

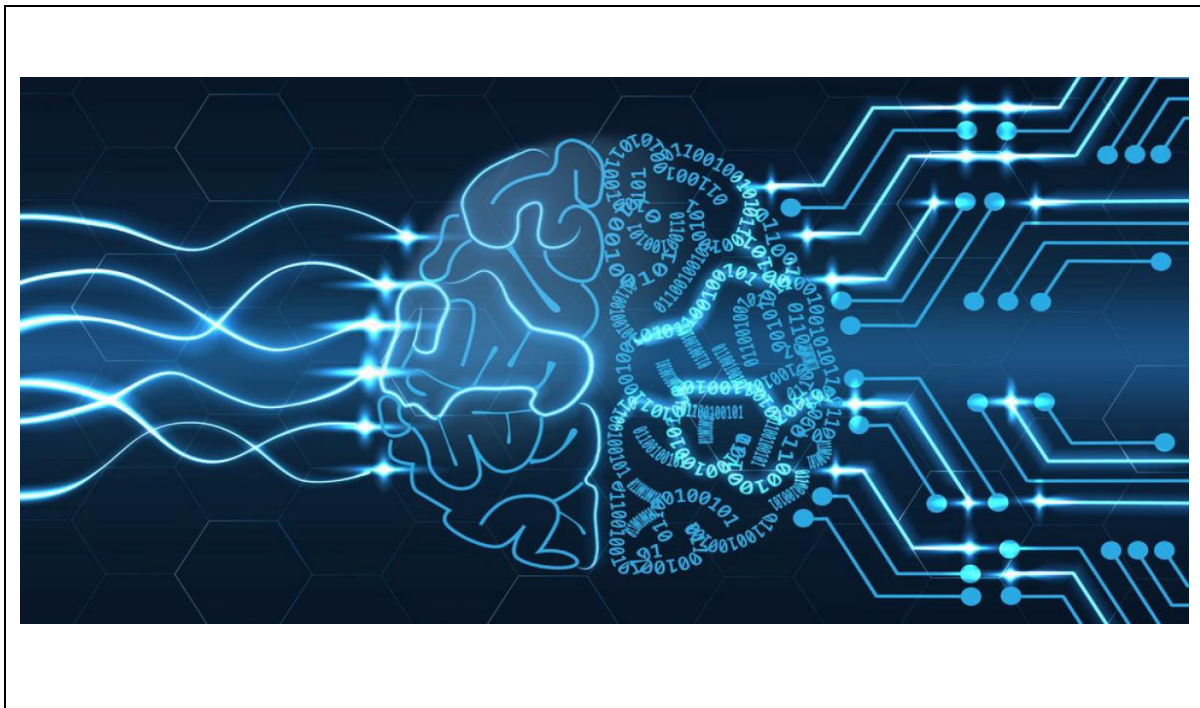


ΔΙΕΘΝΕΣ  
ΠΑΝΕΠΙΣΤΗΜΙΟ  
ΤΗΣ ΕΛΛΑΔΟΣ

ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ  
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ  
ΚΑΙ ΗΛΕΚΤΡΟΝΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

«Αξιοποίηση νευρωνικών δικτύων για τον υπολογισμό  
γενικευμένων αντίστροφων πινάκων»



Του φοιτητή  
Μπαρούτα Βασίλειου-Βαιου  
Αρ. Μητρώου: 2019205

Επιβλέπων  
Τζέκης Παναγιώτης  
Καθηγητής

**Ημερομηνία 9/5/2026**

Τίτλος Δ.Ε. Αξιοποίηση νευρωνικών δικτύων για τον υπολογισμό γενικευμένων αντίστροφων  
πινάκων

Κωδικός Δ.Ε. 25295

Όνοματεπώνυμο φοιτητή Μπαρούτας Βασίλης

Όνοματεπώνυμο εισηγητή Τζέκης Παναγιώτης

Ημερομηνία ανάληψης Δ.Ε. 08-09-2025

Ημερομηνία περάτωσης Δ.Ε. 09-05-2026

*Βεβαιώνω ότι είμαι ο συγγραφέας αυτής της εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, έχω καταγράψει τις όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών, εικόνων και κειμένου, είτε αυτές αναφέρονται ακριβώς είτε παραφρασμένες. Επιπλέον, βεβαιώνω ότι αυτή η εργασία προετοιμάστηκε από εμένα προσωπικά, ειδικά ως διπλωματική εργασία, στο Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων του ΔΙ.ΠΑ.Ε.*

*Η παρούσα εργασία αποτελεί πνευματική ιδιοκτησία του φοιτητή Μπαρούτα Βασιλείου-Βαίου που την εκπόνησε. Στο πλαίσιο της πολιτικής ανοικτής πρόσβασης, ο συγγραφέας/δημιουργός εκχωρεί στο Διεθνές Πανεπιστήμιο της Ελλάδος άδεια χρήσης του δικαιώματος αναπαραγωγής, δανεισμού, παρουσίασης στο κοινό και ψηφιακής διάχυσης της εργασίας διεθνώς, σε ηλεκτρονική μορφή και σε οποιοδήποτε μέσο, για διδακτικούς και ερευνητικούς σκοπούς, άνευ ανταλλάγματος. Η ανοικτή πρόσβαση στο πλήρες κείμενο της εργασίας, δεν σημαίνει καθ' οιονδήποτε τρόπο παραχώρηση δικαιωμάτων διανοητικής ιδιοκτησίας του συγγραφέα/δημιουργού, ούτε επιτρέπει την αναπαραγωγή, αναδημοσίευση, αντιγραφή, πώληση, εμπορική χρήση, διανομή, έκδοση, μεταφόρτωση (downloading), ανάρτηση (uploading), μετάφραση, τροποποίηση με οποιονδήποτε τρόπο, τμηματικά ή περιληπτικά της εργασίας, χωρίς τη ρητή προηγούμενη έγγραφη συναίνεση του συγγραφέα/δημιουργού.*

Η έγκριση της διπλωματικής εργασίας από το Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων του Διεθνούς Πανεπιστημίου της Ελλάδος, δεν υποδηλώνει απαραίτητως και αποδοχή των απόψεων του συγγραφέα, εκ μέρους του Τμήματος.



## Πρόλογος

Η επιλογή της παρούσας διπλωματικής εργασίας από την θέλησή μου να εξελιχθώ προσωπικά και σε νέους τομείς επάνω στην επιστήμη μου. Με γνώμονα την συνεχή εξέλιξη, εστίασα στις προοπτικές αλλά και την γνώση που μπορεί να προσφέρει η μελέτη των γενικευμένων αντίστροφων. Μέσα από αυτή τη διαδικασία, αντιλήφθηκα την αξία του να προσπερνάω ένα καινούργιο εμπόδιο, να το μελετάω και να βρίσκω μια λύση. Η ολοκλήρωση αυτής της προσπάθειας με αφήνει βαθιά κερδισμένο, έχοντας αποκομίσει πολύ σημαντικά εφόδια αλλά και πρωτότυπες γνώσεις που ήταν άγνωστες για εμένα .

## Περίληψη

Ο υπολογισμός γενικευμένων αντιστρόφων αποτελεί ένα ιδιαίτερα σημαντικό πεδίο μελέτης, καθώς βρίσκει εφαρμογές σε ποικίλα προβλήματα της γραμμικής άλγεβρας και της θεωρίας πινάκων. Η παρούσα πτυχιακή εργασία επικεντρώνεται στη χρήση επαναλαμβανόμενων νευρωνικών δικτύων για τον υπολογισμό γενικευμένων αντιστρόφων πινάκων. Στο πρώτο κεφάλαιο παρουσιάζονται οι βασικές έννοιες που σχετίζονται με τη γραμμική άλγεβρα και τη θεωρία των γενικευμένων αντιστρόφων. Στο δεύτερο κεφάλαιο εξετάζεται η εφαρμογή των Gradient Neural Networks (GNNs) στον υπολογισμό γενικευμένου αντιστρόφου για σταθερούς πίνακες. Για περιπτώσεις όπου ο πίνακας χάνει τάξη, εισάγεται η μέθοδος κανονικοποίησης του Tikhonov στη συνάρτηση σφάλματος του μοντέλου ZNN, γεγονός που επιτρέπει τον υπολογισμό του γενικευμένου αντιστρόφου. Στο τελευταίο κεφάλαιο πραγματοποιείται ανασκόπηση της εργασίας και προτείνονται κατευθύνσεις για μελλοντική έρευνα, σε ζητήματα που άπτονται της μεθοδολογίας και των εφαρμογών που παρουσιάστηκαν.

# ««Utilization of Neural Networks for the Computation of Generalized Matrix Inverses»»

«Vasileios-Vaios Baroutas»

## **Abstract**

The computation of generalized inverses constitutes a particularly important field of study, as it finds applications in various problems of linear algebra and matrix theory. This thesis focuses on the use of recurrent neural networks for the computation of generalized matrix inverses. The first chapter presents the fundamental concepts related to linear algebra and the theory of generalized inverses. The second chapter examines the application of Gradient Neural Networks (GNNs) in computing the generalized inverse for constant matrices. Subsequently, the use of the Zhang Neural Network (ZNN) and its variations is studied for the same problem. For cases where the matrix loses rank, the Tikhonov regularization method is introduced into the error function of the ZNN model. The final chapter provides a review of the work and suggests directions for future research, concerning both the methodology and the applications presented.

## Ευχαριστίες

Πρώτα από όλα θα ήθελα να ευχαριστήσω την οικογένεια μου αλλά και τους φίλους μου για την τεράστια στήριξη τους σε ένα άγνωστο βήμα για εμένα που ήταν η διπλωματική. Επίσης θέλω να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή μου κ.Τζεκη για την καθοδήγηση σε όλη την διάρκεια της διπλωματικής καθώς και για τον χρόνο που μου αφιέρωσε. Τέλος να προσθέσω τις ευχαριστίες μου στον κ.Γεροντιτη ο οποίος ήταν βασικός πυλώνας και συμμετείχε με διορθώσεις, παρατηρήσεις αλλά και συμβουλές για να καταφέρω να ολοκληρώσω μια τέτοια εργασία.

# Περιεχόμενα

Πρόλογος.....	v
Περίληψη .....	vi
Abstract .....	vii
Ευχαριστίες .....	viii
Περιεχόμενα .....	ix
Συντομογραφίες.....	x
Κεφάλαιο 1ο: Εισαγωγικές Έννοιες Νευρωνικών Δικτύων.....	1
1.1 Εισαγωγή Στα Νευρωνικά Δίκτυα.....	1
1.2 Τεχνητοί Νευρώνες .....	2
1.3 Επαναλαμβανόμενα Νευρωνικά δίκτυα RNN .....	3
1.4 Νευρωνικα δικτυα βασισμένα στην κλίση.....	4
1.5 Νευρωνικα δικτυα Μηδενισμού.....	4
1.6 Σχέσεις μεταξύ των επαναλαμβανομένων νευρωνικών δικτύων .....	5
1.7 Συναρτήσεις Ενεργοποίησης των Νευρωνικών Δικτύων .....	6
1.8 Υπολογισμός αντίστροφου πίνακα A διάστασης $n \times n$ .....	7
1.9 Γενικευμένοι αντίστροφοι πινάκων .....	7
1.9.1 Moore-Penrose Γενικευμένος αντίστροφος.....	7
1.9.2 Εύρεση γενικευμένου αντίστροφου με πλήρη τάξη στηλών .....	8
1.9.3 Drazin Γενικευμένος αντίστροφος .....	8
1.10 Μοντέλο IENTZNN .....	9
Κεφάλαιο 2ο:.....	11
2.1 Εισαγωγή .....	11
2.2 Βασικές έννοιες, Κίνητρο και Αντικείμενο της μελέτης .....	12
2.3 Χρήση νευρωνικών δικτύων για τον υπολογισμό του Drazin αντίστροφου .....	21
2.4 Μοντέλο GNN για την επίλυση μη γραμμικής εξίσωσης $AXB=D$ και της σύγκλισης του .....	28
2.5 Μη γραμμικά μοντέλα νευρωνικών δικτύων (GNN) για τον υπολογισμό εσωτερικών αντίστροφων .....	34
2.6 Μη Γραμμικά Ενεργοποιημένο Αναδρομικό Νευρωνικό Δίκτυο για τον Υπολογισμό του Αντίστροφου Drazin .....	42
BIBΛΙΟΓΡΑΦΙΑ.....	54

## Συντομογραφίες

Δ.Ε.	Διπλωματική Εργασία
ΔΙΠΑΕ	Διεθνές Πανεπιστήμιο Ελλάδος
Π.Ε.	Πτυχιακή Εργασία

## Κεφάλαιο 1ο: Εισαγωγικές Έννοιες Νευρωνικών Δικτύων

### 1.1 Εισαγωγή Στα Νευρωνικά Δίκτυα

Τα τεχνητά νευρωνικά δίκτυα είναι ένας τρόπος με τον οποίο οι υπολογιστές μπορούν να «μαθαίνουν» και να επεξεργάζονται πληροφορίες, με βάση το πώς δουλεύει ο ανθρώπινος εγκέφαλος. Έχουν βασιστεί στους βιολογικούς νευρώνες, δηλαδή στα κύτταρα που μεταφέρουν και επεξεργάζονται σήματα στον εγκέφαλό μας. Το βασικό κομμάτι ενός τέτοιου συστήματος είναι ο νευρώνας – μόνο που εδώ είναι τεχνητός, δηλαδή φτιαγμένος μέσω προγράμματος.

Ένα τέτοιο δίκτυο αποτελείται από πολλούς τέτοιους νευρώνες που συνδέονται μεταξύ τους και συνεργάζονται για να λύσουν προβλήματα. Αυτό που κάνει τα ΤΝΔ ξεχωριστά είναι ότι μπορούν να μάθουν από παραδείγματα, όπως κάνουμε και εμείς. Μέσα από μια διαδικασία που λέγεται «εκπαίδευση», προσαρμόζουν τον τρόπο με τον οποίο οι νευρώνες τους συνδέονται μεταξύ τους – κάτι σαν το πώς στον εγκέφαλό μας αλλάζουν οι συνδέσεις μεταξύ των κυττάρων όταν μαθαίνουμε κάτι νέο.

Χάρη σε αυτή την εκπαίδευση, τα ΤΝΔ μπορούν να χρησιμοποιηθούν σε διάφορες εφαρμογές, όπως στην αναγνώριση εικόνων ή στη διάκριση διαφορετικών τύπων δεδομένων. Οι αλλαγές στις «συνδέσεις» των νευρώνων του δικτύου – που λέγονται βάρη – είναι αυτές που τελικά κάνουν το δίκτυο πιο «έξυπνο» και ικανό να δίνει σωστά αποτελέσματα [1]–[2].

Η βασική ιδέα πίσω από τα νευρωνικά δίκτυα είναι η ικανότητά τους να μαθαίνουν και να προσαρμόζονται. Μέσα από συνεχή εκπαίδευση και δοκιμές με διάφορα δεδομένα, καταφέρνουν σταδιακά να βελτιώνονται, να γίνονται πιο ακριβή στις προβλέψεις τους και να παίρνουν καλύτερες αποφάσεις – ξεπερνώντας σε πολλές περιπτώσεις τους κλασικούς αλγορίθμους.

Στο επίκεντρο αυτών των συστημάτων βρίσκεται ο τεχνητός νευρώνας, ο οποίος έχει βασιστεί στον τρόπο που λειτουργούν οι νευρώνες του ανθρώπινου εγκεφάλου. Κάθε νευρώνας επεξεργάζεται τις πληροφορίες που λαμβάνει, τις μετασχηματίζει με τη βοήθεια ειδικών μαθηματικών συναρτήσεων και τις προωθεί στους επόμενους. Όλο αυτό συμβαίνει μέσα από ένα πολύπλοκο δίκτυο συνδέσεων, που του επιτρέπει να φτάσει σε μια τελική «έξυπνη» έξοδο.

Αυτό που κάνει τα νευρωνικά δίκτυα τόσο ισχυρά είναι η δυνατότητά τους να προσαρμόζουν συνεχώς αυτές τις συνδέσεις, χρησιμοποιώντας τεχνικές βελτιστοποίησης όπως η μέθοδος της καθοδικής κλίσης. Έτσι, με κάθε κύκλο μάθησης, γίνονται πιο ικανά να αναγνωρίζουν μοτίβα και να «καταλαβαίνουν» τα δεδομένα με μεγαλύτερη ακρίβεια.

Οι εφαρμογές τους είναι πλέον παντού γύρω μας – από την αναγνώριση εικόνων και φωνής, μέχρι τις διαγνώσεις στον χώρο της υγείας, τις προβλέψεις στην οικονομία και τις υπηρεσίες που βασίζονται στην κατανόηση της ανθρώπινης γλώσσας. Τα νευρωνικά δίκτυα έχουν γίνει βασικό εργαλείο σε αυτή τη νέα εποχή της τεχνητής νοημοσύνης, φέρνοντας μαζί τους καινοτομία και λύσεις που δεν ήταν εφικτές παλιότερα.

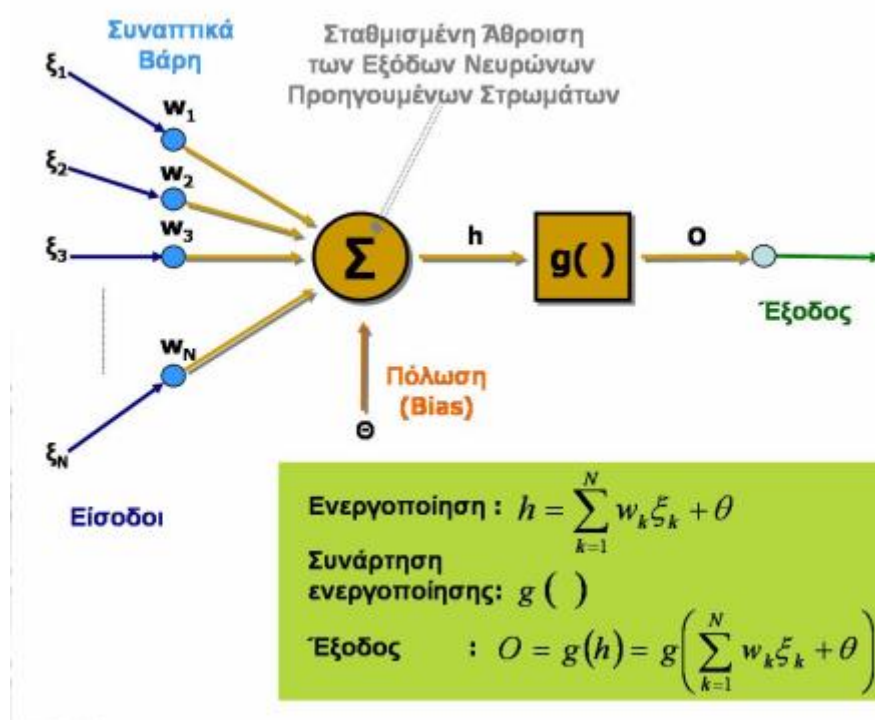
Γι' αυτό, όσο προχωρά η τεχνολογία, είναι σημαντικό να κατανοούμε όχι μόνο πώς λειτουργούν αυτά τα δίκτυα, αλλά και πώς επηρεάζουν τη ζωή μας. Εξερευνώντας τη δομή τους, τον τρόπο εκπαίδευσης τους και τις πραγματικές εφαρμογές τους, μπορούμε να καταλάβουμε καλύτερα τον ρόλο που παίζουν στον σύγχρονο κόσμο – και πώς μπορούν να οδηγήσουν την κοινωνία σε ένα μέλλον γεμάτο νέες δυνατότητες και τεχνολογικές προοπτικές [1]–[2].

