης* $k \times k$ *και zero-padding* p *που ακολουθείται από ReLU ενεργοποίηση. Επίσης, η* $f : \mathbb{R}^{q\times q} \to \mathbb{R}^{m\times m}$ *είναι ένα max-pooling επίπεδο. Τότε*

$$\mathcal{L}\left(g|_{d^2}\right) \leq \sum_{j=0}^{\min(d^2,n^2)} \binom{q^2}{j}$$

και

$$\mathcal{L}\left(f|_{d^2}\right) \leq 2 \sum_{j=0}^{\min(d^2,q^2)} \binom{m^2\binom{\frac{q^2}{m^2}}{2}}{j}$$

Απόδειξη. Αρχικά σημειώνουμε ότι, για να είναι έγκυρη η μελέτη μας, πρέπει να είναι $q = n - k + 2p + 1$ ίσο με τον αριθμό των εφαρμογών του συνελικτικού φίλτρου πάνω στην εικόνα εισόδου.

Το συνελικτικό φίλτρο g ισοδυναμεί με ReLU επίπεδο που έχει διάσταση εισόδου n^2 και εξόδου q^2, όπως προκύπτει με εξήγηση αντίστοιχη της Πρότασης 3.7. Τελικά, σύμφωνα με την Πρόταση 3.9 που αφορά τον περιορισμό στις d^2 διαστάσεις προκύπτει η ζητούμενη σχέση.

Αναφορικά με το max-pooling επίπεδο έχουμε ότι $f = (f_1, ..., f_{m^2})$ με

$$f_i = \max\left\{x_{(i-1)\frac{q^2}{m^2}+1}, ..., x_{i\frac{q^2}{m^2}}\right\} = \max_{(i-1)\frac{q^2}{m^2}+1\leq i\leq i\frac{q^2}{m^2}} \{\mathbf{v}_i^T\mathbf{x} + 0\}$$

όπου $\mathbf{v}_i$ είναι το μοναδιαίο διάνυσμα που έχει 1 στην $i-$οστή θέση και 0 οπουδήποτε αλλού. Παρατηρούμε ότι το max-pooling επίπεδο είναι ειδική περίπτωση ενός MaxOut επιπέδου. Θα υπολογίσουμε το άνω φράγμα των γραμμικών περιοχών, όπως εργαστήκαμε και για MaxOut στην Πρόταση 3.8.

Όπως, ειπώθηκε και στη Πρόταση 3.9 το $\mathcal{L}\left(f|_{d^2}\right)$ ισούται με τον αριθμό των γραμμικών περιοχών ενός max-pooling επιπέδου με $min(d^2, q^2)$ εισόδους και m^2 εξόδους. Το $\text{ENewt}\left(f_i\right)$ αποτελείται από $\frac{q^2}{m^2}$ σημεία, επομένως περιέχει το πολύ $\binom{\frac{q^2}{m^2}}{2}$ μή-παράλληλες ακμές. Επομένως, συνολικά το

$$\bigoplus_{i=1}^{m^2} \text{ENewt}\left(f_i\right)$$

έχει το πολύ $m^2\binom{\frac{q^2}{m^2}}{2}$ μη-παράλληλες ακμές. Το τελικό αποτέλεσμα δίνεται από το πόρισμα 2.3. □

Συνδυάζοντας το αποτέλεσμα τις προηγούμενης πρότασης για όλα τα επίπεδα προκύπτει το ακόλουθο αποτέλεσμα για τις γραμμικές περιοχές ενός βαθιού συνελικτικού νευρωνικού δικτύου.

Πρόταση 3.12. *Το πλήθος των γραμμικών περιοχών ενός Συνελικτικού Νευρωνικού δικτύου που έχει ως είσοδο εικόνες διάστασης $d \times d$ και έξοδο διάστασης $n \times n$ ικανοποιεί*

$$\mathcal{L}(F) \leq \prod_{l=1}^{L} \left(\sum_{j=0}^{\min(d^2, n_l^2)} \binom{q_l^2}{j} \right) \left(2 \sum_{j=0}^{\min(d^2, q_l^2)} \binom{n_{l+1}^2 \binom{\frac{q_l^2}{n_{l+1}^2}}{2}}{j} \right)$$

όπου n_l, q_l είναι οι διαστάσεις εισόδου του $l-$οστού συνελικτικού φίλτρου και φίλτρου max-pooling αντίστοιχα. Κατά σύμβαση θεωρούμε $n_1 = d^2, n_{L+1} = n^2$.

Απόδειξη. Από το λήμμα 3.1 και την προηγούμενη Πρόταση 3.11 έχουμε

$$\mathcal{L}(F) = \mathcal{L}(f_L \circ g_L \circ ... \circ f_1 \circ g_1) \leq \mathcal{L}(f_L|_d)\, \mathcal{L}(g_L|_d)\, \mathcal{L}(f_{L-1} \circ g_{L-1} \circ ... \circ f_1 \circ g_1) \leq$$

$$\leq \left(\prod_{l=2}^{L} \mathcal{L}(f_l|_d)\, \mathcal{L}(g_l|_d) \right) \mathcal{L}(f_1|_d)\, \mathcal{L}(g_1) \Leftrightarrow$$

$$\mathcal{L}(F) \leq \prod_{l=1}^{L} \mathcal{L}(f_l|_d)\, \mathcal{L}(g_l|_d) \leq \prod_{l=1}^{L} \left(\sum_{j=0}^{\min(d^2, n_l^2)} \binom{q_l^2}{j} \right) \left(2 \sum_{j=0}^{\min(d^2, q_l^2)} \binom{n_{l+1}^2 \binom{\frac{q_l^2}{n_{l+1}^2}}{2}}{j} \right)$$

□

3.4.3 Νευρωνικό Δίκτυο ResNet

Εκτελούμε την ίδια διαδικασία για την περίπτωση της αρχιτεκτονικής του ResNet. Η ακόλουθη πρόταση υποδεικνύει ότι για το άνω φράγμα των γραμμικών περιοχών ο υπολογισμός δεν διαφέρει από αυτόν ενός feed-forward νευρωνικού.

Πρόταση 3.13. *Θεωρούμε ένα βασικό residual επίπεδο $F : \mathbb{R}^n \to \mathbb{R}^m$ το οποίο γράφεται ως $\mathcal{F} = f(g(\mathbf{x}) + \mathbf{W}\mathbf{x})$. Τότε το πλήθος των γραμμικών περιοχών έχει άνω φράγμα ίδιο με αυτό που προκύπτει από δύο συνενωμένα ReLU επίπεδα.*

Απόδειξη. Παρατηρούμε ότι $\mathcal{F} = f \circ \tilde{g}$, όπου $\tilde{g}(\mathbf{x}) = g(\mathbf{x}) + \mathbf{W}\mathbf{x}$. Συνεπώς,

$$\mathcal{L}(\mathcal{F}) \leq \mathcal{L}(f|_d)\, \mathcal{L}(\tilde{g})$$

όμως οι $\tilde{g}, g$ είναι γραμμικές ακριβώς στις ίδιες περιοχές. Οπότε προκύπτει και το ζητούμενο, αφού τα f, g είναι ReLU επίπεδα. □

Παρατήρηση. *Ανάλογα με την αρχιτεκτονική του δικτύου, μπορούμε χρησιμοποιώντας τις προηγούμενες προτάσεις να καταλήξουμε σε άνω φράγμα για το πλήθος των γραμμικών περιοχών ενός ReLU activated νευρωνικού.*

Κεφάλαιο 4

Γεωμετρική Συμπίεση Νευρωνικών Δικτύων

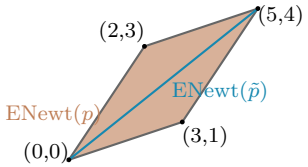
Στα προηγούμενα Κεφάλαια είδαμε ότι η τροπική γεωμετρία καταφέρνει να εξηγήσει αρκετές από τις ιδιότητες των νευρωνικών δικτύων. Είναι, επομένως, λογικό να αναμένουμε ότι μπορεί να αποτελέσει εργαλείο για το πρόβλημα της συμπίεσης τους. Πράγματι, σε αυτό το Κεφάλαιο θα αναδείξουμε την ιδιότητα αυτή της τροπικής γεωμετρίας κατασκευάζοντας αλγορίθμους συμπίεσης νευρωνικών δικτύων [3, 31, 38, 37, 2] βασισμένους στην γεωμετρία των επιπέδων (layers) του δικτύου. Οι αλγόριθμοι αυτοί επενεργούν γεωμετρικά ελαχιστοποιώντας το ζωνότοπο που αναπαριστά το δίκτυο.

Διατύπωση προβλήματος Ας ξεκινήσουμε την ανάλυση μας επιχειρώντας να ορίσουμε το πρόβλημα που επιθυμούμε να επιλύσουμε. Αρχικά, εστιάζουμε στην περίπτωση του δικτύου με ένα κρυφό επίπεδο, όπως στο σχήμα 3.1. Επιθυμούμε να συμπιέσουμε το δίκτυο ώστε από n νευρώνες στο κρυφό επίπεδο να έχει $K < n$. Το είδος συμπίεσης αυτό ονομάζεται δομημένο (structured) [3] διότι αφαιρούμε ολόκληρους νευρώνες από το δίκτυο. Έστω $\tilde{v}(\mathbf{x}) = (\tilde{v}_1(\mathbf{x}), ..., \tilde{v}_m(\mathbf{x}))$ η τροπική ρητή απεικόνιση του τελικού συμπιεσμένου δικτύου. Επιθυμούμε, το δίκτυο αυτό να αποτελεί καλή προσέγγιση του αρχικού, υπό την έννοια ότι η έξοδος του δικτύου είναι ίδια σε κάθε σημείο του πεδίου εισόδου. Γράφουμε το ζητούμενο ως

$$\tilde{v}_j(\mathbf{x}) \approx v_j(\mathbf{x}), \quad \forall\, \mathbf{x} \in \mathcal{B} \tag{4.1}$$

όπου $\mathcal{B}$ είναι η υπερσφαίρα ακτίνας r. Πρακτικά ζητάμε το προσεγγιστικό νευρωνικό να είναι "πιστό αντίγραφο" του αρχικού για κάθε κόμβο εξόδου. Αξίζει να σημειώσουμε ότι η απαίτηση αυτή είναι αρκετά ισχυρή και δεν είναι απαραίτητη ώστε 2 νευρωνικά να είναι ισοδύναμα ως προς τις προβλέψεις τους. Στην πραγματικότητα η αναγκαία συνθήκη ώστε τα δύο νευρωνικά να είναι ισοδύναμα ως προς το σύνολο δεδομένων και ένα classification task, είναι σε κάθε στιγμιότυπο του να έχουν την ίδια πρόβλεψη, δηλαδή ίδιο $\arg\max_{j\in[m]} v_j(\mathbf{x})$. Εμείς, για ευκολία δεν θα ακολουθήσουμε αυτή την εκδοχή, αλλά αυτήν που αφορά το "πιστό αντίγραφο", γεγονός που μας επιτρέπει μάλιστα να αγνοήσουμε τους σταθερούς όρους (biases) του επιπέδου εξόδου. Επιπλέον, η τεχνική αυτή προσφέρει περισσότερες δυνατότητες αφού δεν περιορίζεται σε classification tasks αλλά μπορεί να εφαρμοστεί ώστε να συμπιέσει και ένα νευρωνικό εκπαιδευμένο για regression task.

Θα αναλύσουμε περαιτέρω το πρόβλημα της προσέγγισης ως εξής. Η $j-$οστή έξοδος του προσεγγιστικού δικτύου γράφεται $\tilde{v}_j(\mathbf{x}) = \tilde{p}_j(\mathbf{x}) - \tilde{q}_j(\mathbf{x})$. Επειδή, δεν γνωρίζουμε ισχυρά γεωμετρικά εργαλεία αναφορικά με τις τροπικές ρητές συναρτήσεις, χαλαρώνουμε τις



Σχήμα 4.1: Ζωνότοπο αποτελούμενο από 2 γεννήτορες και προσέγγιση με έναν γεννήτορα.

απαιτήσεις μας ζητώντας την ικανή και όχι αναγκαία συνθήκη

$$\tilde{p}_j(\mathbf{x}) \approx p_j(\mathbf{x}),\ \tilde{q}_j(\mathbf{x}) \approx q_j(\mathbf{x}), \quad \forall\, \mathbf{x} \in \mathcal{B} \tag{4.2}$$

ώστε πράγματι να ικανοποιείται η (4.1).

Η σχέση 4.2 μπορεί να μεταφραστεί σε μία γεωμετρική συνθήκη με την χρήση του Θεωρήματος 2.3. Πράγματι, γνωρίζουμε ότι τα σφάλματα στις προσεγγίσεις των πολυωνύμων $p_j, \tilde{p}_j$ και $q_j, \tilde{q}_j$ φράσσονται από τις αποστάσεις των αντίστοιχων επεκτεταμένων Newton πολυτόπων, δηλαδή τις $\mathcal{H}\left(P_j, \tilde{P}_j\right)$ και $\mathcal{H}\left(Q_j, \tilde{Q}_j\right)$. Επομένως, αναζητούμε ένα νέο νευρωνικό δίκτυο του οποίου τα ζωνότοπα αποτελούν καλή γεωμετρική προσέγγιση των ζωνοτόπων του αρχικού δικτύου.

Πιο συγκεκριμένα, για την προσέγγιση των ζωνοτόπων θα κάνουμε χρήση των γεννητόρων. Η επιλογή αυτή γίνεται διότι όπως εξηγήσαμε σε προηγούμενο κεφάλαιο η πολυπλοκότητα ενός ζωνοτόπου ως προς τις κορυφές επι παραδείγματι αυξάνεται εκθετικά ως προς τους γεννήτορες. Επομένως, αν κοιτάζουμε συνολικά το ζωνότοπο ο αλγόριθμος μας θα είναι λιγότερο αποδοτικός στην ταχύτητα και την μνήμη.

Ο $j-$οστός κόμβος της εξόδου του προσεγγιστικού δικτύου γράφεται

$$\tilde{v}_j(\mathbf{x}) = \sum_{j=1}^{K} \tilde{c}_{ji} \cdot \max\left(\tilde{\mathbf{a}}_i^T \mathbf{x} + \tilde{b}_i, 0\right) = \tilde{p}_j(\mathbf{x}) - \tilde{q}_j(\mathbf{x})$$

Επομένως, απαιτούμε οι γεννήτορες $\tilde{c}_{ji}\left(\tilde{\mathbf{a}}_i^T,\ \tilde{b}_i\right),\ i \in [K]$ των προσεγγιστικών ζωνοτόπων, να επιλέγονται με τέτοιο τρόπο ώστε να ελαχιστοποιούνται οι ζητούμενες αποστάσεις ζωνοτόπων.

Συνοψίζοντας, καταφέραμε να διατυπώσουμε το πρόβλημα της συμπίεσης του νευρωνικού του σχήματος 3.1 με ένα κρυφό επίπεδο, ως πρόβλημα γεωμετρικής προσέγγισης ζωνοτόπων. Αξίζει να αναφέρουμε ότι εφόσον ο αλγόριθμος συμπίεσης δεν παρουσιάζει κάποιο περιορισμό στον αριθμό των κόμβων εξόδου, η τεχνική συμπίεσης του μπορεί να επεκταθεί σε βαθιά νευρωνικά, απλώς επαναλαμβάνοντάς τον διαδοχικά σε όλα τα επίπεδα του δικτύου.

Παράδειγμα 4.1. *Έστω ότι έχουμε ένα δίκτυο με κρυφό επίπεδο με 2 νευρώνες και 1 κόμβο στο επίπεδο εξόδου και επιθυμούμε να το ελαχιστοποιήσουμε. Η συνάρτηση εξόδου του νευρωνικού δίνεται ως*

$$v(x) = \max(2x + 3, 0) + \max(3x + 1, 0)$$

Αναζητούμε $\tilde{v}(\mathbf{x}) = \tilde{p}(\mathbf{x}) - \tilde{q}(\mathbf{x})$, *με* $\text{ENewt}(\tilde{p}) \approx \text{ENewt}(p)$ *και* $\text{ENewt}(\tilde{q}) \approx \text{ENewt}(q)$. *Το δίκτυο αυτό έχει μόνο θετικό ζωνότοπο αφού* $q(x) = 0$. *Μάλιστα το* $\text{ENewt}(p)$ *παρουσιάζεται στο σχήμα 4.1. Τότε, η "καλύτερη" γεωμετρική προσέγγιση του αρχικού νευρωνικού με 1 νευρώνα στο κρυφό επίπεδο είναι η* $\tilde{v}(x) = \max(5x+4, 0)$. *Η προσέγγιση αυτή γίνεται καλύτερη όσο το παραλληλόγραμμο γίνεται πιο στενό, δηλαδή οι γεννήτορες* $(2, 3)$ *και* $(3, 1)$ *είναι πιο κοντά. Η παρατήρηση αυτή μας οδηγεί να χωρίσουμε τους γεννήτορες σε συστάδες και να τους αναπαραστήσουμε με τα κέντρα των συστάδων (αλγόριθμος K-means).*

Μέθοδοι συμπίεσης Γενικά, η προσέγγιση ενός ζωνοτόπου από ένα με λιγότερους γεννήτορες ονομάζεται μείωση τάξης ζωνοτόπου (zonotope order reduction) [19]. Οι αλγόριθμοι που θα χρησιμοποιήσουμε για αυτόν τον σκοπό είναι οι Zonotope K-means, Neural Path K-means και Convolutional Neural Path K-means οι οποίοι κάνουν χρήση του αλγορίθμου K-means. Κάθε ένας από αυτους θα παράγει ένα νέο, μειωμένο ως προς το μέγεθος, σύνολο γεννητόρων που θα προσεγγίζουν τα αρχικά ζωνότοπα. Ιδανικά, επιθυμούμε οι γεννήτορες που προκύπτουν να ικανοποιούν την ζητούμενη προσέγγιση για όλες τις συναρτήσεις εξόδου v_j, $j \in [m]$ αλλά και για το θετικό P_j και το αρνητικό Q_j ζωνότοπο τους. Ωστόσο, αυτό δεν είναι απαραίτητα εύκολο σε κάθε περίπτωση, όπως για παράδειγμα συμβαίνει με τα νευρωνικά δίκτυα που έχουν παραπάνω από μία έξοδο. Σε τέτοιες περιπτώσεις θα κληθούμε να καταφύγουμε σε εναλλακτικές, ευριστικές μεθόδους.

Ο αλγόριθμος Zonotope K-means εφαρμόζει τον αλγόριθμο K-means για να προσδιορίσει το υποσύνολο γεννητόρων που θα αποτελεί καλή προσέγγιση του αρχικού. Ο K-means σε αυτήν την μέθοδο εφαρμόζεται χωριστά στους γεννήτορες του θετικού και του αρνητικού ζωνοτόπου και γι' αυτό έχει τον περιορισμό εφαρμογής μόνο σε δίκτυα μίας εξόδου.

Ο Neural Path K-means γενικεύει τον προηγούμενο αλγόριθμο και εφαρμόζεται σε δίκτυα πολλών εξόδων. Αυτή η μέθοδος εφαρμόζει τον K-means στα διανύσματα που αφορούν τον κάθε κόμβο του κρυφού επιπέδου. Μάλιστα, ονομάζεται Neural Paths K-means διότι τα διανύσματα στα οποία εφαρμόζεται ο K-means σχετίζονται με τα μονοπάτια του δικτύου που περνούν από τους κόμβους του κρυφού επιπέδου. Με αυτόν τον τρόπο επιτυγχάνεται η ταυτόχρονη προσέγγιση των ζωνοτόπων κάθε εξόδου.

Τέλος, ο Convolutional Neural Path K-means εφαρμόζει την ιδέα του Neural Path K-means σε συνελικτικά επίπεδα. Μάλιστα, αποτελεί τον πρώτο αλγόριθμο στην βιβλιογραφία συμπίεσης νευρωνικών δικτύων με τροπική γεωμετρία.

Οι αλγόριθμοι που παρουσιάζουμε σε αυτήν την ενότητα αφορούν γεωμετρική προσέγγιση ζωνοτόπων. Σε επόμενο Κεφάλαιο θα παρουσιάσουμε μεθόδους συμπίεσης οι οποίες θα αφορούν αριθμητική επεξεργασία των πινάκων των γραμμικών επιπέδων.

4.1 Προσέγγιση Ζωνοτόπου

Η πρώτη μέθοδος συμπίεσης που θα παρουσιάσουμε είναι γεωμετρική και κάνει χρήση του αλγορίθμου K-means, ο οποίος είναι εγγενώς ένας αλγόριθμος συμπίεσης δεδομένων που αναπαρίστανται διανυσματικά. Ο αλγόριθμος αυτός θα εφαρμόζεται σε νευρωνικά δίκτυα με 1 κρυφό επίπεδο και έναν μόνο κόμβο στο επίπεδο εξόδου, όπως αυτός στο σχήμα 3.1 αλλά με $m = 1$. Ένα παράδειγμα τέτοιου νευρωνικού φαίνεται στο σχήμα 4.2a. Επιθυμούμε να συμπιέσουμε το νευρωνικό μας ώστε στο κρυφό επίπεδο να έχει K νευρώνες από τους αρχικούς n. Για την περίπτωση της μίας εξόδου θα χρησιμοποιούμε τον συμβολισμό c_i, $i = 1, ..., n$ για τα βάρη του δεύτερου γραμμικού μετασχηματισμού που οδηγεί στην έξοδο. Η μέθοδος συμπίεσης παρουσιάζεται με τον αλγόριθμο 2 και ένα παράδειγμα εκτέλεσής του περιγράφεται με τα σχήματα της εικόνας 4.2.

Algorithm 2 Αλγόριθμος συμπίεσης Zonotope K-means

1. Αρχικά διαχωρίζουμε το σύνολο των γεννητόρων σε θετικούς $\left\{c_i \left(\mathbf{a}_i^T, b_i\right) : c_i > 0\right\}$ και αρνητικούς $\left\{c_i \left(\mathbf{a}_i^T, b_i\right) : c_i < 0\right\}$.

2. Έπειτα εφαρμόζουμε τον αλγόριθμο K-means για $\frac{K}{2}$ κέντρα, ξεχωριστά για τα δύο σύνολα γεννητόρων λαμβάνοντας τα αντιπροσωπευτικά διανύσματα $\left\{\tilde{c}_i \left(\tilde{\mathbf{a}}_i^T, \tilde{b}_i\right) : \tilde{c}_i > 0\right\}$, $\left\{\tilde{c}_i \left(\tilde{\mathbf{a}}_i^T, \tilde{b}_i\right) : \tilde{c}_i < 0\right\}$ ως έξοδο.

3. Τέλος, κατασκευάζουμε τα τελικά βάρη του συμπιεσμένου δικτύου. Για το πρώτο γραμμικό επίπεδο τα βάρη και το bias που αντιστοιχούν στον $i-$οστό νευρώνα γίνονται το διάνυσμα $\tilde{c}_i \left(\tilde{\mathbf{a}}_i^T, \tilde{b}_i\right)$.

4. Τα βάρη του δεύτερου γραμμικού επιπέδου τίθενται ως 1 για κάθε κόμβο του κρυφού επιπέδου όπου ο γεννήτορας $\tilde{c}_i \left(\tilde{\mathbf{a}}_i^T, \tilde{b}_i\right)$ προκύπτει ως κέντρο που αναπαριστά θετικούς γεννήτορες [1] και -1, διαφορετικά.

Πρόταση 4.1. *Ο αλγόριθμος Zonotope K-means παράγει ένα συμπιεσμένο νευρωνικό δίκτυο με συνάρτηση εξόδου $\tilde{v}$ που ικανοποιεί*

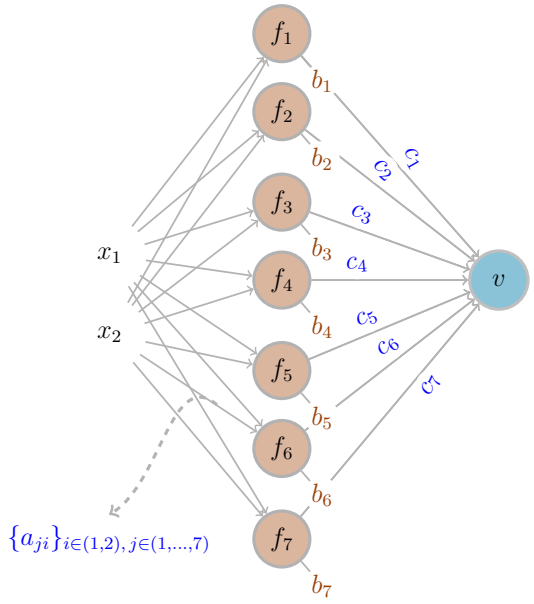
$$\frac{1}{\rho} \cdot \max_{\mathbf{x} \in \mathcal{B}} |v(\mathbf{x}) - \tilde{v}(\mathbf{x})| \leq K \cdot \delta_{max} + \left(1 - \frac{1}{N_{max}}\right) \sum_{i=1}^{n} |c_i| \left\| \left(\mathbf{a}_i^T, b_i\right) \right\|$$

όπου K είναι ο αριθμός των συνολικών κέντρων των δύο K-means εκτελέσεων, δ_{max} είναι η μεγαλύτερη απόσταση ενός γεννήτορα από το πλησιέστερό του κέντρο και N_{max} είναι ο μέγιστος πληθάριθμος μιας συστάδας των K-means.

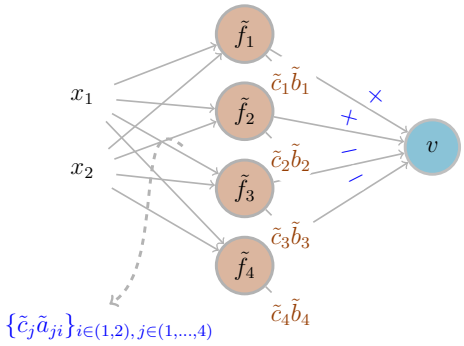
Απόδειξη. Υπενθυμίζουμε ότι η συνάρτηση εξόδου του αρχικού και τελικού νευρωνικού μπορεί να γραφεί ως τροπική ρητή συνάρτηση.

$$v(\mathbf{x}) = p(\mathbf{x}) - q(\mathbf{x}) \ , \tilde{v}(\mathbf{x}) = \tilde{p}(\mathbf{x}) - \tilde{q}(\mathbf{x})$$

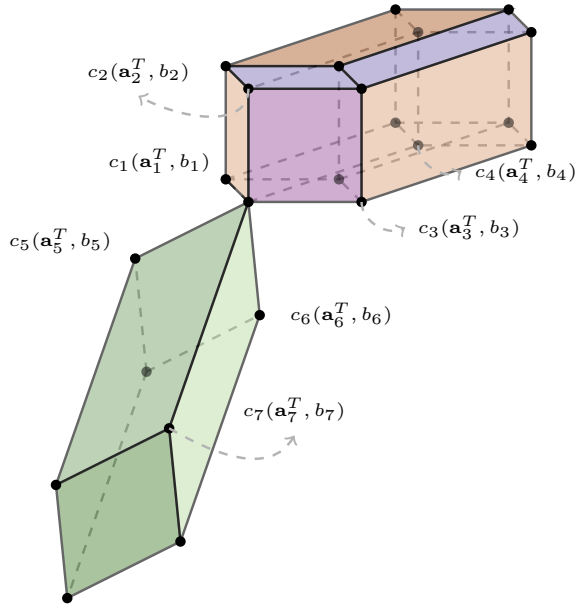
[1] Εδώ θα ήταν λάθος να πούμε θετικό κέντρο αυτό που έχει $\tilde{c}_i > 0$, διότι στην πραγματικότητα ο αλγόριθμος μας επιστρέφει το γινόμενο $\left(\tilde{c}_i \tilde{\mathbf{a}}_i^T, \tilde{c}_i \tilde{b}_i\right)$ και δεν γνωρίζουμε χωριστά το $\tilde{c}_i$.