## 1.2 Τεχνητοί Νευρώνες

Σε κάθε τεχνητό νευρώνα, εκτός από τις συνδέσεις που δέχεται από τους προηγούμενους νευρώνες, υπάρχει συνήθως και ένας επιπλέον όρος που ονομάζεται πόλωση. Ο ρόλος της είναι να δίνει στο δίκτυο μεγαλύτερη ευελιξία, ώστε να μπορεί να αναγνωρίζει και να ξεχωρίζει πιο περίπλοκα πρότυπα. Στις περισσότερες περιπτώσεις η πόλωση περιλαμβάνεται μαζί με τα βάρη, και όταν μιλάμε για βάρη σε ένα νευρωνικό δίκτυο, εννοούμε συνήθως και αυτή.

Η γνώση που αποκτά το ΤΝΔ κατά τη λειτουργία του βρίσκεται αποθηκευμένη στις τιμές των βαρών. Η διαδικασία εκπαίδευσης δεν είναι τίποτα άλλο από μια συνεχή αναπροσαρμογή αυτών των τιμών, μέχρι το δίκτυο να μπορεί να δίνει τις επιθυμητές απαντήσεις. Αυτό γίνεται με αλγορίθμους που καθοδηγούν πώς θα αλλάζουν τα βάρη βήμα-βήμα, μειώνοντας σταδιακά τα λάθη στην έξοδο του δικτύου.

Στο παρακάτω σχήμα φαίνονται τα δομικά στοιχεία ενός νευρώνα. Η είσοδος του νευρώνα είναι ένα διάνυσμα. Σε κάθε στοιχείο του διανύσματος αντιστοιχεί ένα βάρος. Η σταθμισμένη άθροιση των στοιχείων της εισόδου δίνεται ως είσοδος σε μία συνάρτηση μεταφοράς. Η έξοδος της συνάρτησης είναι η έξοδος του νευρώνα.



Σχήμα 1.1: Δομικά στοιχεία ενός νευρώνα [84].

Ο τρόπος με τον οποίο ενώνονται μεταξύ τους οι νευρώνες σε ένα δίκτυο ονομάζεται τοπολογία ή αρχιτεκτονική του ΤΝΔ. Οι νευρώνες χωρίζονται σε επίπεδα, και οι νευρώνες που ανήκουν στο ίδιο επίπεδο μοιράζονται την ίδια συνάρτηση μεταφοράς, πράγμα που σημαίνει ότι συμπεριφέρονται με παρόμοιο τρόπο. Κάθε νευρωνικό δίκτυο έχει τουλάχιστον δύο στρώματα:

το στρώμα εισόδου, που δέχεται τα δεδομένα, και το στρώμα εξόδου, από το οποίο παίρνουμε το τελικό αποτέλεσμα. Ανάμεσα σε αυτά μπορεί να υπάρχουν και άλλα στρώματα, τα λεγόμενα κρυφά στρώματα, οπότε το δίκτυο χαρακτηρίζεται ως πολυστρωματικό.

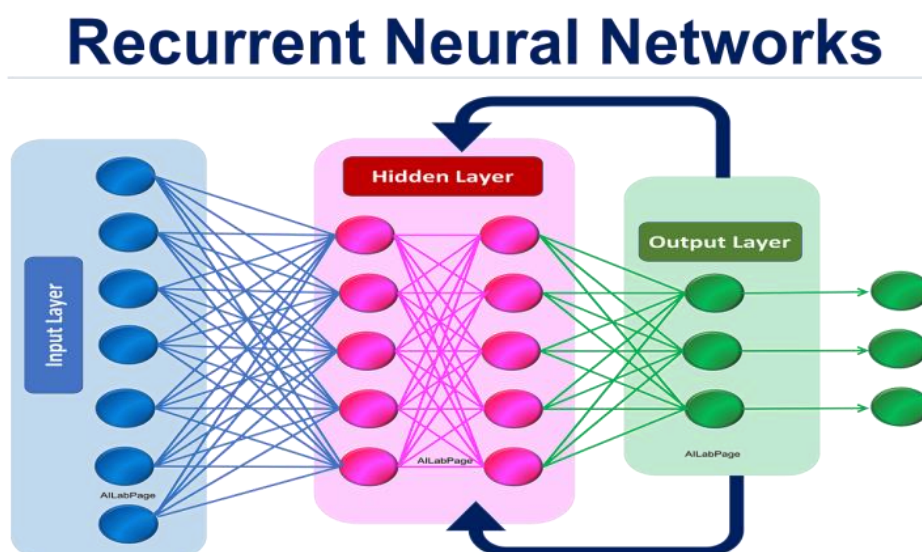
Όσο περισσότερα είναι αυτά τα στρώματα, τόσο πιο δύσκολη γίνεται η εκπαίδευση του δικτύου, αλλά ταυτόχρονα αυξάνεται και η δυνατότητά προσέγγισης του σε πιο σύνθετες σχέσεις και να πλησιάσει καλύτερα την επιθυμητή απόκριση.

Η αρχιτεκτονική ενός ΤΝΔ καθορίζεται από το πώς κυκλοφορούν οι πληροφορίες ανάμεσα στους νευρώνες. Υπάρχουν διαφορετικά είδη δικτύων, όπως τα ΤΝΔ πρόσθιας τροφοδότησης ή τα ΤΝΔ οπίσθιας τροφοδότησης.

Στα ΤΝΔ πρόσθιας τροφοδότησης, το σήμα προχωράει μόνο προς τα εμπρός, χωρίς κάποια έξοδο να επιστρέφει ως είσοδος σε νευρώνες του ίδιου ή προηγούμενου στρώματος. Συνήθως, κάθε νευρώνας ενός στρώματος συνδέεται με όλους τους νευρώνες του επόμενου στρώματος, δημιουργώντας έτσι ένα πλήρως διασυνδεδεμένο πολυστρωματικό δίκτυο πρόσθιας τροφοδότησης [4].

### 1.3 Επαναλαμβανόμενα Νευρωνικά δίκτυα RNN

Στον χώρο της τεχνητής νοημοσύνης και της μηχανικής μάθησης, τα αναδρομικά νευρωνικά δίκτυα (Recurrent Neural Networks - RNN) αποτελούν ένα βασικό εργαλείο, καθώς είναι ιδανικά για την επεξεργασία δεδομένων που εμφανίζονται με τη μορφή ακολουθιών και για την κατανόηση χρονικών εξαρτήσεων. Σε αντίθεση με τα κλασικά νευρωνικά δίκτυα προώθησης (feedforward), που επεξεργάζονται την είσοδο μόνο μία φορά από την αρχή μέχρι το τέλος, τα RNN έχουν τη δυνατότητα να «θυμούνται» πληροφορίες και να τις μεταφέρουν από το ένα χρονικό βήμα στο επόμενο. Έτσι, μπορούν να αναγνωρίζουν πιο σύνθετα μοτίβα και να συλλαμβάνουν σχέσεις που εκτείνονται σε μεγαλύτερα χρονικά διαστήματα. Η βασική τους διαφορά είναι ότι διαθέτουν αναδρομικές συνδέσεις, δηλαδή κυκλικές διαδρομές μέσα στην αρχιτεκτονική τους, οι οποίες επιτρέπουν στις πληροφορίες να παραμένουν ενεργές και να εξελίσσονται με τον χρόνο. Χάρη σε αυτή τη δυνατότητα, τα RNN μπορούν να προσαρμόζονται σε προβλήματα όπου η χρονική αλληλουχία παίζει ρόλο, όπως η πρόβλεψη χρονοσειρών, η ανάλυση και παραγωγή κειμένου, η επεξεργασία φυσικής γλώσσας ή ακόμη και η αναγνώριση φωνής.



Σχήμα 1.2: Αρχιτεκτονική των επαναλαμβανόμενων νευρωνικών δικτύων [5].

Η βασική μονάδα ενός RNN είναι ο αναδρομικός νευρώνας. Ο νευρώνας αυτός διαθέτει μια εσωτερική κατάσταση, κάτι σαν «μνήμη», η οποία ενημερώνεται συνεχώς καθώς φτάνουν νέες εισοδοί. Αυτή η κατάσταση κρατάει πληροφορίες από τα προηγούμενα χρονικά βήματα και επηρεάζει τον τρόπο που το δίκτυο επεξεργάζεται τις επόμενες εισόδους.

Με άλλα λόγια, το RNN δεν βλέπει μόνο το τρέχον δεδομένο, αλλά έχει υπόψη του και το «παρελθόν», γεγονός που του επιτρέπει να κατανοεί ακολουθιακές σχέσεις και να αναγνωρίζει περίπλοκα χρονικά μοτίβα. Παρόλο που τα κλασικά RNN είναι αρκετά εύελικτα και ισχυρά, δεν είναι χωρίς αδυναμίες. Ένα σημαντικό πρόβλημα που εμφανίζεται συχνά κατά την εκπαίδευσή τους είναι το φαινόμενο της εξαφάνισης της κλιμάκωσης (vanishing gradient).

Αυτό δυσκολεύει την εκπαίδευση, επειδή το δίκτυο χάνει την ικανότητα να «θυμάται» μακροπρόθεσμες εξαρτήσεις, περιορίζοντας έτσι την απόδοσή του σε προβλήματα που απαιτούν μεγάλη μνήμη[4].

#### 1.4 Νευρωνικά δίκτυα βασισμένα στην κλίση

Τα νευρωνικά δίκτυα που βασίζονται στη μέθοδο της κλίσης (Gradient-based Neural Networks - GNN) αποτελούν τον πυρήνα των περισσότερων τεχνικών. Η εκπαίδευσή τους γίνεται με την τεχνική της καθόδου κλίσης (gradient-descent), όπου τα βάρη του δικτύου προσαρμόζονται βήμα-βήμα έτσι ώστε να μειώνεται η συνάρτηση απώλειας και να βελτιώνεται η ακρίβεια των προβλέψεων. Η καρδιά της μεθόδου αυτής είναι ο αλγόριθμος της αντίστροφης διάδοσης (backpropagation). Μέσα από αυτόν, το δίκτυο υπολογίζει πώς επηρεάζει κάθε παράμετρος το συνολικό σφάλμα και διορθώνει αναλόγως τα βάρη, ξεκινώντας από την έξοδο και προχωρώντας προς τα πίσω στα προηγούμενα στρώματα. Με αυτό τον τρόπο, το δίκτυο «μαθαίνει» σταδιακά να αναγνωρίζει καλύτερα τα πρότυπα που υπάρχουν στα δεδομένα. Η μεγάλη ευελιξία των GNN έχει ανοίξει τον δρόμο για εντυπωσιακές εφαρμογές σε πολλούς τομείς. Είναι ιδιαίτερα αποτελεσματικά στην κατανόηση και διάκριση σύνθετων μοτίβων και ιεραρχικών σχέσεων, γι' αυτό και χρησιμοποιούνται από την αναγνώριση και ταξινόμηση εικόνων, μέχρι τη μηχανική μετάφραση και τα στρατηγικά παιχνίδια [2] - [4].

Στα GNN μοντέλα ορίζουμε την παρακάτω συνάρτηση σφάλματος :

$$E(X(t), t) = \frac{(\|AX(t) - I\|_F^2)}{2} \quad (1.4.1).$$

Παρατηρούμε ότι λόγω της νόρμας ισχύει  $E(X(t), t) \geq 0$ , συνεπώς το σφάλμα  $E$  παρουσιάζει ελάχιστο στη τιμή  $\bar{X}(t)$  (δηλαδή  $E(X(t), t) \geq E(\bar{X}(t), t) = 0$ ) [5].

#### 1.5 Νευρωνικά δίκτυα Μηδενισμού

Τα νευρωνικά δίκτυα μηδενισμού (Zeroing Neural Networks - ZNN) [ 5 ] αποτελούν μια πιο πρόσφατη κατηγορία δικτύων που εισάγουν μια διαφορετική φιλοσοφία στον τρόπο σύνδεσης των νευρώνων. Σε αντίθεση με τα κλασικά, πλήρως συνδεδεμένα δίκτυα, τα ZNN χρησιμοποιούν πιο αραιές συνδέσεις, «μηδενίζοντας» επιλεκτικά ορισμένους δεσμούς μεταξύ νευρώνων. Με αυτόν τον τρόπο μειώνεται η πολυπλοκότητα του μοντέλου, ενώ ταυτόχρονα βελτιώνεται η δυνατότητα ερμηνείας των αποτελεσμάτων και η ικανότητά του να γενικεύει καλύτερα σε νέα δεδομένα. Ουσιαστικά, με τον στρατηγικό αποκλεισμό συνδέσεων που δεν έχουν μεγάλη σημασία, τα ZNN εστιάζουν στα πιο ουσιαστικά χαρακτηριστικά των δεδομένων και αγνοούν τον περιττό «θόρυβο».

$$\dot{E}(t) = -kG(E(t)) \quad (1.5.1).$$

Η παράμετρος  $k > 0$  είναι μια θετική σταθερά που καθορίζει την ταχύτητα με την οποία το σύστημα συγκλίνει στην εξίσωση. Από την άλλη η  $G(\cdot)$  εκφράζει την κατά στοιχείο εφαρμογή μιας κατάλληλης συνάρτησης ενεργοποίησης  $g(\cdot)$  πάνω στην συνάρτηση σφάλματος  $E(t)$ .

Για να είναι το μοντέλο πιο αποτελεσματικό στην επίλυση χρονο-μεταβαλλόμενων προβλημάτων, πρέπει να ληφθούν υπόψη δύο βασικοί παράγοντες: η παράμετρος ενίσχυσης  $k$  και η συνάρτηση ενεργοποίησης  $G(\cdot)$ . Θεωρητικά, όσο μεγαλύτερη είναι η τιμή του  $k$ , τόσο ταχύτερα επιτυγχάνεται η σύγκλιση του συστήματος.

Αυτό βοηθάει όχι μόνο στην καλύτερη κατανόηση του πώς λειτουργεί το μοντέλο, αλλά και στη μείωση του κινδύνου υπερ-εκπαίδευσης, που αποτελεί μια από τις πιο συνηθισμένες προκλήσεις στα παραδοσιακά δίκτυα. Πέρα όμως από τα πρακτικά τους πλεονεκτήματα, τα ZNN εκφράζουν και μια αλλαγή στη φιλοσοφία σχεδιασμού των νευρωνικών δικτύων. Αντί να στηρίζονται στην υπερβολική πολυπλοκότητα για καλύτερη απόδοση, δίνουν προτεραιότητα στην απλότητα, την αραιότητα και την αποτελεσματικότητα. Έτσι καταφέρνουν να συνδυάζουν υψηλή απόδοση κάτι που συνήθως θεωρείται δύσκολο να επιτευχθεί. Με την αυξανόμενη χρήση της μηχανικής μάθησης σε πολλούς τομείς, η ανάγκη για μοντέλα που είναι ταυτόχρονα αποδοτικά και κατανοητά μεγαλώνει. Σε αυτό το πλαίσιο, τα ZNN εμφανίζονται ως μια ιδιαίτερα ελκυστική λύση, προσφέροντας στους ερευνητές ένα εργαλείο που τους επιτρέπει να δουλεύουν με περισσότερη διαφάνεια και ασφάλεια πάνω σε πραγματικά δεδομένα [2][4].

## 1.6 Σχέσεις μεταξύ των επαναλαμβανομένων νευρωνικών δικτύων

Παρόλο που τα ZNN και GNN έχουν διαφορετικές αρχιτεκτονικές και πεδία εφαρμογών, μοιράζονται ορισμένα βασικά χαρακτηριστικά που τα συνδέουν. Και τα τρία σχετίζονται με την επεξεργασία δεδομένων και με την εξέλιξη εσωτερικών καταστάσεων με την πάροδο του χρόνου. Αυτός ο κοινός άξονας τα τοποθετεί όλα στην ίδια ευρύτερη οικογένεια νευρωνικών δικτύων που ασχολούνται με δυναμικές διαδικασίες. Αντίστοιχα, τα ZNN και τα GNN έχουν επίσης μηχανισμούς που επιτρέπουν την προσαρμογή και εξέλιξη των καταστάσεων ή των παραμέτρων τους μέσα στον χρόνο, με στόχο τη μείωση των λαθών και την καλύτερη απόδοση. Τα ZNN επικεντρώνονται κυρίως στην εξάλειψη σφαλμάτων σε δυναμικά συστήματα, γεγονός που τα καθιστά ιδανικά για εφαρμογές ελέγχου υψηλής ακρίβειας. Από την άλλη, τα GNN βασίζονται στη συνεχή βελτιστοποίηση με τη μέθοδο της καθόδου κλίσης, κάτι που τα καθιστά πολύ χρήσιμα σε εφαρμογές που απαιτούν προσαρμογή σε πραγματικό χρόνο. Σε πιο μαθηματικό επίπεδο, τόσο τα ZNN όσο και τα GNN περιγράφονται συχνά με διαφορικές εξισώσεις, οι οποίες αποτυπώνουν τη συνεχή χρονική εξέλιξη του συστήματος. Αυτή η οπτική συναντάται και σε πιο προχωρημένες εκδοχές των RNN, όπως στα συνεχούς χρόνου αναδρομικά δίκτυα (CTRNN) [ 6 ]. Συνολικά, παρόλο που ZNN και GNN διαφέρουν ως προς τη μεθοδολογία και τις εφαρμογές τους, μπορούν να θεωρηθούν συγγενείς με τα RNN, χάρη στην κοινή έμφαση στη χρονική δυναμική, στη μνήμη κατάστασης και στην προσαρμοστικότητα. Η κατανόηση αυτής της σχέσης μας βοηθά να εκμεταλλευτούμε καλύτερα τα πλεονεκτήματα κάθε προσέγγισης σε εφαρμογές που κυμαίνονται από τον έλεγχο δυναμικών συστημάτων, μέχρι τη βελτιστοποίηση και την ανάλυση δεδομένων σε πραγματικό χρόνο.

## 1.7 Συναρτήσεις Ενεργοποίησης των Νευρωνικών Δικτύων

Οι συναρτήσεις ενεργοποίησης προσθέτουν μη γραμμικότητα στο νευρωνικό δίκτυο, δίνοντάς του τη δυνατότητα να μαθαίνει και να αποτυπώνει πιο σύνθετες σχέσεις ανάμεσα στις εισόδους και τις εξόδους [4]. Με πιο απλά λόγια, του επιτρέπουν να ξεπερνά τους περιορισμούς μιας απλής γραμμικής απεικόνισης. Μερικές από τις πιο γνωστές συναρτήσεις ενεργοποίησης που χρησιμοποιούνται στα νευρωνικά δίκτυα είναι:

Γραμμική Συνάρτηση ενεργοποίησης:

$$g(u) = u \quad (1.7.1).$$

Η γραμμική συνάρτηση είναι η πιο απλή μορφή συνάρτησης που χρησιμοποιείται σε έναν τεχνητό νευρώνα και η έξοδος της είναι ευθέως ανάλογη με την είσοδο και απλώς το νευρωνικό δίκτυο παίρνει το άθροισμα των εισόδων του και το περνάει όπως είναι στο επόμενο επίπεδο.