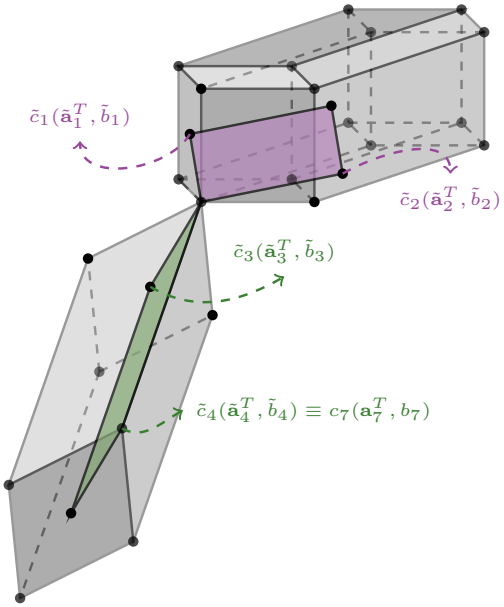
(a) Αρχικό Δίκτυο.



(b) Ελαχιστοποιημένο Δίκτυο.



(c) Αρχικά ζωνότοπα



(d) Ζωνότοπα προσέγγισης αλγορίθμου.

Σχήμα 4.2: Αναπαράσταση της εκτέλεσης του Zonotope K-means. Το αρχικό θετικό ζωνότοπο P παράγεται από τους γεννήτορες $c_i\left(\mathbf{a}_i^T,\, b_i\right)$ με $i = 1, ..., 4$ και το αρνητικό Q από τους εναπομείναντες γεννήτορες για $i = 5, 6, 7$. Το προσεγγιστικό θετικό ζωνότοπο $\tilde{P}$ του P χρωματίζεται με μωβ και παράγεται από τα $\tilde{c}_i\left(\tilde{\mathbf{a}}_i^T,\, \tilde{b}_i\right)$, $i = 1, 2$ όπου ο πρώτος γεννήτορας είναι το κέντρο του K-means που αντιπροσωπεύει τους γεννήτορες $1, 2$ του P ενώ ο δεύτερος αναπαριστά το κέντρο των γεννητόρων $3, 4$. Παρομοίως, το προσεγγιστικό ζωνότοπο $\tilde{Q}$ του Q χρωματίζεται με πράσινο και ορίζεται από τα $\tilde{c}_i\left(\tilde{\mathbf{a}}_i^T,\, \tilde{b}_i\right)$, $i = 3, 4$ που αποτελούν τα αντιπροσωπευτικά κέντρα των γεννητόρων $\{5, 6\}$ και 7 αντίστοιχα.

Από την τριγωνική ανισότητα λαμβάνουμε

$$|v(\mathbf{x}) - \tilde{v}(\mathbf{x})| = |p(\mathbf{x}) - q(\mathbf{x}) - (\tilde{p}(\mathbf{x}) - \tilde{q}(\mathbf{x}))| < |p(\mathbf{x}) - \tilde{p}(\mathbf{x})| + |q(\mathbf{x}) - \tilde{q}(\mathbf{x})|$$

Το Θεώρημα 2.3 φράσσει τις διαφορές $|p(\mathbf{x}) - \tilde{p}(\mathbf{x})|$ και $|q(\mathbf{x}) - \tilde{q}(\mathbf{x})|$ μέσω των αποστάσεων Hausdorff των πολυτόπων $\mathcal{H}\left(P, \tilde{P}\right)$ και $\mathcal{H}\left(Q, \tilde{Q}\right)$ αντίστοιχα. Επομένως, αρκεί να προσδιορίσουμε ένα άνω φράγμα για αυτές τις αποστάσεις, ώστε να φράξουμε τις διαφορές των πολωνύμων. Σύμφωνα με την Πρόταση 3.5 μπορούμε να Θεωρούμε οποιαδήποτε κορυφή του θετικού πολυτόπου P στην μορφή $\mathbf{u} = \sum_{i \in I_+} c_i \left(\mathbf{a}_i^T,\, b_i\right)$.

Για την κορυφή $\mathbf{u} \in P$ μπορούμε να διαλέξουμε μία κορυφή $\mathbf{v} \in \tilde{P}$ η οποία να είναι όσο το δυνατόν κοντά στην $\mathbf{u}$, με στόχο να περιορίσουμε την απόσταση $\text{dist}\left(\mathbf{u}, \tilde{P}\right)$. Μάλιστα, η κορυφή $\mathbf{v}$ θα επιλεγεί ως εξής. Για κάθε γεννήτορα με δείκτη $i \in I_+$ διαλέγουμε k ώστε το κέντρο $\tilde{c}_k \left(\tilde{\mathbf{a}}_k^T,\, \tilde{b}_k\right)$ να είναι το κέντρο της συστάδας στην οποία ανήκει ο εν λόγω γεννήτορας $c_i \left(\mathbf{a}_i^T,\, b_i\right)$. Θα συμβολίζουμε το σύνολο των δεικτών τέτοιων κέντρων με C_+, όπου κάθε δείκτης k στο C_+ εμφανίζεται ακριβώς μία φορά. Τότε, κατασκευάζουμε την κορυφή $\mathbf{v}$ μέσω των επιλεγμένων κέντρων $\mathbf{v} = \sum_{k \in C_+} \tilde{c}_k \left(\tilde{\mathbf{a}}_k^T \quad \tilde{b}_k\right) \in \tilde{P}$. Προκύπτει ότι:

$$\begin{aligned}
\text{dist}\left(\mathbf{u}, \tilde{P}\right) &\leq \left\| \sum_{i \in I_+} c_i \left(\mathbf{a}_i^T,\, b_i\right) - \sum_{k \in C_+} \tilde{c}_k \left(\tilde{\mathbf{a}}_k^T,\, \tilde{b}_k\right) \right\| \\
&\leq \sum_{k \in C_+} \left\| \sum_{i \in I_{k_+}} c_i \left(\mathbf{a}_i^T,\, b_i\right) - \tilde{c}_k \left(\tilde{\mathbf{a}}_k^T,\, \tilde{b}_k\right) \right\| \\
&\leq \sum_{k \in C_+} \sum_{i \in I_{k_+}} \left\| c_i \left(\mathbf{a}_i^T,\, b_i\right) - \frac{\tilde{c}_k \left(\tilde{\mathbf{a}}_k^T,\, \tilde{b}_k\right)}{|I_{k_+}|} \right\| \\
&= \sum_{k \in C_+} \sum_{i \in I_{k_+}} \left\| c_i \left(\mathbf{a}_i^T,\, b_i\right) - \frac{c_i \left(\mathbf{a}_i^T,\, b_i\right) + \varepsilon_i}{|I_{k_+}|} \right\| \\
&\leq \sum_{k \in C_+} \sum_{i \in I_{k_+}} \left[\left(1 - \frac{1}{|I_{k_+}|}\right) |c_i| \left\| \left(\mathbf{a}_i^T,\, b_i\right) \right\| + \frac{\|\varepsilon_i\|}{|I_{k_+}|} \right] \\
&\leq |C_+| \cdot \delta_{\max} + \left(1 - \frac{1}{N_{\max}}\right) \sum_{i \in I_+} |c_i| \left\| \left(\mathbf{a}_i^T,\, b_i\right) \right\|
\end{aligned}$$

όπου με I_{k+} συμβολίζουμε το σύνολο δεικτών $i \in I_+$ που ανήκουν στην συστάδα με κέντρο $k \in C_+$ και $\varepsilon_i = \tilde{c}_k \left(\tilde{\mathbf{a}}_k^T,\, \tilde{b}_k\right) - c_i \left(\mathbf{a}_i^T,\, b_i\right)$ είναι το διάνυσμα διαφοράς του $i-$οστού γεννήτορα με το αντίστοιχο αντιπροσωπευτικό του κέντρο στον K-means.

Το άνω φράγμα μεγιστοποιείται όταν το σύνολο I_+ περιέχει όλους τους δείκτες των γεννητόρων που αφορούν το θετικό ζωνότοπο, δηλαδή όλα τα i ώστε $c_i > 0$. Πράγματι, όλοι οι όροι στο παραπάνω άνω φράγμα είναι θετικοί και όσο προσθέτουμε γεννήτορες η τιμή του αυξάνεται. Με αυτόν τον τρόπο βρίσκουμε ένα άνω φράγμα για την απόσταση $\max_{\mathbf{u} \in P} d(\mathbf{u}, \tilde{P})$. Για να φράξουμε την Hausdorff απόσταση μένει ακόμα να υπολογίσουμε άνω φράγμα για την $\max_{\mathbf{v} \in \mathcal{V}_{\tilde{P}}} d(P, \mathbf{v})$. Για αυτόν τον σκοπό θεωρούμε $\mathbf{v} = \sum_{k \in C_+} \tilde{c}_k \left(\tilde{\mathbf{a}}_k^T \quad \tilde{b}_k\right)$ και επιλέγουμε την κορυφή $\sum_{i \in I_+} c_i \left(\mathbf{a}_i^T,\, b_i\right) \in P$ του αρχικού πολυτόπου, με I_+ να είναι το σύνολο των δεικτών των θετικών γεννητόρων που αντιστοιχούν σε

όλους τους γεννήτορες που προκύπτουν από τις συστάδες με κέντρα στο C_+. Παρατηρούμε ότι η απόσταση που προκύπτει

$$\left\| \sum_{i \in I_+} c_i \left(\mathbf{a}_i^T,\, b_i \right) - \sum_{k \in C_+} \tilde{c}_k \left(\tilde{\mathbf{a}}_i^T,\, \tilde{b}_i \right) \right\|$$

έχει ήδη υπολογιστεί ως άνω φράγμα για την τιμή της απόστασης $\max_{\mathbf{u} \in \mathcal{V}_P} d(\mathbf{u}, \tilde{P})$, οπότε και οι δύο αποστάσεις επιδέχονται το ίδιο άνω φράγμα. Συνεπώς,

$$\mathcal{H}\left(P, \tilde{P}\right) \leq K_+ \cdot \delta_{\max} + \left(1 - \frac{1}{N_{\max}}\right) \sum_{i \in I_+} |c_i| \left\| \left(\mathbf{a}_i^T,\, b_i \right) \right\|$$

όπου K_+ είναι ο αριθμός των κέντρων του K-means που αντιστοιχούν στο $\tilde{P}$ και I_+ οι δείκτες όλων των θετικών γεννητόρων του αρχικού πολυτόπου P. Όμοια,

$$\mathcal{H}\left(Q, \tilde{Q}\right) \leq K_- \cdot \delta_{\max} + \left(1 - \frac{1}{N_{\max}}\right) \sum_{i \in I_-} |c_i| \left\| \left(\mathbf{a}_i^T,\, b_i \right) \right\|$$

όπου τα K_-, I_- ορίζονται αντίστοιχα για το αρνητικό ζωνότοπο. Συνδυάζοντας τις προηγούμενες σχέσεις προκύπτει το ζητούμενο φράγμα.

$$\frac{1}{\rho} \cdot |v(\mathbf{x}) - \tilde{v}(\mathbf{x})| \leq \mathcal{H}\left(P, \tilde{P}\right) + \mathcal{H}\left(Q, \tilde{Q}\right) \leq K \cdot \delta_{\max} + \left(1 - \frac{1}{N_{\max}}\right) \sum_{i=1}^{n} |c_i| \left\| \left(\mathbf{a}_i^T,\, b_i \right) \right\|$$

□

Με την παραπάνω πρόταση προκύπτει ένα άνω φράγμα στο σφάλμα της προσέγγισης που έχει το δίκτυο που προκύπτει από τον Zonotope K-means με το αρχικό. Συμπεραίνουμε ότι ο αλγόριθμος παράγει ζωνότοπα τα οποία αποτελούν καλή προσέγγιση των αρχικών. Το σφάλμα της προσέγγισης αυξάνεται όσο χρησιμοποιούμε λιγότερα κέντρα στον K-means, δηλαδή όσο αυξάνουμε το ποσοστό συμπίεσης. Πράγματι, με $K \approx n$ κέντρα το σφάλμα είναι σχεδόν μηδενικό, διότι $\delta_{\max} \to 0, N_{\max} \to 1$, ενώ για $K \approx 0$ το σφάλμα παίρνει μία σταθερή τιμή που εξαρτάται στην απόλυτη τιμή και νόρμα των βαρών του δικτύου.

4.2 Πολλαπλή Προσέγγιση Ζωνοτόπων

Η ακριβής προσέγγιση ζωνοτόπων που πραγματοποιεί ο αλγόριθμος Zonotope K-means έχει το μειονέκτημα ότι δεν μπορεί να εφαρμοστεί σε νευρωνικά δίκτυα με πολλές εξόδους. Αυτό συμβαίνει για τον εξής λόγο. Έστω, λοιπόν, ότι θέλουμε να κάνουμε ταυτόχρονη προσέγγιση των θετικών και αρνητικών ζωνοτόπων για κάθε μία από τις m εξόδους ενός νευρωνικού δικτύου. Αυτό θα απαιτούσε την εκτέλεση $2m$ διαφορετικών K-means αλγορίθμων οι οποίοι δεν είναι απαραίτητα συμβατοί μεταξύ τους. Δηλαδή, δεν γίνεται να προκύψουν τα βάρη του τελικού δικτύου με κάποιον άμεσο τρόπο. Πράγματι, μπορούμε να υποθέσουμε ότι $c_{j_1 i} > 0$ και $c_{j_2 i} < 0$ για δύο κόμβους εξόδου v_{j_1}, v_{j_2}. Τότε αυτό σημαίνει ότι ο $i-$οστός γεννήτορας $c_{j_1 i}\,(\mathbf{a}_i,\, b_i)$ συμβάλλει στο θετικό ζωνότοπο του v_{j_1}, ενώ ο $i-$οστός $c_{j_2 i}\,(\mathbf{a}_i,\, b_i)$ συμβάλλει στο v_{j_2}. Επομένως, με αυτόν τον τρόπο το διάνυσμα $(\mathbf{a}_i,\, b_i)$ συμβάλλει σε διαφορετικά ζωνότοπα, και άρα δεν υπάρχει ξεκάθαρος τρόπος ώστε να επιλέξουμε σε ποιο ζωνότοπο θα συμβάλλει ο K-means αντιπρόσωπός του.

Ένας τρόπος επίλυσης αυτού του προβλήματος θα ήταν να έχουμε ένα αντίγραφο του πρώτου γραμμικού επιπέδου για κάθε κόμβο εξόδου και να εφαρμόζαμε ξεχωριστά τον K-means σε αυτά. Βέβαια, αυτό θα είχε ως αποτέλεσμα να προκύπτουν mn κόμβοι στο κρυφό επίπεδο που θα έκανε την συμπίεση πιο απαιτητική εως και αδύνατη. Γι' αυτόν τον λόγο επιλέγουμε να εφαρμόσουμε μία διαφορετική τεχνική για τα νευρωνικά πολλών εξόδων η οποία κάνει ταυτόχρονη προσέγγιση των ζωνοτόπων με ευριστική τεχνική.

Για να αντιμετωπίσουμε αυτό το πρόβλημα θα εφαρμόσουμε τον αλγόριθμο K-means ώστε να έχουμε με ευριστικό τρόπο ταυτόχρονη προσέγγιση των ζωνοτόπων. Η μέθοδος αυτή καλείται *Neural Path K-means* και εφαρμόζει τον K-means απευθείας στα διανύσματα βαρών $\left(\mathbf{a}_i^T,\, b_i,\, c_{1i},\, ...,\, c_{mi}\right)$ τα οποία κατασκευάζονται από όλα τα βάρη που αφορούν τον $i-$οστό νευρώνα του κρυφού επιπέδου. Αυτό είναι λογικό να το σκεφτεί κανείς, αφού θέλουμε να μειώσουμε σε πλήθος τους νευρώνες του κρυφού επιπέδου και κάθε ένας από αυτούς αναπαρίσταται μοναδικά από τα βάρη εισόδου και εξόδου του κόμβου. Η ακριβής διαδικασία παρουσιάζεται στον αλγόριθμο 3.

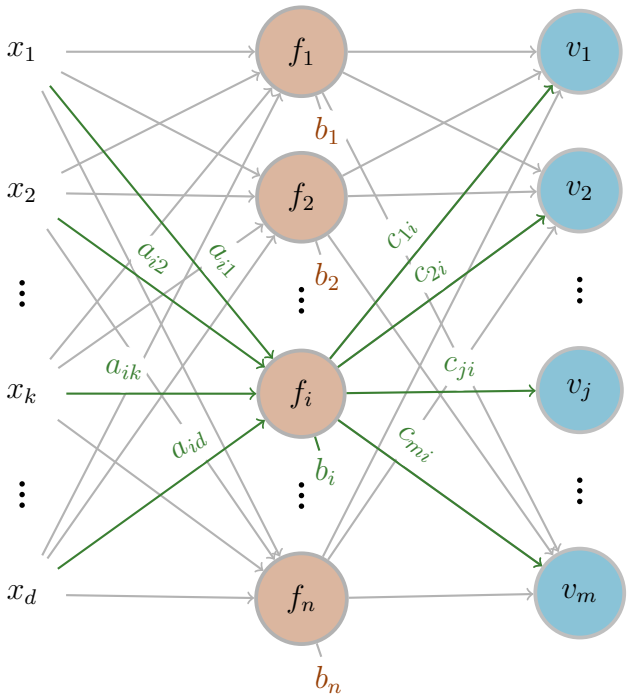
Παρακάτω παρουσιάζουμε το σχήμα 4.3 στο οποίο δίνουμε έμφαση στο διάνυσμα που αφορά τον $i-$οστό κόμβο του κρυφού επιπέδου. Το διάνυσμα αποτελείται από όλα τα βάρη που αφορούν τον $i-$οστό κόμβο. Το όνομα του αλγορίθμου προέρχεται από το γεγονός ότι τα διανύσματα που εφαρμόζονται στον K-means περιέχουν όλα τα μονοπάτια που ξεκινούν από την είσοδο, τερματίζουν στην έξοδο και ορίζονται από τις ακμές του δικτύου.

Algorithm 3 Neural Path K-means Compression

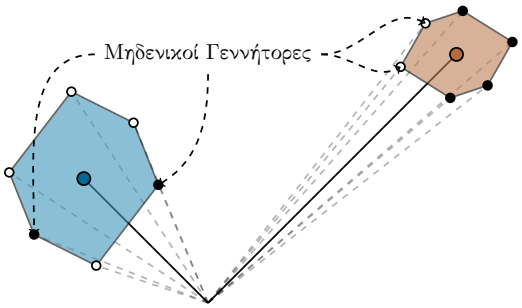
1. Εφαρμόζουμε τον αλγόριθμο K-means με K κέντρα για τα διανύσματα $\left(\mathbf{a}_i^T,\, b_i,\, C_{:,i}^T\right)$, $i = 1, ..., n$, και λαμβάνουμε ως αποτέλεσμα τα κέντρα $\left(\tilde{\mathbf{a}}_i^T,\, \tilde{b}_i,\, \tilde{C}_{:,i}^T\right)$, $i = 1, ..., K$.

2. Κατασκευάζουμε τα τελικά βάρη του δικτύου ως εξής. Για το πρώτο γραμμικό επίπεδο η $i-$οστή γραμμή τίθενται ίση με $\left(\tilde{\mathbf{a}}_i^T,\, \tilde{b}_i\right)$, ενώ για το δεύτερο γραμμικό επίπεδο, θέτουμε την $i-$οστή στήλη ίση με $\tilde{C}_{:,i}$.

Ο αλγόριθμος Neural Path K-means δεν εφαρμόζει απευθείας συμπίεση στα ζωνότοπα που αφορούν το δίκτυο αλλά αποτελεί μία ευριστική μέθοδο για αυτό το ζήτημα. Συγκεκριμένα, αν εστιάσουμε στους γεννήτορες των ζωνοτόπων της $j-$οστής εξόδου, ο Neural Path K-means μπορεί να αναμίξει μαζί θετικούς και αρνητικούς γεννήτορες στην ίδια



Σχήμα 4.3: Συμπίεση νευρωνικού δικτύου πολλών εξόδων με τον αλγόριθμο Neural Path K-means. Με πράσινο χρώμα διακρίνουμε τα βάρη που αντιστοιχούν στο διάνυσμα που αναπαριστά τον $i-$οστό κόμβο στην εκτέλεση του αλγορίθμου K-means.



Σχήμα 4.4: Οπτικοποίηση του K-means στον πολυδιάστατο χώρο $\mathbb{R}^{d+1+n}$, όπου d είναι η διάσταση εισόδου του νευρωνικού και n το μέγεθος του κρυφού επιπέδου. Χρωματίζουμε τα σημεία αναφορικά με την $j-$οστή έξοδο του δικτύου. Μαύρα και άσπρα σημεία αντιστοιχούν σε γεννήτορες των P_j and Q_j αντίστοιχα. Άσπρα σημεία που βρίσκονται σε θετικές (καφέ) συστάδες ή μαύρα σημεία σε αρνητικές (μπλε) συστάδες είναι μηδενικοί γεννήτορες αναφορικά με την $j-$οστή έξοδο.

συστάδα του K-means. Για παράδειγμα, ας υποθέσουμε ότι το διάνυσμα $\left(\tilde{\mathbf{a}}_k^T,\, \tilde{b}_k,\, \tilde{C}_{:,k}^T\right)$ είναι το κέντρο της συστάδας που αποτελείται από τα διανύσματα $\left(\mathbf{a}_i^T,\, b_i,\, C_{:,i}^T\right)$ με $i \in I$. Τότε, αναφορικά με την έξοδο j, δεν είναι απαραίτητο ότι για κάθε $i \in I$ όλα τα c_{ji} θα έχουνε το ίδιο πρόσημο. Επομένως, μπορεί το τελικό συμπιεσμένο θετικό ζωνότοπο $\tilde{P}_j$ της $j-$οστής εξόδου να περιέχει γεννήτορες από το αρνητικό ζωνότοπο Q_j του αρχικού δικτύου, γεγονός που προφανώς χαλάει την αποτελεσματικότητα της συμπίεσης. Το ίδιο μπορεί να συμβεί και αντιστρόφως, δηλαδή αρνητικοί γεννήτορες να συμβάλλουν στο τελικό θετικό ζωνότοπο. Θα καλούμε τους γεννήτορες $c_{ji}\left(\mathbf{a}_i^T,\, b_i\right)$ που συμβάλλουν στο αντίθετο ζωνότοπο *μηδενικούς γεννήτορες*. Παρακάτω περιλαμβάνεται ένα σχήμα για την οπτικοποίηση των μηδενικών γεννητόρων.

Πρόταση 4.2. *Ο αλγόριθμος Neural Path K-means παράγει ένα συμπιεσμένο νευρωνικό δίκτυο με συνάρτηση εξόδου $\tilde{v}$ η οποία ικανοποιεί*

$$\frac{1}{\rho}\cdot\sum_{j=1}^{m}\max_{\mathbf{x}\in\mathcal{B}}|v_j(\mathbf{x})-\tilde{v}_j(\mathbf{x})| \le \sqrt{m}K\delta_{max}^2+\sqrt{m}\left(1-\frac{1}{N_{max}}\right)\sum_{i=1}^{n}\|C_{:,i}\|\left\|\left(\mathbf{a}_i^T,\, b_i\right)\right\| + \frac{\sqrt{m}\delta_{max}}{N_{min}}\sum_{i=1}^{n}\left(\left\|\left(\mathbf{a}_i^T,\, b_i\right)\right\|+\|C_{:,i}\|\right)+\sum_{j=1}^{m}\sum_{i\in\mathcal{N}_j}|c_{ji}|\left\|\left(\mathbf{a}_i^T,\, b_i\right)\right\|$$

όπου K είναι ο αριθμός των κέντρων του K-means, δ_{max} είναι η μεγαλύτερη απόσταση ενός σημείου από το αντίστοιχο κέντρο της συστάδας στην οποία ανήκει, N_{max}, N_{min} ο μέγιστος και ο ελάχιστος πληθάριθμος μίας συστάδας και $\mathcal{N}_j$ είναι το σύνολο των μηδενικών γεννητόρων αναφορικά με την έξοδο j.

Απόδειξη. Για να αποδείξουμε το ζητούμενο αρχικά θα εστιάσουμε στο σφάλμα που έχει η έξοδος $\tilde{v}_j$ του συμπιεσμένου δικτύου από την v_j του αρχικού. Όπως και στην απόδειξη του Zonotope K-means θα φράξουμε τις αποστάσεις $\mathcal{H}\left(P_j, \tilde{P}_j\right), \mathcal{H}\left(Q_j, \tilde{Q}_j\right)$ για κάθε έξοδο $j \in [m]$. Από την τριγωνική ανισότητα βρίσκουμε

$$|v_j(\mathbf{x})-\tilde{v}_j(\mathbf{x})| \le |p_j(\mathbf{x})-\tilde{p}_j(\mathbf{x})| + |q_j(\mathbf{x})-\tilde{q}_j(\mathbf{x})|$$

Κάθε κορυφή $\mathbf{u} \in P_j$ μπορεί να γραφεί ως $\mathbf{u} = \sum_{i\in I_{j+}} c_{ji}\left(\mathbf{a}_i^T,\, b_i\right)$ όπου το σύνολο δεικτών I_{j+} ικανοποιούν $c_{ji} > 0, \forall i \in I_{j+}$, δηλαδή αντιστοιχεί σε ένα υποσύνολο θετικών γεννητόρων. Εμείς μένει να επιλέξουμε για την κορυφή $\mathbf{u}$ μία κοντινή κορυφή στο $\tilde{P}_j$. Η επιλογή γίνεται με τον εξής τρόπο. Για κάθε $i \in I_{j+}$ διαλέγουμε το κέντρο $\begin{pmatrix}\tilde{\mathbf{a}}_k^T & \tilde{b}_k & \tilde{C}^{(k)T}\end{pmatrix}$ της συστάδας στην οποία ανήκει το σημείο $\begin{pmatrix}\mathbf{a}_i^T & b_i & C^{(i)T}\end{pmatrix}$, μόνο στην περίπτωση όπου $\tilde{c}_{jk} > 0$. Η συνθήκη αυτή δεν είναι απαραίτητο να ικανοποιείται. Μάλιστα, αυτό συμβαίνει μόνο όταν ο $c_{ji}\left(\mathbf{a}_i^T,\, b_i\right)$ δεν είναι μηδενικός γεννήτορας. Διαφορετικά, διαλέγουμε σαν κορυφή το σημείο $\mathbf{0}$, που είναι πράγματι κορυφή του ζωνοτόπου. Κατά αυτόν τον τρόπο, κάθε κέντρο συστάδας ή το μηδενικό διάνυσμα $\mathbf{0}$, λαμβάνεται υπόψιν ακριβώς μία φορά και συντίθεται η κορυφή $\sum_{k\in C_{j+}}\tilde{c}_{jk}\left(\tilde{\mathbf{a}}_k,\, \tilde{b}_k\right) \in \tilde{P}_j$. Με αυτήν την επιλογή λαμβάνουμε:

$$
\begin{aligned}
\max_{u\in\mathcal{V}_{P_j}} \operatorname{dist}\left(u, \tilde{P}_j\right) &\leq \left\| \sum_{i\in I_{j+}} c_{ji}\left(\mathbf{a}_i^T,\, b_i\right) - \sum_{k\in C_{j+}} \tilde{c}_{jk}\left(\tilde{\mathbf{a}}_k^T,\, \tilde{b}_k\right) \right\| \\
&\leq \sum_{k\in C_{j+}} \left\| \sum_{i\in I_{jk+}} c_{ji}\left(\mathbf{a}_i^T,\, b_i\right) - \tilde{c}_{jk}\left(\tilde{\mathbf{a}}_k^T,\, \tilde{b}_k\right) \right\| + \sum_{i\in N_{j+}} |c_{ji}| \left\| \left(\mathbf{a}_i^T,\, b_i\right) \right\| \\
&\leq \sum_{k\in C_{j+}} \sum_{i\in I_{jk+}} \left\| c_{ji}\left(\mathbf{a}_i^T,\, b_i\right) - \frac{\tilde{c}_{jk}\left(\tilde{\mathbf{a}}_k^T,\, \tilde{b}_k\right)}{|I_{jk+}|} \right\| + \sum_{i\in N_{j+}} |c_{ji}| \left\| \left(\mathbf{a}_i^T,\, b_i\right) \right\| \\
&\leq \sum_{k\in C} \sum_{i\in I_{jk+}} \left\| c_{ji}\left(\mathbf{a}_i^T,\, b_i\right) - \frac{(c_{ji}+\varepsilon_{ji})\left[\left(\mathbf{a}_i^T,\, b_i\right) + \lambda_i\right]}{|I_{jk+}|} \right\| + \sum_{i\in N_{j+}} |c_{ji}| \left\| \left(\mathbf{a}_i^T,\, b_i\right) \right\| \\
&\leq \sum_{k\in C_{j+}} \sum_{i\in I_{jk+}} \left[\frac{|\varepsilon_{ji}| \|\lambda_i\|}{|I_{jk+}|} + \left(1 - \frac{1}{|I_{jk+}|}\right) |c_{ji}| \left\| \left(\mathbf{a}_i^T,\, b_i\right) \right\| \right] + \\
&+ \sum_{k\in C_{j+}} \sum_{i\in I_{jk+}} \left[\frac{|\varepsilon_{ji}| \left\| \left(\mathbf{a}_i^T,\, b_i\right) \right\| + |c_{ji}| \|\lambda_i\|}{|I_{jk+}|} \right] + \sum_{i\in N_{j+}} |c_{ji}| \left\| \left(\mathbf{a}_i^T,\, b_i\right) \right\|
\end{aligned}
$$

όπου κάθε $i \in I_{jk+}$ είναι το $i-$th διάνυσμα του K-means που ανήκει στην συστάδα που ορίζεται από το $k-$οστό K-means κέντρο $k \in C_{j+}$. Επίσης, τα ε_{ij} και λ_i ορίζονται ως $\tilde{C}_{:,i} = C_{:,i} + \varepsilon_{:,i} \Rightarrow \tilde{c}_{ji} = c_{ji} + \varepsilon_{ji}$ και $\left(\tilde{\mathbf{a}}_i^T,\, \tilde{b}_i\right) = \left(\mathbf{a}_i^T,\, b_i\right) + \lambda_i$.

Όπως και στην απόδειξη του Zonotope K-means, η μέγιστη τιμή του άνω φράγματος προκύπτει όταν το I_{j+} περιέχει όλους τους δείκτες i με $c_{ji} > 0$, δηλαδή αυτούς που αφορούν τους θετικούς γεννήτορες της $j-$οστής εξόδου. Για να υπολογίσουμε ένα άνω φράγμα για το $\max_{\mathbf{v}\in\mathcal{V}_{\tilde{P}_j}} \operatorname{dist}\left(P_j, \mathbf{v}\right)$ γράφουμε την κορυφή $\mathbf{v}$ στην μορφή $\mathbf{v} = \sum_{k\in C_{j+}} \tilde{c}_{jk}\left(\tilde{\mathbf{a}}_k^T \quad \tilde{b}_k\right) \in \tilde{P}_j$ και διαλέγουμε την κορυφή $\mathbf{u} = \sum_{i\in I_{j+}} c_{ji}\left(\mathbf{a}_i^T,\, b_i\right)$ του P_j όπου το I_{j+} είναι οι δείκτες των όλων των γεννητόρων που ανήκουν στις συστάδες που καθορίζονται από τα $k \in C_{j+}$. Ωστόσο, αυτή η απόσταση είχε ληφθεί υπόψιν όταν υπολογιζόταν άνω φράγμα για την απόσταση $\max_{\mathbf{u}\in\mathcal{V}_{P_j}} \operatorname{dist}\left(\mathbf{u}, \tilde{P}_j\right)$. Συνεπώς, και οι δύο αποστάσεις λαμβάνουν το ίδιο άνω φράγμα οπότε με το ίδιο άνω φράγμα φράσσεται και η Hausdorff απόσταση

$$
\begin{aligned}
\mathcal{H}\left(P_j, \tilde{P}_j\right) \leq &\sum_{k\in C_{j+}} \sum_{i\in I_{jk+}} \left[\frac{|\varepsilon_{ji}| \|\lambda_i\|}{|I_{jk+}|} + \left(1 - \frac{1}{|I_{jk+}|}\right) |c_{ji}| \left\| \left(\mathbf{a}_i^T,\, b_i\right) \right\| \right] + \\
&+ \sum_{k\in C_{j+}} \sum_{i\in I_{jk+}} \left[\frac{|\varepsilon_{ji}| \left\| \left(\mathbf{a}_i^T,\, b_i\right) \right\| + |c_{ji}| \|\lambda_i\|}{|I_{jk+}|} \right] + \sum_{i\in N_{j+}} |c_{ji}| \left\| \left(\mathbf{a}_i^T,\, b_i\right) \right\|
\end{aligned}
$$

όπου το I_{j+} περιέχει όλους τους δείκτες i που αντιστοιχούν σε $c_{ji} > 0$. Όμοια λαμβάνουμε

$$
\begin{aligned}
\mathcal{H}\left(Q_j, \tilde{Q}_j\right) \leq &\sum_{k\in C_{j-}} \sum_{i\in I_{jk-}} \left[\frac{|\varepsilon_{ji}| \|\lambda_i\|}{|I_{jk-}|} + \left(1 - \frac{1}{|I_{jk-}|}\right) |c_{ji}| \left\| \left(\mathbf{a}_i^T,\, b_i\right) \right\| \right] + \\
&+ \sum_{k\in C_{j-}} \sum_{i\in I_{jk-}} \left[\frac{|\varepsilon_{ji}| \left\| \left(\mathbf{a}_i^T,\, b_i\right) \right\| |c_{ji}| \|\lambda_i\|}{|I_{jk-}|} \right] + \sum_{i\in N_{j-}} |c_{ji}| \left\| \left(\mathbf{a}_i^T,\, b_i\right) \right\|
\end{aligned}
$$

όπου αντίστοιχα το I_{j-} περιέχει όλα τα i ώστε $c_{ji} < 0$. Τα δύο άνω φράγματα σε συνδυασμό με το Θεώρημα 2.3 δίνουν

$$\frac{1}{\rho} \cdot \max_{\mathbf{x}\in\mathcal{B}} |v_j(\mathbf{x}) - \tilde{v}_j(\mathbf{x})| \leq \sum_{k\in C_j} \sum_{i\in I_{jk}} \left[\frac{|\varepsilon_{ji}| \|\lambda_i\|}{|I_{jk}|} + \left(1 - \frac{1}{|I_{jk}|}\right) |c_{ji}| \left\| \left(\mathbf{a}_i^T,\, b_i\right) \right\| \right] +$$
$$+ \sum_{k\in C_j} \sum_{i\in I_{jk}} \left[\frac{|\varepsilon_{ji}| \left\| \left(\mathbf{a}_i^T,\, b_i\right) \right\| + |c_{ji}| \|\lambda_i\|}{|I_{jk}|} \right] + \sum_{i\in N_j} |c_{ji}| \left\| \left(\mathbf{a}_i^T,\, b_i\right) \right\|$$

Στις παραπάνω σχέσεις χρησιμοποιήσαμε τους συμβολισμούς $C_j = C_{j+} \cup C_{j-} = \{1, 2, ..., K\}$ και I_{jk} είναι είτε ίσο με το I_{jk+} είτε I_{jk-} ανάλογα τι κέντρο αναπαριστά το $k \in C_j$ (θετικό ή αρνητικό). Σημειώνουμε ότι $\{i | i \in I_{jk}, k \in C_j\} = \{1, 2, ..., n\} \setminus N_j \subseteq \{1, 2, ..., n\}$, αφού κάθε μη-μηδενικός γεννήτορας αντιστοιχεί σε κάποιο κέντρο συστάδας που έχει ίδιο πρόσημο. Επιπλέον, χρησιμοποιώντας την σχέση $N_{\max} \geq |I_{jk}| \geq N_{\min}$, προκύπτει ότι

$$\frac{1}{\rho} \cdot \max_{\mathbf{x}\in\mathcal{B}} |v_j(\mathbf{x}) - \tilde{v}_j(\mathbf{x})| \leq \sum_{i=1}^{n} \left[\frac{|\varepsilon_{ji}| \|\lambda_i\|}{N_{\min}} + \left(1 - \frac{1}{N_{\max}}\right) |c_{ji}| \left\| \left(\mathbf{a}_i^T,\, b_i\right) \right\| \right] +$$
$$+ \sum_{i=1}^{n} \left[\frac{|\varepsilon_{ji}| \left\| \left(\mathbf{a}_i^T,\, b_i\right) \right\| + |c_{ji}| \|\lambda_i\|}{N_{\min}} \right] + \sum_{i\in N_j} |c_{ji}| \left\| \left(\mathbf{a}_i^T,\, b_i\right) \right\|$$

Υπολογίζουμε ένα άνω φράγμα για το συνολικό κόστος που συνδυάζει όλες τις εξόδους εφαρμόζοντας την ανισότητα

$$\left(\sum_{j=1}^{m} |u_j| \right)^2 \leq m \left(\sum_{j=1}^{m} |u_j|^2 \right) \Leftrightarrow \sum_{j=1}^{m} |u_j| \leq \sqrt{m}\, \|(u_1, ..., u_m)\|$$

η οποία αποτελεί άμεση εφαρμογή της Cauchy-Schwartz ανισότητας. Χρησιμοποιώντας και τις σχέσεις $\|\varepsilon_{:,i}\| < \delta_{\max}, \quad \|\lambda_i\| < \delta_{\max}$, βρίσκουμε

$$\begin{aligned}
\frac{1}{\rho} \cdot \sum_{j=1}^{m} \max_{\mathbf{x}\in\mathcal{B}} |v_j(\mathbf{x}) - \tilde{v}_j(\mathbf{x})| &\leq \sum_{j=1}^{m} \sum_{i=1}^{n} \left[\frac{|\varepsilon_{ji}| \|\lambda_i\|}{N_{\min}} + \left(1 - \frac{1}{N_{\max}}\right) |c_{ji}| \left\| \left(\mathbf{a}_i^T,\, b_i\right) \right\| \right] + \\
&+ \sum_{j=1}^{m} \sum_{i=1}^{n} \left[\frac{|\varepsilon_{ji}| \left\| \left(\mathbf{a}_i^T,\, b_i\right) \right\| + |c_{ji}| \|\lambda_i\|}{N_{\min}} \right] + \sum_{j=1}^{m} \sum_{i\in N_j} |c_{ji}| \left\| \left(\mathbf{a}_i^T,\, b_i\right) \right\| \\
&\leq \sum_{i=1}^{n} \left[\frac{\sqrt{m}\, \|\varepsilon_{:,i}\| \|\lambda_i\|}{N_{\min}} + \left(1 - \frac{1}{N_{\max}}\right) \sqrt{m}\, \|C_{:,i}\| \left\| \left(\mathbf{a}_i^T,\, b_i\right) \right\| \right] + \\
&+ \sum_{i=1}^{n} \left[\frac{\sqrt{m}\, \|\varepsilon_{:,i}\| \left\| \left(\mathbf{a}_i^T,\, b_i\right) \right\| + \sqrt{m}\, \|C_{:,i}\| \|\lambda_i\|}{N_{\min}} \right] + \sum_{j=1}^{m} \sum_{i\in N_j} |c_{ji}| \left\| \left(\mathbf{a}_i^T,\, b_i\right) \right\| \\
&\leq \sqrt{m} K \delta_{\max}^2 + \sqrt{m} \left(1 - \frac{1}{N_{\max}}\right) \sum_{i=1}^{n} \|C_{:,i}\| \left\| \left(\mathbf{a}_i^T,\, b_i\right) \right\| + \\
&\frac{\sqrt{m}\delta_{\max}}{N_{\min}} \sum_{i=1}^{n} \left(\left\| \left(\mathbf{a}_i^T,\, b_i\right) \right\| + \|C_{:,i}\| \right) + \sum_{j=1}^{m} \sum_{i\in N_j} |c_{ji}| \left\| \left(\mathbf{a}_i^T,\, b_i\right) \right\|
\end{aligned}$$

που δίνει το επιθυμητό αποτέλεσμα. □

Με την Πρόταση 4.2 αξιολογούμε την επίδοση του Neural Path K-means στο ζήτημα της συμπίεσης νευρωνικών δικτύων πολλών εξόδων. Το αποτέλεσμα που λαμβάνουμε είναι

αντίστοιχο αυτού που βρήκαμε για τον Zonotope K-means. Το σφάλμα της προσέγγισης ελαττώνεται όσο το πλήθος των κέντρων K πλησιάζει το μέγεθος του κρυφού επιπέδου n. Ωστόσο, όσο μειώνεται το K, το σφάλμα της προσέγγισης χειροτερεύει προσεγγίζοντας μία ανώτατη τιμή η οποία εξαρτάται από το μέγεθος των βαρών στα γραμμικά επίπεδα του δικτύου μαζί με τα μεγέθη των μηδενικών γεννητόρων.

4.3 Συμπίεση Συνελικτικών Δικτύων

Εμπνευσμένοι από την μέθοδο συμπίεσης Neural Path K-means που αφορά feed-forward νευρωνικά δίκτυα θα αναπτύξουμε αλγόριθμο συμπίεσης για δίκτυα με συνελικτικά επίπεδα. Η ιδέα της συμπίεσης στην προηγούμενη ενότητα ήταν να μειώνουμε σε πλήθος τους νευρώνες του κρυφού επιπέδου εφαρμόζοντας K-means. Τα διανύσματα στα οποία εφαρμοζόταν ο K-means ήταν ένα για κάθε νευρώνα του κρυφού επιπέδου που περιέχει όλα τα βάρη που αφορούν τον εν λόγω νευρώνα. Αντίστοιχα, στην περίπτωση των συνελικτικών δικτύων σκοπεύουμε για δύο διαδοχικά συνελικτικά επίπεδα να επιτύχουμε μείωση του πλήθους των νευρώνων στο κρυφό επίπεδο. Η μείωση αυτή επιλέγεται να πραγματοποιηθεί μειώνοντας το πλήθος των καναλιών του κρυφού επιπέδου. Ο αλγόριθμος συμπίεσης που χρησιμοποιούμε ονομάζεται Convolutional Neural Path K-means και περιγράφεται ως εξής. Για κάθε κανάλι του κρυφού επιπέδου δημιουργούμε ένα διάνυσμα που περιέχει όλα τα βάρη των φίλτρων που αλληλεπιδρούν με το συγκεκριμένο κανάλι. Έπειτα, εφαρμόζουμε K-means στα διανύσματα που προκύπτουν και τελικά το συμπιεσμένο δίκτυο προκύπτει από τα κανάλια που αφορούν τα κέντρα του K-means. Την διαδικασία αυτή θα αναλύσουμε με λεπτομέρεια παρακάτω.

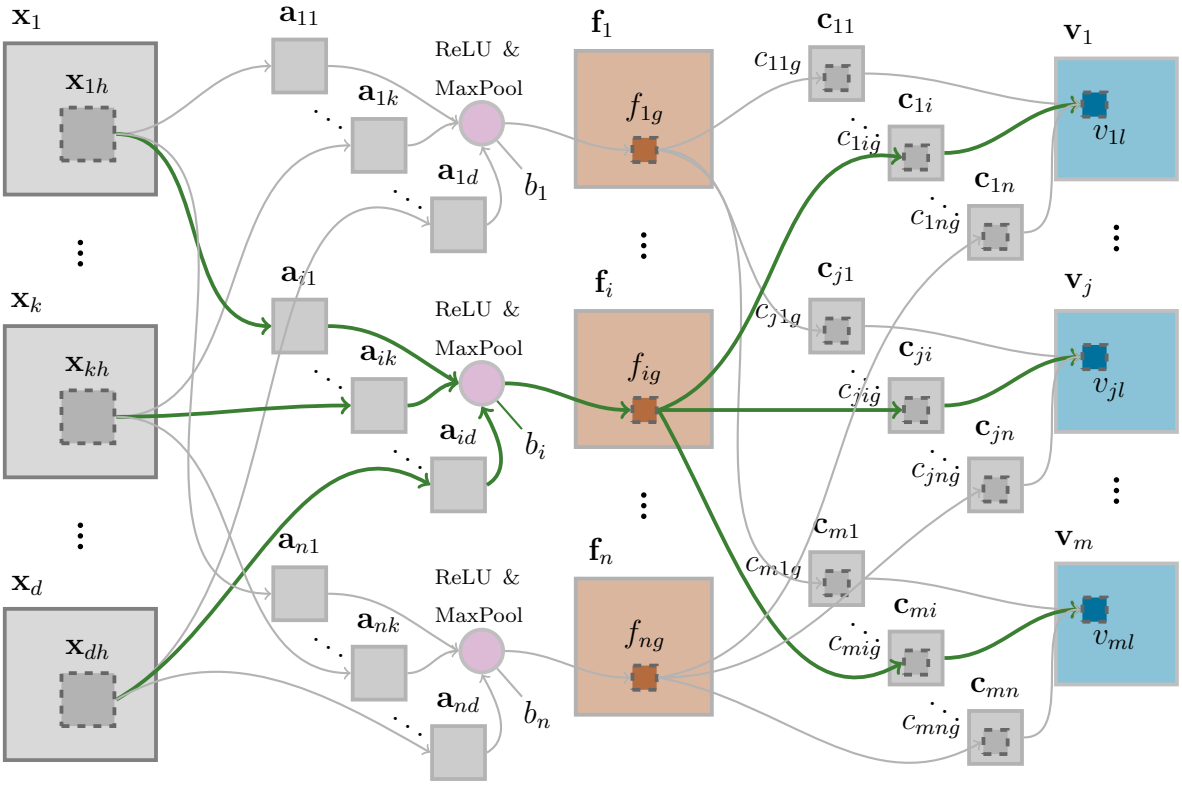
Υποθέτουμε ότι εργαζόμαστε με το συνελικτικό νευρωνικό δίκτυο της εικόνας 4.5. Αυτό αποτελείται από δύο συνελικτικά επίπεδα μεταξύ των οποίων παρεμβάλλονται ReLU και MaxPooling επίπεδα. Το πρώτο συνελικτικό επιπέδο έχει ως είσοδο την εικόνα $\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 & \dots & \mathbf{x}_d \end{pmatrix}$ αποτελούμενη από d κανάλια, και στην έξοδό του δίνει εικόνα αποτελούμενη από n κανάλια. Υποθέτουμε ότι η έξοδος έπειτα από το ReLU και MaxPooling επίπεδα είναι η εικόνα $\mathbf{f} = \begin{pmatrix} \mathbf{f}_1 & \dots & \mathbf{f}_n \end{pmatrix}$. Εν συνεχεία, το δεύτερο συνελικτικό επίπεδο δέχεται ως είσοδο την εικόνα $\mathbf{f}$ και δίνει ως έξοδο την $\mathbf{v} = \begin{pmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_m \end{pmatrix}$ εικόνα αποτελούμενη από m κανάλια. Θεωρούμε πως το πρώτο συνελικτικό επίπεδο έχει τα φίλτρα $\mathbf{a}_{i1}, ..., \mathbf{a}_{id}$ και σταθερό όρο b_i για το $i-$οστό κανάλι του κρυφού επιπέδου όπου $i = 1, ..., n$.

Παρατήρηση. *Σημειώνουμε ότι με boldface γράμματα θα δηλώνουμε διανύσματα στήλες ακόμα και αν αυτά αναφέρονται σε διδιάστατα αντικείμενα. Για παράδειγμα, τα φίλτρα $\mathbf{a}_{ik}$ αποτελούν διδιάστατες δομές, αλλά εμείς θα εργαζόμαστε με αυτά με την ξεδιπλωμένη μονοδιάστατη (flattened) εκδοχή τους.*

Θα υπολογίσουμε ένα pixel στην έξοδο του $i-$οστού καναλιού του πρώτου συνελικτικού επιπέδου. Έστω, λοιπόν ότι υπολογίζουμε το $w-$οστό pixel. Τότε θα υπάρχει ένας δείκτης h ο οποίος καθορίζει τις υποεικόνες $\mathbf{x}_{1h}, ..., \mathbf{x}_{dh}$, όπως φαίνεται στην εικόνα 4.5, ο οποίος είναι κατάλληλος ώστε το εσωτερικό γινόμενο των φίλτρων με αυτές να δίνει τον τρέχοντα υπολογισμό της συνέλιξης που αφορά το επιθυμητό pixel εξόδου. Η έξοδος στο $w-$οστό pixel του $i-$οστού καναλιού, επομένως, θα είναι

$$\sum_{k=1}^{d} \mathbf{a}_{ik}^T \mathbf{x}_{kh} + b_i \tag{4.3}$$

Επεκτείνουμε τους υπολογισμούς μας επιθυμώντας να βρούμε το $g-$οστό pixel του $i-$οστού καναλιού των εικόνων, όταν αυτές έχουν περάσει από τα ReLU και MaxPooling επίπεδα. Το



Σχήμα 4.5: Συνελικτικό Νευρωνικό Δίκτυο αποτελούμενο από 2 συνελικτικά επίπεδα. Μεταξύ του πρώτου και του δεύτερου επιπέδου παρεμβάλλονται ReLU και MaxPooling επίπεδα. Με πράσινο χρώμα σημειώνουμε όλα τα μονοπάτια του δικτύου (Neural Paths) τα οποία αφορούν το $i-$οστό κανάλι του κρυφού επιπέδου.

pixel αυτό ανήκει στην εικόνα $\mathbf{f}_i(\mathbf{x})$ και συμβολίζεται ως η συνάρτηση $f_{ig}(\mathbf{x})$. Παρατηρούμε, ότι έπειτα από την έξοδο του ReLU επιπέδου, τα pixel των εικόνων θα έχουν συγκριθεί με το 0 και θα έχει επιλεγεί η μέγιστη τιμή. Οι τιμές αυτές ομαδοποιούνται με συγκεκριμένο τρόπο που ορίζει το MaxPooling επίπεδο, και σε κάθε ομάδα επιλέγεται το μέγιστο στοιχείο. Συνεπώς, θα υπάρχουν δείκτες $h : h(g)$ που εξαρτώνται από την επιλογή του g και καθορίζονται από το εύρος του MaxPooling, ούτως ώστε

$$f_{ig}(\mathbf{x}) = \max_{h:h(g)} \left\{ \max \left\{ \sum_{k=1}^{d} \mathbf{a}_{ik}^T \mathbf{x}_{kh} + b_i, 0 \right\} \right\}$$

Στο δεύτερο συνελικτικό επίπεδο έχουμε τα φίλτρα $\mathbf{c}_{j1}, ..., \mathbf{c}_{jn}$ για κάθε $j = 1, ..., m$. Για τον υπολογισμό του $l-$οστού pixel εξόδου του $j-$οστού καναλιού $v_{jl}(\mathbf{x})$ εργαζόμαστε ως εξής. Χρησιμοποιώντας τον συμβολισμό c_{jig} για το $g-$οστό pixel του φίλτρου $\mathbf{c}_{ji}$ προκύπτει ότι

$$v_{jl}(\mathbf{x}) = \sum_{i=1}^{n} \sum_{g:g(l)} c_{jig} f_{ig}(\mathbf{x})$$

όπου το g είναι ένας δείκτης που εξαρτάται από το l και οι τιμές του διατρέχουν όλα τα pixel του φίλτρου $\mathbf{c}_{ji}$ καθώς και τις κατάλληλες τιμές $f_{ig}(\mathbf{x})$ της εικόνας $\mathbf{f}_i$.

Αναφορικά με την τροπική γεωμετρία των συνελικτικών δικτύων, παρατηρούμε ότι οι συναρτήσεις εξόδου $v_{jl}(\mathbf{x})$ αποτελούν τροπικές ρητές συναρτήσεις, αφού τα $f_{ig}(\mathbf{x})$ είναι τροπικά πολυώνυμα. Αυτό, άλλωστε, ήταν αναμενόμενο δεδομένου ότι ένα συνελικτικό επίπεδο με ReLU ενεργοποιήσεις ισοδυναμεί με ένα πλήρως συνδεδεμένο γραμμικό επίπεδο. Επομένως μπορούμε να γράψουμε

$$v_{jl}(\mathbf{x}) = p_{jl}(\mathbf{x}) - q_{jl}(\mathbf{x})$$

όπου $p_{jl}(\mathbf{x}), q_{jl}(\mathbf{x})$ τροπικά πολυώνυμα. Αξίζει να παρατηρήσουμε ότι τα ENewt (p_{jl}), ENewt (q_{jl}) δεν είναι ζωνότοπα, αφού τα $f_{ig}(\mathbf{x})$ έχουν παραπάνω από 2 κορυφές. Ωστόσο, αποτελούν μία ιδιαίτερη δομή την οποία ονομάζουμε γενικευμένο ζωνότοπο.