Διπολική σιγμοειδή συνάρτηση ενεργοποίησης :

$$\frac{1 + \exp(-\rho)}{1 + \exp(-\rho)} \cdot \frac{1 - \exp(-\rho u)}{1 + \exp(-\rho u)}, \quad \rho > 2(4) \quad (1.7.2).$$

Η σιγμοειδής συνάρτηση έχει τις τιμές της στο εύρος (0,1), μεταφέρει στο κέντρο της το 0 και βάζει τις τιμές εξόδου στο εύρος (-1,1).

Συνάρτηση δύναμης:

$$g(u) = u^p \quad (1.7.3).$$

όπου το  $p$  είναι περιττός ακέραιος με  $p \geq 3$ .

Πρόκειται για μια συνάρτηση η οποία υψώνει την είσοδο σε μια συγκεκριμένη δύναμη.

Δυναμική σιγμοειδή συνάρτηση ενεργοποίησης:

$$g(u) = \begin{cases} u^p, & \text{αν } |u| \geq 1 \\ \frac{1 + \exp(-\hat{\rho})}{1 - \exp(-\hat{\rho})} \cdot \frac{1 - \exp(-\hat{\rho}u)}{1 + \exp(-\hat{\rho}u)}, & \text{αλλιώς, (6)} \end{cases} \quad (1.7.4).$$

οπού  $\hat{\rho} > 2$  και  $p \geq 3$

Αυτή η συνάρτηση δεν έχει σταθερό σχήμα και η κλίση της μαθαίνεται από το ίδιο το δίκτυο κατά την ώρα της εκπαίδευσης.

ομαλή δυναμο-σιγμοειδής συνάρτηση ενεργοποίησης:

$$g(u) = \frac{1}{2}(u^p) + \frac{1 + \exp(-\hat{\rho})}{1 - \exp(-\hat{\rho})} \cdot \frac{1 - \exp(-\hat{\rho}u)}{1 + \exp(-\hat{\rho}u)}, \quad p \geq 3 \quad \hat{\rho} > 2 \quad (1.7.5).$$

Είναι μια συνάρτηση η οποία συνδυάζει τις ιδιότητες των πολυωνύμων μαζί με την σιγμοειδής συνάρτηση.

Συνάρτηση ενεργοποίησης αθροίσματος δυνάμεων:

$$g(u) = \sum_{k=1}^N u^{2k-1} \quad (1.7.6).$$

οπού  $N > 1$

1. Όταν  $N=2$  τότε ισχύει ότι  $g(u)=u+u^3$

2. Όταν  $N=3$  τότε ισχύει ότι  $g(u) = u+u^3+u^5$

Σε αυτή την συνάρτηση ο νευρώνας δεν κάνει απλώς μια πράξη αλλά αξιολογεί ένα ολόκληρο πολυώνυμο.

Συνάρτηση ενεργοποίησης υπερβολικού ημίτονου:

$$g(u) = \sinh(mu) = \frac{\exp(mu) - \exp(-mu)}{2} \quad (1.7.7).$$

Αυτή η συνάρτηση είναι τριγωνομετρική και βασίζεται σε εκθετικά μοντέλα.

## 1.8 Υπολογισμός αντίστροφου πίνακα A διάστασης n x n

Κάθε τετραγωνικός πίνακας  $A \in R^{n \times n}$  είναι αντιστρέψιμος όταν η ορίζουσα του είναι διάφορη του μηδενός δηλαδή  $\det(A) \neq 0$ . Σε κάθε πίνακα αν  $\det(A) = 0$  ονομάζεται μη αντιστρέψιμος η ιδιάζων.

Ο ανάστροφος πίνακας ( $A^T$ ) ενός πίνακα  $A$ , είναι ο πίνακας που προκύπτει από τον  $A$  αν οι σειρές γίνουν στήλες και οι στήλες σειρές. Δηλαδή η πρώτη σειρά να γίνει πρώτη στήλη, η δεύτερη σειρά δεύτερη στήλη.

Ο αντίστροφος πίνακας ( $A^{-1}$ ) υπολογίζεται από τον τυπο  $A^{-1} = \frac{1}{\det(A)} \text{adj}(A)$ .

Για να γίνει η κατασκευή του προσαρτημένου πίνακα  $\text{adj}(A)$  δημιουργούμε τα αλγεβρικά συμπληρώματα του πίνακα  $A$  από τις σχέσεις  $A_{ij} = (-1)^{i+j} \det(c_{ij}A)$ , όπου  $c_{ij}$  είναι η ορίζουσα του πίνακα  $A$  διαγράφοντας την  $i$  γραμμή και  $j$  στήλη [6]. Ο ανάστροφος πίνακας των αλγεβρικών συμπληρωμάτων θα είναι ο προσαρτημένος πίνακας του  $A$ . Δηλαδή,  $\text{adj}(A) = ((-1)^{i+j} |A_{ij}|)^T$

Ο αντίστροφος του πίνακα δίνεται από την σχέση  $A^{-1} = \frac{1}{\det(A(t))} \text{adj}(A)$ . Γενικεύοντας για πολυωνύμικους πίνακες ο αντίστροφος ενός πολυωνύμου πίνακα θα δίνεται από την παρακάτω σχέση

$$A^{-1}(t) = \frac{1}{\det(A(t))} \text{adj}(A(t)) \quad (1.8.1).$$

Για να υπολογιστεί ο αντίστροφος ενός πολυωνυμικού πίνακα διαχωρίζουμε το πρόβλημα σε δύο μέρη:

1° μέρος: τον υπολογισμό της ορίζουσας.

2° μέρος: του προσαρτημένου πίνακα  $A$  ( $\text{adj}(A)$ ).

## 1.9 Γενικευμένοι αντίστροφοι πινάκων

Γενικευμένος αντίστροφος ενός πίνακα  $A \in R^{m \times n}$  ορίζεται ο πίνακας  $A^g \in R^{n \times m}$  που ικανοποιεί την συνθήκη  $AA^gA = A$  [ 6 ].

Υπάρχουν πολλοί τύποι γενικευμένων αντιστρόφων. Ωστόσο οι περισσότεροι δεν είναι μοναδικοί. Ο γενικευμένος αντίστροφος για ορθογώνιο πίνακα συνδέεται με την λύση του γραμμικού συστήματος  $Ax=b$ . Η λύση της εξίσωσης είναι  $x = (A^T A)^{-1} A^T b$  η οποία είναι ιση με  $x = A^g b$  [6].

### 1.9.1 Moore-Penrose Γενικευμένος αντίστροφος

Για κάθε πίνακα  $A \in R^{m \times n}$  υπάρχει ένας μοναδικός πίνακας που θα συμβολίζεται  $A^\dagger \in R^{n \times m}$  και ονομάζεται Moore-Penrose γενικευμένος αντίστροφος και ικανοποιεί τις παρακάτω εξισώσεις [ 6 ].

$$1. AA^\dagger A = A$$

$$2. A^\dagger AA^\dagger = A^\dagger$$

$$3. (AA^\dagger)^* = AA^\dagger$$

$$4. (A^\dagger A)^* = A^\dagger A \text{ (1.9.1.1).}$$

Οπού με  $A^*$  είναι ο ανάστροφος συζυγής του  $A$ . Στην ειδική περίπτωση που ο  $A$  είναι τετραγωνικός και αντιστρέψιμος τότε ο γενικευμένος αντίστροφος συμπίπτει με τον συνηθισμένο αντίστροφο.

Η εκτενέστερη μελέτη των δυναμικών συστημάτων που υλοποιείται με διαφορικές εξισώσεις ήταν ο λόγος εμφάνισης του Moore-Penrose γενικευμένου αντίστροφου ενός πίνακα. Ανάλογοι ορισμοί ισχύουν και στην περίπτωση των πολυωνυμικών πινάκων

Για κάθε πίνακα  $A(t) \in R^{m \times n}$  υπάρχει ένας μοναδιαίος πίνακας που θα συμβολίζεται  $A(t)^\dagger \in R^{n \times m}$  και ονομάζεται Moore-Penrose και ικανοποιεί τις παρακάτω εξισώσεις

$$1. A(t)A^\dagger(t)A(t) = A(t)$$

$$2. A^\dagger(t)A(t)A^\dagger(t) = A^\dagger(t)$$

$$3. (A(t)A^\dagger(t))^* = A(t)A^\dagger(t)$$

$$4. (A^\dagger(t)A(t))^* = A^\dagger(t)A(t) \text{ (1.9.1.2).}$$

Οπού  $A(t)^*$  με συμβολίζεται ο ανάστροφος συζυγής του  $A(t)$ . Στην ειδική περίπτωση που ο  $A(t)$  είναι τετράγωνος και αντιστρέψιμος τότε ο γενικευμένος αντίστροφος συμπίπτει με τον συνηθισμένο αντίστροφο [6].

## 1.9.2 Εύρεση γενικευμένου αντίστροφου με πλήρη τάξη στηλών

Ο γενικευμένος αντίστροφος Moore-Penrose υπολογίζεται από την σχέση

$$A^\dagger(t) = (A^T(t)A(t))^{-1}A^T(t) \text{ (1.9.2.1).}$$

Μετάπειτα εφαρμόζω την τεχνική υπολογισμού παρεμβολής ως εξής:

$$A^\dagger(t) = (A^T(t) \cdot A(t))^{-1} \cdot A^T(t) = \frac{1}{\det(A^T(t) \cdot A(t))} \cdot [\text{adj}(A^T(t) \cdot A(t)) \cdot A^T(t)] \text{ (1.9.2.2).}$$

Θέτουμε  $B(t) = A^T(t) \cdot A(t)$

$$A^\dagger(t) = \frac{1}{\det(B(t))} \cdot [\text{adj}(B(t)) \cdot A^T(t)] \text{ (1.9.2.3).}$$

Για να υπολογίσουμε τον γενικευμένο αντίστροφο ενός πολυωνυμικού πίνακα χωρίζουμε τον υπολογισμό σε 2 μέρη.

1° μέρος : Υπολογισμός του  $B(t)$  και της ορίζουσας του.

2° μέρος: Γινόμενο προσαρτημένου πίνακα  $\text{adj}(B)$  με τον ανάστροφο  $A^T(t)$  [6].

## 1.9.3 Drazin Γενικευμένος αντίστροφος

Ένας ακόμη γενικευμένος αντίστροφος που έχει πολύ ενδιαφέρον είναι ο Drazin και ορίζεται μόνο για τετράγωνους πίνακες .

Για κάθε τετράγωνο πίνακα  $A \in R^{m \times m}$  υπάρχει ένας μοναδικός πίνακας που συμβολίζεται  $A^D \in R^{m \times m}$  και ονομάζεται Drazin γενικευμένος αντίστροφος που ικανοποιεί τις συνθήκες παρακάτω [6]:

$$\begin{aligned} A^{k+1}A^D &= A^k \text{ για } k := \text{ind}(A) := \min(k \in \mathbb{N}^+ : \text{rank}(A^k) = \text{rank}(A^{k+1})) \\ A^D A A^D &= A^D \\ A A^D &= A^D A \quad (1.9.3.1). \end{aligned}$$

Για κάθε τετραγωνο πίνακα  $A(t) \in R^{m \times m}$  υπάρχει ένας μοναδικός πίνακας που συμβολίζεται  $A^D(t) \in R^{m \times m}$  και ονομάζεται Drazin γενικευμένος αντίστροφος που ικανοποιεί τις παρακάτω συνθήκες:

$$\begin{aligned} A(t)^{k+1}A^D(t) &= A(t)^k \\ \text{για } k &:= \text{ind}(A(t)) := \min(k \in \mathbb{N}^+ : \text{rank}(A(t)^k) = \text{rank}(A(t)^{k+1})) \\ A^D(t)A(t)A(t)^D &= A^D(t) \\ A(t)A^D(t) &= A^D(t)A(t) \quad (1.9.3.2). \end{aligned}$$

### 1.10 Μοντέλο IENTZNN

Λόγω της προσθήκης του ολοκληρώματος, το συγκεκριμένο μοντέλο ονομάζεται ενισχυμένο ZNN (IEZNN) και μπορεί να θεωρηθεί ως μια δεύτερης τάξης διαφορική εξίσωση. Με το ολοκλήρωμα, το IEZNN αποκτά τη δυνατότητα να λειτουργεί ικανοποιητικά ακόμα και με παρουσία θορύβου, ενώ με τη βοήθεια των θετικών παραμέτρων  $\eta > 0$  και  $\kappa > 0$ , εξασφαλίζεται εκθετική σύγκλιση [7]-[8].

Η σύγκλιση και η ανθεκτικότητα είναι δύο βασικά χαρακτηριστικά που επιδιώκουμε σε ένα νευρωνικό δίκτυο. Αν και το μοντέλο IEZNN μπορεί να επιτύχει εκθετική σύγκλιση, αυτό δεν ισχύει όταν το σύστημα επηρεάζεται από θόρυβο. Σε τέτοιες περιπτώσεις, το σφάλμα του IEZNN ενδέχεται να μην μηδενίζεται, ακόμη και μετά από άπειρο χρόνο.

Με βάση την ανάλυση που κάναμε παραπάνω η συνάρτηση σφάλματος IEZNN μπορεί να επιλεγθεί να είναι η ίδια με αυτή του OZNN. Έτσι το ακόλουθο μοντέλο  $A(t) = \min(m,n)$ , όταν μπορεί να ληφθεί με βάση τον κανόνα εξέλιξης.

A. Μοντέλο IEZNN

$$\begin{aligned} \dot{X}(t)A(t)A^T(t) &= -X(t)(A(t)A^T(t) + A(t)\dot{A}^T(t)) + \dot{A}^T(t) \\ &\quad - \eta(X(t)A(t)A^T(t) - A^T(t)) \\ &\quad - \kappa \int_0^t (X(\tau)A(\tau)A^T(\tau) - A^T(\tau))d\tau \quad (1.10.1). \end{aligned}$$

Ενώ αν το  $\text{rank}(A(t)) = n = \min(m,n)$  θα έχουμε

$$\begin{aligned} A^T(t)A(t)\dot{X}(t) &= -(\dot{A}^T(t)A(t) + A(t)^T \dot{A}(t))X(t) + \dot{A}^T(t) \\ &\quad - \eta(A^T(t)A(t)X(t) - A^T(t)) \\ &\quad - \kappa \int_0^t (A^T(\tau)A(\tau)X(\tau) - A^T(\tau))d\tau \quad (1.10.2). \end{aligned}$$

B. Μοντέλο OZNN

## 1° Κεφάλαιο

Ορίζουμε αρχικά μια συνάρτηση σφάλματος . Για την περίπτωση δεξιού αντίστροφου πίνακα ,  $\text{rank}(A(t))=m= \min(m,n)$  και ικανοποιείται η εξής σχέση  $A^\dagger(t)A(t)A^T(t) - A^T(t) = 0$  και έτσι η συνάρτηση σφάλματος μπορεί να γραφτεί ως εξής

$$\varepsilon(t) = X(t)A(t)A^T(t) - A^T(t) \quad (1.10.3).$$

Όπου  $X(t)$  είναι η λύση του ΟΖΝΝ. Έπειτα με βάση τον κανόνα εξέλιξης το μοντέλο γράφεται και έτσι

$$\dot{X}(t)A(t)A^T(t) = -X(t)(\dot{A}(t)A^T(t) + A(t)\dot{A}^T(t)) + \dot{A}^T(t) - \eta(X(t)A(t)A^T(t) - A^T(t)) \quad (1.10.4).$$

Αντίστοιχα αν  $\text{rank}(A(t)) = \{n \min(m,n)\}$  δηλαδή του αριστερού αντίστροφου πίνακα η συνάρτηση σφάλματος θα είναι

$$\varepsilon(t) = A^T(t)A(t)X(t) - A^T(t) \quad (1.10.5).$$

Και το μοντέλο μας θα γράφεται ως εξής

$$A^T(t)A(t)\dot{X}(t) = -(\dot{A}^T(t)A(t) + A^T(t)\dot{A}(t))X(t) + \dot{A}^T(t) - \eta(A^T(t)A(t)X(t) - A^T(t)) \quad (1.10.6).$$

## Κεφάλαιο 2ο:

### 2.1 Εισαγωγή

Τελευταία, τα νευρωνικά δίκτυα έχουν επιδείξει τεράστιες δυνατότητες ως παράλληλα κατανεμημένα υπολογιστικά μοντέλα για την επίλυση πολλών υπολογιστικά απαιτητικών προβλημάτων, όπως διάφορα είδη προβλημάτων βελτιστοποίησης με περιορισμούς [9]- [10].

Πρόσφατα, σημαντικός αριθμός αποτελεσμάτων σχετικών με την εφαρμογή νευρωνικών δικτύων στην επίλυση ποικίλων προβλημάτων γραμμικής άλγεβρας έχει δημοσιευθεί. Έχουν εισαχθεί διαφορετικοί τύποι νευρωνικών δικτύων για την επίλυση συστημάτων γραμμικών αλγεβρικών εξισώσεων.

Οι Cichocki και Unbehauen [11] και οι Wang και Li [12] σχεδίασαν επαναλαμβανόμενα νευρωνικά δίκτυα (RNNs) για την επίλυση ταυτόχρονων γραμμικών αλγεβρικών εξισώσεων.

Ο Wang [13]–[15] πρότεινε ένα νευρωνικό δίκτυο βαθμίδας για την επίλυση ταυτόχρονων γραμμικών εξισώσεων.

Στο [15] επαληθεύθηκε ότι τα προτεινόμενα RNNs είναι σταθερά υπό ευρείες συνθήκες και ικανά να υπολογίζουν αντίστροφους πίνακες και να επιλύουν εξισώσεις Lyapunov. Δύο τριδιάστατα δομημένα δίκτυα για την επίλυση γραμμικών εξισώσεων και της εξίσωσης Lyapunov αναπτύχθηκαν στο [16].

Πλήθος άλλων προβλημάτων γραμμικής άλγεβρας έχουν ήδη λυθεί χρησιμοποιώντας νευρωνικά δίκτυα. Πολλά μη γραμμικά και γραμμικά μοντέλα RNN έχουν αναπτυχθεί για την αντιστροφή και τη γενικευμένη αντιστροφή τετραγωνικών πινάκων ([15], [16]–[18]).

Ένας νέος τύπος σύνθετης τιμής (ZNN), που βασίζεται σε μια μιγαδική συνάρτηση Zhang, στο [19]. Μοντέλα ZNN για την μεταβολή του Moore–Penrose αντιστρόφου χρονικά μεταβαλλόμενων πινάκων πλήρους βαθμίδας έχουν αναλυθεί στο [20].