Ορισμός. Ονομάζουμε γενικευμένο ζωνότοπο $P \subset \mathbb{R}^d$ κάθε Minkowski άθροισμα της μορφής

$$P = P_1 \oplus P_2 \oplus ... \oplus P_k$$

όπου τα P_i είναι πολύτοπα που έχουν τον ίδιο αριθμό κορυφών.

Ο παραπάνω ορισμός αποτελεί πράγματι γενίκευση των ζωνοτόπων, αφού στην περίπτωση όπου τα P_i έχουν 2 κορυφές, δηλαδή είναι ευθύγραμμα τμήματα, τότε το P αποτελεί ζωνότοπο. Σε αναλογία με τα κλασσικά νευρωνικά, χρησιμοποιούμε τους συμβολισμούς

$$P_{jl} = \text{ENewt}\left(p_{jl}\right), \quad Q_{jl} = \text{ENewt}\left(q_{jl}\right)$$

Για τα πολύτοπα αυτά έχουμε την παρακάτω πρόταση.

Πρόταση 4.3. *Τα πολύτοπα P_{jl}, Q_{jl} είναι γενικευμένα ζωνότοπα.*

Απόδειξη. Πράγματι, για το θετικό πολυώνυμο p_{jl} έχουμε

$$p_{jl}(\mathbf{x}) = \sum_{i=1}^{n} \sum_{g:g(l)\wedge c_{jig}>0} c_{jig} f_{ig}(\mathbf{x}) =$$
$$= \sum_{i=1}^{n} \sum_{g:g(l)\wedge c_{jig}>0} c_{jig} \max_{h:h(g)} \left\{ \max \left\{ \sum_{k=1}^{d} \mathbf{a}_{ik}^T \mathbf{x}_{kh} + b_i, 0 \right\} \right\} \xrightarrow{Prop.2.1}$$
$$P_{jl} = \bigoplus_{i,g:g(l)\wedge c_{jig}>0} \text{ENewt} \left(\max_{h:h(g)} \left\{ \max \left\{ \sum_{k=1}^{d} c_{jig} \mathbf{a}_{ik}^T \mathbf{x}_{kh} + c_{jig} b_i, 0 \right\} \right\} \right)$$

Επομένως, το P_{jl} γράφεται σαν Minkowski άθροισμα πολυτόπων. Κάθε ένα από αυτά τα πολύτοπα περιέχει τόσες κορυφές όσες και οι δυνατές τιμές του h που καθορίζονται από το εύρος του MaxPooling συν μία, την μηδενική κορυφή $\mathbf{0}$. Επομένως, όλα τα πολύτοπα έχουν τον ίδιο αριθμό κορυφών και το P_{jl} αποτελεί γενικευμένο ζωνότοπο. Όμοια προκύπτει το αποτέλεσμα για το αρνητικό γενικευμένο ζωνότοπο Q_{jl}. □

Θα συνεχίσουμε την ανάλυση του συνελικτικού νευρωνικού ούτως ώστε να μπορέσουμε να αξιολογήσουμε την απόδοση του αλγορίθμου συμπίεσης Convolutional Neural Path K-means. Για να γίνει αυτό θα πρέπει, όπως και στα feed-forward νευρωνικά, να υπολογίσουμε τις κορυφές των θετικών και αρνητικών πολυτόπων κάθε εξόδου. Θέλουμε να βρούμε έναν πίνακα ο οποίος να αναπαριστά την κορυφή του επεκτεταμένου πολυτόπου Newton του πολυωνύμου $\sum_{k=1}^{d} \mathbf{a}_{ik}^T \mathbf{x}_{kh} + b_i$. Για αυτόν τον σκοπό ορίζουμε

$$A_{ih} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{a}_{i3} & \dots & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{a}_{i1} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{a}_{i2} & \mathbf{0} & \ddots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \mathbf{a}_{ik} & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} & \mathbf{a}_{id} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \end{bmatrix}$$

Ο παραπάνω πίνακας έχει τόσες γραμμές όσες και η διάσταση του διανύσματος εισόδου $\mathbf{x}$ και d στήλες. Για την κατασκευή του A_{ih} αρχικά έχουμε όλες τις στήλες μηδενικές και εν συνεχεία στην κατάλληλη θέση της στήλης k τοποθετείται το διάνυσμα $\mathbf{a}_{ik}$ για κάθε $k = 1, ..., d$. Τα $\mathbf{a}_{ik}$ είναι καταλλήλως τοποθετημένα ούτως ώστε το $\mathbf{a}_{ik}$ να αντιστοιχεί στα entries του διανύσματος εισόδου $\mathbf{x}$ που αντιστοιχούν στο $\mathbf{x}_{kh}$. Σημειώνουμε ότι για ευκολία στην αναπαράσταση του πίνακα A_{ih} τα μηδενικά έχουν τοποθετηθεί με τυχαίο τρόπο. Επίσης, με $\mathbb{1}$ συμβολίζουμε το διάνυσμα που έχει με d entries όλα ίσα με 1. Πράγματι, με αυτόν τον ορισμό βρίσκουμε ότι το γραμμικό πολυώνυμο $\sum_{k=1}^{d} \mathbf{a}_{ik}^T \mathbf{x}_{kh} + b_i$ αντιστοιχεί στην κορυφή $\big(A_{ih}\mathbb{1}, b_i\big)$. Κατ' επέκταση, λαμβάνουμε το ακόλουθο αποτέλεσμα.

Πρόταση 4.4. *Έστω P_{jl}, Q_{jl} τα γενικευμένα ζωνότοπα της τροπικής ρητής συνάρτησης που αντιστοιχεί στο $l-$οστό pixel του $j-$οστού καναλιού της εξόδου. Αν* **v** *κορυφή του θετικού και* **u** *κορυφή του αρνητικού γενικευμένου ζωνοτόπου, τότε μπορούμε να γράψουμε*

$$\mathbf{v} = \sum_{(i,g)\in I_+} c_{jig}\left(A_{ih}\mathbb{1},\, b_i\right), \quad \mathbf{u} = \sum_{(i,g)\in I_-} c_{jig}\left(A_{ih}\mathbb{1},\, b_i\right)$$

όπου το I_+ είναι ένα σύνολο ζευγών δεικτών (i,g) τέτοιο ώστε $c_{jig} > 0$, $1 \le i \le n$, το g επιλέγεται καταλλήλως εξαρτώμενο από το l, δηλαδή, $g : g(l)$ ώστε να γίνεται η συνέλιξη με τα $f_{ig}(\mathbf{x})$ και $h = h(g)$ κατάλληλα επιλεγμένος δείκτης σύμφωνα με την επιλογή του g. Αντίστοιχα, ορίζουμε το I_- για το Q_{jl}.

Απόδειξη. Γράφουμε

$$v_{jl}(\mathbf{x}) = \sum_{i=1}^{n} \sum_{g:g(l)} c_{jig} f_{ig}(\mathbf{x})$$

όπου, όπως έχουμε εξηγήσει, ισχύει

$$f_{ig}(\mathbf{x}) = \max_{h:h(g)} \left\{ \max\left\{ \sum_{k=1}^{d} \mathbf{a}_{ik}^T \mathbf{x}_{kh} + b_i, 0 \right\} \right\}$$

Τα f_{ig} είναι τροπικά πολυώνυμα και τα επεκτεταμένα Newton πολύτοπά τους έχουν κορυφές

$$\{\mathbf{0}\} \cup \{(A_{ih}\mathbb{1},\, b_i),\, h : h(g)\}$$

Όπως είδαμε στην Πρόταση 4.3 το θετικό μέρος p_{jl} γράφεται σαν γραμμικός συνδυασμός με συντελεστές $c_{jig} > 0$ των τροπικών πολυωνύμων f_{ig}, οπότε και το επεκτεταμένο Newton πολύτοπό του είναι το Minkowski άθροισμα των $\mathrm{ENewt}\,(c_{jig} f_{ig})$. Επομένως, κάθε κορυφή $v \in P_{jl}$ κατασκευάζεται από την επιλογή για κάθε (i,g) κορυφή $\mathbf{0}$ ή δείκτη $h = h(g)$ και τελικά προκύπτει

$$\mathbf{v} = \sum_{(i,g)\in I_+} c_{jig}\left(A_{ih}\mathbb{1},\, b_i\right)$$

όπως ήταν επιθυμητό. Αντίστοιχα προκύπτει το αποτέλεσμα για τις κορυφές του αρνητικού γενικευμένου ζωνοτόπου Q_{jl}. □

Παρακάτω παρουσιάζουμε τον αλγόριθμο Convolutional Neural Path K-means, ο οποίος αξιοποιείται για την συμπίεση ενός νευρωνικού δικτύου με δύο συνελικτικά επίπεδα στα οποία παρεμβάλλονται ReLU ενεργοποιήσεις και MaxPooling. Σημειώνουμε, ότι όπως και με τον αλγόριθμο Neural Path K-means, η διαδοχική εφαρμογή της συμπίεσης στα επίπεδα ενός βαθιού νευρωνικού μας επιτρέπει να το συμπιέσουμε εξ' ολοκλήρου.

Εν συνεχεία ακολουθεί η θεωρητική ανάλυση της απόδοσης του αλγορίθμου 4 αναφορικά με το σφάλμα που έχει η συνάρτηση εξόδου του συμπιεσμένου δικτύου σε σχέση με το αρχικό. Για την απόδειξη θα χρησιμοποιήσουμε τους συμβολισμούς

$$\mathbf{a}_i = \begin{pmatrix} \mathbf{a}_{i1} \\ \vdots \\ \mathbf{a}_{id} \end{pmatrix}, \quad \mathbf{c}_i = \begin{pmatrix} \mathbf{c}_{1i} \\ \vdots \\ \mathbf{c}_{mi} \end{pmatrix}$$

Algorithm 4 Convolutional Neural Path K-means Compression

1. Κατασκευάζουμε τα διανύσματα $\left(\mathbf{a}_{i1}^T \quad \dots \mathbf{a}_{id}^T \quad b_i \quad \mathbf{c}_{1i}^T \quad \dots \mathbf{c}_{mi}^T\right)$ για κάθε $i = 1, ..., n$ και εκτελούμε σε αυτά τον αλγόριθμο K-means για K κέντρα.

2. Από τον αλγόριθμο K-means λαμβάνουμε τα κέντρα $\left(\tilde{\mathbf{a}}_{i1}^T \quad \dots \tilde{\mathbf{a}}_{id}^T \quad \tilde{b}_i \quad \tilde{\mathbf{c}}_{1i}^T \quad \dots \tilde{\mathbf{c}}_{mi}\right)^T$ για κάθε $i = 1, ..., K$.

3. Κατασκευάζουμε τα τελικά φίλτρα του δικτύου ως εξής. Το κρυφό επίπεδο έχει K κανάλια, όπου για το $i-$οστό κανάλι έχουμε τα φίλτρα $\left(\tilde{\mathbf{a}}_{i1}, \dots \tilde{\mathbf{a}}_{id}, \tilde{b}_i\right)$ ενώ για το δεύτερο συνελικτικό επίπεδο τα φίλτρα του $j-$οστού καναλιού είναι τα $(\tilde{\mathbf{c}}_{j1}, \dots, \tilde{\mathbf{c}}_{jK})$.

Αξίζει να σημειώσουμε ότι και σε αυτήν την περίπτωση, λόγω της φύσης του αλγορίθμου συμπίεσης, θα υπάρχουν γεννήτορες $c_{jig}\left(A_{ih}\mathbb{1}, b_i\right)$ των πολυτόπων οι οποίοι είναι μηδενικοί, δηλαδή συνδράμουν σε cluster του K-means το οποίο εν τέλει έχει αντίθετο πρόσημο.

Πρόταση 4.5. *Ο αλγόριθμος Convolutional Neural Path K-means παράγει ένα συμπιεσμένο νευρωνικό δίκτυο με συνάρτηση εξόδου $\tilde{v}$ η οποία ικανοποιεί*

$$\frac{1}{\rho}\cdot\sum_{j=1}^{m}\sum_{l}\max_{\mathbf{x}\in\mathcal{B}}|v_{jl}(\mathbf{x})-\tilde{v}_{jl}(\mathbf{x})| \leq N_2\sqrt{(d+1)mF_2}\cdot\left[K\delta_{max}^2+\left(1-\frac{1}{N_{max}}\right)\sum_{i=1}^{n}\|\mathbf{c}_i\|\left\|\left(\mathbf{a}_i^T, b_i\right)\right\|+\right.$$
$$\left.+\frac{\delta_{max}}{N_{min}}\sum_{i=1}^{n}\left(\|\mathbf{c}_i\|+\left\|\left(\mathbf{a}_i^T, b_i\right)\right\|\right)\right]+\sum_{j,l}\sum_{(i,g)\in\mathcal{N}_{jl}}|c_{jig}|\left\|\left(\mathbf{a}_i^T, b_i\right)\right\|$$

όπου K είναι ο αριθμός των κέντρων του K-means, δ_{max} είναι η μεγαλύτερη απόσταση ενός σημείου από το αντίστοιχο κέντρο της συστάδας στην οποία ανήκει, N_{max}, N_{min} ο μέγιστος και ο ελάχιστος πληθάριθμος μίας συστάδας, $\mathcal{N}_{jl}$ το σύνολο των μηδενικών γεννητόρων αναφορικά με την l έξοδο του καναλιού j και F_2, N_2 το πλήθος των pixel στα συνελικτικά φίλτρα και στις εικόνες εξόδου, αντίστοιχα, στο δεύτερο συνελικτικό επίπεδο.

Απόδειξη. Γράφουμε τις τροπικές ρητές συναρτήσεις εξόδου $v_{jl}, \tilde{v}_{jl}$ σαν διαφορά τροπικών πολυωνύμων, οπότε με χρήση τριγωνικής ανισότητας λαμβάνουμε:

$$\begin{aligned}|v_{jl}(\mathbf{x})-\tilde{v}_{jl}(\mathbf{x})| &\leq |p_{jl}(\mathbf{x})-\tilde{p}_{jl}(\mathbf{x})|+|q_{jl}(\mathbf{x})-\tilde{q}_{jl}(\mathbf{x})|\\ &\leq \rho\cdot\left(\mathcal{H}\left(P_{jl},\tilde{P}_{jl}\right)+\mathcal{H}\left(Q_{jl},\tilde{Q}_{jl}\right)\right)\end{aligned}$$

όπου στην τελευταία ανισότητα χρησιμοποιήσαμε το Θεώρημα 2.3. Μένει, λοιπόν να φράξουμε τις αποστάσεις Hausdorff των αρχικών πολυτόπων με τα συμπιεσμένα. Θα εργαστούμε με το θετικό πολύτοπο αφού η διαδικασία για το αρνητικό είναι εντελώς όμοια. Ακολουθούμε τεχνική όμοια με αυτήν στις προηγούμενες αποδείξεις. Αρχικά, θεωρούμε μία κορυφή $\mathbf{u} \in P_{jl}$ και προσδιορίζουμε κατάλληλα κορυφή στο $\tilde{P}_{jl}$ ώστε να βρίσκεται κοντά στην $\mathbf{u}$. Ο τρόπος επιλογής έχει επαναληφθεί στις προηγούμενες αποδείξεις. Για κάθε γεννήτορα $c_{jig}\left(A_{ih}\mathbb{1}, b_i\right)$ διαλέγουμε το αντίστοιχο κέντρο $\tilde{c}_{jkg}\left(\tilde{A}_{kh}\mathbb{1}, \tilde{b}_k\right)$ αν ο γεννήτορας δεν είναι μηδενικός, αλλιώς διαλέγουμε το $\mathbf{0}$. Χρησιμοποιώντας την Πρόταση 4.4 για τον υπολογισμό κορυφών προκύπτει ότι

$$
\begin{aligned}
\max_{\mathbf{u}\in\mathcal{V}_{P_{jl}}} \operatorname{dist}\left(\mathbf{u}, \tilde{P}_{jl}\right) &\leq \left\| \sum_{(i,g)\in I_{j+}} c_{jig}\left(A_{ih}\mathbb{1},\, b_i\right) - \sum_{(k,g)\in C_{j+}} \tilde{c}_{jkg}\left(\tilde{A}_{kh}\mathbb{1},\, \tilde{b}_k\right) \right\| \\
&\leq \sum_{(k,g)\in C_{j+}} \left\| \sum_{(i,g)\in I_{jk+}} c_{jig}\left(A_{ih}\mathbb{1},\, b_i\right) - \tilde{c}_{jkg}\left(\tilde{A}_{kh}\mathbb{1},\, \tilde{b}_k\right) \right\| + \sum_{(i,g)\in\mathcal{N}_{jl}} |c_{jig}|\, \|(A_{ih}\mathbb{1},\, b_i)\| \\
&\leq \sum_{(k,g)\in C_{j+}} \sum_{(i,g)\in I_{jk+}} \left\| c_{jig}\left(A_{ih}\mathbb{1},\, b_i\right) - \frac{\tilde{c}_{jkg}}{|I_{jk+}|}\left(\tilde{A}_{kh}\mathbb{1},\, \tilde{b}_k\right) \right\| + \sum_{(i,g)\in\mathcal{N}_{jl}} |c_{jig}|\, \|(A_{ih}\mathbb{1},\, b_i)\| \\
&\leq \sum_{(k,g)\in C_{j+}} \sum_{(i,g)\in I_{jk+}} \left\| c_{jig}\left(A_{ih}\mathbb{1},\, b_i\right) - \frac{c_{jig}+\varepsilon_{jig}}{|I_{jk+}|}\left(A_{ih}\mathbb{1} + M_{ih}\mathbb{1},\, b_i + \lambda_i\right) \right\| + \sum_{(i,g)\in\mathcal{N}_{jl}} |c_{jig}|\, \|(A_{ih}\mathbb{1},\, b_i)\| \\
&\leq \sum_{(k,g)\in C_{j+}} \sum_{(i,g)\in I_{jk+}} \left[\frac{|\varepsilon_{jig}|}{|I_{jk+}|}\, \|(M_{ih}\mathbb{1},\, \lambda_i)\| + \left(1 - \frac{1}{|I_{jk+}|}\right) |c_{jig}|\, \|(A_{ih}\mathbb{1},\, b_i)\| + \right. \\
&\quad \left. + \frac{|c_{jig}|}{|I_{jk+}|}\, \|(M_{ih}\mathbb{1},\, \lambda_i)\| + \frac{|\varepsilon_{jig}|}{|I_{jk+}|}\, \|(A_{ih}\mathbb{1},\, b_i)\| \right] + \sum_{(i,g)\in\mathcal{N}_{jl}} |c_{jig}|\, \|(A_{ih}\mathbb{1},\, b_i)\|
\end{aligned}
$$

όπου με C_{j+} συμβολίζουμε τα κέντρα (k, g) του K-means τα οποία έχουν επιλεγεί όπως περιγράψαμε. Επίσης, με I_{jk+} ορίζουμε το σύνολο των ζευγών (i, g) στα οποία οι γεννήτορες $c_{jig}\left(A_{ih}\mathbb{1},\, b_i\right)$ αντιστοιχούν στην συστάδα με κέντρο $\tilde{c}_{jkg}\left(\tilde{A}_{kh}\mathbb{1},\, \tilde{b}_k\right)$. Τέλος, τα σύμβολα $\varepsilon, M, \lambda, \mu$ αναπαριστούν τις διαφορές των διανυσμάτων από τα εκάστοτε κέντρα του K-means. Συγκεκριμένα ορίζουμε

$$\tilde{\mathbf{c}}_k = \mathbf{c}_i + \boldsymbol{\varepsilon}_i \Leftrightarrow \tilde{c}_{jkg} = c_{jig} + \varepsilon_{jig}$$

$$\tilde{\mathbf{a}}_k = \mathbf{a}_i + \boldsymbol{\mu}_i \Leftrightarrow \tilde{A}_{kh} = A_{ih} + M_{ih}$$

$$\tilde{b}_k = b_i + \lambda_i$$

Η μέγιστη τιμή του άνω φράγματος που προέκυψε, λαμβάνεται όταν το I_{j+} περιέχει όλους τους δείκτες (i, g) με $c_{jig} > 0$, αφού λόγω της απόλυτης τιμής, όσο περισσότεροι οι γεννήτορες, τόσο μεγαλύτερο και το άνω φράγμα.

Για να υπολογίσουμε ένα άνω φράγμα για το $\max_{\mathbf{v}\in\mathcal{V}_{\tilde{P}_{jl}}} \operatorname{dist}\left(P_{jl}, \mathbf{v}\right)$ γράφουμε την κορυφή $\mathbf{v}$ στην μορφή $\mathbf{v} = \sum_{k\in C_{j+}} \tilde{c}_{jkg}\left(\tilde{A}_{kh}\mathbb{1},\, \tilde{b}_k\right) \in \tilde{P}_{jl}$ και διαλέγουμε την κορυφή $\mathbf{u} = \sum_{i\in I_{j+}} c_{jig}\left(A_{ih}\mathbb{1},\, b_i\right)$ του P όπου το I_{j+} είναι οι δείκτες των όλων των γεννητόρων που ανήκουν στις συστάδες που καθορίζονται από τα $k \in C_{j+}$. Ωστόσο, αυτή η απόσταση είχε ληφθεί υπόψιν όταν υπολογιζόταν άνω φράγμα για την απόσταση $\max_{\mathbf{u}\in\mathcal{V}_{P_{jl}}} \operatorname{dist}\left(\mathbf{u}, \tilde{P}_{jl}\right)$. Συνεπώς, και οι δύο αποστάσεις λαμβάνουν το ίδιο άνω φράγμα οπότε με αυτό φράσσεται και η Hausdorff απόσταση

$$
\begin{aligned}
\mathcal{H}\left(P_j, \tilde{P}_j\right) \leq \sum_{(k,g)\in C_{j+}} \sum_{(i,g)\in I_{jk+}} &\left[\frac{|\varepsilon_{jig}|}{|I_{jk+}|}\, \|(M_{ih}\mathbb{1},\, \lambda_i)\| + \left(1 - \frac{1}{|I_{jk+}|}\right) |c_{jig}|\, \|(A_{ih}\mathbb{1},\, b_i)\| + \right. \\
&\left. + \frac{|c_{jig}|}{|I_{jk+}|}\, \|(M_{ih}\mathbb{1},\, \lambda_i)\| + \frac{|\varepsilon_{jig}|}{|I_{jk+}|}\, \|(A_{ih}\mathbb{1},\, b_i)\| \right] + \sum_{(i,g)\in\mathcal{N}_{jl}} |c_{jig}|\, \|(A_{ih}\mathbb{1},\, b_i)\|
\end{aligned}
$$

όπου το I_{jk+} περιέχει όλους τους δείκτες (i, g) με $c_{jig} > 0$ που αντιστοιχούν στο cluster $(k, g) \in C_{j+}$. Όμοια λαμβάνουμε

$$\mathcal{H}\left(Q_j, \tilde{Q}_j\right) \leq \sum_{(k,g)\in C_{j-}} \sum_{(i,g)\in I_{jk-}} \left[\frac{|\varepsilon_{jig}|}{|I_{jk-}|} \|(M_{ih}\mathbb{1},\, \lambda_i)\| + \left(1 - \frac{1}{|I_{jk-}|}\right) |c_{jig}|\, \|(A_{ih}\mathbb{1},\, b_i)\| + \right.$$
$$\left. + \frac{|c_{jig}|}{|I_{jk-}|} \|(M_{ih}\mathbb{1},\, \lambda_i)\| + \frac{|\varepsilon_{jig}|}{|I_{jk-}|} \|(A_{ih}\mathbb{1},\, b_i)\|\right] + \sum_{(i,g)\in\mathcal{N}_{jl}} |c_{jig}|\, \|(A_{ih}\mathbb{1},\, b_i)\|$$

όπου αντίστοιχα το I_{jk-} περιέχει όλα τα $c_{jig} < 0$ του cluster (k, g). Τα δύο άνω φράγματα με πρόσθεση δίνουν

$$\frac{1}{\rho} \cdot \max_{\mathbf{x}\in\mathcal{B}} |v_{jl}(\mathbf{x}) - \tilde{v}_{jl}(\mathbf{x})| \leq \sum_{(k,g)\in C_j} \sum_{(i,g)\in I_{jk}} \left[\frac{|\varepsilon_{jig}|}{|I_{jk}|} \|(M_{ih}\mathbb{1},\, \lambda_i)\| + \left(1 - \frac{1}{|I_{jk}|}\right) |c_{jig}|\, \|(A_{ih}\mathbb{1},\, b_i)\| + \right.$$
$$\left. + \frac{|c_{jig}|}{|I_{jk}|} \|(M_{ih}\mathbb{1},\, \lambda_i)\| + \frac{|\varepsilon_{jig}|}{|I_{jk}|} \|(A_{ih}\mathbb{1},\, b_i)\|\right] + \sum_{(i,g)\in\mathcal{N}_{jl}} |c_{jig}|\, \|(A_{ih}\mathbb{1},\, b_i)\|$$

Στις παραπάνω σχέσεις χρησιμοποιήσαμε τους συμβολισμούς $C_j = C_{j+} \cup C_{j-}$ και I_{jk} που είναι είτε ίσο με το I_{jk+} είτε με το I_{jk-} ανάλογα με το πρόσημο του κέντρου $(k, g) \in C_j$ ($\tilde{c}_{jkg} > 0$ ή < 0). Σημειώνουμε ότι το άθροισμα $\sum_{(k,g)\in C_j} \sum_{(i,g)\in I_{jk}}$ διατρέχει το πολύ, όλες τις δυνατές τιμές των (i, g), αφού κάθε μη-μηδενικός γεννήτορας αντιστοιχεί σε κάποιο κέντρο συστάδας που έχει ίδιο πρόσημο. Επομένως, έναντι αυτού μπορούμε να χρησιμοποιήσουμε για το άνω φράγμα το άθροισμα $\sum_{i=1}^{n} \sum_{g}$, όπου για το g θεωρούμε ότι λαμβάνει όλες τις πιθανές τιμές του, που αντιστοιχούν σε όλα τα pixel ενός φίλτρου $\mathbf{c}_{ji}$. Με αυτό το σκεπτικό γράφουμε

$$\frac{1}{\rho} \cdot \max_{\mathbf{x}\in\mathcal{B}} |v_{jl}(\mathbf{x}) - \tilde{v}_{jl}(\mathbf{x})| \leq \sum_{i=1}^{n} \sum_{g} \left[\frac{|\varepsilon_{jig}|}{N_{\min}} \|(M_{ih}\mathbb{1},\, \lambda_i)\| + \left(1 - \frac{1}{N_{\max}}\right) |c_{jig}|\, \|(A_{ih}\mathbb{1},\, b_i)\| + \right.$$
$$\left. + \frac{|c_{jig}|}{N_{\min}} \|(M_{ih}\mathbb{1},\, \lambda_i)\| + \frac{|\varepsilon_{jig}|}{N_{\min}} \|(A_{ih}\mathbb{1},\, b_i)\|\right] + \sum_{(i,g)\in\mathcal{N}_{jl}} |c_{jig}|\, \|(A_{ih}\mathbb{1},\, b_i)\|$$