Το RNN για την αντιστροφή πινάκων, που έχει αναπτυχθεί στο [18], αποτελούταν από έναν αριθμό ανεξάρτητων υποδικτύων που αντιστοιχούσαν στις στήλες του αντίστροφου πίνακα. Αντίστοιχα, και το νευρωνικό δίκτυο αποτελείται από έναν αριθμό ανεξάρτητων υποδικτύων που αντιστοιχούν στις στήλες του Drazin αντιστρόφου. Συνοπτικά συγκρίνω την προσέγγιση RNN για τον υπολογισμό του Drazin αντιστρόφου με τις επαναληπτικές μεθόδους που στοχεύουν στον ίδιο σκοπό. Το βασικό πλεονέκτημα του RNN είναι στην εύρεση της αρχικής συνθήκης. Συγκεκριμένα, οι επαναληπτικές μέθοδοι για τον υπολογισμό του Drazin αντιστρόφου επιβάλλουν ορισμένες συνθήκες αναγκαίες για τη σύγκλιση της επαναληπτικής διαδικασίας [23].

Οι συνθήκες του [23] είναι αρκετά αυστηρές και συχνά δεν μπορούν να ικανοποιηθούν — π.χ. όταν η νόρμα του υπολοίπου που απαιτείται για την αρχική επανάληψη είναι  $<1$ . Αντιθέτως, προτείνονται πιο βολικές συνθήκες, που εξασφαλίζουν σύγκλιση και σταθερότητα του προτεινόμενου δικτύου. Αυτές οι συνθήκες επιτυγχάνονται απλά μέσω κατάλληλης επιλογής δύναμης του πίνακα εισόδου. Επιπλέον, το μοντέλο μας διασπά ολόκληρη τη διαδικασία σε ανεξάρτητες υποδιαδικασίες, που εφαρμόζονται στις στήλες των προσεγγίσεων.

Με αυτόν τον τρόπο, είναι δυνατή η αξιοποίηση υποδικτύων που προσεγγίζουν τις στήλες του Drazin αντιστρόφου ανεξάρτητα και τα οποία μπορούν να λειτουργούν ταυτόχρονα σε πραγματικό χρόνο. Πιο συγκεκριμένα, η  $i$ -οστή στήλη ενός πίνακα που δίνεται  $A$ ,  $1 \leq i \leq n$ , μετασχηματίζεται σε μια προσέγγιση της  $i$ -οστής στήλης του Drazin αντιστρόφου χρησιμοποιώντας το αντίστοιχο  $i$ -οστό υποδίκτυο, ανεξάρτητα από τα υπόλοιπα.

Αυτή η προσέγγιση με την κατάλληλη υλοποίηση υλικού υλοποιείται με τρεις τελεστικούς ενισχυτές: 1) αθροιστικό, 2) ολοκληρωτικό, 3) αναστρέφοντα [18] και μπορεί να προσφέρει σημαντικά υπολογιστικά πλεονεκτήματα έναντι των υφιστάμενων σειριακών αλγορίθμων στην επίλυση προβλημάτων Drazin αντιστρόφου μεγάλης κλίμακας σε εφαρμογές πραγματικού χρόνου.

Αυτό δεν είναι εφικτό με τους συνηθισμένους επαναληπτικούς αλγορίθμους λόγω της φύσης τους ως σειριακές διαδικασίες. Στο [23] προτείνεται μια αρχιτεκτονική νευρωνικού δικτύου προώθησης προς τα εμπρός για τον υπολογισμό του Drazin αντιστρόφου. Σε σύγκριση με την προσέγγισή αυτή, η συγκεκριμένη αρχιτεκτονική δεν επιτρέπει πλήρη παραλληλοποίηση της διαδικασίας, αφού όλα τα διανύσματα και οι μεταβλητές πινάκων πρέπει να είναι διαθέσιμα κατά τη διαδικασία εκπαίδευσης.

Αυτό μπορεί να οδηγήσει σε δυσκολίες όταν αντιμετωπίζονται προβλήματα μεγάλης κλίμακας που απαιτούν μεγάλη υπολογιστική ισχύ και μνήμη. Προτείνεται ένα RNN με γραμμικές συναρτήσεις ενεργοποίησης.

Η δυναμική εξίσωση που στηρίζει αυτό το νευρωνικό δίκτυο δείχνει ότι κάθε υποδίκτυο είναι ουσιαστικά όμοιο με το RNN που παρουσιάστηκε στο [12].

Οι Wang και Li [12] απέδειξαν ότι τα νευρωνικά δίκτυα με γραμμικές συναρτήσεις ενεργοποίησης είναι ικανά να επιλύουν γραμμικές εξισώσεις, είναι ταχύτερα στη σύγκλιση και ευκολότερα στην υλοποίηση σε σχέση με τα νευρωνικά δίκτυα με μη γραμμικές συναρτήσεις ενεργοποίησης. Επιπλέον, η καθιερωμένη θεωρία δυναμικών γραμμικών συστημάτων καθιστά την ανάλυση των γραμμικών νευρωνικών δικτύων πολύ ευκολότερη.

## 2.2 Βασικές έννοιες, Κίνητρο και Αντικείμενο της μελέτης

Για κάθε πίνακα  $A \in \mathbb{C}^{n \times n}$ , οι στήλες και ο μηδενικός χώρος συμβολίζονται με  $R(A)$  και  $N(A)$ , αντίστοιχα. Ο δείκτης του πίνακα συμβολίζεται με  $\text{Ind}(A)$  και ορίζεται ως ο μικρότερος μη αρνητικός ακέραιος  $k$  για τον οποίο ισχύει η σχέση:

$$\text{rank}(A^k) = \text{rank}(A^{k+1}) \quad (2.2.1).$$

οπού  $\text{rank}(A)$  δηλώνει τον βαθμό του πίνακα  $A$ .

Έστω  $\sigma(A)$  το φάσμα του πίνακα  $A$  και

$$s(A) = \text{Re}(\sigma(A)) = \{\text{Re}(\lambda) : \lambda \in \sigma(A)\} \quad (2.2.2).$$

τότε η συνθήκη  $s(A) \geq 0$  θα σημαίνει ότι  $\text{Re}(\lambda_i) \geq 0$ ,  $i=1,2,\dots,n$

Ένας τετραγωνικός πίνακας λέγεται Drazin αντίστροφος ενός πίνακα  $A \in \mathbb{C}^{n \times n}$  όταν ικανοποιεί τις ακόλουθες τρεις εξισώσεις:

$$A^{l+1}X = A^l, l \geq \text{Ind}(A), XAX = X, AX = XA \quad (1) \quad (2.2.3).$$

Συνήθως χρησιμοποιούμε την σημείωση  $X = A^D$  για να συμβολίσει τον Drazin αντίστροφο του πίνακα  $A$ . Επιπλέον όταν έχουμε  $k = 1$ , ο Drazin αντίστροφος ταυτίζεται με τον αντίστροφο  $A^\#$ .

Πρόσφατα με βάση την Ji [24] αποδείχτηκε μια απλή αναπαράσταση του Drazin αντίστροφου  $A^D$  για έναν πίνακα  $A \in \mathbb{C}^{n \times n}$  ως εξής:

**Λήμμα (2.2.1.1) [24]**

Έστω  $A \in \mathbb{C}^{n \times n}$   $l \geq k$  τότε

$$A^D = \lim_{\lambda \rightarrow 0} (A + \lambda I)^{(-1+1)} A^l, -\lambda \notin \sigma(A) \quad (2.2.4).$$

Ωστόσο μια εναλλακτική έκφραση όσον αφορά τον αντίστροφο δίνεται στο Λήμμα (2.2.2.2) και μπορούμε να πάρουμε τα αποτελέσματα αυτού χρησιμοποιώντας το [25].

**Λήμμα (2.2.2.2)[26]**

Η κλειστή μορφή του  $A^D$  για έναν τετραγωνικό πίνακα  $A \in \mathbb{C}^{n \times n}$  είναι:

$$A^D = \lim_{\lambda \rightarrow 0} (A^l (A^{2l+1})^H A^{l+1} + \lambda I)^{-1} A^l (A^{2l+1})^H A^l, l \geq k \quad (2.2.5).$$

Προχωρώντας αναπτυχθήκαν μοντέλα GNN για τον υπολογισμό του Drazin αντίστροφου ενός σταθερού πίνακα, βασισμένα στις αναπαραστάσεις των Λημμάτων 2.2.2.1 και 2.2.2.2

Διάφορες συναρτήσεις παρακολούθησης σφάλματος και τα αντίστοιχα GNN με μη γραμμική ενεργοποίηση διαμορφώνονται και αναλύονται.

Αυτή η προσέγγιση πρόσφερε ένα σημαντικό πλεονέκτημα σε σχέση με τα προηγούμενα GNN μοντέλα για τον υπολογισμό του Drazin αντίστροφου και συγκεκριμένα η σταθερότητα του RNN που προτείνεται στο [26] εξασφαλίζεται μόνο όταν τα πραγματικά μέρη των ιδιοτιμών του  $A^{l+1} \in \mathbb{R}^{n \times n}$  είναι θετικά δηλαδή :

$$\text{Re}(\sigma(A^{l+1})) \geq 0, l \geq k \quad (2.2.6).$$

Η οποία είναι μια διαδικασία η οποία μας διευκολύνει στην εκτίμηση μας όσον αφορά την σωστή βαθμίδα δύναμης του Drazin που προτάθηκε στο [26]. Μια εναλλακτική λύση η οποία προτάθηκε στο [21] και βασίζεται στη δυναμική εξίσωση κατάστασης:

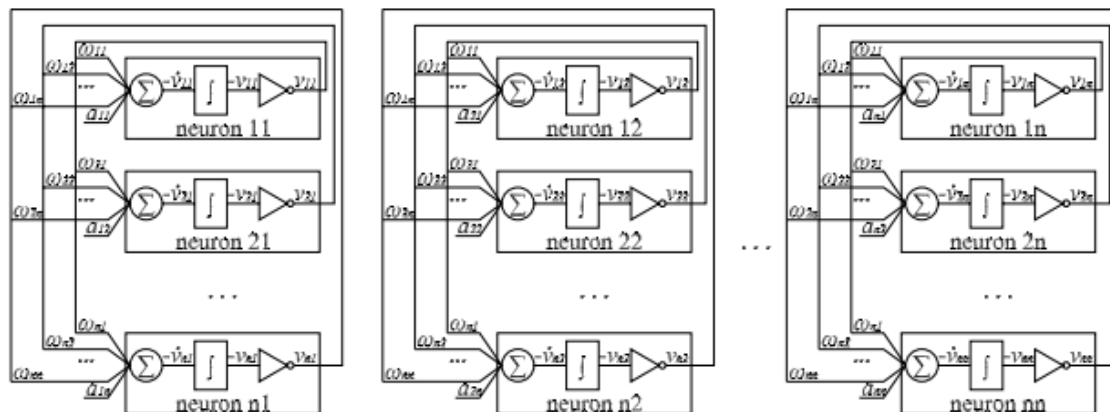
$$\frac{dV}{dt} = -\beta G(AV(t) - I), V(0) = 0, G = A^k (A^{2k+1})^T A^k, k = \text{ind}(A) \quad (2.2.7).$$

Ένα μείζον μειονέκτημα της δυναμικής εξίσωσης είναι ο μεγάλος αριθμός πράξεων πινάκων που απαιτεί. Επιπλέον, τα στοιχεία που εμπεριέχονται στο γινόμενο πινάκων  $G$  αυξάνονται, κάτι που μπορεί να προκαλέσει αριθμητική αστάθεια.

Στην παρούσα εργασία υπάρχει μια φυσική και υπολογιστικά οικονομικότερη λύση, βασισμένη στις οριακές αναπαραστάσεις του Drazin αντίστροφου.

Πιο συγκεκριμένα, επιλέγουμε μια μικρή πραγματική παράμετρο  $\lambda$  και έναν όρο μετατόπισης  $\lambda I$ , ο οποίος εξασφαλίζει την αντιστρεψιμότητα των υποκείμενων πινάκων σύμφωνα με την αρχή της κανονικοποίησης Tikhonov.

Εφόσον η συνάρτηση ενεργοποίησης αποτελεί θεμελιώδες τμήμα ενός νευρωνικού δικτύου, η δεύτερη επιδίωξή μας είναι να εξετάσουμε θεωρητικά και αριθμητικά την επίδραση διαφορετικών συναρτήσεων ενεργοποίησης στα GNN μοντέλα που ορίζονται στην παρούσα εργασία.



Σχήμα 2.1 Αρχιτεκτονική δικτύου RNN για τον υπολογισμό του Drazin αντίστροφου [ 79 ]

Με άλλα λόγια, η επιρροή διαφορετικών ενεργοποιήσεων στα GNN μοντέλα μπορεί να μας προσφέρει μια εναλλακτική απάντηση στα προβλήματα και τις παρενέργειες που προκύπτουν από ατελείς υλοποιήσεις και μη γραμμικότητες που εμφανίζονται σε γραμμικές ενεργοποιήσεις [28].

Η ενεργοποίηση σε μορφή πίνακα  $F_e = (e_{ij})$  ορίζεται ως πίνακας  $(f(e_{ij}))$ ,  $i, j = 1, 2, \dots, n$  όπου  $f(\bullet)$  είναι μια αύξουσα και περιττή βαθμωτή συνάρτηση.

Γενικά οποιαδήποτε μονοτονικά αύξουσα και περιττή συνάρτηση ενεργοποίησης  $f(\bullet)$  και να έχουμε μπορεί να χρησιμοποιηθεί για κατασκευή νευρωσικού δικτύου.

Όπως έχει φανεί ο ρυθμός σύγκλισης βελτιώνεται σημαντικά όταν γίνεται κατάλληλη επιλογή της συνάρτησης ενεργοποίησης.

Έτσι αποδεικνύεται θεωρητικά ότι οι μη γραμμικές ενεργοποιήσεις των GNN μοντέλων συγκλίνουν στον θεωρητικό Drazin αντίστροφο. Επιπλέον παρουσιάζεται σύγκλιση τόσο θεωρητικά όσο και από αριθμητικά παραδείγματα στην περίπτωση όπου χρησιμοποιείται η Δυναμική σιγμοειδής συνάρτηση ενεργοποίησης συνάρτηση ενεργοποίησης σε σχέση με το γραμμικό μοντέλο. Επιπλέον η αποτελεσματικότητα των μοντέλων για τον υπολογισμό του Drazin αντιστρόφου επαληθεύεται πλήρως μέσω παραδειγμάτων προσομοίωσης στον υπολογιστή.

Το αρχικό στάδιο της έρευνας ήταν η επαναληπτική μέθοδος που προκύπτει από τον αλγόριθμο βαθμιαίας καθόδου για τη μοναδική λύση ελάχιστης Frobenius νόρμας της αρχικής εξίσωσης  $AXA=A$  η οποία ορίστηκε και μελετήθηκε όπως είπαμε και παραπάνω στο [34]. Η επαναληπτική μέθοδος παράγει το  $A^\dagger$  εάν η αρχική επανάληψη επιλεγεί ως  $X_0 = A^*$ .

Η βασική ιδέα του βαθμωτού αλγορίθμου είναι η αναζήτηση του ελαχίστου της ακόλουθης συνάρτησης που ορίζεται με τη νόρμα Frobenius:

$$J(X) = \frac{1}{2} \|A - AXA\|_F^2 \quad (2.2.8).$$

Εάν

$$\frac{\partial J(t)}{\partial X} = -A^*(A - AXA)A^* \quad (2.2.9).$$

οι επαναληπτικές σχέσεις που έγιναν στο [34] οριστούν και από τον κανόνα

$$X_{k+1} = X_k - \gamma \frac{\partial J(t)}{\partial X} = X_k + \gamma A^*(A - AX_k A) A^* \quad (2.2.10).$$

όπου  $\gamma$  είναι η πραγματική παράμετρος που ονομάζεται παράγοντας σύγκλισης.

Γενικά, η σύγκλιση των επαναληπτικών μεθόδων προκαλείται και καθορίζεται από την επιλογή της αρχικής προσέγγισης  $X_0$ . Πιο συγκεκριμένα, μια κατάλληλη  $X_0$  είναι αναγκαία για να εξασφαλιστεί η σύγκλιση της μεθόδου και να οδηγήσει στη λύση ενός συγκεκριμένου γενικευμένου αντιστρόφου.

### Θεώρημα 2.2.1 [51]

Έστω  $A \in R^{n \times n}$  ένας ιδιάζων πίνακας που μας έχει δοθεί και έστω  $\sigma(A) = \{\lambda_1, \dots, \lambda_n\}$  το φάσμα του  $A$ . Τότε, το νευρωνικό δίκτυο κλίσης που περιγράφεται στο 2.2.2 είναι ευσταθές κατά Lyapunov.

**Παράδειγμα 2.2.1** Δανειζόμαστε τον πίνακα από το [51]

$$A = \begin{bmatrix} 2 & -1.6 & 5.6 & -5.6 & 0 & 5.6 \\ 0 & 1 & 6 & -6 & 0 & 6 \\ 0 & 0 & 4 & -4 & 0.1 & 3.9 \\ 0 & 0 & 0 & 0 & 0.1 & -0.1 \\ 0 & 0 & 0 & 0 & 0.1 & 0.1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

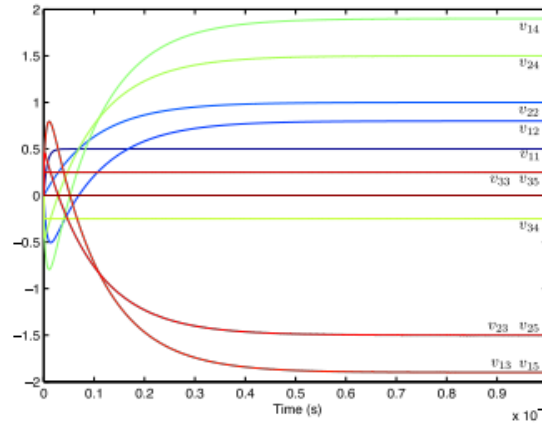
με  $\text{ind}(A)=3$ . Ο ακριβής Drazin αντίστροφος του  $A$  είναι ίσος με

$$A^D = \begin{bmatrix} 0.5000 & 0.8000 & 1.9000 & 1.9000 & 0 & -1.9000 \\ 0 & 1.0000 & 1.5001 & 1.5001 & 0 & -1.5001 \\ 0 & 0 & -0.2500 & -0.2500 & 0 & 0.2500 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Οι ιδιοτιμές του πίνακα  $A$  είναι  $\{2, 1, 4, 0, 0.1, 0\}$ . Δεδομένου ότι ο πίνακας  $A^4$  έχει όλες τις ιδιοτιμές μη αρνητικές μπορούμε να επιλέξουμε  $k=\text{ind}(A)=3$ .

Επιλέγουμε  $\beta=10^7$  αρά ο πίνακας βαρών σύνδεσης  $W = -10^7 A^4$  έχει όλες τις ιδιοτιμές μη θετικές. Οι τροχιές για μηδενικές αρχικές συνθήκες παρουσιάζονται στο Σχήμα και καταδεικνύουν συμπεριφορά σύγκλισης σε χρόνο. Ο πίνακας που παράγεται από τον αλγόριθμο είναι

$$A^d = \begin{bmatrix} 0.5000 & 0.8000 & 1.9000 & 1.9000 & 0 & -1.9000 \\ 0 & 1.0000 & 1.5001 & 1.5001 & 0 & -1.5001 \\ 0 & 0 & -0.2500 & -0.2500 & 0 & 0.2500 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$



Σχήμα 2.2: Συμπεριφορά σύγκλισης σε  $10^{-6}$  δευτερόλεπτα από παράδειγμα 2.2.1 [80]

**Θεώρημα 2.2.2(α) [61]**

Έστω ένας δεδομένος πίνακας  $A \in \mathbb{R}^{n \times n}$  με δείκτη  $k = \text{ind}(A)$  με ιδιοτιμές να είναι  $\sigma(A) = \{\lambda_j, j = 1, 2, \dots, n\}$  και για κάθε  $\lambda_j$  θεωρούμε ότι οι πιθανές τιμές του εκθέτη  $m+1$  που εξασφαλίζουν το  $\text{Re}(\lambda_j^{m+1}) \geq 0$ . Το σύνολο  $M_2^j$  ορίζεται ως  $M_2^j = [l, +\infty]$ ,  $l \geq \text{ind}(A)$  για κάθε  $j = 1, 2, \dots, n$ .

Έστω  $M_2^j$  το σύνολο τιμών της παραμέτρου  $m$  που ορίζονται με βάση του περιορισμούς από την γνωστή τιμή  $\lambda_j \in \sigma(A)$ . Όμοια με  $M_3^j, M_4^j$  θα συμβολίσουμε τις αποδέκτες τιμές της παραμέτρου  $m$ .

Επομένως:

$$M_2^j = \{m + 1 = 2l : l \in \mathbb{N}^+\},$$

$$M_3^j = \{m + 1 = 4l : l \in \mathbb{N}^+\},$$

$$M_4^j = \{m : \frac{4l - 1}{2} \frac{\pi}{|\varphi_j|} < m + 1 < \frac{4l + 1}{2} \frac{\pi}{|\varphi_j|}, l \in \mathbb{N}_0\}$$

Στην συνέχεια οι θετικοί αριθμοί που αντιστοιχούν στο  $\lambda_j$  τους συμβολίζουμε με  $U_j$ . Έτσι εύκολα συμπεραίνουμε ότι  $U_j = M_2^j$  ή  $U_j = M_3^j$  ή  $U_j = M_4^j$ . Το RNN που ορίζεται είναι σταθερό και ο πίνακας μόνιμης κατάστασης είναι ίσος με  $A^D$ :

$$\lim_{t \rightarrow \infty} V(t) = A^D \quad (2.2.11).$$

**Παράδειγμα 2.2.2 (α)** Περιλαμβάνοντας τον πίνακα από το [61]

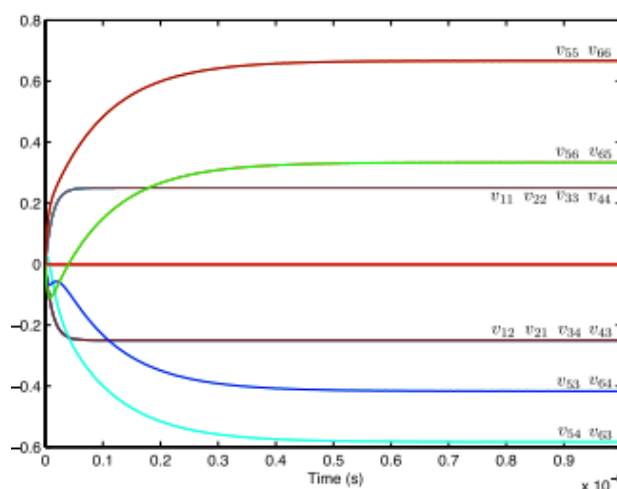
$$A = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ -1 & -1 & 1 & -1 & 0 & 0 \\ -1 & -1 & -1 & 1 & 0 & 0 \\ -1 & -1 & -1 & 0 & 2 & -1 \\ -1 & -1 & 0 & -1 & -1 & 2 \end{bmatrix}$$

Το φάσμα του  $A$  είναι  $\sigma(A) = \{3, 2, 2, 1, 0, 0\}$  και χρησιμοποιώντας το  $m = \text{ind}(A) = 2$  και  $\beta = 10^7$ , η μέθοδος θα παράγει την προσέγγιση του Drazin αντίστροφου μετά από  $10^{-7}$ s:

$$A^d = \begin{bmatrix} 0.2500 & 0.2500 & 0 & 0 & 0 & 0 \\ 0.2500 & 0.2500 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.2500 & -0.2500 & 0 & 0 \\ 0 & 0 & 0.2500 & 0.2500 & 0 & 0 \\ 0 & 0 & -0.4166 & -0.5833 & 0.6666 & 0.3333 \\ 0 & 0 & -0.5833 & -0.4166 & 0.3333 & 0.6666 \end{bmatrix}$$

Η συμπεριφορά σύγκλισης του δικτύου σε χρόνο  $10^{-7}$ s φαίνεται γραφικά στο σχήμα 2.3

$$A^D = \begin{bmatrix} \frac{1}{4} & -\frac{1}{4} & 0 & 0 & 0 & 0 \\ -\frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} & -\frac{1}{4} & 0 & 0 \\ 0 & 0 & -\frac{1}{4} & \frac{1}{4} & 0 & 0 \\ 0 & 0 & -\frac{5}{12} & -\frac{7}{12} & \frac{2}{3} & \frac{1}{3} \\ 0 & 0 & -\frac{7}{12} & -\frac{5}{12} & \frac{1}{3} & \frac{2}{3} \end{bmatrix}$$



Σχήμα 2.3: Συμπεριφορά σύγκλισης σε  $10^{-6}$  δευτερόλεπτα από παράδειγμα 2.2.2(α) [81]

### Θεώρημα 2.2.2(β) [61]

Έστω ένας δεδομένος πίνακας  $A \in R^{n \times n}$  με δείκτη  $k = \text{ind}(A)$  με ιδιοτιμές να είναι  $\sigma(A) = \{\lambda_j, j = 1, 2, \dots, n\}$  και για κάθε  $\lambda_j$  θεωρούμε ότι οι πιθανές τιμές του εκθέτη  $m+1$  που εξασφαλίζουν το  $\text{Re}(\lambda_j^{m+1}) \geq 0$ . Το σύνολο  $M_2^j$  ορίζεται ως  $M_2^j = [l, +\infty]$ ,  $l \geq \text{ind}(A)$  για κάθε  $j = 1, 2, \dots, n$ .

Έστω  $M_2^j$  το σύνολο τιμών της παραμέτρου  $m$  που ορίζονται με βάση του περιορισμούς από την γνωστή τιμή  $\lambda_j \in \sigma(A)$ . Ομοια με  $M_3^j, M_4^j$  θα συμβολίσουμε τις αποδέκτες τιμές της παραμέτρου  $m$ .

Επομένως:

$$M_2^j = \{m + 1 = 2l : l \in \mathbb{N}^+\}$$

$$M_3^j = \{m + 1 = 4l : l \in \mathbb{N}^+\}$$

$$M_4^j = \{m: \frac{4l-1}{2} \frac{\pi}{|\varphi_j|} < m+1 < \frac{4l-1}{2} \frac{\pi}{|\varphi_j|}, l \in N_0\}$$

Στην συνέχεια οι θετικοί αριθμοί που αντιστοιχούν στο  $\lambda_j$  τους συμβολίζουμε με  $U_j$ . Έτσι ευκολά συμπεραίνουμε ότι  $U_j = M_2^j$  ή  $U_j = M_3^j$  ή  $U_j = M_4^j$ . Το RNN που ορίζεται είναι σταθερό και ο πίνακας μόνιμης κατάστασης είναι ίσος με  $A^D$ :

$$\lim_{t \rightarrow \infty} V(t) = A^D \quad (2.2.12).$$

Περίπτωση 1:  $\sigma(A) \subset \mathbb{R}, \lambda_j < 0$  για τουλάχιστον έναν δείκτη  $j$  και το  $m$  να ικανοποιεί την σχέση  $m+1 \geq \text{ind}(A)$ , ο  $m+1$  είναι άρτιος.

**Παράδειγμα 2.2.2 (β) [61]**

Θεωρώντας έναν διαφορετικό ιδιάζων πίνακα από το [33] με δείκτη  $\text{ind}(A)=4$

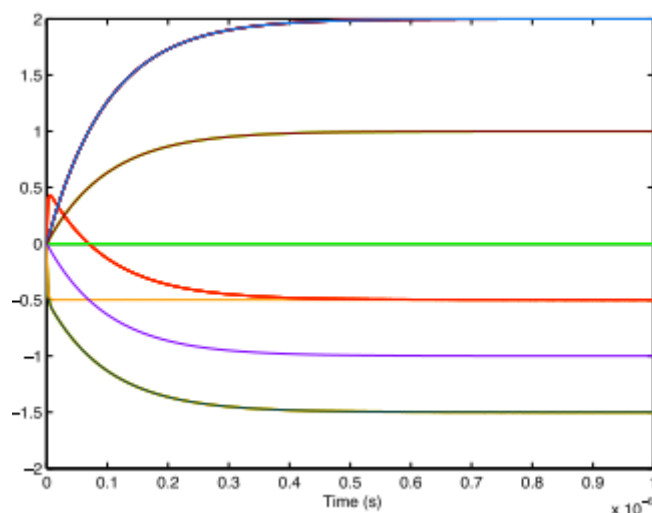
$$A = \begin{bmatrix} 3 & -1 & -1 & -1 & 1 & -3 & 1 & 2 \\ 3 & -1 & -1 & -1 & 1 & -3 & 1 & 2 \\ 2 & 0 & -1 & -1 & 1 & -3 & 1 & 2 \\ 1 & 0 & 0 & -1 & 1 & -3 & 1 & 2 \\ 0 & 0 & 0 & 0 & 1 & -3 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & -2 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Υπάρχει το  $\sigma(A) = \{0,0,0,0,1,-2,-1,1\}$ . Επειδή ο  $A$  έχει αρνητικές ιδιοτιμές μπορούμε να πάρουμε το  $m = 5$  επειδή ο πίνακας  $A^{m+1} = A^{5+1} = A^6$  και έχει μόνο μη αρνητικές ιδιοτιμές. Έτσι χρησιμοποιώντας  $\beta = 10^7$  το νευρωνικό δίκτυο παράγει την λύση

$$A^d = \begin{bmatrix} 0 & 0 & 0 & 0 & 1.0000 & -1.4999 & -0.5000 & 1.9999 \\ 0 & 0 & 0 & 0 & 1.0000 & -1.4999 & -0.5000 & 1.9999 \\ 0 & 0 & 0 & 0 & 1.0000 & -1.4999 & -0.5000 & 1.9999 \\ 0 & 0 & 0 & 0 & 1.0000 & -1.4999 & -0.5000 & 1.9999 \\ 0 & 0 & 0 & 0 & 1.0000 & -1.4999 & -0.5000 & 1.9999 \\ 0 & 0 & 0 & 0 & 1.0000 & -0.5000 & -0.5000 & 1.9999 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1.0000 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1.0000 \end{bmatrix}$$

Ο ακριβής αντίστροφος Drazin του  $A$  είναι:

$$A^D = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & -\frac{3}{2} & -\frac{1}{2} & 2 \\ 0 & 0 & 0 & 0 & 1 & -\frac{3}{2} & -\frac{1}{2} & 2 \\ 0 & 0 & 0 & 0 & 1 & -\frac{3}{2} & -\frac{1}{2} & 2 \\ 0 & 0 & 0 & 0 & 1 & -\frac{3}{2} & -\frac{1}{2} & 2 \\ 0 & 0 & 0 & 0 & 1 & -\frac{3}{2} & -\frac{1}{2} & 2 \\ 0 & 0 & 0 & 0 & 0 & -\frac{1}{2} & -\frac{1}{2} & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$



Σχήμα 2.4: Συμπεριφορά σύγκλισης σε  $10^{-6}$  δευτερόλεπτα από παράδειγμα 2.2.2 [82]

### Θεώρημα 2.2.4

Μερικές φορές είναι δύσκολο να βρεθεί μια αρχική τιμή που να ικανοποιεί τις απαιτούμενες συνθήκες. Επιπλέον, η σύγκλιση των επαναληπτικών μεθόδων μπορεί να είναι αργή όταν οι αρχικές επαναλήψεις απέχουν πολύ από τη λύση [35].

Επίσης, η σύγκλιση επαναληπτικών μεθόδων για τον υπολογισμό του Drazin αντιστρόφου είναι μερικές φορές αργή και επηρεάζεται αρνητικά από τη χρήση της δύναμης του πίνακα [36]. Η πιο σωστή κίνηση είναι να γίνει ανάπτυξη και διερεύνηση ενός GNN μοντέλου για την διαδικτυακή κατασκευή του Moore–Penrose αντιστρόφου ενός σταθερού πίνακα, ως συνεχούς χρόνου αναλογία της επαναληπτικής μεθόδου .

Επιπλέον, ορίζεται ένα ανάλογο GNN μοντέλο για τον υπολογισμό του Drazin αντιστρόφου. Το κύριο πλεονέκτημα των GNN μοντέλων που ορίζονται είναι η παγκόσμια και ταχεία σύγκλισή τους.

Η δυναμική των GNN μοντέλων για τον υπολογισμό γενικευμένων αντιστρόφων βασίζεται στη χρήση της βαθμωτής συνάρτησης σφάλματος βασισμένης σε νόρμα

$$e(t) = e(V(t)) = \frac{1}{2} \|E(t)\|_F^2 \quad (2.2.13).$$

### Απόδειξη

Οπού  $E(t)$  είναι ένας πίνακας σφάλματος και  $\|A\|_F := \sqrt{\text{Tr}(A^T A)}$  δηλώνει την Frobenius νόρμα του πίνακα  $A$  και  $\text{Tr}(\bullet)$  δηλώνει το ίχνος του πίνακα. Ο γενικός τύπος σχεδίασης ορίζεται κατά μήκος του αρνητικού όρου  $-\partial e(V(t))/\partial V$  του  $e(V(t))$  ως εξής:

$$\frac{dV(t)}{dt} = -\gamma \frac{\partial e(V(t))}{\partial V} \quad (2.2.14).$$

Εδώ, το  $V(t)$  είναι ο πίνακας των μεταβλητών κατάστασης ενεργοποίησης, το  $t \in [0, +\infty)$  είναι μια χρονική παράμετρος και το  $\gamma$  είναι θετική σταθερά κλιμάκωσης, της οποίας οι τιμές πρέπει να εναρμονίζονται με το επιλεγμένο χρονικό διάστημα.

Η δυναμική εξίσωση του (RNN) για την αντιστροφή ενός πίνακα προκύπτει από τον πίνακα σφάλματος  $E(t) = AV(t) - I$  και προτάθηκε στο [37]. Η ίδια αρχή επεκτάθηκε για τον υπολογισμό του

Moore–Penrose ενός ορθογώνιου πίνακα πλήρους στήλης  $A \in R_n^{m \times n}$  η ενός ορθογώνιου πίνακα πλήρους γραμμής  $A \in R_m^{m \times n}$  και αυτό οδήγησε στην δημιουργία του μοντέλου παρακάτω:

$$\dot{V}(t) = \frac{dV(t)}{dt} = \begin{cases} -\gamma A^T (AV(t) - I) = -\frac{\gamma}{2} \frac{\partial \|AV(t) - I\|_F^2}{\partial V}, m \leq n, V(0) = V_0 \\ -\gamma (V(t)A - I)A^T = -\frac{\gamma}{2} \frac{\partial \|V(t)A - I\|_F^2}{\partial V}, m > n, V(0) = V_0 \end{cases} \quad (2.2.15).$$

Στο [42], αποδείχτηκε η σφαιρική εκθετική σύγκλιση και η σταθερότητα του νευρωνικού μοντέλου βασισμένου στη μεθοδολογία βαθμίδωσης της εξίσωσης για τον διαδιδυακό υπολογισμό της αντίστροφης μήτρας.

Ο Wang έδειξε στο [38] ότι το μοντέλο μπορεί επίσης να χρησιμοποιηθεί για τον υπολογισμό του Moore–Penrose για πίνακες ελλιπούς βαθμού, υπό την προϋπόθεση ότι η αρχική κατάσταση είναι μηδενική, δηλαδή  $V(0)=0$ .

Η δυναμική εξίσωση κατάστασης του επαναληπτικού νευρωνικού δικτύου (RNN) για τον υπολογισμό της Drazin αντίστροφης  $A^D$  προτάθηκε στο [39]. Βασίζεται στη χρήση του Frobenius norm του πίνακα σφάλματος  $E(t) = A^{m+1}V(t) - A^m, m \geq \text{ind}(A)$  χρησιμοποιώντας:

$$\frac{\partial (\frac{1}{2} \|A^{m+1}V(t) - A^m\|_F^2)}{\partial V} = (A^{m+1})^T (A^{m+1}V(t) - A^m) \quad (2.2.16).$$

Και παραλείποντας τον σταθερό όρο  $(A^{m+1})^T$  το γενικό δυναμικό σχήμα οδηγεί στην εξίσωση

$$\frac{dV(t)}{dt} = -\gamma (A^{m+1}V(t) - A^m), m \geq \text{ind}(A), v(0) = 0 \quad (2.2.17).$$

Η εφαρμογή του μοντέλου προϋποθέτει ότι

$$\text{Re}(\lambda_j^{m+1}) \geq 0, j = 1, \dots, n, \quad (2.2.18).$$

όπου  $\sigma(A) = \{\lambda_1, \dots, \lambda_n\}$  είναι το φάσμα του A, και  $m \geq \text{ind}(A)$  [39]. Ένα από τα αποτελέσματα του [39] ήταν ο αλγόριθμος για την εκτίμηση του εκθέτη στο  $A^{m+1}$ , ο οποίος εξασφαλίζει την ιδιότητα .

Μία δυνατότητα για να αντιμετωπιστεί ο περιορισμός προτάθηκε στο [40], και βασίζεται στη χρήση του πίνακα  $A^k (A^{2k+1})^T A^k, k = \text{ind}(A)$ .

Το μοντέλο ZNN για online αντιστροφή χρονικά αμετάβλητου πίνακα βασίζεται στον πίνακα σφάλματος  $E(t) = AV(t) - I$ , και ορίζεται εφαρμόζοντας τον γενικό κανόνα σχεδίασης:

$$\frac{dE(t)}{dt} = -\gamma E(t) = -\gamma (AV(t) - I) \quad (2.2.19).$$

Το οποίο με την σειρά του οδηγεί στην δυναμική

$$A\dot{V}(t) = -\gamma (AV(t) - I) \quad (2.2.20).$$

Η έμμεση δυναμική προτάθηκε αρχικά στο [41] για την αναστροφή ενός χρονικά μεταβαλλόμενου πίνακα  $A(t)$ .

Στο [41] αποδείχθηκε ότι η δυναμική Zhang συγκλίνει καθολικά και εκθετικά προς τη θεωρητική αντίστροφη  $A^{-1}$  ξεκινώντας από οποιαδήποτε αρχική κατάσταση  $X(0)$ , με ρυθμό εκθετικής σύγκλισης ίσο με  $\gamma$ .