Τονίζουμε ότι χρησιμοποιήσαμε τη σχέση $N_{\max} \geq |I_{jk}| \geq N_{\min}$. Εν συνεχεία, Θα αποδείξουμε μία ανισότητα με την βοήθεια της ανισότητας Cauchy-Schwartz, που θα μας φανεί χρήσιμη. Θεωρούμε $\mathbf{u}_i = \begin{pmatrix} u_{i1} \\ u_{i2} \\ \vdots \\ u_{id} \end{pmatrix}$ για κάθε $i = 1, ..., k$ και $\mathbf{u} = \begin{pmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_k \end{pmatrix}$ να είναι η κάθετη παράθεση των k διανυσμάτων. Τότε ισχύει ότι:

$$\|\mathbf{u}_1 + ... + \mathbf{u}_k\| = \sqrt{\sum_{j=1}^{d} \left(\sum_{i=1}^{k} u_{ij}\right)^2} \leq \sqrt{\sum_{j=1}^{d} k \left(\sum_{i=1}^{k} u_{ij}^2\right)} = \sqrt{k}\|\mathbf{u}\|$$

Η παραπάνω ανισότητα πρακτικά αποδεικνύει ότι η νόρμα του αθροίσματος k διανυσμάτων είναι μικρότερη από την νόρμα του διανύσματος που αποτελείται από όλα τα entries των k διανυσμάτων, επί έναν παράγοντα $\sqrt{k}$. Μπορούμε να χρησιμοποιήσουμε την ανισότητα αυτή για την νόρμα $\|(M_{ih}\mathbb{1},\, \lambda_i)\|$ που αποτελείται από το άθροισμα $d + 1$ διανυσμάτων $M_{ih}\mathbb{1}$ και $(\mathbf{0},\, \lambda_i)$. Το άθροισμα $M_{ih}\mathbb{1}$ προκύπτει από το άθροισμα των διανυσμάτων στηλών του M_{ih} του οποίου τα μη μηδενικά entries είναι τα διανύσματα $\boldsymbol{\mu}_{ik}$ για $k = 1, ..., d$. Αν συμβολίσουμε

με μ_{ikl} τα entries του $\boldsymbol{\mu}_{ik}$ προκύπτει ότι

$$\|(M_{ih}\mathbb{1},\, \lambda_i)\| \leq \sqrt{(d+1)\left(\sum_{k=1}^{d}\sum_{l}\boldsymbol{\mu}_{ikl}^2 + \lambda_i^2\right)} = \sqrt{d+1}\,\left\|\left(\boldsymbol{\mu}_i^T,\, \lambda_i\right)\right\| \leq \sqrt{d+1}\cdot\delta_{\max}$$

Όμοια προκύπτει ότι $\|(A_{ih}\mathbb{1},\, b_i)\| \leq \sqrt{d+1}\,\left\|\left(\mathbf{a}_i^T,\, b_i\right)\right\|$. Με αυτές τις ανισότητες απλοποιούμε περαιτέρω το άνω φράγμα του σφάλματος ως εξής.

$$\begin{aligned}\frac{1}{\rho}\cdot\max_{\mathbf{x}\in\mathcal{B}}|v_{jl}(\mathbf{x})-\tilde{v}_{jl}(\mathbf{x})| \leq \sqrt{d+1}\cdot\sum_{i=1}^{n}\sum_{g}\Bigg[&\frac{\delta_{\max}|\varepsilon_{jig}|}{N_{\min}} + \left(1-\frac{1}{N_{\max}}\right)|c_{jig}|\,\left\|\left(\mathbf{a}_i^T,\, b_i\right)\right\| + \\ &+ \frac{\delta_{\max}|c_{jig}|}{N_{\min}} + \frac{|\varepsilon_{jig}|}{N_{\min}}\left\|\left(\mathbf{a}_i^T,\, b_i\right)\right\|\Bigg] + \sum_{(i,g)\in\mathcal{N}_{jl}}|c_{jig}|\,\left\|\left(\mathbf{a}_i^T,\, b_i\right)\right\|\end{aligned}$$

Θα χρησιμοποιήσουμε, επίσης, την ανισότητα που είχαμε δει και στην απόδειξη της Πρότασης 4.2.

$$\left(\sum_{j=1}^{m}|u_j|\right)^2 \leq m\left(\sum_{j=1}^{m}|u_j|^2\right) \Leftrightarrow \sum_{j=1}^{m}|u_j| \leq \sqrt{m}\;\|(u_1,...,u_m)\| \Rightarrow$$

$$\sum_{j=1}^{m}\sum_{g}|\varepsilon_{jig}| \leq \sqrt{mF_2}\|\boldsymbol{\varepsilon}_i\|, \quad \sum_{j=1}^{m}\sum_{g}|c_{jig}| \leq \sqrt{mF_2}\|\mathbf{c}_i\|$$

Σε συνδυασμό με την σχέση $\|\boldsymbol{\varepsilon}_i\| < \delta_{\max}$ βρίσκουμε, τελικά, ότι

$$\begin{aligned}\frac{1}{\rho}\cdot\sum_{j=1}^{m}\sum_{l}\max_{\mathbf{x}\in\mathcal{B}}|v_{jl}(\mathbf{x})-\tilde{v}_{jl}(\mathbf{x})| &\leq \sqrt{(d+1)mF_2}\cdot\sum_{l}\sum_{i=1}^{n}\Bigg[\frac{\delta_{\max}\|\boldsymbol{\varepsilon}_i\|}{N_{\min}} + \left(1-\frac{1}{N_{\max}}\right)\|\mathbf{c}_i\|\,\left\|\left(\mathbf{a}_i^T,\, b_i\right)\right\| + \\ &+ \frac{\delta_{\max}\|\mathbf{c}_i\|}{N_{\min}} + \frac{\|\boldsymbol{\varepsilon}_i\|}{N_{\min}}\left\|\left(\mathbf{a}_i^T,\, b_i\right)\right\|\Bigg] + \sum_{j,l}\sum_{(i,g)\in\mathcal{N}_{jl}}|c_{jig}|\,\left\|\left(\mathbf{a}_i^T,\, b_i\right)\right\| \\ &\leq N_2\sqrt{(d+1)mF_2}\cdot\Bigg[K\delta_{\max}^2 + \left(1-\frac{1}{N_{\max}}\right)\sum_{i=1}^{n}\|\mathbf{c}_i\|\,\left\|\left(\mathbf{a}_i^T,\, b_i\right)\right\| + \\ &+ \frac{\delta_{\max}}{N_{\min}}\sum_{i=1}^{n}\left(\|\mathbf{c}_i\| + \left\|\left(\mathbf{a}_i^T,\, b_i\right)\right\|\right)\Bigg] + \sum_{j,l}\sum_{(i,g)\in\mathcal{N}_{jl}}|c_{jig}|\,\left\|\left(\mathbf{a}_i^T,\, b_i\right)\right\|\end{aligned}$$

που δίνει το επιθυμητό αποτέλεσμα. □

Η παραπάνω θεωρητική ανάλυση, όπως και στους προηγούμενους αλγορίθμους συμπίεσης βασίζεται στο Θεώρημα προσέγγισης πολυτόπων 2.3. Το συμπέρασμα που εξάγουμε για την απόδοση του αλγορίθμου 4 είναι παρεμφερές με αυτά των αλγορίθμων που παρουσιάστηκαν στην αρχή του Κεφαλαίου. Ο αλγόριθμος δείχνει να έχει καλύτερη επίδοση όσο τα κέντρα του K-means είναι περισσότερα. Το σφάλμα του φράσσεται από έναν παράγοντα ο οποίος εξαρτάται από τα βάρη των συνελικτικών φίλτρων και τις παραμέτρους που προκύπτουν από την εκτέλεση του K-means.

Παρατήρηση. *Σε αυτήν την ενότητα αναλύσαμε την συμπίεση ενός νευρωνικού αποτελούμενο από Conv - ReLU - MaxPooling - Conv επίπεδα (layers). Δεδομένου ότι ένα συνελικτικό layer μπορεί να αναπαρασταθεί ισοδύναμα γραμμικό επίπεδο (fully connected layer) μέσω Toeplitz πίνακα, για την συμπίεση ενός νευρωνικού με Conv - ReLU - MaxPooling*

- Flattening - Linear Layers, μπορούμε να χρησιμοποιήσουμε σχεδόν ίδιο αλγόριθμο με αντίστοιχη θεωρητική ανάλυση σφάλματος. Ωστόσο, η επακριβής αντιμετώπιση αυτού του ζητήματος αφήνεται για μελλοντική δουλειά.

Κεφάλαιο 5

Αριθμητική Συμπίεση Νευρωνικών Δικτύων

Στο προηγούμενο Κεφάλαιο παρουσιάσαμε τεχνικές συμπίεσης Νευρωνικών Δικτύων οι οποίες βασίζονταν στον αλγόριθμο K-means. Παρουσιάσαμε 3 διαφορετικούς αλγορίθμους οι οποίοι συμπιέζαν δίκτυα με ReLU ενεργοποιήσεις μίας ή πολλών εξόδων και εφαρμόζονταν είτε σε γραμμικά είτε σε συνελικτικά επίπεδα. Η ιδέα στην οποία βασίστηκε η συμπίεση σε κάθε περίπτωση ήταν η γεωμετρική αναπαράσταση των ζωνοτόπων της εξόδου με λιγότερους γεννήτορες.

Σε αυτό το Κεφάλαιο θα επανεξετάσουμε την περίπτωση της συμπίεσης των feed-forward ReLU νευρωνικών, χρησιμοποιώντας δύο αλγορίθμους, τον AMM και semi-NMF οι οποίοι είναι αριθμητικής φύσεως. Οι αλγόριθμοι αυτοί θα εφαρμοστούν για την συμπίεση γραμμικών επιπέδων νευρωνικών δικτύων. Ο αλγόριθμος AMM που θα χρησιμοποιήσουμε βασίζεται σε Προσέγγιση Γινομένου Πινάκων η οποία γενικότερα βρίσκει εφαρμογή στην επιτάχυνση και μείωση απαιτήσεων σε μνήμη υπολογιστικών συστημάτων. Θα εξετάσουμε τον αλγόριθμο αυτή ως προς την απόδοσή του με χρήση τροπικής γεωμετρίας και, επιπλέον, θα τον εντάξουμε σε ένα γενικότερο πλαίσιο πιθανοτικών αλγορίθμων PAC Learning, που υποδεικνύει ότι σαν μέθοδος μας επιτρέπει να "μάθουμε" την συνάρτηση εξόδου ενός νευρωνικού. Τέλος, θα παρουσιάσουμε και έναν εναλλακτικό αλγόριθμο semi-NMF που βασίζεται σε μη-αρνητική παραγοντοποίηση πινάκων, του οποίου η θεωρητική μελέτη με τροπική γεωμετρία αφήνεται ως θεωρητική δουλειά.

5.1 Προσέγγιση Γινομένου Πινάκων AMM

Η Προσέγγιση Γινομένου Πινάκων [10] (Approximate Matrix Multiplication ή AMM) αποτελεί έναν πιθανοτικό αλγόριθμο ο οποίος δοθέντων δύο πινάκων A, B βρίσκει δύο πίνακες $\tilde{A}, \tilde{B}$, με μικρότερη κοινή διάσταση, ούτως ώστε $\tilde{A}\tilde{B} \approx AB$. Ο AMM έχει εφαρμογές στην επιτάχυνση και μείωση απαιτήσεων σε μνήμη υπολογιστικών εφαρμογών. Για παράδειγμα στο [30] ο AMM χρησιμοποιείται για την αύξηση στην ταχύτητα υπολογισμού της εξόδου νευρωνικών δικτύων και Distributed Simultaneous Localization and Mapping (SLAM). Εμείς θα χρησιμοποιήσουμε τον αλγόριθμο AMM προσαρμοσμένο στα νευρωνικά δίκτυα, με τρόπο ώστε να προσεγγίζει το γινόμενο των πινάκων που αντιστοιχούν σε δύο διαδοχικά επίπεδα. Ο αλγόριθμος θα μας επιστρέφει τους πίνακες των γραμμικών επιπέδων του συμπιεσμένου δικτύου. Αξίζει να σημειώσουμε ότι και αυτή η μέθοδος συμπίεσης αναφέρεται σε δίκτυο με ένα κρυφό επίπεδο (Σχήμα 3.1) αλλά μπορεί με διαδοχικές εκτελέσεις να εφαρμοστεί για την συμπίεση ολόκληρου του δικτύου.

Algorithm 5 Προσέγγιση Γινομένου Πινάκων (AMM)

1. Δίνεται ως είσοδος A, C οι πίνακες των γραμμικών επιπέδων του πρώτου και δεύτερου επιπέδου του νευρωνικού αντίστοιχα.
2. Υπολογίζουμε την πιθανότητα επιλογής του $i-$οστού νευρώνα του κρυφού επιπέδου $p_i = \frac{\|C_{:,i}\|\|A_{i,:}\|}{\sum_{j=1}^{n}\|C_{:,j}\|\|A_{j,:}\|}$, $\forall i \in \{1, 2, ..., n\}$.
3. Επαναλαμβάνουμε για $t \in \{1, ..., K\}$: επιλέγουμε τυχαία $i \in \{1, 2, ..., n\}$ με πιθανότητα p_i και θέτουμε $\tilde{A}_{(t)} = \frac{A_{i,:}}{\sqrt{Kp_i}}, \tilde{C}^{(t)} = \frac{C_{:,i}}{\sqrt{Kp_i}}$.
4. Το συμπιεσμένο νευρωνικό προκύπτει από γραμμικά επίπεδα με πίνακες $\tilde{A}, \tilde{C}$.

Η μέθοδος συμπίεσης AMM παρουσιάζεται στον Αλγόριθμο 5. Υποθέτουμε ότι έχουμε ένα νευρωνικό δίκτυο με ένα κρυφό επίπεδο, όπως αυτό του σχήματος 3.1. Στο δίκτυο αυτό συμβολίζουμε με A τον πίνακα του πρώτου γραμμικού επιπέδου, με την $i-$οστή γραμμή του A να περιέχει και τον σταθερό όρο $A_{i,:} = \left(\mathbf{a}_i^T, b_i\right)$ και C ο πίνακας του δεύτερου γραμμικού επιπέδου, αγνοώντας τους σταθερούς όρους αφού δεν συμμετέχουν στην ανάλυσή μας. Ως είσοδος του αλγορίθμου εισάγονται οι πίνακες A, C. Ως έξοδος προκύπτου οι πίνακες $\tilde{A}, \tilde{C}$ που αντιστοιχούν στο πρώτο και το δεύτερο πλήρως συνδεδεμένα επίπεδα του τελικού συμπιεσμένου νευρωνικού δικτύου. Οι πίνακες $\tilde{A}, \tilde{C}$ έχουν την ιδιότητα ότι $\tilde{C}\tilde{A} \approx CA$. Συγκρίνοντας την υπολογιστική πολυπλοκότητα του AMM με τους αλγορίθμους συμπίεσης που έκαναν χρήση του K-means, συμπεραίνουμε ότι ο AMM έχει λιγότερες υπολογιστικές απαιτήσεις με αποτέλεσμα να το κάνουν πιο γρήγορο και του δίνουν την δυνατότητα να εφαρμοστεί σε νευρωνικά με μεγάλα γραμμικά επίπεδα.

Ο αλγόριθμος 5 μπορεί διαισθητικά να ερμηνευτεί με δύο τρόπους. Ο πρώτος είναι ότι αγνοούμε το ReLU επίπεδο μεταξύ των δύο γραμμικών επιπέδων του νευρωνικού, οπότε το συνολικό νευρωνικό προκύπτει ως την σύνθεση δύο γραμμικών επιπέδων τα οποία ισοδυναμούν με ένα γραμμικό επίπεδο που έχει ως πίνακα αναπαράστασης το γινόμενο των επιμέρους πινάκων. Συνεπώς, η προσέγγιση του γινομένου είναι ένα εύλογο ζητούμενο. Η δεύτερη ερμηνεία του αλγορίθμου είναι ότι κάνουμε τυχαίο pruning του νευρωνικού δημιουργώντας ένα νέο συμπιεσμένο νευρωνικό με τον εξής τρόπο. Για το νέο νευρωνικό κάνουμε K ανεξάρτητες τυχαίες επιλογές όπου στην $j-$οστή επιλογή επιλέγουμε με πιθανότητα p_i τον $i-$οστό νευρώνα του κρυφού επιπέδου του αρχικού νευρωνικού. Τονίζουμε ότι με αυτόν τον τρόπο μπορεί να προκύψουν στο τελικό νευρωνικό δύο ή παραπάνω ίδιοι νευρώνες, δηλαδή στην τυχαία επιλογή ο νευρώνας i του κρυφού επιπέδου να έχει επιλεγεί μία η παραπάνω φορές (re-sampling).

Εν συνεχεία υπολογίζουμε ένα άνω φράγμα για τη μέση τιμή του σφάλματος του αλγορίθμου AMM, όπως και στους προηγούμενους αλγορίθμους, βασισμένο στο Θεώρημα 2.3.

Πρόταση 5.1. *Ο αλγόριθμος 5 υπολογίζει ένα συμπιεσμένο νευρωνικό με έξοδο $\tilde{v}$ η οποία ικανοποιεί*

$$\frac{1}{\rho^2}\cdot\mathbb{E}\left[\max_{\mathbf{x}\in\mathcal{B}}\|v(\mathbf{x})-\tilde{v}(\mathbf{x})\|^2\right] \leq \frac{2n-2}{K}\left(\sum_{i=1}^{n}\|C_{:,i}\|\|A_{i,:}\|\right)^2 + 2\mathbb{E}\left[\left\|CA-\tilde{C}\tilde{A}\right\|_F^2\right]$$

Απόδειξη. Αρχικά παρατηρούμε ότι

$$\max_{\mathbf{x}\in\mathcal{B}} \|v(\mathbf{x}) - \tilde{v}(\mathbf{x})\|^2 = \max_{\mathbf{x}\in\mathcal{B}} \sum_{j=1}^{m} |v_j(\mathbf{x}) - \tilde{v}_j(\mathbf{x})|^2 \leq \sum_{j=1}^{m} \max_{\mathbf{x}\in\mathcal{B}} |v_j(\mathbf{x}) - \tilde{v}_j(\mathbf{x})|^2$$

Επίσης, από την τριγωνική ανισότητα συνδυασμένη με την απλή ανισότητα $(a+b)^2 \leq 2(a^2+b^2)$, λαμβάνουμε

$$\begin{aligned}
\frac{1}{\rho^2} \cdot \|v_j(\mathbf{x}) - \tilde{v}_j(\mathbf{x})\|^2 &\leq \frac{1}{\rho^2} \cdot \left(|p_j(\mathbf{x}) - \tilde{p}_j(\mathbf{x})| + |q_j(\mathbf{x}) - \tilde{q}_j(\mathbf{x})|\right)^2 \\
&\leq \frac{2}{\rho^2} \cdot \left(|p_j(\mathbf{x}) - \tilde{p}_j(\mathbf{x})|^2 + |q_j(\mathbf{x}) - \tilde{q}_j(\mathbf{x})|^2\right) \\
&\leq 2\left(\mathcal{H}\left(P_j, \tilde{P}_j\right)^2 + \mathcal{H}\left(Q_j, \tilde{Q}_j\right)^2\right)
\end{aligned}$$

Θεωρούμε τις ακόλουθες τυχαίες μεταβλητές.

$$X^i_{jk} = x_i \frac{c_{ji}}{\sqrt{Kp_i}} \frac{A_{ik}}{\sqrt{Kp_i}} = x_i \frac{c_{ji}A_{ik}}{Kp_i},\ \forall i \in [n],\ j \in [m],\ k \in [d+1]$$

όπου x_i είναι ο συνολικός αριθμός που έχει επιλεχθεί ο i-οστός νευρώνας κατά την τυχαία διαδικασία του αλγορίθμου. Σημειώνουμε ότι, εάν ένας νευρώνας έχει επιλεγεί παραπάνω από μία φορά τότε αυτοί οι νευρώνες είναι ταυτόσημοι και δίνουν την ίδια έξοδο σε κάθε σημείο της εισόδου. Επομένως, μπορούν να αναπαρασταθούν από έναν μόνο νευρώνα, του οποίου τα βάρη είναι πολλαπλασιασμένα με x_i.

Η $j-$οστή γραμμή του X^i αντιστοιχεί στον γεννήτορα του ζωνοτόπου της $j-$οστής συνάρτησης εξόδου του συμπιεσμένου νευρωνικού $\tilde{v}_j(\mathbf{x})$ που προέρχεται από τον $i-$οστό κόμβο του κρυφού επιπέδου. Αξίζει να παρατηρήσουμε ότι οι x_i είναι διωνυμικές τυχαίες μεταβλητές $\sim \mathcal{B}(K,\, p_i)$, οπότε μπορούμε να υπολογίσουμε την μέση τιμή και διασπορά των X^i_{jk} ως:

$$\mathbb{E}\left[X^i_{jk}\right] = \frac{c_{ji}A_{ik}}{Kp_i}\mathbb{E}\left[x_i\right] = c_{ji}A_{ik}, \quad \mathrm{Var}\left(X^i_{jk}\right) = \left(\frac{c_{ji}A_{ik}}{Kp_i}\right)^2 \cdot \mathrm{Var}\left(x_i\right) = \frac{c^2_{ji}A^2_{ik}}{Kp_i}(1-p_i)$$

Τα διανύσματα θα έχουν τον ίδιο ρόλο στην ανάλυσή μας όπως και τα κέντρα του K-means στις προηγούμενες αποδείξεις. Δηλαδή, κάθε $X^i_{j,:}$ θα έχει τον ρόλο του γεννήτορα που αναπαριστά ένα σύνολο άλλων γεννητόρων. Συγκεκριμένα, κάθε γεννήτορας $c_{ji}\left(\mathbf{a}_i^T,\, b_i\right)$ θα αναπαρίσταται από τον γεννήτορα $X^i_{j,:}$ του συμπιεσμένου νευρωνικού. Η διαφορά, λοιπόν, σε αυτήν την απόδειξη, λόγω της πιθανοτικής φύσης της, είναι ότι κάθε γεννήτορας θα αντιστοιχεί σε διαφορετικό αντιπροσωπευτικό γεννήτορα (ο οποίος μπορεί να τυχαίνει να είναι $\mathbf{0}$). Με αυτήν την επιλογή λαμβάνουμε:

$$\mathcal{H}\left(P_j, \tilde{P}_j\right)^2 + \mathcal{H}\left(Q_j, \tilde{Q}_j\right)^2 \leq \left\|\sum_{i\in I_+} \left(c_{ji}\left(\mathbf{a}_i^T,\, b_i\right) - X^i_{j,:}\right)\right\|^2 + \left\|\sum_{i\in I_-} \left(c_{ji}\left(\mathbf{a}_i^T,\, b_i\right) - X^i_{j,:}\right)\right\|^2$$

Στην παραπάνω σχέση γραψαμε απευθείας ότι είχαμε αναλύσει σε προηγούμενες αποδείξεις. Για να φράξουμε το $\mathcal{H}\left(P_j, \tilde{P}_j\right)$ πρέπει να φράξουμε τα $\text{dist}\left(\mathbf{u}, \tilde{P}_j\right)$ και $\text{dist}\left(P_j, \mathbf{v}\right)$ για οποιεσδήποτε κορυφές $\mathbf{u} \in P_j$, $\mathbf{v} \in \tilde{P}_j$. Η πρώτη απόσταση φράσσεται επιλέγοντας τους αντιπροσώπους $X^i_{j,:}$ για τους γεννήτορες $c_{ji}\left(\mathbf{a}_i^T,\, b_i\right)$ που κατασκευάζουν την κορυφή $\mathbf{u}$. Η δεύτερη φράσσεται επιλέγοντας για τους γεννήτορες $X^i_{j,:}$ της κορυφής $\mathbf{v}$ τους αντίστοιχους αντιπροσώπους $c_{ji}\left(\mathbf{a}_i^T,\, b_i\right)$, οπότε οι δύο ποσότητες λαμβάνουν το ίδιο φράγμα όπως παρουσιάζεται παραπάνω. Επαναλαμβάνουμε το ίδιο για το $\mathcal{H}\left(Q_j, \tilde{Q}_j\right)$. Σημειώνουμε ότι το γεγονός ότι οι $X^i_{j,:}$ είναι οι γεννήτορες του συμπιεσμένου νευρωνικού δεν είναι προφανές και εξηγείται σε προηγούμενη παράγραφο με την παρατήρηση ότι μπορούμε να αντικαταστήσουμε τους ίδιους νευρώνες που έχουν επιλεγεί πολλές φορές με έναν ενιαίο.