Μια νέα μορφή αναδρομικής έμμεσης δυναμικής για πραγματικού χρόνου (real-time) αναστροφή πινάκων προτάθηκε και μελετήθηκε παραπάνω. Η δυναμική αυτή βασίζεται στη χρήση του θετικά ορισμένου πίνακα  $AA^T + I$  στην σχέση ως εξής:

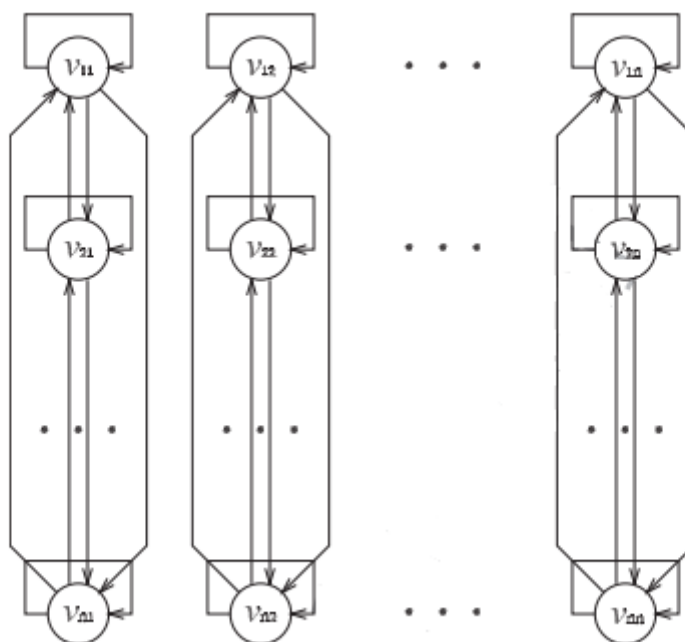
$$A\dot{V}(t) = -\gamma(AA^T + I)(AV(t) - I) \quad (2.2.21).$$

### 2.3 Χρήση νευρωνικών δικτύων για τον υπολογισμό του Drazin αντίστροφου

Η γενική ιδιότητα των ορισμένων νευρωνικών δικτύων για τον υπολογισμό γενικευμένων αντιστρόφων βασίζεται σε ένα πρότυπο: απαιτεί την επίλυση μιας εξίσωσης πίνακα χρησιμοποιώντας τη δυναμική προσέγγιση συστημάτων, η οποία ορίζεται μέσω μιας βαθμωτής συνάρτησης σφάλματος  $E(t)$

Οι προτεινόμενες μέθοδοι προσπαθούν να βρουν το ελάχιστο της συνάρτησης υπολοίπου  $E(t)$  χρησιμοποιώντας τον σχεδιαστικό τύπο:  $dV(t)/dt = -\beta \partial E(t)/\partial V$  [42]-[43].

Οι βασικές έννοιες που θα χρησιμοποιηθούν είναι οι ολοκληρωτικές αναπαραστάσεις του Drazin αντιστρόφου σε συνδυασμό με την εξίσωση δυναμικού συστήματος, η οποία ορίζεται μέσω μιας κατάλληλα ορισμένης συνάρτησης σφάλματος  $E(t)$  και θέτουμε  $G = A^k (A^{2k+1})^T A^k$ ,  $k = \text{ind}(A)$ .



Σχήμα 2.5: Αρχιτεκτονική του RNN για τον υπολογισμό του Drazin αντίστροφου[54]

#### Πρόταση 1

Έστω ότι  $A \in C^{n \times n}$  και  $k = \text{ind}A$  τότε:

$$A^D = \int_0^\infty \exp(-A^k (A^{2k+1})^T A^{k+1} \tau) A^k (A^{2k+1})^T A^k d\tau \quad (2.3.1).$$

Η γενική ιδιότητα των νευρωνικών δικτύων για τον υπολογισμό γενικευμένων αντιστρόφων βασίζεται στο ότι απαιτεί την επίλυση μιας αντιπροσωπευτικής εξίσωσης πινάκων χρησιμοποιώντας τη

δυναμική προσέγγιση συστημάτων, η οποία ορίζεται μέσω μιας βαθμωτής συνάρτησης σφάλματος βασισμένης σε νόρμα,  $E(t)$ .

Οι προτεινόμενες μέθοδοι βρίσκουν το ελάχιστο της συνάρτησης υπολοίπου  $E(t)$  χρησιμοποιώντας τον σχεδιαστικό τύπο  $dV(t)/dt = -\beta \partial E(t)/\partial V$ . Συνεχίζοντας αυτήν την προσέγγιση σε αυτή την εργασία, οι κύριες έννοιες που θα χρησιμοποιηθούν είναι οι ολοκληρωτικές αναπαραστάσεις του αντιστρόφου Drazin σε συνδυασμό με την εξίσωση δυναμικού συστήματος, ορισμένη ως κατάλληλη συνάρτηση σφάλματος βασισμένη σε νόρμα  $E(t)$ . Για απλοποίηση των σημείων εισάγουμε τον πίνακα  $G = A^k (A^{2k+1})^T A^k, k = \text{ind}(A)$ . Προφανώς

$$GAA^D = G \quad (2.3.2).$$

ικανοποιείται και μπορεί να θεωρηθεί ως η εξίσωση πινάκων ως προς τον άγνωστο πίνακα  $V$ :

$$GAV - G = 0 \quad (2.3.3).$$

Η βαθμωτή συνάρτηση σφάλματος βασισμένη σε νόρμα που αντιστοιχεί είναι η εξής:

$$E(t) = \frac{\|GAV(t) - G\|_F^2}{2} \quad (2.3.4).$$

Η ελάχιστη τιμή  $E(t) = 0$  του  $E(t)$  ισχύει αν και μόνο αν ο πίνακας  $V(t)$  αποτελεί την ακριβή λύση της εξίσωσης. Η κατεύθυνση της βαθμίδας για τη  $E(t)$  ορίζεται από τον όρο  $-\partial E(t)/\partial V$

$$\frac{\partial E(t)}{\partial V} = (GA)^T (GAV(t) - G) \quad (2.3.5).$$

Αν αφαιρέσουμε τον σταθερό παράγοντα, η εξίσωση που αντιστοιχεί στην ολοκληρωτική αναπαράσταση θα διατυπωθεί ως εξής:

$$\frac{dV(t)}{dt} = -\beta G(AV(t) - I) = V_0, G = A^k (A^{2k+1})^T A^k, k = \text{ind}(A) \quad (2.3.6).$$

Εδώ έχουμε έναν πίνακα  $V(t)$  που είναι μεταβλητών καταστάσεων ενεργοποίησης  $t \in [0, +\infty)$  και  $\beta$  μια θετική σταθερά κλιμάκωσης.

Το παραπάνω RNN είναι γραμμικά δυναμικό σύστημα σε μορφή πίνακα. Σύμφωνα με την θεωρία γραμμικών συστημάτων, η λύση σε κλειστή μορφή του πίνακα κατάστασης είναι:

$$V(t) = \exp(-\beta GAT) V(0) + \beta \exp(-\beta GAT) \int_0^t \exp(\beta GAT) G d\tau \quad (2.3.7).$$

Εάν ο πίνακας  $A$  είναι θετικά ορισμένος δηλαδή όλες οι τιμές του είναι πραγματικές και θετικές, τότε όλοι οι πίνακες  $A^m, m \in N$  είναι θετικά ορισμένοι.

Επομένως ο πίνακας  $G_A = A^k (A^{2k+1})^T A^k, k = \text{ind}(A)$  είναι επίσης θετικά ορισμένος σε αυτή την περίπτωση και έτσι όλες του οι ιδιοτιμές είναι πραγματικές και θετικές.

Δεδομένου ότι  $\beta > 0$ , όλες οι τιμές του  $-\beta G_A$  είναι πραγματικές και αρνητικές, και σύμφωνα με τη θεωρία γραμμικών δυναμικών το RNN είναι συμπτωτικά σταθερό. Επομένως ο πρώτος όρος του πίνακα όσον αφορά το δεξί μέρος στην εξίσωση τείνει στον πίνακα μηδενικών στοιχείων του ίδιου μεγέθους καθώς ο χρόνος τείνει στο άπειρο για οποιοδήποτε  $V(0)$

$$\lim_{t \rightarrow \infty} \exp(-\beta GAT) V(0) = 0 \quad (2.3.8).$$

**ΘΕΩΡΗΜΑ 2.3.3.1[54]**

Θέτω  $A \in R^{n \times n}$  και  $k = \text{ind } A$ . Βάζοντας την εξίσωση  $\bar{V} = \lim_{t \rightarrow \infty} \exp(-\beta GA\tau) \int_0^t \exp(\beta GA\tau) G d\tau$  (2.4.2)

και παράγει τον αντίστροφο του Drazin  $A^D$ :

$$\bar{V} = A^D \quad (2.3.9).$$

**Παράδειγμα 2.3.3.1[54]**

Περιλαμβάνοντας τον ακόλουθο μη αναστρέψιμο πίνακα από [85]:

$$A = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & -1 & 1 & -1 & 0 & 0 & 0 & 0 \\ -1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & -1 & -1 \\ 0 & 0 & 0 & 0 & -1 & 1 & -1 & -1 \\ 0 & 0 & 0 & -1 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix}$$

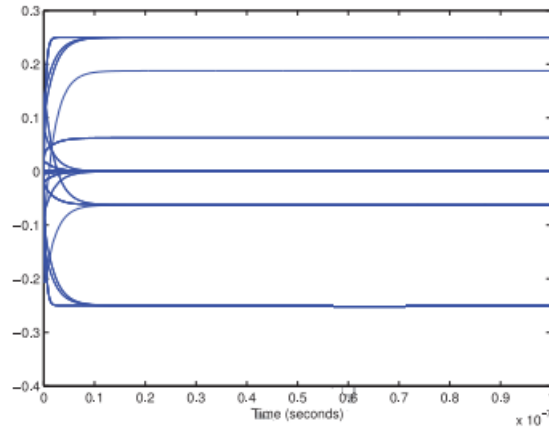
Έχουμε ότι το  $\sigma(A) = \{0,0,0,0,2,2,2,2\}$ . Χρησιμοποιώντας  $k = \text{ind}(A) = 4$  και  $\beta = 10^4$  λαμβάνουμε μετα από χρόνο  $10^{-7}$  δευτερόλεπτα.

$A^d =$

$$\begin{bmatrix} 0.2500 & -0.2500 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.2500 & 0.2500 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.2500 & -0.2500 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0.2500 & 0.2500 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0.0625 & 0.0625 & 0.2500 & -0.2500 & 0 & 0 \\ 0 & 0 & -0.0625 & 0.0625 & -0.2500 & 0.2500 & 0 & 0 \\ 0.0625 & 0.0625 & -0.0625 & -0.0625 & 0 & 0 & 0.2500 & -0.2500 \\ -0.0625 & -0.0625 & 0.1875 & -0.0625 & 0 & 0 & -0.2500 & 0.2500 \end{bmatrix}$$

Ο ακριβής Drazin αντίστροφος είναι ίσος με

$$A^D = \begin{bmatrix} \frac{1}{4} & -\frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} & -\frac{1}{4} & 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{16} & \frac{1}{16} & \frac{1}{4} & -\frac{1}{4} & 0 & 0 \\ 0 & 0 & -\frac{1}{16} & \frac{1}{16} & -\frac{1}{4} & \frac{1}{4} & 0 & 0 \\ \frac{1}{16} & \frac{1}{16} & -\frac{1}{16} & -\frac{1}{16} & 0 & 0 & \frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{16} & -\frac{1}{16} & \frac{3}{16} & -\frac{1}{16} & 0 & 0 & -\frac{1}{4} & \frac{1}{4} \end{bmatrix}$$



Σχήμα 2.5: Συμπεριφορά σύγκλισης του δικτύου από παράδειγμα 2.3.3.1 [54]

**Θεώρημα 2.3.3.2 [55]**

Έστω ένα δίκτυο του οποίου η κατάσταση περιγράφεται από ένα διάνυσμα  $V(t) \in \mathbb{R}^n$  την χρονική στιγμή  $t$ . Έτσι ορίζεται το παρακάτω διάνυσμα.

Για την παρακολούθηση της σύγκλισης του δικτύου, ενδεικνύται η χρήση της νορμας του σφαλματος  $\|V(t) - A^D(t)\|$ .

**Παράδειγμα 2.3.3.2 [55]**

Περιλαμβάνοντας των παρακάτω πίνακα:

$$A = \begin{bmatrix} \frac{3}{2} & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{4} & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & -1 & \frac{3}{4} & -\frac{3}{4} & 0 & 0 & 0 & 0 \\ -1 & -1 & -\frac{3}{4} & \frac{3}{4} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{3}{4} & -\frac{3}{4} & -1 & -1 \\ 0 & 0 & -1 & 0 & -\frac{3}{4} & \frac{3}{4} & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -\frac{1}{4} \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{3}{2} \end{bmatrix}$$

Οι ιδιοτιμές του πίνακα  $A$  περιλαμβάνονται στο σύνολο  $\sigma(A) = \{1.5, 1.5, 1.25 + 0.1443i, 1.25 - 0.1443i, 1.25 - 0.1443i, 1.25 + 0.1443i\}$ .

Χρησιμοποιώντας το  $k = \text{Ind}(A) = 2$  και  $\beta = 10^8$  λαμβάνουμε μετά από  $10^{-7}$  δευτερόλεπτα την ακόλουθη προσέγγιση του Drazin αντίστροφου:

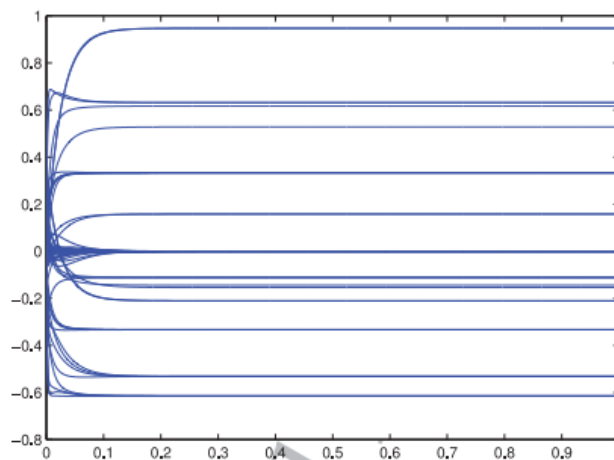
$A^d =$

$$\begin{bmatrix} 0.6316 & -0.2105 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.1579 & 0.9474 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.6150 & -0.5319 & 0.3333 & -0.3333 & 0 & 0 & 0 & 0 \\ -0.6150 & -0.5319 & -0.3333 & 0.3333 & 0 & 0 & 0 & 0 \\ 0.6166 & 0.5282 & -0.1111 & 0.3333 & 0.3333 & -0.3333 & -0.5319 & -0.6150 \\ -0.1442 & -0.1538 & -0.1111 & -0.1111 & -0.3333 & 0.3333 & -0.5319 & -0.6150 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.9474 & 0.1579 \\ 0 & 0 & 0 & 0 & 0 & 0 & -0.2105 & 0.6316 \end{bmatrix}$$

Η συμπεριφορά σύγκλισης του δικτύου που χρησιμοποιήθηκε στο Παράδειγμα 2 σε χρόνο  $10^{-7}$  δευτερολέπτων απεικονίζεται στο Σχήμα .

Ο ακριβής Drazin αντίστροφος είναι :

$$A^D = \begin{bmatrix} \frac{12}{19} & -\frac{4}{19} & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{3}{19} & \frac{18}{19} & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{222}{361} & -\frac{192}{361} & \frac{1}{3} & -\frac{1}{3} & 0 & 0 & 0 & 0 \\ -\frac{222}{361} & -\frac{192}{361} & -\frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 \\ \frac{12688}{20577} & \frac{32608}{61731} & -\frac{1}{9} & \frac{1}{3} & \frac{1}{3} & -\frac{1}{3} & -\frac{192}{361} & -\frac{222}{361} \\ -\frac{2968}{20577} & -\frac{9496}{61731} & -\frac{1}{9} & -\frac{1}{9} & -\frac{1}{3} & \frac{1}{3} & -\frac{192}{361} & -\frac{222}{361} \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{18}{19} & -\frac{3}{19} \\ 0 & 0 & 0 & 0 & 0 & 0 & -\frac{4}{19} & \frac{12}{19} \end{bmatrix}$$



Σχήμα 2.6: Συμπεριφορά σύγκλισης του δικτύου από παράδειγμα 2.3.3.2 [55]

## ΑΠΟΔΕΙΞΗ

Σύμφωνα με την βασική προϋπόθεση του Drazin αντίστροφου, αμέσως μετά ακολουθεί το εξής:

$$G = A^k (A^{2k1})^T A^k = A^k (A^{2k1})^T A^{k+1} A^D = G A A^D \quad (2.3.10)$$

Το οποίο σε συνδυασμό με την εξίσωση (2.3.10) συνεπάγεται ότι :

$$\bar{V} = \lim_{t \rightarrow \infty} \exp(-\beta G A \tau) \int_0^t \exp(\beta G A \tau) \beta G A A^D \tau$$

$$\begin{aligned}
 &= \left[ \lim_{t \rightarrow \infty} \exp(-\beta GA \tau) \int_0^t d(\exp(\beta GA \tau)) \right] A^D \\
 &= \left[ \lim_{t \rightarrow \infty} \exp(-\beta GA \tau) \exp(\beta GA \tau) \right] A^D \\
 &= \left\{ \lim_{t \rightarrow \infty} \exp(-\beta GA \tau) [\exp(\beta GA \tau) - I] \right\} A^D \\
 &= \left[ I - \lim_{t \rightarrow \infty} \exp(-\beta GA \tau) \right] A^D \quad (2.3.11)
 \end{aligned}$$

Η εξίσωση

$$V(t) = \exp(-\beta GA \tau)V(0) + \beta \exp(-\beta GA \tau) \int_0^t \beta GA \tau \exp(2.3.12)$$

μπορεί να απλοποιηθεί αν μηδενίσουμε τον πρώτο όρο στο δεξί μέλος, επιλέγοντας κατάλληλες συνθήκες. Για τον σκοπό αυτό, είναι χρήσιμο να ορίσουμε  $V(0) = 0$  μέσα στον πρώτο όρο του δεξιού μέλους της εξίσωσης, κάτι που επίσης επιταχύνει τη σύγκλιση. Έτσι, η δυναμική εξίσωση κατάστασης του RNN για τον υπολογισμό του αντιστρόφου Drazin μπορεί να γραφτεί ως:

$$\frac{dV}{dt} = -\beta G(AV(t) - I), V(0) = 0, G = A^k(A^{2k1})^T A^k, k = \text{ind}(A) \quad (2.3.13)$$

### Θεώρημα 2.3.3.3 [52]

Έστω  $A \in R^{n \times n}$  και  $k = \text{ind}(A)$  τότε η εξίσωση που παράγει τον αντίστροφο του Drazin  $A^D$  θα είναι :

$$\bar{V} = A^D$$

### Παράδειγμα 2.3.3.3 [52]

Εδώ φαίνεται με βάση το παράδειγμα ότι το νευρωνικό μας δίκτυο παράγει έναν κανονικό αντίστροφο στην περίπτωση που ο πίνακας εισόδου είναι αντιστρέψιμος