Επιπλέον, χρησιμοποιώντας την ανισότητα

$$\left\|\sum_{i=1}^{n} \mathbf{u}_i\right\|^2 \underset{\text{Triangle Ineq.}}{\leq} \left(\sum_{i=1}^{n} \|\mathbf{u}_i\|\right)^2 \underset{\text{C-S}}{\leq} n\left(\sum_{i=1}^{n} \|\mathbf{u}_i\|^2\right)$$

παίρνουμε ότι

$$\mathcal{H}\left(P_j, \tilde{P}_j\right)^2 + \mathcal{H}\left(Q_j, \tilde{Q}_j\right)^2 \leq n \sum_{i=1}^{n} \left\|c_{ji}\left(\mathbf{a}_i^T,\, b_i\right) - X^i_{j,:}\right\|^2$$

Επιπλέον, χρησιμοποιώντας την γραμμικότητα της μέσης τιμής προκύπτει ότι

$$\begin{aligned}
\frac{1}{\rho^2} \cdot \mathbb{E}\left[\max_{\mathbf{x}\in\mathcal{B}} \|v(\mathbf{x}) - \tilde{v}(\mathbf{x})\|^2\right] &\leq \sum_{j=1}^{m} 2n \sum_{i=1}^{n} \mathbb{E}\left[\left\|c_{ji}\left(\mathbf{a}_i^T,\, b_i\right) - X^i_{j,:}\right\|^2\right] \\
&= 2n \sum_{j=1}^{m}\sum_{i=1}^{n}\sum_{k=1}^{d+1} \mathbb{E}\left[\left\|c_{ji}A_{ik} - X^i_{jk}\right\|^2\right] \\
&= 2n \sum_{j=1}^{m}\sum_{i=1}^{n}\sum_{k=1}^{d+1} \text{Var}\left(X^i_{jk}\right) \\
&= 2n \sum_{i=1}^{n}\left(\sum_{j=1}^{m}\sum_{k=1}^{d+1} \frac{c_{ji}^2 A_{ik}^2}{K p_i}(1-p_i)\right) \\
&= 2n \sum_{i=1}^{n} \frac{\|C_{:,i}\|^2 \|A_{i,:}\|^2}{K p_i}(1-p_i) \\
&= 2n \sum_{i=1}^{n} \frac{\|C_{:,i}\|^2 \|A_{i,:}\|^2}{K p_i} - \frac{2n}{K}\sum_{i=1}^{n} \|C_{:,i}\|^2 \|A_{i,:}\|^2
\end{aligned}$$

Από το ([10], Λήμμα 4) υπολογίζουμε ότι

$$\mathbb{E}\left[\left\|CA - \tilde{C}\tilde{A}\right\|_F^2\right] = \sum_{i=1}^{n} \frac{\|C_{:,i}\|^2 \|A_{i,:}\|^2}{K p_i} - \frac{1}{K}\|CA\|_F^2$$

το οποίο μάλιστα, δίνει την παρακάτω ανισότητα

$$\sum_{i=1}^{n} \frac{\|C_{:,i}\|^2 \|A_{i,:}\|^2}{p_i} \geq \|CA\|_F^2, \, \forall \{p_i\}_{i=1}^n \Rightarrow$$
$$n \sum_{i=1}^{n} \|C_{:,i}\|^2 \|A_{i,:}\|^2 \geq \|CA\|_F^2$$

Η τελευταία ανισότητα προκύπτει αν θεωρήσουμε ομοιόμορφη κατανομή $p_i = \frac{1}{n}, \forall i = \{1, 2, ..., n\}$, δεδομένου ότι η πρώτη ανισότητα ισχύει για οποιαδήποτε κατανομή. Εν συνεχεία βρίσκουμε,

$$\begin{aligned}
\mathbb{E}\left[\max_{\mathbf{x}\in\mathcal{B}} \|v(\mathbf{x}) - \tilde{v}(\mathbf{x})\|^2\right] &\leq 2n \sum_{i=1}^{n} \frac{\|C_{:,i}\|^2 \|A_{i,:}\|^2}{K p_i} - \frac{2n}{K} \sum_{i=1}^{n} \|C_{:,i}\|^2 \|A_{i,:}\|^2 \\
&\leq 2n \sum_{i=1}^{n} \frac{\|C_{:,i}\|^2 \|A_{i,:}\|^2}{K p_i} - \frac{2}{K} \|CA\|_F^2 \\
&= (2n-2) \sum_{i=1}^{n} \frac{\|C_{:,i}\|^2 \|A_{i,:}\|^2}{K p_i} + 2\mathbb{E}\left[\left\|CA - \tilde{C}\tilde{A}\right\|_F^2\right]
\end{aligned}$$

Επιλέγοντας $p_i = \frac{\|C_{:,i}\| \|A_{i,:}\|}{\sum_{j=1}^{n} \|C_{:,j}\| \|A_{j,:}\|}$ προκύπτει

$$\frac{1}{\rho^2} \cdot \mathbb{E}\left[\max_{\mathbf{x}\in\mathcal{B}} \|v(\mathbf{x}) - \tilde{v}(\mathbf{x})\|^2\right] \leq \frac{2n-2}{K} \left(\sum_{i=1}^{n} \|C_{:,i}\| \|A_{i,:}\|\right)^2 + 2\mathbb{E}\left[\left\|CA - \tilde{C}\tilde{A}\right\|_F^2\right]$$

όπως ήταν και ζητούμενο. □

Αξίζει να παρατηρήσουμε ότι το άνω φράγμα που υπολογίσαμε εξαρτάται στο εγγενές σφάλμα που έχει ο αλγόριθμος προσέγγισης γινομένου πινάκων AMM αλλά και σε ένα σφάλμα το οποίο για κάθε τιμή του K είναι μη μηδενικό και εξαρτάται από τη νόρμα των πινάκων των γραμμικών επιπέδων. Ωστόσο, όπως θα δούμε πειραματικά, το σφάλμα αυτό είναι αμελιτέο και το πιο σημαντικό είναι το σφάλμα που προκύπτει από την προσέγγιση του γινομένου πινάκων.

5.1.1 Μελέτη AMM με PAC-Learning

Ο αλγόριθμος AMM όπως αποδείξαμε προσφέρει έναν εναλλακτικό τρόπο συμπίεσης ενός feed-forward νευρωνικού δικτύου. Στο σημείο αυτό θα επεκτείνουμε θεωρητικά την συμβολή του AMM στην μελέτη των νευρωνικών υπό το πρίσμα του Probably Approximately Correct Learning (PAC-Learning) [42, 17, 18, 44]. Για τον σκοπό αυτής της μελέτης, υποθέτουμε ότι έχουμε ένα feed-forward νευρωνικό με ένα κρυφό επίπεδο για το οποίο δεν γνωρίζουμε τα βάρη του και το θεωρούμε σαν black-box. Η μόνη διαδικασία που επιτρέπεται να εφαρμόσουμε είναι αυτή του AMM. Δηλαδή, επιτρέπεται να κάνουμε sample K νευρώνες από το κρυφό επίπεδο μαζί με τα βάρη τους, ούτως ώστε να κατασκευάσουμε ένα νευρωνικό που προσεγγίζει ικανοποιητικά το black-box δοσμένο νευρωνικό. Όπως αποδεικνύουμε με το παρακάτω Θεώρημα, η δειγματοληπτική διαδικασία αυτή είναι αποτελεσματική στην εκμάθηση του black-box νευρωνικού από το προσεγγιστικό. Δηλαδή η κλάση αυτή των νευρωνικών χαρακτηρίζεται PAC-learnable.

Θεώρημα 5.1. *Η κλάση των νευρωνικών δικτύων με ένα κρυφό επίπεδο και ReLU ενεργοποιήσεις είναι PAC-learnable.*

Απόδειξη. Έστω $v(\mathbf{x}) = (v_1(\mathbf{x}), ..., v_m(\mathbf{x}))$ η συνάρτηση εξόδου του νευρωνικού. Δειγματοληπτούμε με επανάληψη K νευρώνες από το κρυφό επίπεδο, όπου ο $i-$οστός νευρώνας επιλέγεται με πιθανότητα $p_i = \frac{\|C_{:,i}\|\|A_{i,:}\|}{\sum_{i=1}^{n}\|C_{:,i}\|\|A_{i,:}\|}$. Με αυτόν τον τρόπο από το black-box νευρωνικό προκύπτει ένα sampled νευρωνικό με συνάρτηση εξόδου $\tilde{v}(\mathbf{x}) = (\tilde{v}_1(\mathbf{x}), ..., \tilde{v}_m(\mathbf{x}))$. Για να είναι η κλάση των νευρωνικών με ένα κρυφό επίπεδο PAC-learnable πρέπει να αποδείξουμε ότι η πιθανότητα η $\tilde{v}(\mathbf{x})$ να αποκλίνει από την $v(\mathbf{x})$ παραπάνω από $\varepsilon > 0$ να είναι πολύ μικρή.

Έχουμε ότι

$$\begin{aligned}
\mathbb{P}\left(\|v(\mathbf{x}) - \tilde{v}(\mathbf{x})\| > \varepsilon\right) &= \mathbb{P}\left(\sum_{j=1}^{m} |v_j(\mathbf{x}) - \tilde{v}_j(\mathbf{x})|^2 > \varepsilon^2\right) \\
&\leq \mathbb{P}\left(\bigcup_{j=1}^{m}\left\{|v_j(\mathbf{x}) - \tilde{v}_j(\mathbf{x})| > \frac{\varepsilon}{\sqrt{m}}\right\}\right) \\
&\leq \sum_{j=1}^{m} \mathbb{P}\left(|v_j(\mathbf{x}) - \tilde{v}_j(\mathbf{x})| > \frac{\varepsilon}{\sqrt{m}}\right)
\end{aligned}$$

Για έναν κόμβο εξόδου υπολογίζουμε ένα άνω φράγμα στην πιθανότητα ως εξής. Θα χρησιμοποιήσουμε το Θεώρημα 2.3 για αυτόν τον σκοπό.

$$\begin{aligned}
\mathbb{P}\left(|v_j(\mathbf{x}) - \tilde{v}_j(\mathbf{x})| > \frac{\varepsilon}{\sqrt{m}}\right) &\leq \mathbb{P}\left(|p_j(\mathbf{x}) - \tilde{p}_j(\mathbf{x})| + |q_j(\mathbf{x}) - \tilde{q}_j(\mathbf{x})| > \frac{\varepsilon}{\sqrt{m}}\right) \\
&\leq \mathbb{P}\left(\mathcal{H}\left(P_j, \tilde{P}_j\right) + \mathcal{H}\left(Q_j, \tilde{Q}_j\right) > \frac{\varepsilon}{\rho\sqrt{m}}\right)
\end{aligned}$$

Όμως ισχύει

$$\mathcal{H}\left(P_j, \tilde{P}_j\right) + \mathcal{H}\left(Q_j, \tilde{Q}_j\right) \leq \sum_{i=1}^{n} \left\|c_{ji}\left(\mathbf{a}_i^T, b_i\right) - X_{j,:}^{i}\right\|$$

για λόγους που αναλύσαμε στην προηγούμενη απόδειξη. Συνοπτικά, αναφέρουμε ότι ο

$X^i_{j,:} = x_i \frac{c_{ji}\left(\mathbf{a}_i^T, b_i\right)}{Kp_i}$ είναι ο γεννήτορας του συμπιεσμένου νευρωνικού που αντιπροσωπεύει τον γεννήτορα $c_{ji}\left(\mathbf{a}_i^T,\, b_i\right)$ του αρχικού νευρωνικού, με x_i να είναι ο αριθμός των φορών που έχει επιλεχθεί ο νευρώνας i του κρυφού επιπέδου κατά την τυχαία διαδικασία.

Κατα συνέπεια προκύπτει ότι

$$\begin{aligned}
\mathbb{P}\left(|v_j(\mathbf{x}) - \tilde{v}_j(\mathbf{x})| > \frac{\varepsilon}{\sqrt{m}}\right) &\leq \mathbb{P}\left(\mathcal{H}\left(P_j, \tilde{P}_j\right) + \mathcal{H}\left(Q_j, \tilde{Q}_j\right) > \frac{\varepsilon}{\rho\sqrt{m}}\right) \\
&\leq \mathbb{P}\left(\sum_{i=1}^{n} \left\| c_{ji}\left(\mathbf{a}_i^T,\, b_i\right) - X^i_{j,:}\right\| > \frac{\varepsilon}{\rho\sqrt{m}}\right) \\
&\leq \sum_{i=1}^{n} \mathbb{P}\left(\left\| c_{ji}\left(\mathbf{a}_i^T,\, b_i\right) - X^i_{j,:}\right\| > \frac{\varepsilon}{\rho n\sqrt{m}}\right) \\
&\leq \sum_{i=1}^{n} \mathbb{P}\left(\sum_{k=1}^{d+1} \left| c_{ji}A_{ik} - X^i_{jk}\right| > \frac{\varepsilon}{\rho n\sqrt{m}}\right) \\
&\leq \sum_{i=1}^{n} \mathbb{P}\left(\sum_{k=1}^{d+1} |c_{ji}A_{ik}| \left|1 - \frac{x_i}{Kp_i}\right| > \frac{\varepsilon}{\rho n\sqrt{m}}\right) \\
&\leq \sum_{i=1}^{n} \mathbb{P}\left(\left|1 - \frac{x_i}{Kp_i}\right| > \frac{\varepsilon}{\rho n\sqrt{m}|c_{ji}|\| \left(\mathbf{a}_i^T,\, b_i\right) \|_1}\right)
\end{aligned}$$

Χρησιμοποιώντας Chernoff-Hoeffding για $\frac{x_i}{p_i} = Y_1 + ... + Y_K = Y$, όπου Y_i τυχαίες μεταβλητές Bernoulli προκύπτει

$$\begin{aligned}
\mathbb{P}\left(|v_j(\mathbf{x}) - \tilde{v}_j(\mathbf{x})| > \frac{\varepsilon}{\sqrt{m}}\right) &\leq \sum_{i=1}^{n} \mathbb{P}\left(\left|1 - \frac{x_i}{Kp_i}\right| > \frac{\varepsilon}{\rho n\sqrt{m}|c_{ji}|\| \left(\mathbf{a}_i^T,\, b_i\right) \|_1}\right) \\
&\leq \sum_{i=1}^{n} \mathbb{P}\left(\left|\bar{Y} - \mathbb{E}\left[\bar{Y}\right]\right| > \frac{\varepsilon}{\rho n\sqrt{m}|c_{ji}|\| \left(\mathbf{a}_i^T,\, b_i\right) \|_1}\right) \\
&\leq \sum_{i=1}^{n} 2 \cdot \exp\left(-2K^2 \frac{\varepsilon^2}{\rho^2 n^2 m|c_{ji}|^2\| \left(\mathbf{a}_i^T,\, b_i\right) \|_1^2 \frac{K}{p_i^2}}\right) \\
&= \sum_{i=1}^{n} 2 \cdot \exp\left(\frac{-2Kp_i^2\varepsilon^2}{\rho^2 n^2 m|c_{ji}|^2\| \left(\mathbf{a}_i^T,\, b_i\right) \|_1^2}\right) \\
&= 2 \cdot \exp\left(\frac{-2K\varepsilon^2}{\rho^2 n^2 m} \sum_{i=1}^{n} \frac{p_i^2}{|c_{ji}|^2\| \left(\mathbf{a}_i^T,\, b_i\right) \|_1^2}\right) \sim 2e^{-\omega\varepsilon^2}
\end{aligned}$$

όπου με ω αντικαθιστούμε την ποσότητα $\frac{2K}{\rho^2 n^2 m} \sum_{i=1}^{n} \frac{p_i^2}{|c_{ji}|^2\|\left(\mathbf{a}_i^T, b_i\right)\|_1^2}$. Τελικά, καταλήγουμε στην σχέση

$$\mathbb{P}\left(\|v(\mathbf{x}) - \tilde{v}(\mathbf{x})\| > \varepsilon\right) \leq 2me^{-\omega\varepsilon^2}$$

που αποδεικνύει τον ισχυρισμό μας, διότι εάν επιβάλλουμε $2me^{-\omega\varepsilon^2} < \delta$ τότε προκύπτει ότι

$$K \geq \frac{1}{\varepsilon^2} \ln\left(\frac{2m}{\delta}\right) \frac{\rho^2 n^2 m}{\sum_{i=1}^{n} \frac{p_i^2}{|c_{ji}|^2\|\left(\mathbf{a}_i^T, b_i\right)\|_1^2}}$$

δηλαδή το K αρκεί να είναι πολυωνυμικό ως προς $\frac{1}{\varepsilon}$, $\frac{1}{\delta}$ για να δώσει το επιθυμητό φράγμα στην πιθανότητα. □

5.2 Μη-αρνητική Παραγοντοποίηση Πίνακα semi-NMF

Στην ενότητα αυτή θα επιχειρήσουμε να χρησιμοποιήσουμε έναν αριθμητικό αλγόριθμο ο οποίος αφορά την παραγοντοποίηση πινάκων για την συμπίεση γραμμικών επιπέδων νευρωνικού δικτύου. Ο αλγόριθμος που θα χρησιμοποιήσουμε ονομάζεται Μη-αρνητική Παραγοντοποίηση Πίνακα (Non-negative Matrix Factorization ή NMF) και δοθέντος ενός πίνακα X βρίσκει πίνακες F, G με μή-αρνητικά στοιχεία, ούτως ώστε $X = FG^T$. Θα εκμεταλλευτούμε το γεγονός ότι μπορούμε να επιλέξουμε την κοινή διάσταση των πινάκων F, G ώστε να μειώσουμε την διάσταση των πινάκων των γραμμικών επιπέδων. Ωστόσο, μία τέτοια παραγοντοποίηση δεν είναι εφικτή σε οποιονδήποτε πίνακα X. Επίσης, για τους πίνακες των γραμμικών επιπέδων του νευρωνικού δικτύου η μη-αρνητικότητα είναι αρκετά περιοριστική και γι' αυτό θα προτιμήσουμε να εφαρμόσουμε μία εναλλακτική εκδοχή του αλγορίθμου που ονομάζεται semi-NMF [9]. Σε αυτήν την εκδοχή δεν υπάρχει περιορισμός για τα πρόσημα των στοιχείων του πίνακα F. Αξίζει να σημειώσουμε πως στην πράξη συνήθως χρησιμοποιείται προσεγγιστική εκδοχή του αλγορίθμου, δηλαδή τα F, G επιλέγονται ώστε $X \approx FG^T$, ενώ η ποιότητα της παραγοντοποίησης χαρακτηρίζεται από κάποια συνάρτηση σφάλματος. Μάλιστα, όταν η συνάρτηση σφάλματος είναι η νόρμα Frobenius $\|X - FG^T\|^2$, τότε ο semi-NMF είναι ισοδύναμος με συσταδοποίηση K-means [9].

Η διαδικασία που θα ακολουθήσουμε για την προσάρτηση του semi-NMF στην διαδικασία συμπίεσης του νευρωνικού δικτύου είναι η εξής. Υποθέτουμε ότι εργαζόμαστε με το δίκτυο του σχήματος 3.1. Θεωρούμε $\mathbf{x}_1, ..., \mathbf{x}_D$ όλα τα διανύσματα εισόδου του δικτύου. Για κάθε δείγμα εισόδου $\mathbf{x}_l$ η έξοδος του πρώτου γραμμικού επιπέδου έπειτα από την ενεργοποίηση ReLU είναι

$$\mathbf{f}_l = \begin{pmatrix} f_1(\mathbf{x_l}) \\ f_2(\mathbf{x_l}) \\ \vdots \\ f_n(\mathbf{x_l}) \end{pmatrix} = \begin{pmatrix} \max\{\mathbf{a}_1^T\mathbf{x}_l + b_1,\, 0\} \\ \max\{\mathbf{a}_2^T\mathbf{x}_l + b_2,\, 0\} \\ \vdots \\ \max\{\mathbf{a}_n^T\mathbf{x}_l + b_n,\, 0\} \end{pmatrix}$$

Αν γράψουμε τα D διανύσματα αυτά σε έναν πίνακα $F = \begin{bmatrix} \mathbf{f}_1 & \mathbf{f}_2 & ... & \mathbf{f}_D \end{bmatrix}$ τότε οι πιθανές έξοδοι του δικτύου για κάθε δείγμα εισόδου είναι οι στήλες του πίνακα CF.

Ο semi-NMF υπεισέρχεται στον αλγόριθμο συμπίεσης ως εξής. Επιθυμούμε να βρούμε πίνακες $\tilde{C}, \tilde{F}$ κοινής διαστάσεων $m \times K$ και $K \times D$ με $K < n$ ώστε $\tilde{C}\tilde{F} \approx CF$. Επομένως, εάν καταφέρουμε να υπολογίσουμε πίνακα $\tilde{A}$ διαστάσεων $K \times d$ ώστε $\tilde{\mathbf{a}}_i^T\mathbf{x}_l + b_i = \tilde{F}_{il}$ για κάθε $i = 1, ..., K$ και δείγμα $l = 1, ..., D$, τότε πράγματι επιτύχαμε συμπίεση του δικτύου. Με λιγότερους νευρώνες K στο κρυφό επίπεδο, το συμπιεσμένο δίκτυο επιτυγχάνει προσέγγιση της απόκρισης του αρχικού για κάθε δείγμα εισόδου στο σύνολο εκπαίδευσης. Αναμένουμε, λοιπόν, το ίδιο να συμβαίνει και στο σύνολο εξέτασης.

Από τον πίνακα $\tilde{F} = \begin{bmatrix} \tilde{\mathbf{f}}_1 & \tilde{\mathbf{f}}_2 & ... & \tilde{\mathbf{f}}_D \end{bmatrix}$ θα υπολογίσουμε τον πίνακα A ως εξής. Πρέπει να ισχύει

$$\forall\, i \in [K],\, l \in [D] : \begin{cases} \tilde{\mathbf{a}}_i^T\mathbf{x}_l + b_i \le 0, & \tilde{F}_{il} < 0 \\ \tilde{\mathbf{a}}_i^T\mathbf{x}_l + b_i = \tilde{F}_{il}, & \tilde{F}_{il} \ge 0 \end{cases}$$

Το σύνολο των παραπάνω εξισώσεων δεν είναι απαραίτητο ότι έχει λύση και γι' αυτό επιθυμούμε να βρούμε μία προσεγγιστική, η οποία θα ελαχιστοποιεί το τετραγωνικό σφάλμα

$$\sum_{i\in[K]}\sum_{l\in[D]}\left|\tilde{F}_{il}-\left(\tilde{\mathbf{a}}_i^T\mathbf{x}_l+b_i\right)\right|^2=\sum_{i\in[K]}\left|\tilde{F}_{i,:}-\left(\tilde{\mathbf{a}}_i^T,\,b_i\right)X\right|^2=\|F-AX\|^2$$

όπου $X=\begin{bmatrix}\mathbf{x}_1 & \mathbf{x}_2 & ... & \mathbf{x}_d\\ 1 & 1 & ... & 1\end{bmatrix}$. Για την υλοποίησή μας επιλέγουμε την λύση μέσω του ψευδοαντιστρόφου

$$\tilde{A}^T=X^{T\dagger}F=(XX^T)^{-1}X\tilde{F}$$

Ο αλγόριθμος semi-NMF 6 για την συμπίεση του δικτύου με ένα κρυφό επίπεδο τελικά συνοψίζεται ως εξής.

Algorithm 6 Μη-αρνητική Παραγοντοποίηση Πίνακα (semi-NMF)

1. Δίνεται ως είσοδος A, C οι πίνακες των γραμμικών επιπέδων του δικτύου.
2. Για κάθε δείγμα εισόδου $\mathbf{x}_l$, $l=1,...,D$ υπολογίζουμε την έξοδο του κρυφού επιπέδου $\mathbf{f}_l$ και σχηματίζουμε τον πίνακα F με στήλες $\mathbf{f}_l$, $l=1,...,D$.
3. Εφαρμόζουμε τον αλγόριθμο semi-NMF για τον πίνακα CF και λαμβάνουμε τους πίνακες $\tilde{C}$, $\tilde{F}$ με διαστάσεις $m\times K$ και $K\times D$ αντίστοιχα.
4. Υπολογίζουμε $\tilde{A}=F^TX^T(X^TX)^{-1}$, όπου $X=\begin{bmatrix}\mathbf{x}_1 & \mathbf{x}_2 & ... & \mathbf{x}_d\\ 1 & 1 & ... & 1\end{bmatrix}$.
5. Το συμπιεσμένο νευρωνικό προκύπτει από γραμμικά επίπεδα με πίνακες $\tilde{A}, \tilde{C}$.

Ο αλγόριθμος αυτός αν και σχετίζεται με τον AMM, δεν έχει ως τώρα κάποια ανάλυση μέσω τροπικής γεωμετρίας. Ως εκ τούτου, η περαιτέρω μελέτη αυτού αφήνεται ως μελλοντική εργασία.

Κεφάλαιο 6

Πειραματικά Αποτελέσματα Συμπίεσης Νευρωνικών Δικτύων

Στην ενότητα αυτή θα εξετάσουμε πειραματικά την απόδοση των αλγορίθμων που αναλύσαμε. Οι αλγόριθμοι Zonotope K-means, Neural Path K-means, AMM και semi-NMF αφορούν συμπίεση γραμμικών επιπέδων, ενώ ο Convolutional Neural Path K-means αφορά συμπίεση συνελικτικών επιπέδων. Ο κάθε αλγόριθμος θα εκτελεστεί σε συνελικτικά νευρωνικά δίκτυα συμπιέζοντας είτε τα γραμμικά επίπεδα του δικτύου, που βρίσκονται προς την έξοδο, είτε τα συνελικτικά που βρίσκονται πιο κοντά στην είσοδο, ανάλογα με τον είδος συμπίεσης που εφαρμόζει.

Η επίδοση των αλγορίθμων θα αξιολογηθεί συγκριτικά με άλλες μεθόδους ελαχιστοποίησης νευρωνικών δικτύων. Οι αλγόριθμοι που θα χρησιμοποιηθούν για σύγκριση θα επιλεχθούν ώστε να είναι ομοειδείς και η σύγκριση να έχει υπόσταση και θεωρητική σημασία. Στο σημείο αυτό αξίζει να περιγράψουμε εν συντομία το είδος συμπίεσης που ακολουθούμε. Συγκεκριμένα, οι αλγόριθμοι μας στο επίπεδο που συμπιέζουν διαλέγουν ένα υποσύνολο νευρώνων (ή καναλιών) τα οποία αφαιρούν από το δίκτυο, ενώ στα εναπομείναντα τροποποιούν τα βάρη των συνδέσεών τους. Τέτοιοι αλγορίθμοι συμπίεσης που αφαιρούν ολόκληρους νευρώνες ονομάζονται δομημένοι (structured). Η συμπίεση πραγματοποιείται μία φορά στο εκπαιδευμένο δίκτυο και σε όλο του το βάθος, δηλαδή σε όλα τα γραμμικά ή συνελικτικά επίπεδα, εφαρμόζοντας τον ίδιο αλγόριθμο διαδοχικά, με κατεύθυνση από την είσοδο προς την έξοδο. Τέλος, αξίζει να σημειώσουμε ότι όλοι οι αλγόριθμοι μας αποδίδουν ένα συμπιεσμένο νευρωνικό δίκτυο το οποίο προσεγγίζει συναρτησιακά το αρχικό, δηλαδή ισχύει $\tilde{v}(\mathbf{x}) \approx v(\mathbf{x})$, $\forall \mathbf{x} \in \mathcal{B}$ όπου $v, \tilde{v}$ οι συναρτήσεις εξόδου του αρχικού και του τελικού δικτύου αντίστοιχα και $\mathcal{B}$ οι υπερσφαίρα που περιέχει όλα τα στιγμιότυπα εισόδου του δικτύου. Το γεγονός αυτό κάνει πιο ισχυρή την προσέγγιση μας δίνοντας την δυνατότητα να εφαρμοστεί όχι μόνο σε προβλήματα ταξινόμησης (classification tasks), αλλά και σε προβλήματα παλινδρόμησης (regression tasks).