Έστω ο παρακάτω πίνακας ότι είναι ένας τυχαίος πίνακας:

$$A = \begin{bmatrix} 0.0046 & 0.8001 & 0.8693 & 0.5132 & 0.4173 & 0.9001 & 0.0965 \\ 0.7749 & 0.4314 & 0.5797 & 0.4018 & 0.04965 & 0.3692 & 0.1320 \\ 0.8173 & 0.9106 & 0.5499 & 0.0760 & 0.9027 & 0.1112 & 0.9421 \\ 0.8687 & 0.1818 & 0.1450 & 0.2399 & 0.9448 & 0.7803 & 0.9561 \\ 0.0844 & 0.2638 & 0.8530 & 0.1233 & 0.4909 & 0.3897 & 0.5752 \\ 0.3998 & 0.1455 & 0.6221 & 0.1839 & 0.4893 & 0.2417 & 0.05978 \\ -0.2599 & 0.1361 & 0.3510 & 0.2400 & 0.3377 & 0.4039 & 0.2348 \end{bmatrix}$$

Ο πίνακας ικανοποιεί την ιδιότητα  $\text{ind}(A)=0$ . Το φάσμα αυτού του πίνακα είναι

$$\sigma(A) = \{3.0953, -0.9778, -0.5571, 0.2904+0.3749i, 0.2904-0.3749i, 0.2354, -0.1833\}.$$

Χρησιμοποιώντας ότι  $k = \text{Ind}(A)=0$  και  $\beta=10^{11}$  το νευρωνικό δίκτυο παράγει την ακόλουθη προσέγγιση του Drazin αντίστροφου μετά από  $10^{-7}$  δευτερόλεπτα:

$A^d =$

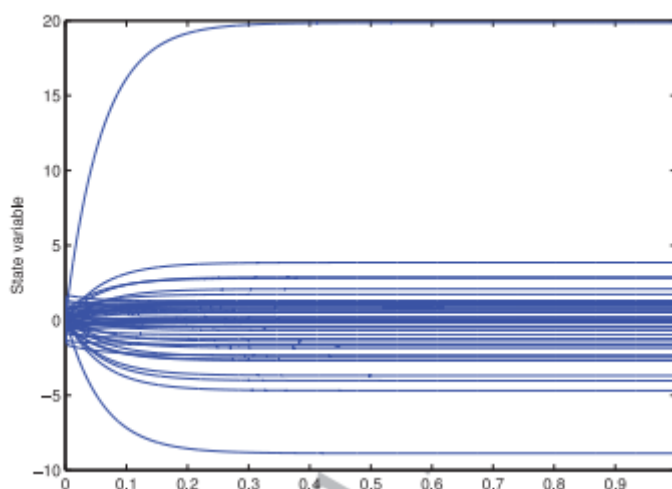
$$\begin{bmatrix} 0.00439 & 1.2022 & -0.4308 & 1.1517 & 0.0312 & 1.1600 & -4.0275 \\ 1.0869 & 0.1738 & 0.5595 & 0.2298 & -0.4034 & -0.1539 & -2.6975 \\ -0.0877 & 0.5401 & -0.4067 & -0.0575 & 1.3195 & 0.8319 & -1.8463 \\ -2.3972 & -1.4086 & 2.0950 & -4.6892 & -2.6342 & -3.6889 & 19.8601 \\ -0.0888 & -1.6265 & 0.6749 & -0.4814 & -1.2210 & 1.1998 & 2.8891 \\ 1.7409 & 0.8704 & -1.6648 & 2.8154 & 1.0611 & 1.1242 & -8.8760 \\ -0.9648 & -0.0431 & 0.5126 & -0.6807 & 0.8499 & -2.3282 & 3.8575 \end{bmatrix}$$

Η συμπεριφορά σύγκλισης του δικτύου που χρησιμοποιήθηκε στο Παράδειγμα 2.2.1 σε χρόνο  $10^{-7}$  δευτερολέπτων απεικονίζεται στο Σχήμα 2.2 .

Ο ακριβής Drazin αντίστροφος είναι :

$$A^D = A^{-1} =$$

$$\begin{bmatrix} 0.00440 & 1.2022 & -0.4308 & 1.1517 & 0.0312 & 1.1600 & -4.0275 \\ 1.0869 & 0.1738 & 0.5595 & 0.2298 & -0.4034 & -0.1539 & -2.6975 \\ -0.0877 & 0.5401 & -0.4067 & -0.0575 & 1.3195 & 0.8319 & -1.8463 \\ -2.3972 & -1.4086 & 2.0950 & -4.6892 & -2.6342 & -3.6889 & 19.8601 \\ -0.0888 & -1.6265 & 0.6749 & -0.4814 & -1.2210 & 1.1998 & 2.8891 \\ 1.7409 & 0.8704 & -1.6648 & 2.8154 & 1.0611 & 1.1242 & -8.8760 \\ -0.9648 & -0.0431 & 0.5126 & -0.6807 & 0.8499 & -2.3282 & 3.8575 \end{bmatrix}$$



Σχήμα 2.2: Συμπεριφορά σύγκλισης Drazin αντίστροφου από παράδειγμα 2.3.3.3 [52]

### ΘΕΩΡΗΜΑ 2.3.3.4

Το Νευρωνικό δίκτυο περιγράφεται στην εξίσωση με την δυναμική σταθερότητα του Lyapunov.

### ΑΠΟΔΕΙΞΗ

Η μέθοδος του Lyapunov χρησιμοποιείται για να δείξει σταθερότητα. Μπορούμε να αποδώσουμε την συνάρτηση Lyapunov ως εξής:

$$E(t) = \frac{\|GAV - G\|_F^2}{2} = \frac{Tr((GAV - G)^T(GAV - G))}{2} \quad (2.3.14)$$

Είναι προφανές ότι ισχύει η ανισότητα  $E(t) \geq 0$ . Εφαρμόζοντας λογισμό πινάκων μπορεί να επαληθευτεί ότι:

$$\frac{\partial E(t)}{\partial v} = (GA)^T (GAV - G) \quad (2.3.15)$$

Έτσι η χρονική παράγωγος του  $E(t)$  θα είναι:

$$\begin{aligned} \frac{dE(t)}{dt} &= Tr\left(\left(\frac{\partial E}{\partial v}\right)^T \frac{dv}{dt}\right) \\ &= Tr((GAV - G)^T GA(-\beta(GAV - G))) \end{aligned}$$

$$= -\beta \text{Tr}(GA(GAV - G)(GAV - G)^T) \quad (2.3.16)$$

Στο υπόλοιπο της απόδειξης, εξετάζουμε το Ερμιτιανό μέρος ενός τετραγωνικού πραγματικού πίνακα  $GA$ , το οποίο ορίζεται ως:

$$H(GA) = \frac{1}{2}(GA + (GA)^T) \quad (2.3.17)$$

Από την στιγμή που οι πίνακες  $(GAV - G)(GAV - G)^T$  και  $H(GA)$  είναι συμμετρικοί, μετά από τους καταλλήλους μετασχηματισμούς μπορούμε να επαληθεύσουμε ότι :

$$\frac{dE(t)}{dt} = -\beta \text{Tr}(H(GA)(GAV - G)(GAV - G)^T) \quad (2.3.18)$$

Όλες οι ιδιοτιμές ενός τετραγωνικού πίνακα είναι ίσες με τις ιδιοτιμές του αντιστρόφου του. Επιπλέον, επειδή ο  $H(GA)$  είναι συμμετρικός, είναι και θετικά ορισμένος. Τέλος, αφού και οι δύο πίνακες  $H(GA)$  και  $(GAV - G)(GAV - G)^T$  είναι θετικά ορισμένοι, και αυτό οδηγεί άμεσα στο συμπέρασμα ότι:

$$\frac{dE(t)}{dt} \leq 0 \quad (2.3.19)$$

## 2.4 Μοντέλο GNN για την επίλυση μη γραμμικής εξίσωσης $AXB=D$ και της σύγκλισης του

Σύμφωνα με την προσέγγιση της νευρωνικής δυναμικής βασισμένης σε βαθμίδες ορίζουμε τη συνάρτηση σφάλματος πίνακα ως  $E(t)=D-AV(t)B$ , η οποία σχετίζεται με τη μαθηματική εξίσωση  $AXB=D$ . Στην εξίσωση αυτή, το  $V(t)$  αντιπροσωπεύει έναν πίνακα μεταβλητών ενεργοποίησης, ενώ το  $t \in [0, +\infty)$  δηλώνει τον χρόνο.

Γενικεύοντας το μοντέλο GNN-MP που έχει προταθεί στη βιβλιογραφία, προκύπτει μια πιο γενικευμένη μορφή νευρωνικού μοντέλου, ικανή να περιγράψει τη δυναμική συμπεριφορά των συστημάτων αυτών.

Σύμφωνα με τη μελέτη [45], από τη θεώρηση αυτή προκύπτει το ακόλουθο νευροδυναμικό μοντέλο βασισμένο σε βαθμίδες για την επίλυση της εξίσωσης  $AXB=D$ . Το μοντέλο αυτό ονομάζεται GNN(A, B, D) και περιγράφεται από τη σχέση:

$$\frac{dV(t)}{dt} = \dot{V}(t) = \gamma A^T F(D - AV(t)B)B^T. \quad (2.4.1)$$

Η παράμετρος  $\gamma$  λειτουργεί ως παράγοντας κλιμάκωσης, που μπορεί να ερμηνευτεί είτε ως παράμετρος επαγωγής είτε ως το αντίστροφο μιας παραμέτρου χωρητικότητας. Η τιμή της μπορεί να επιλεγεί όσο το δυνατόν μεγαλύτερη, ώστε να επιταχύνει τη διαδικασία σύγκλισης του μοντέλου. Η συνάρτηση  $F(C)$  εφαρμόζεται στοιχείο προς στοιχείο σε έναν πίνακα

$$C = (c_{ij}) \in R^{m \times n}, \text{ i.e. } F(C) = (f(c_{ij})), i = 1, \dots, m, j = 1, \dots, n \quad (2.4.2)$$

όπου η συνάρτηση  $f(\cdot)$  είναι περιττή και μονότονα αύξουσα. Ουσιαστικά, οποιαδήποτε συνάρτηση ενεργοποίησης με αυτά τα χαρακτηριστικά μπορεί να χρησιμοποιηθεί για την υλοποίηση του νευρωνικού δικτύου [46].

Η γενικευμένη εκδοχή του μοντέλου αυτού, που περιλαμβάνει μη γραμμικές συναρτήσεις ενεργοποίησης (γνωστή ως GGNN model), μπορεί να εφαρμοστεί τόσο σε χρονικά μεταβαλλόμενες όσο και σε χρονικά σταθερές περιπτώσεις.



Αν χρησιμοποιηθεί μια συνάρτηση ενεργοποίησης  $F(\cdot)$ , η οποία βασίζεται σε περιττή και μονότονα αύξουσα στοιχειοσυνάρτηση  $f(\cdot)$ , τότε ο νευρωνικός πίνακας κατάστασης  $V(t) \in \mathbb{R}^{n \times p}$  του μοντέλου  $GNN(A, B, D)$ , όπως αυτό ορίζεται στην εξίσωση (2.1), συγκλίνει ασυμπτωτικά προς τη λύση της εξίσωσης  $AXB=D$  δηλαδή ισχύει ότι  $AV(t)B \rightarrow D$  όταν  $t \rightarrow +\infty$ , ανεξάρτητα από την αρχική κατάσταση του συστήματος  $V(0)$ .

### ΑΠΟΔΕΙΞΗ

Αρχικά, το ότι η συνθήκη  $AA^1DB^1B = D$  ισχύει για κάποιους εσωτερικούς αντίστροφους  $A^{(1)} \in A\{1\}$  και  $B^{(1)} \in B\{1\}$  εξασφαλίζει ότι η εξίσωση  $AVB=D$  είναι πράγματι επιλύσιμη, συνεχίζοντας κάνουμε την αντικατάσταση  $V(t) = \tilde{V}(t) + A^{(1)}DB^{(1)}$ , η οποία μετασχηματίζει τη νευροδυναμική εξίσωση σε μια ισοδύναμη μορφή:

$$\frac{d\tilde{V}(t)}{dt} = \frac{dV(t)}{dt} = \gamma A^T F(D - AV(t)B) B^T = \gamma A^T F(D - A(\tilde{V}(t) + ADB)B) B^T \quad (2.4.6)$$

Σύμφωνα με το (2.4.6) ακολουθεί

$$\frac{d\tilde{V}(t)}{dt} = \gamma A^T F(D - AADB - A\tilde{V}(t)B) B^T = -\gamma A^T F(A\tilde{V}(t)B) B^T \quad (2.4.7)$$

Στη συνέχεια βάζουμε την υποψήφια συνάρτηση Lyapunov, η οποία χρησιμοποιείται για να αξιολογήσει το πόσο καλά συγκλίνει το σύστημα

$$L(\tilde{V}(t), t) = \frac{1}{2} \|\tilde{V}(t)\|_F^2 = \frac{1}{2} Tr(\tilde{V}(t)^T \tilde{V}(t)) \quad (2.4.8)$$

Είναι προφανές ότι η ανισότητα  $L(V(t), t) \geq 0$  ισχύει για κάθε  $V(t) \neq 0$ . Στη συνέχεια, με βάση την παραπάνω σχέση, κάνοντας την υπόθεση και χρησιμοποιώντας το γνωστό αποτέλεσμα:

$$dTr(X^T X) = 2Tr(X^T dX) \quad (2.4.9)$$

μαζί με βασικές ιδιότητες του τελεστή ίχνους, μπορούμε να γράψουμε την παράγωγο της συνάρτησης  $L(V(t), t)$  ως προς τον χρόνο με την ακόλουθη μορφή

$$\begin{aligned} \frac{dL(\tilde{V}(t), t)}{dt} &= Tr(\tilde{V}(t)^T \frac{d\tilde{V}(t)}{dt}) = \\ &= -\gamma Tr[\tilde{V}(t)^T A^T F(A\tilde{V}(t)B) B^T] = \\ &= -\gamma Tr[(A\tilde{V}(t)B)^T F(A\tilde{V}(t)B)]. \end{aligned} \quad (2.4.10)$$

Επειδή η βαθμωτή συνάρτηση  $F(\cdot)$ , είναι περιττή και γνησίως αύξουσα, μπορούμε γενικεύοντας την τεχνική που χρησιμοποιήθηκε στο [45], να επαληθεύσουμε το ακόλουθο αποτέλεσμα

$$\frac{dL(\tilde{V}(t), t)}{dt} = -\gamma Tr[(W^T F(W))] = -\gamma \sum_{i=1}^m \sum_{j=1}^n w_{ij} f(w_{ij}) \begin{cases} < 0 & \text{αν } A\tilde{V}(t)B \neq 0 \\ = 0 & \text{αν } A\tilde{V}(t)B = 0 \end{cases} \quad (2.4.11)$$

Αφού ισχύει ότι  $W(t) = A\tilde{V}(t)B = A(V(t) - A^{(1)}DB^{(1)})B = AV(t)B - D$  μπορούμε να επιβεβαιώσουμε ότι:

$$\frac{dL(\tilde{V}(t), t)}{dt} \begin{cases} < 0 & \text{αν } W(t) \neq 0 \\ = 0 & \text{αν } W(t) = 0 \end{cases} \quad (2.4.12)$$

Σύμφωνα με τη θεωρία ευστάθειας Lyapunov, το  $W(t) = AV(t)B - D$  συγκλίνει παγκόσμια στο μηδενικό πίνακα, ανεξάρτητα από την αρχική τιμή του συστήματος.

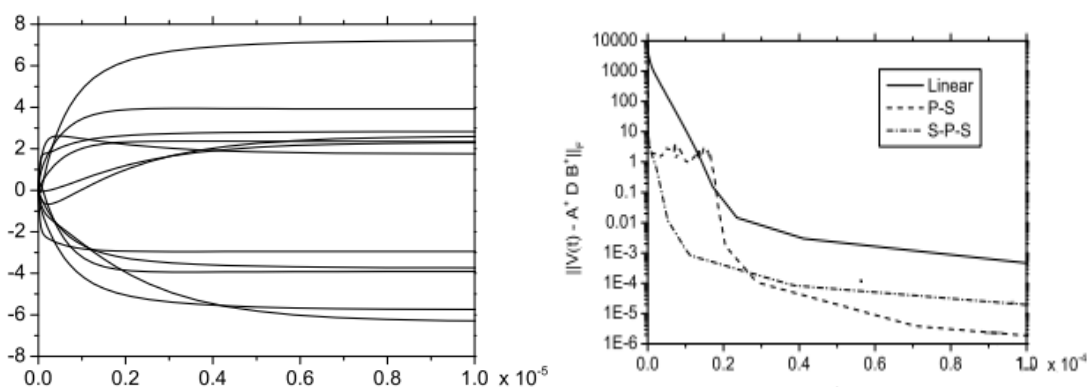
**Παράδειγμα 2.4.4.1 [57]** Συμπεριλαμβάνοντας τους παρακάτω πίνακες:

$$A = \begin{bmatrix} -8 & 8 & -4 \\ 11 & 4 & -7 \\ 1 & -4 & 3 \\ 0 & 12 & -10 \\ 6 & 12 & -12 \end{bmatrix}, B = \begin{bmatrix} 12 & -4 & -16 \\ -35 & -10 & 45 \\ 2 & -5 & -3 \\ 14 & 17 & -17 \end{bmatrix}, D = \begin{bmatrix} -84 & 2524 & 304 \\ -2252 & -623 & 2897 \\ 484 & -885 & -701 \\ -1894 & 2278 & 2652 \\ -2278 & 1524 & 3750 \end{bmatrix},$$

Και οι 3 έχουν βαθμό  $r=2$  και ικανοποιούν την συνθήκη  $AA^{\dagger}DB^{\dagger}B=D$ . Χρησιμοποιούμε το μοντέλο  $GNN(A,B,D)$  για τον υπολογισμό της λύσης ελάχιστου νόρμας ελάχιστων τετράγωνων

$$X = A^{\dagger}DB^{\dagger} \begin{bmatrix} 2.6143 & 1.7536 & 2.3115 & -6.3287 \\ -5.7458 & 2.8326 & -3.7428 & 7.2225 \\ 3.9167 & -2.9451 & 2.3485 & -3.9092 \end{bmatrix}$$

της εξίσωσης της μήτρας  $AXB=D$ . Η παράμετρος κέρδους του μοντέλου είναι  $\gamma = 10$  και  $V(0) = 0$ . Χρησιμοποιώντας τη γραμμική συνάρτηση ενεργοποίησης  $f(x) = x$ , λαμβάνουμε τη λύση με ακρίβεια  $10^{-12}$  σε χρόνο  $t_{\text{tot}}=2 \cdot 10^{-3}$ . Οι κατά στοιχείο τροχιές του  $V(t)$  παρουσιάζονται στο Σχήμα 2.8. Είναι προφανές ότι οι τροχιές υποδεικνύουν συνήθη συμπεριφορά σύγκλισης, συνεπώς το σύστημα είναι καθολικά ασυμπτωτικά σταθερό. Μπορεί να συναχθεί ότι οι μη γραμμικές συναρτήσεις ενεργοποίησης επιταχύνουν τη σύγκλιση. Οι διακυμάνσεις της νόρμας του σφάλματος όταν χρησιμοποιείται η συνάρτηση ενεργοποίησης  $f_{\text{ps}}$  πιθανόν οφείλονται στη μη συνέχεια της παραγωγού της στα σημεία  $x=\pm 1$ .