Σύμφωνα με τα προαναφερθέντα, οι αλγόριθμοι που θα χρησιμοποιήσουμε για σύγκριση θα είναι οι εξής. Αρχικά, επιλέγουμε τις μεθόδους [38, 37] οι οποίες βασίζονται και αυτές σε τροπική γεωμετρία και την καινοτόμα οπτική της τροπικής διαίρεσης προσεγγίζοντας νευρωνικά δίκτυα μίας και πολλών εξόδων αντίστοιχα. Η ιδέα στην οποία βασίζονται είναι η επιλογή των κορυφών στο ζωνότοπο του δικτύου, οι οποίες είναι πιο συχνά "ενεργοποιημένες" από τις εισόδους του δικτύου. Πέρα των τροπικών μεθόδων θα συγκρίνουμε τις μεθόδους Neural Path K-means, AMM και semi-NMF για νευρωνικά πολλών εξόδων με τις βασικές μεθόδους συμπίεσης (baseline pruning methods) Random και $L1$ στην structured εκδοχή τους. Η Random μέθοδος αφαιρεί νευρώνες από το κρυφό επίπεδο με ομοιόμορφη

πιθανότητα, ενώ η L1 αφαιρεί το ποσοστό των νευρώνων με την μικρότερη $L1$ νόρμα των βαρών που τους αντιστοιχούν. Τέλος, επεκτείνουμε τις συγκρίσεις μας εφαρμόζοντας την γνωστή και σπουδαία μέθοδο ThiNet [22] η οποία εφαρμόζεται σε συνελικτικά επίπεδα, αλλά θα την προσαρμόσουμε και για σύγκριση σε γραμμικά.

Συνολικά θα πραγματοποιήσουμε 4 πειράματα. Το πρώτο θα χρησιμοποιεί τις μεθόδους Zonotope K-means, Neural Path K-means και AMM για γραμμικά επίπεδα, εφαρμόζοντας συμπίεση σε δίκτυο εκπαιδευμένο για binary classification task και θα γίνει σύγκριση με την μέθοδο [38]. Το δεύτερο πείραμα θα πραγματοποιηθεί σε multiclass classification task, οπότε μόνο οι αλγόριθμοι Neural Path K-means, AMM και semi-NMF θα χρησιμοποιηθούν. Η σύγκριση σε αυτήν την περίπτωση θα αφορά την τροπική μέθοδο [37], τις baseline pruning μεθόδους Random και L1 καθώς και την τροποποιημένη έκδοση της μεθόδου ThiNet. Στο τρίτο πείραμα θα επεκταθούμε σε συμπίεση μεγάλων νευρωνικών δικτύων στο CIFAR dataset, όπου θα εφαρμόσουμε τις μεθόδους Neural Path K-means και AMM σε σύγκριση με τις Random και L1. Τέλος, το τέταρτο πείραμα θα αφορά αποκλειστικά την συμπίεση συνελικτικών επιπέδων για δίκτυα εκπαιδευμένα σε multiclass datasets. Η μέθοδος που θα χρησιμοποιήσουμε είναι η Convolutional Neural Path K-means και θα συγκριθεί με τις ThiNet, Random και L1.

Αξίζει να σημειώσουμε ότι τα πειράματά μας ακολουθούν proof-of-concept λογική. Πιο συγκεκριμένα, επιχειρούμε να αναδείξουμε ότι η θεωρητική μας δουλειά έχει πρακτική εφαρμογή στον τομέα της συμπίεσης νευρωνικών δικτύων. Για σύγκριση με μοντέρνες state-of-the-art μεθόδους απαιτείται περαιτέρω έρευνα και χρήση εξειδικευμένων τεχνικών. Για παράδειγμα οι σύγχρονες μέθοδοι εφαρμόζουν "fine-tuning", δηλαδή εκπαίδευση για λίγες εποχές έπειτα από την συμπίεση για την ανάκαμψη της επίδοσης του δικτύου, ενώ άλλες συμπιέζουν "on the fly", καθώς δηλαδή το δίκτυο εκπαιδεύεται. Θεωρούμε πως η κύρια συμβολή της εργασίας μας έγκειται στο θεωρητικό τμήμα, όπου παρουσιάζουμε μία καινοτόμο τροπική γεωμετρική προσέγγιση νευρωνικών δικτύων τόσο για γραμμικά όσο και για συνελικτικά επίπεδα.

6.1 Συμπίεση Δικτύων μίας εξόδου

Το πρώτο πείραμα που θα πραγματοποιήσουμε εφαρμόζεται στα binary classification tasks των ζευγών 3/5 και 4/9, του συνόλου δεδομένων χειρόγραφων ψηφίων MNIST[1]. Το συγκεκριμένο classification αφορά την εκπαίδευση νευρωνικών δικτύων για τον διαχωρισμό του ψηφίου 3 από το 5 και του 4 από το 9. Σε αυτό το πείραμα θα χρησιμοποιήσουμε όλες μας τις μεθόδους αφού το output του δικτύου θα έχει έναν κόμβο, οπότε οι μέθοδοι θα χρησιμοποιηθούν για να συμπιέσουν τα δύο τελευταία γραμμικά επίπεδα.

Στους Πίνακες 6.1, 6.2 παρουσιάζουμε τα αποτελέσματα της σύγκρισης των μεθόδων μας με την τροπική μέθοδο [38]. Για την εκτέλεση του πειράματος χρησιμοποιούμε το ίδιο νευρωνικό δίκτυο που παρουσιάζεται στο [38], το οποίο ονομάζουμε CNN2D και διαθέτει δύο συνελικτικά και δύο γραμμικά επίπεδα. Τα γραμμικά επίπεδα, που συμπιέζουμε στην προκειμένη περίπτωση, έχουν 1000 κόμβους στο κρυφό επίπεδο. Στους πίνακες καταγράφονται τα ποσοστά accuracy του δικτύου για τα διάφορα ποσοστά συμπίεσης.

Πίνακας 6.1: Σύγκριση Zonotope K-means, Neural Path K-means και AMM με την τροπική μέθοδο [38] στο task MNIST 3/5.

Ποσοστό εναπομείναντων νευρώνων	MNIST 3/5			
	Smyrnis et al. [38]	Zonotope K-means	Neural Path K-means	AMM
100% (Original)	99.18 ± 0.27	99.47 ± 0.14	99.47 ± 0.14	99.47 ± 0.14
50%	99.11 ± 0.44	99.45 ± 0.18	99.47 ± 0.12	99.54 ± 0.08
25%	99.12 ± 0.37	99.39 ± 0.16	99.45 ± 0.14	99.56 ± 0.08
10%	99.11 ± 0.36	99.38 ± 0.18	99.46 ± 0.14	99.46 ± 0.16
1%	99.18 ± 0.36	99.41 ± 0.19	99.46 ± 0.16	99.24 ± 0.11

Πίνακας 6.2: Σύγκριση Zonotope K-means, Neural Path K-means και AMM με την τροπική μέθοδο [38] στο task MNIST 4/9.

Ποσοστό εναπομείναντων νευρώνων	MNIST 4/9			
	Smyrnis et al. [38]	Zonotope K-means	Neural Path K-means	AMM
100% (Original)	99.05 ± 0.27	99.57 ± 0.09	99.57 ± 0.09	99.57 ± 0.09
50%	99.05 ± 0.28	99.51 ± 0.09	99.56 ± 0.07	99.55 ± 0.07
25%	98.99 ± 0.34	99.50 ± 0.09	99.55 ± 0.09	99.56 ± 0.10
10%	99.01 ± 0.31	99.54 ± 0.06	99.55 ± 0.08	99.46 ± 0.09
1%	98.81 ± 0.37	99.40 ± 0.31	99.47 ± 0.14	99.28 ± 0.22

[1] Διάκριση των χειρόγραφων ψηφίων 3 και 5 ή των 4 και 9

Πίνακας 6.4: Πειραματικός υπολογισμός θεωρητικών άνω φραγμάτων για Zonotope K-means, Neural Path K-means και AMM στο σύνολο δεδομένων MNIST 4/9.

Ποσοστό εναπομείναντων νευρώνων	MNIST 4/9		
	Zonotope K-means	Neural Path K-means Bound	AMM Bound
100%	0.00	0.00	2054.57
10%	18.85	229.37	20545.66
5%	17.72	101.48	41091.31
2.5%	17.02	63.37	82182.62
1%	16.44	45.58	205456.55
0.5%	16.20	39.71	410913.10

Πίνακας 6.3: Πειραματικός υπολογισμός θεωρητικών άνω φραγμάτων των σφαλμάτων για Zonotope K-means, Neural Path K-means και AMM στο σύνολο δεδομένων MNIST 3/5.

Ποσοστό εναπομείναντων νευρώνων	MNIST 3/5		
	Zonotope K-means	Neural Path K-means Bound	AMM Bound
100%	0.00	0.00	1652.77
10%	17.07	246.74	16527.75
5%	16.03	99.89	33055.50
2.5%	15.35	59.42	66111.00
1%	14.79	42.22	165277.50
0.5%	14.57	36.47	330555.00

Παρατηρούμε ότι οι μέθοδοι μας επιτυγχάνουν αντίστοιχη επίδοση με την τροπική μέθοδο [38]. Πράγματι, επιτυγχάνουν διατήρηση του ποσοστού ακρίβειας (accuracy), ακόμα και όταν μειώνεται το ποσοστό των νευρώνων του κρυφού επιπέδου στο 1%.

Στους Πίνακες 6.3, 6.4 παρουσιάζονται τα θεωρητικά άνω φράγματα των προτάσεων 4.1, 4.2 και 5.1 για τα διάφορα ποσοστά συμπίεσης του δικτύου CNN2D στα προαναφερθέντα σύνολα δεδομένων. Παρατηρούμε, ότι τα αποτελέσματα είναι εν μέρει αναμενόμενα. Αρχικά, για πλήρη 100% διατήρηση νευρώνων έχουμε μηδενικό άνω φράγμα, αφού με $K = n$ κέντρα τα θεωρητικά άνω φράγματα 2.6, 4.2 μηδενίζονται και ο K-means αποδίδει το αρχικό δίκτυο. Ο λόγος που τα άνω φράγματα μηδενίζονται είναι επειδή η απόσταση $\delta_{\max}$ μηδενίζεται, το μέγιστο πλήθος συστάδας $N_{\max}$ γίνεται 1 και το σύνολο των μηδενικών γεννητόρων γίνεται κενό, αφού κάθε διάνυσμα στον K-means γίνεται αποτελεί και από ένα κέντρο. Επίσης, για τον AMM παρατηρούμε ότι το άνω φράγμα αυξάνεται όσο αυξάνεται και η συμπίεση. Αυτό είναι αναμενόμενο διότι όσο μεγαλύτερη η συμπίεση, τόσο χειρότερη αναμένεται να είναι και η προσέγγιση. Αντίθετα αυτό δεν συμβαίνει στις περιπτώσεις των K-means αλγορίθμων. Πειραματικά αποδεικνύεται ότι αυτό οφείλεται στην τάξη μεγέθους των διαφόρων ποσοτήτων, αφού ο όρος που περιέχει το K είναι συγκριτικά μεγαλύτερος από τους υπόλοιπους. Συγκεκριμένα, όσο αυξάνουμε την συμπίεση, ο αριθμός των κέντρων K μειώνεται, τα $N_{\max}, N_{\min}$ αυξάνονται και το $\delta_{\max}$ αυξάνεται, εξακολουθώντας να είναι μικρό σε σχέση με το K. Με δεδομένο ότι τα βάρη του αρχικού δικτύου είναι σταθερές ποσότητες, το άνω φράγμα πράγματι μικραίνει.

6.2 Συμπίεση Δικτύων πολλών εξόδων

Στο δεύτερο πείραμα ασχολούμαστε με την συμπίεση δικτύων εκπαιδευμένα για multiclass classification tasks και επομένως εφαρμόζουμε τις μεθόδους Neural Path K-means, AMM και semi-NMF. Οι δύο πρώτες μέθοδοι παρουσιάζονται στους Πίνακες 6.5, 6.6 σε σύγκριση με την τροπική μέθοδο [37] που αφορά δίκτυα πολλών εξόδων. Επίσης, και οι 3 μεθόδοι μαζί παρουσιάζονται στα διαγράμματα 6.1a, 6.1b συγκριτικά με τις βασικές μεθόδους ελαχιστοποιήσης Random και L1 και μία τροποποιημένη έκδοχη της ThiNet για γραμμικά επίπεδα. Τα σύνολα δεδομένων που θα χρησιμοποιήσουμε για την εκπαίδευση είναι τα MNIST και Fashion-MNIST. Τα δίκτυα που θα συμπιέσουμε είναι 2. Αρχικά, για την σύγκριση με την [37] συμπιέζουμε το ίδιο δίκτυο CNN2D όπως και πρίν με την τροποποίηση ότι έχει 500 νευρώνες για την περίπτωση του MNIST συνόλου και 1000 στην περίπτωση του Fashion-MNIST. Έπειτα, για την σύγκριση με Random, L1 και ThiNet εκτελούμε συμπίεση στο μικρό δίκτυο LeNet5 [20] τόσο για το MNIST όσο και για το Fashion-MNIST.

Πίνακας 6.5: Σύγκριση Neural Path K-means και AMM με την τροπική μέθοδο [37] στο σύνολο δεδομένων MNIST.

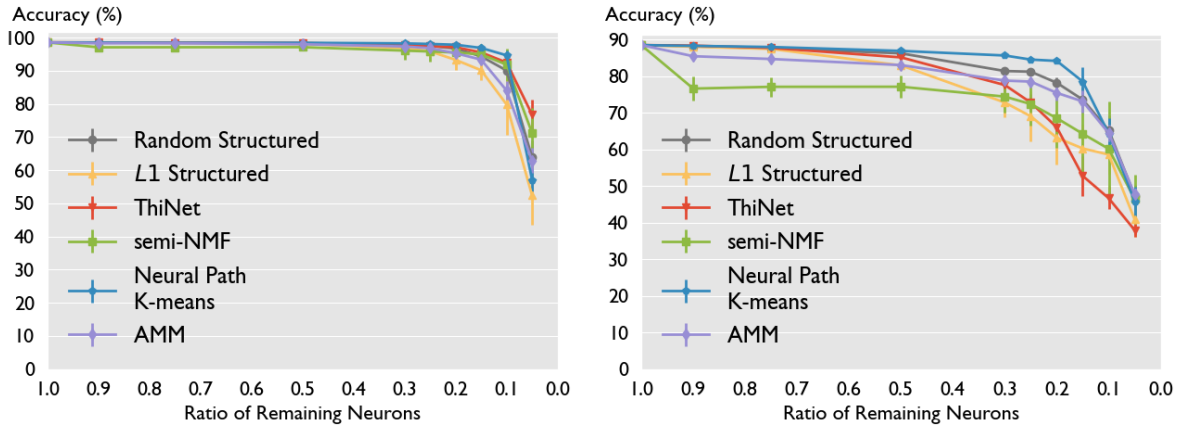
Ποσοστό εναπομείναντων νευρώνων	MNIST		
	Smyrnis et al. [37]	Neural Path K-means	AMM
100% (Original)	98.60 ± 0.03	98.61 ± 0.11	98.61 ± 0.11
50%	96.39 ± 1.18	98.13 ± 0.28	98.31 ± 0.08
25%	95.15 ± 2.36	98.42 ± 0.42	97.72 ± 0.42
10%	93.48 ± 2.57	96.89 ± 0.55	96.36 ± 0.67
5%	92.93 ± 2.59	96.31 ± 1.29	89.21 ± 3.02

Πίνακας 6.6: Σύγκριση Neural Path K-means και AMM με την τροπική μέθοδο [37] στο σύνολο δεδομένων Fashion-MNIST.

Ποσοστό εναπομείναντων νευρώνων	Fashion-MNIST		
	Smyrnis et al. [37]	Neural Path K-means	AMM
100% (Original)	88.66 ± 0.54	89.52 ± 0.19	89.52 ± 0.19
50%	83.30 ± 2.80	88.22 ± 0.32	89.06 ± 0.07
25%	82.22 ± 2.85	86.67 ± 1.12	88.37 ± 0.14
10%	80.43 ± 3.27	86.04 ± 0.94	85.94 ± 0.44
5%	–	83.68 ± 1.06	81.21 ± 1.45

Από τους Πίνακες 6.5, 6.6 εξάγουμε παρόμοια συμπεράσματα με προηγουμένως. Οι μέθοδοι μας δείχνουν να αποδίδουν ελαφρώς καλύτερα από την τροπική μέθοδο [37].

Επιπλέον στα διαγράμματα 6.1a, 6.1b που αφορούν την συμπίεση του δικτύου LeNet5 φαίνεται Neural Path K-means να αποδίδει ελαφρώς καλύτερα έναντι των υπολοίπων. Το γεγονός αυτό αναδεικνύει την ποιότητα των γεωμετρικών μεθόδων συμπίεσης. Ο Neural Path K-means συμπεριφέρεται χειρότερα μόνο στις περιπτώσεις με το ελάχιστο ποσοστό εναπομείναντων νευρώνων. Αυτό πιθανότατα να συμβαίνει λόγω κακής ποιότητας συμπίεσης με μικρό αριθμό κέντρων στον K-means. Αξίζει να σημειώσουμε ότι το δίκτυο LeNet5 συμπιέζεται στο κρυφό επίπεδο μεταξύ των δύο γραμμικών του επιπέδων, όπου διαθέτει



(a) LeNet5 on MNIST

(b) LeNet5 on Fashion-MNIST

Σχήμα 6.1: Μέθοδοι Neural Path K-means, AMM και semi-NMF σε σύγκριση με τις baseline pruning μεθόδους Random και L1, και την τροποποιημένη εκδοχή της ThiNet. Ο οριζόντιος άξονας των διαγραμμάτων αφορά το ποσοστό των εναπομεινάντων νευρώνων σε κάθε κρυφό επίπεδο στο πλήρως συνδεδεμένο (fully connected part) του δικτύου.

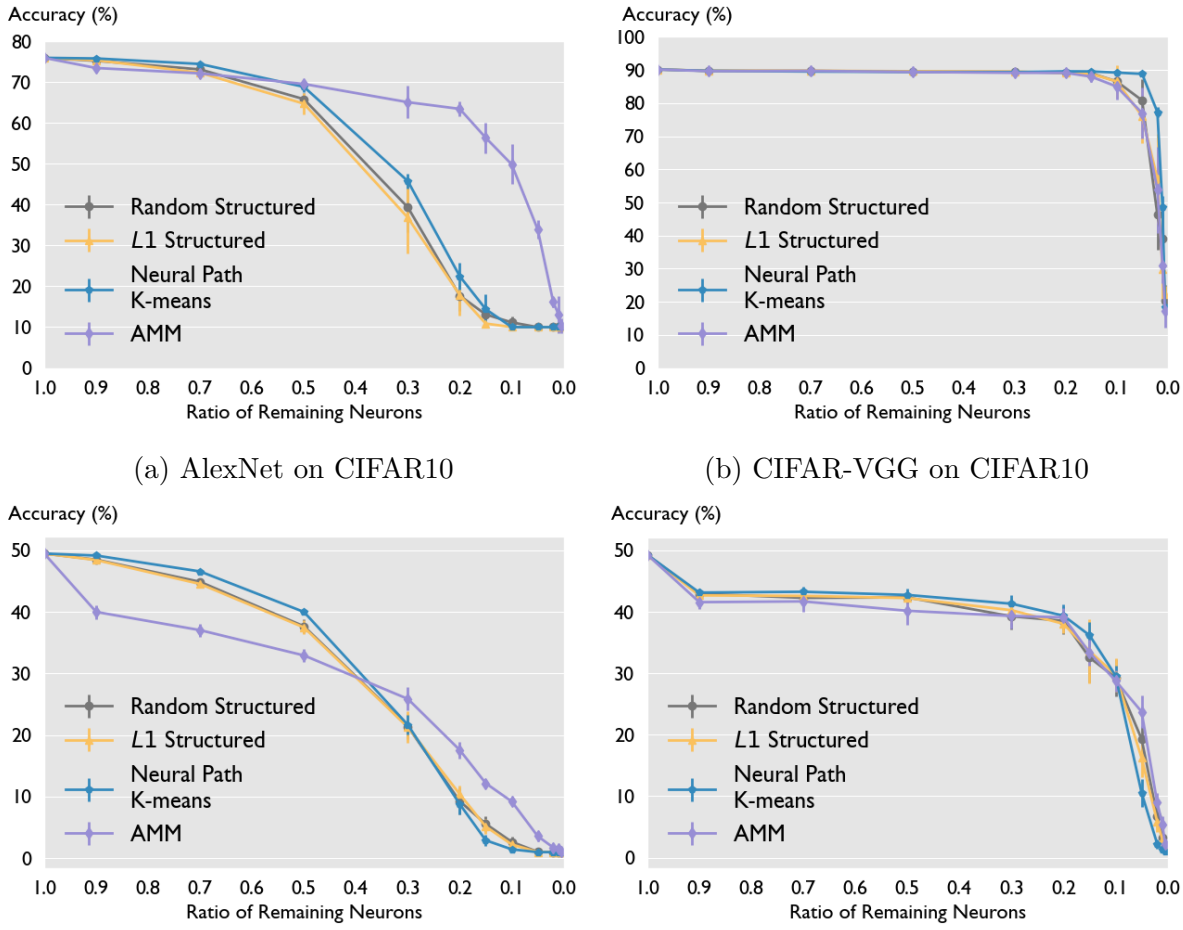
εξαρχής μόνο 84 νευρώνες. Επομένως, στα υψηλά ποσοστά συμπίεσης είναι εύλογη η πτώση της επίδοσης.

6.3 Συμπίεση μεγάλων δικτύων

Το επόμενο πείραμα που θα εκτελέσουμε αφορά την δοκιμή των αλγορίθμων μας σε δυσκολότερα dataset και νευρωνικά. Συγκεκριμένα, θα εκτελέσουμε τις μεθόδους Neural Path K-means και AMM στα δίκτυα CIFAR-VGG [3] και σε μία τροποποιημένη έκδοση του AlexNet αναφορικά με τα δύνολα δεδομένων CIFAR10 και CIFAR100. Το δίκτυο CIFAR-VGG [3] είναι μία εγκεκριμένη παραλλαγή του δικτύου VGG ειδικά σχεδιασμένη για να έχει υψηλή επίδοση στο CIFAR10. Αντίστοιχα, εμείς παρουσιάζουμε μία παραλλαγή του δικτύου AlexNet ούτως ώστε να προσαρμόζεται καλύτερα στα σύνολα δεδομένων CIFAR. Σημειώνουμε, ότι το AlexNet έχει στην έξοδο του 3 γραμμικά επίπεδα που διαθέτουν δύο κρυφά επίπεδα, επομένως θα χρειαστεί διπλή εφαρμογή των αλγορίθμων συμπίεσης για την ελαχιστοποίησή τους. Με αυτό το πείραμα, λοιπόν, θα διαφανεί και η επίδοση των αλγορίθμων μας όταν αυτοί εφαρμόζονται σε παραπάνω από ένα κρυφό επίπεδο του ίδιου δικτύου.

Τα πειραματικά αποτελέσματα σε αυτήν την περίπτωση παρατίθενται στα διαγράμματα 6.2a-6.2d. Παρατηρούμε ότι οι μέθοδοι μας παρουσιάζουν παρόμοια ή και καλύτερη επίδοση αναφορικά με το ποσοστό ακρίβειας και σε ορισμένες περιπτώσεις έχουν μικρότερη διασπορά. Ο Neural Path K-means σε σύγκριση με τον AMM αποδίδει καλύτερα στο δίκτυο CIFAR-VGG, ενώ ο AMM έχει καλύτερη επίδοση όταν διατηρούμε λίγους νευρώνες στο δίκτυο AlexNet. Η καλύτερη επίδοση του AMM πιθανότατα συμβαίνει διότι όσο μειώνουμε τα κέντρα του K-means τόσο χαλάει και η ποιότητα του clustering. Επίσης, για τον AMM παρατηρούμε ένα μεγάλο ποσοστό πτώσης της ακρίβειας στην αρχή του κάθε διαγράμματος. Αυτό επιβεβαιώνει την θεωρητική μας παρατήρηση στην Πρόταση 5.1, όπου το άνω φράγμα δεν μηδενίζεται ακόμα και για μηδενικό ποσοστό συμπίεσης. Ωστόσο, η πτώση αυτή δεν χρήζει ιδιαίτερης σημασίας, αφού αφορά χαμηλά ποσοστά και η πραγματική αξία του αλγορίθμου διαφαίνεται στα υψηλότερα ποσοστά συμπίεσης.

Επίσης, ο Neural Path K-means συγκριτικά με τους baseline αλγορίθμους παρουσιάζει



(a) AlexNet on CIFAR10

(b) CIFAR-VGG on CIFAR10

(c) AlexNet on CIFAR100

(d) CIFAR-VGG on CIFAR100

Σχήμα 6.2: Μέθοδοι Neural Path K-means και AMM σε σύγκριση με τις baseline pruning μεθόδους Random και L1 σε μεγαλύτερα νευρωνικά στα σύνολα δεδομένων CIFAR10 και CIFAR100. Ο οριζόντιος άξονας των διαγραμμάτων αφορά το ποσοστό των εναπομεινάντων νευρώνων σε κάθε κρυφό επίπεδο στο πλήρως συνδεδεμένο (fully connected part) του δικτύου.

ελαφριά βελτίωση σε κάθε διάγραμμα. Ωστόσο, δείχνει να έχει ελαφρώς χειρότερη επίδοση όταν χρησιμοποιούμε μικρό αριθμό νευρώνων. Το γεγονός αυτό οφείλεται στην αδυναμία να πραγματοποιήσει σωστό clustering. Αντίθετα ο AMM με λίγους νευρώνες δείχνει να συμπεριφέρεται καλύτερα στο δίκτυο AlexNet, ενώ στο CIFAR-VGG έχει παρόμοια επίδοση με τους baselines.

6.4 Συμπίεση συνελικτικών επιπέδων

Στο τελευταίο πείραμα θα παρουσιάσουμε τον αλγόριθμο Convolutional Neural Path K-means ο οποίος συμπιέζει κρυφά επίπεδα μεταξύ συνελικτικών φίλτρων. Τον αλγόριθμο αυτό θα συγκρίνουμε με την μέθοδο ThiNet [22] καθώς και τις βασικές μεθόδους Random και L1. Οι αλγόριθμοί προς σύγκριση εκτελούν όλοι μείωση του κρυφού επιπέδου ως προς το πλήθος των καναλιών, δηλαδή εφαρμόζουν structured pruning.

Για να είναι δίκαιη η σύγκριση των μεθόδων θα εφαρμόσουμε μία τεχνική τροποποίησης των συμπιεσμένων βαρών, η οποία υιοθετείται στο ThiNet και επιτυγχάνει την βελτίωση της απόδοσης του συμπιεσμένου δικτύου. Η τροποποίηση αυτή ονομάζεται weight rescale και πραγματοποιείται έπειτα από την συμπίεση ενός κρυφού επιπέδου τροποποιώντας τα βάρη του δεύτερου συνελικτικού επιπέδου. Θα περιγράψουμε την διαδικασία του rescaling του δεύτερου συνελικτικού επιπέδου χρησιμοποιώντας τους συμβολισμούς της ενότητας 4.3.

Συγκεκριμένα, υποθέτουμε ότι $\mathbf{v}$ είναι η έξοδος του δεύτερου συνελικτικού επιπέδου του αρχικού δικτύου και για κάθε δείγμα του συνόλου δεδομένων, επιλέγουμε τυχαία ένα pixel v_l από αυτήν. Υποθέτουμε ότι το pixel αυτό μπορεί να γραφεί ως

$$v_l = \sum_{i=1}^{n} \sum_{g} c_{jig} f_{ig}(\mathbf{x}) = \sum_{i=1}^{n} \hat{f}_{li}$$

δηλαδή γράφουμε το pixel σαν άθροισμα των συμβολών $\hat{f}_{li}$ που προκύπτουν από το κάθε κανάλι στο κρυφό επίπεδο. Σημειώνουμε ότι η περιοχή που διατρέχει ο δείκτης g εξαρτάται από το l, οπότε για αυτόν τον λόγο για κάθε δείγμα v_l τοποθετούμε δείκτη l στα $\hat{f}_{li}$. Στο συμπιεσμένο νευρωνικό δίκτυο η σχέση αυτή ισχύει υπό την μορφή

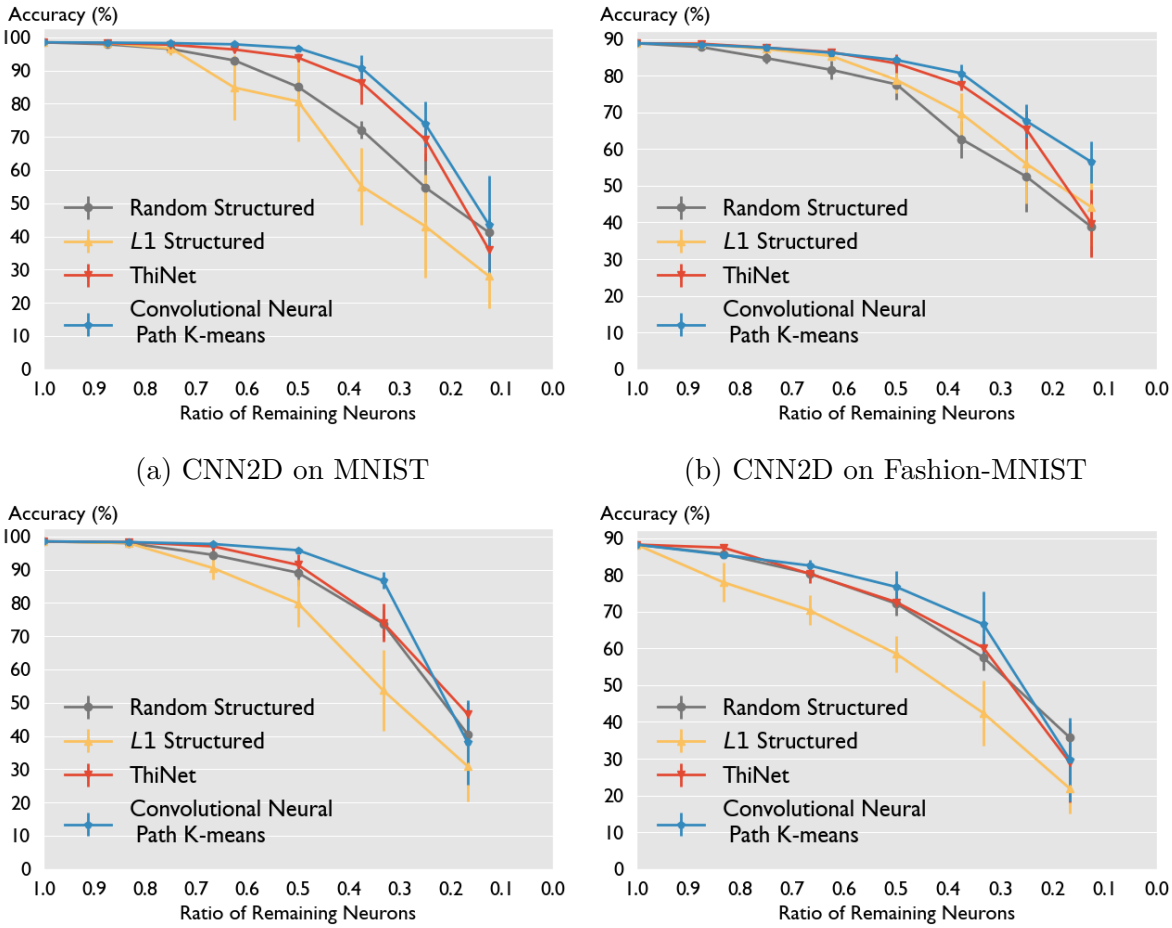
$$\tilde{v}_l = \sum_{i=1}^{n} \sum_{g} \tilde{\mathbf{c}}_{jig} \tilde{f}_{ig} = \sum_{i=1}^{n} \hat{f}^*_{li}$$

Επιθυμούμε για κάθε δείγμα εισόδου $l = 1, ..., D$ να ισχύει $v_l \approx \tilde{v}_l$ και εφαρμόζουμε ένα rescale $\mathbf{w}$ στις συνιστώσες $\hat{f}^*_{li}$ που να ελαχιστοποιούν την συνάρτηση τετραγωνικού σφάλματος

$$\sum_{l=1}^{D} (v_l - \tilde{v}_l)^2 = \sum_{l=1}^{D} \left(v_l - \mathbf{w}^T \hat{\mathbf{f}}^*_l \right)^2$$

όπου το διάνυσμα $\hat{\mathbf{f}}^*_l$ περιέχει όλα τα entries $\hat{f}^*_{li}$, $i = 1, ..., n$. Η λύση που ελαχιστοποιεί την παραπάνω ποσότητα δίνεται από τον ψευδοαντίστροφο του $F = \begin{bmatrix} \hat{\mathbf{f}}_1^{*T} \\ \vdots \\ \tilde{\mathbf{f}}_D^{*T} \end{bmatrix}$.

$$\mathbf{w} = (F^T F)^{-1} F^T \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_D \end{bmatrix}$$



(a) CNN2D on MNIST

(b) CNN2D on Fashion-MNIST

(c) LeNet5 on MNIST

(d) LeNet5 on Fashion-MNIST

Σχήμα 6.3: Μέθοδος Convolutional Neural Path K-means σε σύγκριση με ThiNet και baseline pruning μεθόδους. Ο οριζόντιος άξονας των διαγραμμάτων αφορά το ποσοστό των εναπομεινάντων καναλιών σε κάθε κρυφό επίπεδο στο συνελικτικό τμήμα (features) του δικτύου.

Η διαδικασία αυτή του rescaling μπορεί να ειδωθεί σαν ένα βέλτιστο fine-tuning το οποίο εφαρμόζεται έπειτα από την συμπίεση. Αξίζει να σημειωθεί πως το rescaling εφαρμόζεται σε όλες τις μεθόδους που παρουσιάζονται στα διαγράμματα του σχήματος 6.3.

Από τα διαγράμματα της εικόνας 6.3 διαπιστώνουμε πως η μέθοδος Convolutional Neural Path K-means εφαρμόζεται με επιτυχία και παρουσιάζει στις περισσότερες περιπτώσεις ελαφρώς καλύτερες επιδόσεις από την ανταγωνιστική μέθοδο ThiNet και από τις βασικές μεθόδους Random και L1. Η καλύτερη επίδοση του δικού μας αλγορίθμου έναντι του ThiNet μπορεί να ερμηνευτεί ως εξής. Ο αλγόριθμος K-means δίνει ως αποτέλεσμα νευρώνες οι οποίοι έχουν προκύψει από γραμμικό συνδυασμό των νευρώνων του αρχικού δικτύου, με αποτέλεσμα να διατηρείται σημαντικό μέρος της πληροφορίας του δικτύου, το οποίο σε συνδυασμό με το rescaling γίνεται ισχυρότερο και επιτυγχάνει υψηλές επιδόσεις. Αντίθετα ο ThiNet αφαιρεί ολόκληρα κανάλια με αποτέλεσμα να χάνει πληροφορία. Ωστόσο, σε ορισμένες περιπτώσεις όπου το ποσοστό συμπίεσης είναι υψηλό, ο Convolutional Neural Path K-means είναι πιο αδύναμος και αυτό μπορεί να οφείλεται σε κακή ποιότητα συμπίεσης των διανυσμάτων, λόγω μικρού διαθέσιμου αριθμού κέντρων συστάδων.

6.5 Τεχνικές Λεπτομέρειες

Στα πειράματα που παρουσιάσαμε υπεισέρχονται οι εξής λεπτομέρειες που αφορούν την προγραμματιστική υλοποίηση. Τα μικρά συνελικτικά δίκτυα, δηλαδή το CNN2D και το LeNet5 [20] εκπαιδεύονται με χρήση Adam optimizer με learning rate 10^{-3} για 20 εποχές. Τα μεγαλύτερα νευρωνικά CIFAR-VGG και το τροποποιημένο AlexNet εκπαιδεύονται με παγωμένα τα συνελικτικά φίλτρα με χρήση SGD σε learning rate 10^{-2}, momentum 0.9 και weight decay $5 \cdot 10^{-4}$ για 28 εποχές. Σε κάθε πείραμα εκπαιδεύουμε από την αρχή 5 μοντέλα με ίδια αρχιτεκτονική. Η αναφερόμενη ακρίβεια λαμβάνεται ως μέσος όρος των ακριβειών που παρουσιάζουν τα 5 μοντέλα. Επίσης, μαζί με τη μέση ακρίβεια παρατίθεται και ένα εύρος σφάλματος που προκύπτει από την τυπική απόκλιση των τιμών της. Η τυπική απόκλιση περιλαμβάνεται στα διαγράμματα υπό την μορφή κατακόρυφης μπάρας σφάλματος (error bar).

Επίσης, αναφέρουμε ότι οι αλγόριθμοί μας δεν είναι ντετερμινιστικοί, δεδομένου ότι ο K-means έχει τυχαία αρχικοποίηση (k-means++), ο AMM είναι αμιγώς πιθανοτικός και ο semi-NMF έχει τυχαία αρχικοποίηση λόγω χρήσης K-means. Για να εξασφαλίσουμε καλύτερη επίδοση επαναλαμβάνουμε έναν αριθμό φορών την συμπίεση στο μοντέλο και από αυτές διαλέγουμε την καλύτερη επίδοση. Συγκεκριμένα, για το πείραμα 2 στους Πίνακες 6.5, 6.6 οι επαναλήψεις των αλγορίθμων είναι 20, στα διαγράμματα 6.2 είναι 1, ενώ στα υπόλοιπα πειράματα 5. Το καλύτερο μοντέλο λαμβάνεται σε κάθε περίπτωση σύμφωνα με την ακρίβεια στο σύνολο δεδομένων επαλήθευσης (validation set).

Κεφάλαιο 7

Επίλογος

7.1 Αποτελέσματα και Συνεισφορές

Στην εργασία αυτή παρουσιάσαμε ένα καινοτόμο θεωρητικό πλαίσιο προσέγγισης τροπικών πολυωνύμων μέσω της Hausdorff απόστασης των πολυτόπων τους. Η προσέγγιση αυτή εφαρμοσμένη στα ζωνότοπα νευρωνικών δικτύων αποτελεί την βάση για την κατασκευή αλγορίθμων συμπίεσης. Κατά αυτόν τον τρόπο προτείνουμε γεωμετρικούς αλγορίθμους συμπίεσης γραμμικών ή συνελικτικών επιπέδων νευρωνικών δικτύων οι οποίοι έχουν αξιόλογη επίδοση η οποία βελτιώνει βασικές μεθόδους pruning αλλά και σε ορισμένες περιπτώσεις την πιο εξεζητημένη μέθοδο ThiNet. Συνοπτικά η συμβολή μας συγκεντρώνεται στα εξής στοιχεία:

- Αποδείξαμε το θεώρημα προσέγγισης τροπικών πολυωνύμων. Με βάση αυτό, η μέγιστη διαφορά δύο τροπικών πολυωνύμων σε μία φραγμένη υπερσφαίρα δεν είναι μεγαλύτερη από την απόσταση Hausdorff των πολυτόπων τους επί έναν παράγοντα ρ που αφορά την ακτίνα της υπερσφαίρας. Το αποτέλεσμα αυτό γενικεύει την αμφιμονοσήμαντη αντιστοιχία μεταξύ των γραμμικών περιοχών ενός τροπικού πολυωνύμου και των κορυφών του άνω φλοιού των ζωνοτόπων του.

- Το θεώρημα αυτό υποδεικνύει ότι η συμπίεση των ζωνοτόπων που αφορούν ένα δίκτυο μπορεί να οδηγήσει στην συμπίεση του δικτύου. Με αυτήν την σκέψη κατασκευάζουμε 3 γεωμετρικούς αλγορίθμους Zonotope K-means, Neural Path K-means και Convolutional Neural Path K-means. Η ποιότητα συμπίεσης των ζωνοτόπων του δικτύου μέσω του θεωρήματος προσέγγισης τροπικών πολυωνύμων, παρέχει ένα άνω φράγμα για το σφάλμα που έχει το τελικό δίκτυο σε σχέση με το αρχικό. Με αυτόν τον τρόπο μελετούμε θεωρητικά μέσω τροπικής γεωμετρίας τους αλγορίθμους αυτούς. Αξίζει να σημειωθεί ότι οι Zonotope K-means και Neural Path K-means αφορούν συμπίεση γραμμικών επιπέδων ενώ ο Convolutional Neural Path K-means αποτελεί τον πρώτο αλγόριθμο τροπικης γεωμετρίας για συμπίεση συνελικτικών επιπέδων.

- Εξετάζουμε 2 μη-τροπικούς αλγόριθμους AMM και semi-NMF συμπίεσης γραμμικών επιπέδων νευρωνικών δικτύων. Ο AMM υλοποιείται με βάση την προσέγγιση γινομένου πινάκων ενώ ο semi-NMF με μη-αρνητική παραγοντοποίηση πίνακα. Μάλιστα, μέσω του θεωρήματος προσέγγισης τροπικών πολυωνύμων υπολογίζουμε άνω φράγμα για το μέσο σφάλμα του AMM.

- Εκτελούμε πειράματα συμπίεσης νευρωνικών δικτύων για την αξιολόγηση της επίδοσης των αλγορίθμων. Τα πειράματα αναδεικνύουν την ικανότητα των τροπικών μεθόδων

για συμπίεση. Ο αλγόριθμος Neural Path K-means δείχνει βελτίωση έναντι άλλων τροπικών μεθόδων αλλά και των βασικών μεθόδων Random και L1. Επίσης, ο Convolutional Neural Path K-means σημειώνει ανταγωνιστική επίδοση σε σχέση με την μέθοδο συμπίεσης συνελικτικών επιπέδων ThiNet.

7.2 Μελλοντικές κατευθύνσεις

Η εργασία αυτή δίνει έναυσμα για την συνέχιση της μελέτης της τροπικής γεωμετρίας με εφαρμογές στα νευρωνικά δίκτυα. Οι επεκτάσεις που προτείνουμε είναι οι εξής:

Θεωρητικές

- Μελέτη του αλγορίθμου semi-NMF με τροπική γεωμετρία. Αυτό μπορεί να προσεγγισθεί με το θεώρημα προσέγγισης τροπικών πολυωνύμων. Συγκεκριμένα, μπορεί κάποιος να βρει την συσχέτιση των ζωνοτόπων που παράγονται από τον αλγόριθμο με τα τελικά ζωνότοπα του δικτύου και επομένως να βρει άνω φράγμα στο σφάλμα της προσέγγισης του αλγορίθμου.

- Εν συνεχεία της προηγούμενης κατεύθυνσης, πιθανότατα υπάρχει η δυνατότητα γενίκευσης της τροπικής μελέτης των αλγορίθμων pruning. Γίνεται με κάποιον τρόπο μέσω τροπικής γεωμετρίας να μπορούμε να μελετήσουμε και να αξιολογήσουμε θεωρητικά οποιονδήποτε αλγόριθμο pruning; Για αρχή μπορεί κάποιος να δοκιμάσει να αποδείξει άνω φράγματα για τις μεθόδους Random και L1.

- Στην εργασία αυτή ασχοληθήκαμε με structured αλγορίθμους. Αντίστοιχα, κάποιος μπορεί να διερευνήσει μέσω τροπικής γεωμετρίας την περίπτωση των unstructured αλγορίθμων, δηλαδή αυτών που δεν αφαιρούν ολόκληρους νευρώνες αλλά μηδενίζουν entries στους πίνακες των βαρών.

- Επίσης μπορεί να γίνει επέκταση του θεωρήματος προσέγγισης τροπικών πολυωνύμων. Για παράδειγμα μπορεί κάποιος να μελετήσει πόσο σφιχτό είναι το φράγμα και εαν είναι εφικτό να βελτιωθεί. Επιπλέον, ενδιαφέρον θα είχε αν θα μπορούσε να αποδωθεί ένα κάτω φράγμα στην μέγιστη διαφορά δύο τροπικών πολυωνύμων. Αυτό το θεωρητικό αποτέλεσμα θα μπορούσε να δώσει περισσότερες πληροφορίες για την μελέτη αλγορίθμων συμπίεσης.

- Το θεώρημα προσέγγισης τροπικών πολυωνύμων αν εφαρμοστεί για τα πολυώνυμα p, q που καθορίζονται από την έξοδο ενός νευρωνικού $v = p - q$, μας δίνει άνω φράγμα για το μέγεθος της εξόδους του νευρωνικού. Κατ' επέκταση, μπορεί να διερευνηθεί εάν υπάρχουν άλλες θεωρητικές ή πρακτικές εφαρμογές αυτού του θεωρήματος.

Πειραματικές

- Μία πρώτη επέκταση που θα μπορούσε να γίνει είναι η εκτέλεση των αλγορίθμων συμπίεσης σε μεγαλύτερα συνελικτικά δίκτυα (π.χ ResNet50) αλλά και σε μεγαλύτερα και δυσκολότερα dataset (π.χ. ImageNet).

- Για καθαρά πειραματική κατεύθυνση μπορεί κάποιος να μελετήσει περισσότερο τις τεχνικές pruning που προτείνουμε σε συνδυασμό με σύγχρονες τεχνικές όπως fine tuning έπειτα από την συμπίεση ή και συμπίεση κατά την διάρκεια της εκπαίδευσης (on the fly).

- Σε σχέση με την συμπίεση κατά την διάρκεια της εκπαίδευσης, μπορεί κάποιος να τροποποιήσει τον αλγόριθμο συμπίεσης K-means και να συμπιέζει νευρώνες κατά την διάρκεια εκπαίδευσης, λίγους την φορά, με βάση κάποιο κριτήριο που θα καθορίζει ότι εκείνη την στιγμή οι νευρώνες σχηματίζουν κάποια συστάδα.

Bibliography

[1] M. Akian, S. Gaubert, and A. Guterman. Tropical polyhedra are equivalent to mean payoff games. *International Journal of Algebra and Computation*, 22(01):1250001, 2012.

[2] M. Alfarra, A. Bibi, H. Hammoud, M. Gaafar, and B. Ghanem. On the Decision Boundaries of Deep Neural Networks: A Tropical Geometry Perspective. *arXiv preprint arXiv:2002.08838*, 2020.

[3] D. Blalock, J. J. G. Ortiz, J. Frankle, and J. Guttag. What is the state of neural network pruning? *arXiv preprint arXiv:2003.03033*, 2020.

[4] P. Butkovič. *Max-linear systems: theory and algorithms*. Springer monographs in mathematics. Springer, 2010.

[5] V. Charisopoulos and P. Maragos. Morphological perceptrons: geometry and training algorithms. In *International Symposium on Mathematical Morphology and Its Applications to Signal and Image Processing*, pages 3–15. Springer, 2017.

[6] V. Charisopoulos and P. Maragos. A tropical approach to neural networks with piecewise linear activations. *arXiv preprint arXiv:1805.08749*, 2018.

[7] R. A. Cuninghame-Green. *Minimax algebra*, volume 166. Springer Science & Business Media, 2012.

[8] N. Dimitriadis and P. Maragos. Advances in Morphological Neural Networks: Training, Pruning and Enforcing Shape Constraints. In *Proc. 46th IEEE Int'l Conf. Acoustics, Speech and Signal Processing (ICASSP-2021)*, Toronto, June 2021.

[9] C. H. Ding, T. Li, and M. I. Jordan. Convex and Semi-Nonnegative Matrix Factorizations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(1):45–55, 2010.

[10] P. Drineas, R. Kannan, and M. W. Mahoney. Fast Monte Carlo algorithms for matrices I: Approximating matrix multiplication. *SIAM Journal on Computing*, 36(1):132–157, 2006.

[11] R. J. Duffin and E. L. Peterson. Geometric programming with signomials. *Journal of Optimization Theory and Applications*, 11(1):3–35, 1973.

[12] J. Frankle and M. Carbin. The lottery ticket hypothesis: Finding sparse, trainable neural networks. *arXiv preprint arXiv:1803.03635*, 2018.

[13] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[14] N. Grigg and N. Manwaring. An elementary proof of the fundamental theorem of tropical algebra. *arXiv preprint arXiv:0707.2591*, 2007.

[15] P. Gritzmann and B. Sturmfels. Minkowski addition of polytopes: computational complexity and applications to Gröbner bases. *SIAM Journal on Discrete Mathematics*, 6(2):246–269, 1993.

[16] B. Grünbaum. *Convex polytopes*, volume 221. Springer Science & Business Media, 2013.

[17] M. Kearns and L. Valiant. Cryptographic limitations on learning boolean formulae and finite automata. *Journal of the ACM (JACM)*, 41(1):67–95, 1994.

[18] M. J. Kearns, U. V. Vazirani, and U. Vazirani. *An introduction to computational learning theory*. 1994.

[19] A.-K. Kopetzki, B. Schürmann, and M. Althoff. Methods for order reduction of zonotopes. In *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, pages 5626–5633. IEEE, 2017.

[20] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[21] B. Lin and N. M. Tran. Linear and rational factorization of tropical polynomials. *arXiv preprint arXiv:1707.03332*, 2017.

[22] J.-H. Luo, J. Wu, and W. Lin. ThiNet: A Filter Level Pruning Method for Deep Neural Network Compression. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 5068–5076, 2017.

[23] D. Maclagan and B. Sturmfels. *Introduction to tropical geometry*, volume 161. American Mathematical Soc., 2015.

[24] P. Maragos. Dynamical systems on weighted lattices: general theory. *Mathematics of Control, Signals, and Systems*, 29(4):1–49, 2017.

[25] P. Maragos, V. Charisopoulos, and E. Theodosis. Tropical Geometry and Machine Learning. *Proceedings of the IEEE*, 109(5):728–755, 2021.

[26] P. Maragos and E. Theodosis. Tropical geometry and piecewise-linear approximation of curves and surfaces on weighted lattices. *arXiv preprint arXiv:1912.03891*, 2019.

[27] P. Maragos and E. Theodosis. Multivariate Tropical Regression and Piecewise-Linear Surface Fitting. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3822–3826. IEEE, 2020.

[28] D. Mellouli, T. M. Hamdani, J. J. Sanchez-Medina, M. B. Ayed, and A. M. Alimi. Morphological convolutional neural network architecture for digit recognition. *IEEE transactions on neural networks and learning systems*, 30(9):2876–2885, 2019.

[29] G. Montúfar, R. Pascanu, K. Cho, and Y. Bengio. On the number of linear regions of deep neural networks. *arXiv preprint arXiv:1402.1869*, 2014.

[30] B. Plancher, C. D. Brumar, I. Brumar, L. Pentecost, S. Rama, and D. Brooks. Application of Approximate Matrix Multiplication to Neural Networks and Distributed SLAM. In *2019 IEEE High Performance Extreme Computing Conference (HPEC)*, pages 1–7. IEEE, 2019.

[31] G. Retsinas, A. Elafrou, G. I. Goumas, and P. Maragos. Weight Pruning via Adaptive Sparsity Loss. *CoRR*, abs/2006.02768, 2020.

[32] G. Ritter and G. Urcid. Lattice algebra approach to single-neuron computation. *IEEE Transactions on Neural Networks*, 14(2):282–295, 2003.

[33] G. X. Ritter and P. Sussner. An introduction to morphological neural networks. In *Proceedings of 13th International Conference on Pattern Recognition*, volume 4, pages 709–717. IEEE, 1996.

[34] G. X. Ritter, P. Sussner, and J. Diza-de Leon. Morphological associative memories. *IEEE Transactions on neural networks*, 9(2):281–293, 1998.

[35] R. Schneider. *Convex bodies: the Brunn - Minkowski theory*. Number 151. Cambridge University Press, 2014.

[36] G. Smyrnis and P. Maragos. Tropical polynomial division and neural networks. *arXiv preprint arXiv:1911.12922*, 2019.

[37] G. Smyrnis and P. Maragos. Multiclass Neural Network Minimization via Tropical Newton Polytope Approximation. In *Proc. Int'l Conf. on Machine Learning, PMLR*, 2020.

[38] G. Smyrnis, P. Maragos, and G. Retsinas. Maxpolynomial Division with Application To Neural Network Simplification. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4192–4196. IEEE, 2020.

[39] E. Theodosis and P. Maragos. Analysis of the Viterbi algorithm using tropical algebra and geometry. In *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pages 1–5. IEEE, 2018.

[40] E. Theodosis and P. Maragos. Tropical Modeling of Weighted Transducer Algorithms on Graphs. In *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8653–8657, 2019.

[41] N. Tsilivis, A. Tsiamis, and P. Maragos. Sparsity in Max-Plus Algebra and Applications in Multivariate Convex Regression. In *Proc. 46th IEEE Int'l Conf. Acoustics, Speech and Signal Processing (ICASSP-2021)*, Toronto, June 2021.

[42] L. G. Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, 1984.

[43] H. Xiong, L. Huang, M. Yu, L. Liu, F. Zhu, and L. Shao. On the number of linear regions of convolutional neural networks. In *International Conference on Machine Learning*, pages 10514–10523. PMLR, 2020.

[44] P.-F. Yang and P. Maragos. Min-max classifiers: Learnability, design and application. *Pattern Recognition*, 28(6):879–899, 1995.

[45] L. Zhang, G. Naitzat, and L.-H. Lim. Tropical geometry of deep neural networks. In *International Conference on Machine Learning*, pages 5824–5832. PMLR, 2018.

[46] G. M. Ziegler. *Lectures on polytopes*, volume 152. Springer Science & Business Media, 2012.

[47] Γ. Σμυρνής. *Διαίρεση τροπικών πολυωνύμων και ελαχιστοποίηση νευρωνικών δικτύων*. ΣΗΜΜΥ ΕΜΠ, 2020. http://artemis.cslab.ece.ntua.gr:8080/jspui/handle/123456789/17590.