Σχήμα 2.8: Τροχιές σύγκλισης των στοιχείων του δικτύου  $GNN(A,B,D)$  με την γραμμική συνάρτηση ενεργοποίησης (αριστερά σχήμα) και η τροχιά του σφάλματος για διαφορετικές συναρτήσεις ενεργοποίησης (δεξιά σχήμα) από παράδειγμα 2.4.4.1[57]

Τα τελικά υπολειπόμενα σφάλματα μετά τον συνολικό χρόνο  $t_{\text{tot}} = 2 \times 10^{-3}$  είναι  $3.34 \times 10^{-12}$ ,

$1.85 \times 10^{-12}$  αντίστοιχα, για τη γραμμική, τη δυναμική σιγμοειδής και την ομαλή δυναμική-σιγμοειδή συνάρτηση ενεργοποίησης. Τα υπολειπόμενα σφάλματα από διαφορετικές επιλογές παραμέτρων  $\gamma$ ,  $t_{\text{tot}}$  και του  $V(0)$  που μας δίνεται και του επίλυση ODE, που φαίνονται στον πίνακα 1.

Activation	$\gamma$	$\ AV(t)B - D\ _F$	Activation	$\gamma$	$\ AV(t)B - D\ _F$
$(2 \cdot 10^{-3}, ode15s), V(0) = 0$			$(2 \cdot 10^{-3}, ode15s), V(0) = B^T$		
lin	1	3.5951e-13	lin	1	2.0714e-12
ps	1	2.8422e-13	ps	1	<b>2.2737e-13</b>
sps	1	<b>1.6078e-13</b>	sps	1	1.2085e-12
$(2 \cdot 10^{-4}, ode15s), V(0) = 0$			$(2 \cdot 10^{-4}, ode15s), V(0) = B^T$		
lin	1	0.083172	lin	1	0.020686
ps	1	<b>1.054869e-06</b>	ps	1	<b>1.14258e-06</b>
sps	1	0.000366	sps	1	1.88123e-06
$(10^{-4}, ode15s), V(0) = 0$			$(10^{-4}, ode15s), V(0) = B^T$		
lin	1	12.23908	lin	1	3.774600
ps	1	0.017396	ps	1	1.366278
sps	1	<b>0.000146</b>	sps	1	<b>0.031365</b>
$(10^{-6}, ode15s), V(0) = eye(3, 2) \cdot eye(2, 4)$			$(10^{-5}, ode15s), V(0) = magic(3) \cdot eye(3, 4)$		
lin	$10^3$	7.668596e-13	lin	$10^3$	<b>1.59972e-12</b>
ps	$10^3$	<b>5.11591e-13</b>	ps	$10^3$	6.996003e-07
sps	$10^3$	0.000369	sps	$10^3$	1.64158e-12

Πίνακας 1: Η απόδοση του μοντέλου GNN(A,B,D) από το παράδειγμα 2.4.4.1[57]

### ΘΕΩΡΗΜΑ 2.4.4.2[57]

Ας υποθέσουμε ότι οι πραγματικοί πίνακες  $A \in R^{m \times n}$ ,  $B \in R^{m \times n}$  και  $D \in R^{m \times n}$  ικανοποιούν τις απαιτούμενες προϋποθέσεις.

$$AA^\dagger DB^\dagger B = D \quad (2.4.13)$$

Τότε, ο πίνακας μεταβλητών ενεργοποίησης  $V(t)$  του μοντέλου GNN(A, B, D), όπως ορίζεται στην εξίσωση, συγκλίνει καθώς  $t \rightarrow +\infty$  και καταλήγει σε μία κατάσταση ισορροπίας.

$$V(t) \rightarrow \bar{V} = A^\dagger DB^\dagger + V(0) - A^\dagger AV(0) B B^\dagger \quad (2.4.14)$$

### ΑΠΟΔΕΙΞΗ

Αυτό συμβαίνει για οποιονδήποτε αρχικό πίνακα  $V(0) \in R^{n \times p}$

$$\frac{dV_1(t)}{dt} = \gamma A^\dagger A A^T \frac{dV(t)}{dt} B^T B B^\dagger = \gamma A^\dagger A A^T F(E(t)) B^T B B^\dagger \quad (2.4.15)$$

Στη συνέχεια, χρησιμοποιώντας ορισμένες βασικές ιδιότητες του Moore-Penrose,  $B^T B B^\dagger = B^T$ ,  $A^\dagger A A^T = A^T$  προκύπτουν οι σχέσεις:

$$\frac{dV_1(t)}{dt} = \gamma A^T F(E(t)) B^T = \frac{dV(t)}{dt} \quad (2.4.16)$$

Αυτό έχει σαν αποτέλεσμα όταν ορίζουμε  $V_2(t) = V(t) - V_1(t)$  να προκύπτει ότι ο  $V_2(t)$  ικανοποιεί  $\frac{dV_2(t)}{dt} = 0$ .

$$V_2(t) = V_2(0) = V(0) - V_1(0) = V(0) - A^\dagger AV(0) B B^\dagger, t \geq 0 \quad (2.4.17)$$

Επιπλέον σύμφωνα με το  $AV(t)B \rightarrow D$   $V_1(t)$  τείνει στο άπειρο

$$V_1(t) = A^\dagger (AV(t) B) B^\dagger \rightarrow A^\dagger DB^\dagger, t \rightarrow +\infty. \quad (2.4.18)$$

Και λαμβάνοντας υπόψιν το παραπάνω προκύπτει ότι συγκλίνει στο άπειρο

$$V(t) = V_1(t) + V_2(t) \rightarrow \bar{V} = A^\dagger DB^\dagger + V_2(0) = A^\dagger DB^\dagger + V(0) - A^\dagger V(0)BB^\dagger. \quad (2.4.19)$$

Είναι γνωστό ότι ισχύει η ανισότητα

$$\|AXB - D\|^2 \geq \|AA^\dagger DB^\dagger B - D\|^2 \quad (2.4.20)$$

και επιτυγχάνεται μόνο αν ο πίνακας  $X$  έχει την μορφή

$$X = A^\dagger DB^\dagger + Y - A^\dagger AYBB^\dagger \quad (2.4.21)$$

όπου  $Y$  είναι πίνακας κατάλληλων διαστάσεων.

Επίσης ο πίνακας  $A^\dagger DB^\dagger$  είναι ο μοναδικός ελαχιστοποίησής του κανόνα  $\|\cdot\|$  αναμεσα σε όλες τις λύσεις ελάχιστων τετράγωνων:

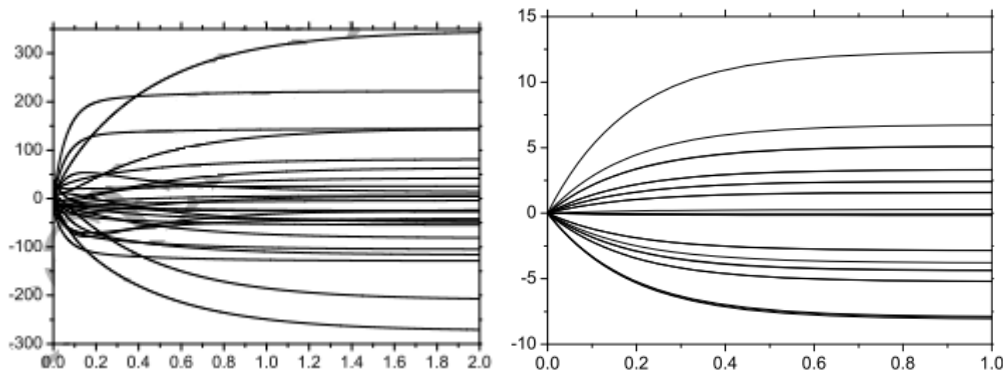
$$\|A^\dagger DB^\dagger\|^2 \leq \|A^\dagger DB^\dagger + Y - A^\dagger AYBB^\dagger\|^2 \quad (2.4.22)$$

#### Παραδειγμα 2.4.4.2[57]

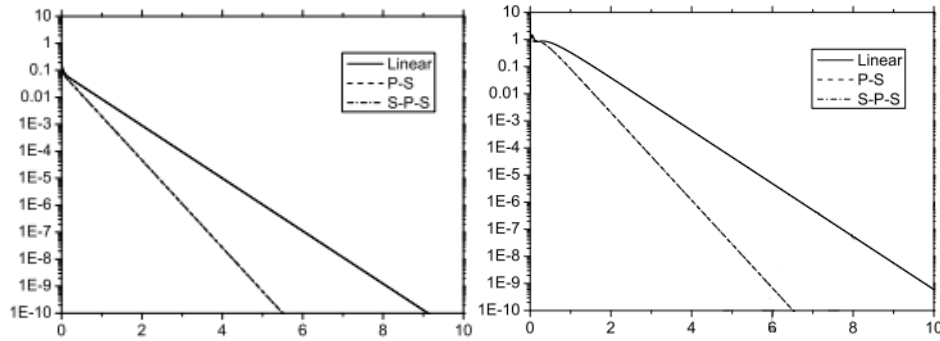
Το παραδειγμα δειχνει την εφαρμογη του μοντελου GNN ( $A_p, A_q, A_k$ ) για τον υπολογισμο του Drazin αντιστροφου:

$$A = \frac{1}{788} \begin{bmatrix} 448 & 574 & 98 & 392 & -826 \\ 574 & -58 & -198 & -280 & 206 \\ 320 & 788 & 396 & 560 & -1084 \\ -144 & -377 & -55 & -28 & 555 \\ 32 & 146 & -50 & 56 & -86 \end{bmatrix}$$

Ο βαθμος και ο δεικτης του  $A$  είναι ισοι με  $r = \text{rank}(A) = 3$  και  $k = \text{ind}(A) = 2$  αντιστοιχα. Ετσι θεωρουμε ότι το μοντελο GNN( $A, B, D$ ) με τις ακολουθες παραμετρους:  $p = 2k = 4$ ,  $q = k+1 = 3$ ,  $\gamma = 108$  και  $V(0)=0$ . Ετσι χρησιμοποιωντας την συναρτηση ενεργοποιησης ,λαμβανουμε την λυση  $\tilde{V}$  της εξισωσης πινακων  $A^p V A^q = A^k$



Σχήμα 2.9: Τροχιά σύγκλισης κατά στοιχείο μεταβλητής κατάστασης του  $V(t)$  από το GNN( $A_p, A_q, A_k$ ) με την γραμμική συνάρτηση ενεργοποίησης (αριστερά σχήμα) και το  $X(t)=A_k V(t) A_k$  που συγκλίνει στο αντίστροφο Drazin  $A^D$  (δεξιά σχήμα) από παράδειγμα 2.4.4.2[57]



Σχήμα 2.10:Νορμα υπόλοιπου  $\|A^k X(t)A - A^k\|_F$  (αριστερά σχήμα) και  $\|X(t)AX(t) - X(t)\|_F$  (δεξιά σχήμα) από παράδειγμα 2.4.4.2[57]

## 2.5 Μη γραμμικά μοντέλα νευρωνικών δικτύων (GNN) για τον υπολογισμό εσωτερικών αντιστροφών

Σε αυτή την ενότητα αναπτύσσεται ένα μοντέλο νευρωνικού δικτύου (GNN) που έχει ως στόχο τον υπολογισμό του εσωτερικού αντιστρόφου και του αντίστροφου Moore–Penrose ενός πίνακα. Το μοντέλο αυτό στηρίζεται στην εξίσωση  $AXA = A$ , η οποία αποτελεί τη θεωρητική του βάση. Παράλληλα, μελετάται το πώς η αρχική τιμή του συστήματος, η παράμετρος σχεδιασμού  $\gamma$  και οι διαφορετικές μη γραμμικές συναρτήσεις ενεργοποίησης επηρεάζουν τη σύγκλιση και τη συμπεριφορά του μοντέλου κατά τη διάρκεια της εκπαίδευσης ή της εξέλιξής του στο χρόνο.

Στην παρούσα ανάλυση επικεντρωνόμαστε σε πραγματικούς πίνακες, θεωρώντας ότι  $A \in R^{m \times n}$

Έστω ότι  $V(t) \in R^{n \times m}$  είναι ένας πίνακας που μεταβάλλεται με το χρόνο και αποτελεί το ζητούμενο της διαδικασίας.

Στόχος είναι να βρεθεί ο πίνακας  $V(t)$  που ικανοποιεί την εξίσωση:

$$AV(t)A = A \quad (2.5.1)$$

Για την περιγραφή του σφάλματος, ορίζεται η συνάρτηση σφάλματος σε μορφή πίνακα ως:

$$E(t) = A - AV(t)A. \quad (2.5.2)$$

Για να υπολογίσουμε το μέτρο αυτού του σφάλματος, χρησιμοποιείται μια συνάρτηση σφάλματος βαθμωτής μορφής, βασισμένη στον κανόνα Frobenius :

$$\epsilon(t) = \epsilon(V(t)) = \frac{1}{2} \|E(t)\|_F^2 = \frac{\|A - AV(t)A\|_F^2}{2} \quad (2.5.3)$$

Η συνάρτηση αυτή επιτρέπει την επίλυση της εξίσωσης με τη βοήθεια μιας δυναμικής μεθόδου, όπου η διαδικασία μάθησης καθοδηγείται από την αρνητική βαθμίδα της  $\epsilon(V(t))$  αν ληφθεί υπόψη ότι:

$$\frac{\partial \epsilon(V(t))}{\partial V} = -A^T (A - AV(t)A) A^T = -A^T E(t) A^T \quad (2.5.4)$$

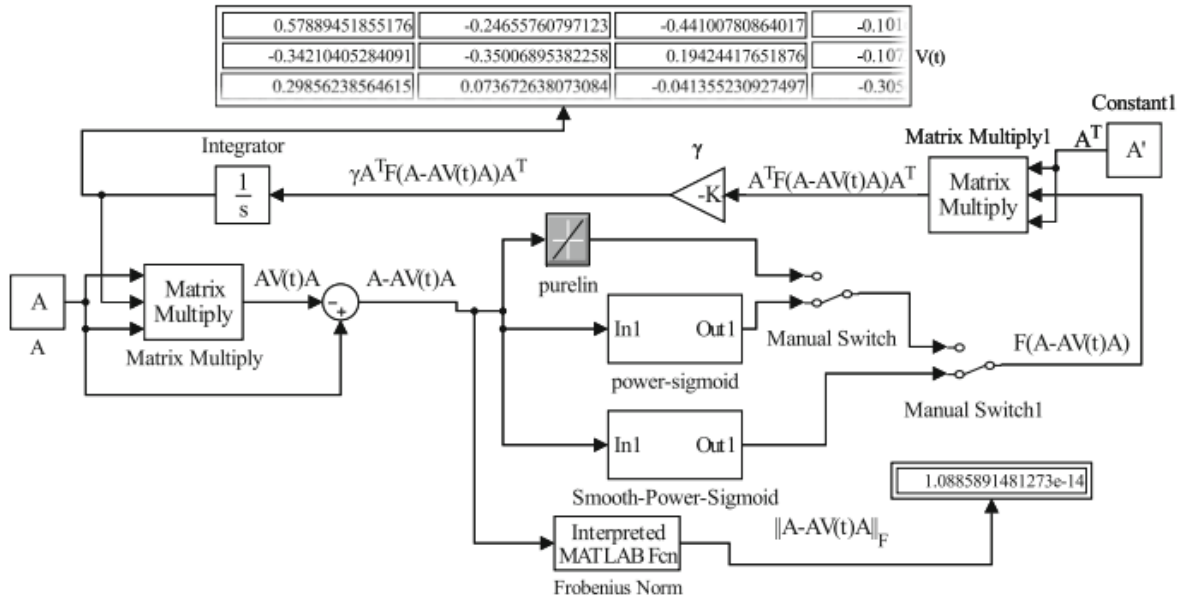
και χρησιμοποιηθεί ο γενικός σχεδιαστικός τύπος  $\frac{dV(t)}{dt} = \gamma \frac{\partial \epsilon(V(t))}{\partial V}$ , προκύπτει η εξίσωση δυναμικής κατάστασης του συστήματος:

$$\frac{dV(t)}{dt} = \gamma A^T (A - AV(t)A) A^T, V(0) = V_0 \quad (2.5.5)$$

Η παράμετρος  $\gamma$  είναι μια θετική σταθερά κλιμάκωσης, η οποία επιλέγεται όσο το δυνατόν μεγαλύτερη ώστε να επιταχύνει τη διαδικασία σύγκλισης.

Ακολουθώντας την αρχή της μη γραμμικής ενεργοποίησης που έχει χρησιμοποιηθεί στα GNN μοντέλα της [48] για τον υπολογισμό του αντίστροφου Drazin, το γραμμικό GNN μοντέλο μπορεί να επεκταθεί, ενσωματώνοντας μια μη γραμμική συνάρτηση ενεργοποίησης σε μορφή πίνακα  $F(\cdot)$ . Έτσι, προκύπτει η μη γραμμική εκδοχή του GNN μοντέλου, η οποία περιγράφεται από την εξίσωση:

$$\frac{dV(t)}{dt} = \gamma A^T F(A - AV(t)A) A^T, V(0) = V_0 \quad (2.5.6)$$



Σχήμα 2.12: Αναπαράσταση Simulink του μοντέλου GNN-MP[57]

### ΘΕΩΡΗΜΑ 2.5.5.1[57]

Θεωρούμε έναν πραγματικό πίνακα  $A \in \mathbb{R}^{n \times n}$  με δείκτη  $\text{Ind}(A)=k$ . Αν χρησιμοποιηθεί μια περιττή και αύξουσα συνάρτηση  $f(\cdot)$  για τον ορισμό της συνάρτησης ενεργοποίησης σε μορφή πίνακα  $F(\cdot)$ , τότε ο πίνακας κατάστασης  $V(t) \in \mathbb{R}^{n \times n}$  του νευρωνικού μοντέλου θα συγκλίνει προς τον Drazin αντίστροφο  $A^D$ , για οποιαδήποτε αρχική τιμή του  $V(0)$ .

### ΑΠΟΔΕΙΞΗ

Με βάση την απόδειξη θέτουμε  $\bar{V}(t) = V(t) - A^D$ , αρά ορίζουμε τον πίνακα που εκφράζει τη διαφορά ανάμεσα στη λύση  $V(t)$ , η οποία παράγεται από το νευρωνικό μοντέλο, και τον θεωρητικό Drazin αντίστροφο  $A^D$ .

$$(A + \lambda I)^{l+1} A^D - A^l \approx 0 \quad (2.5.7)$$

Με την αντικατάσταση του τύπου  $V(t) = \bar{V}(t) - A^D$  στην δυναμική εξίσωση βγάζουμε συμπέρασμα ότι:

$$\frac{d\bar{V}(t)}{dt} = \frac{dV(t)}{dt} = -\beta((A + \lambda I)^{\lambda+1})^T F((A + \lambda I)^{\lambda+1} \bar{V}(t)). \quad (2.5.8)$$

Η συνάρτηση Lyapunov ορίζεται ως:

$$L(\bar{V}(t), t) = \frac{1}{2} \|\bar{V}(t)\|_F^2 = \frac{1}{2} \text{Tr}(\bar{V}(t)^T \bar{V}(t)). \quad (2.5.9)$$

Η ανισότητα  $L(\bar{V}(t), t) > 0$  ισχύει για  $\bar{V}(t) \neq 0$  και αν λάβουμε υπόψιν την (2.5.9) τότε η χρονική παράγωγος  $L(\bar{V}(t), t)$  ικανοποιεί

$$\begin{aligned} \frac{dL(\bar{V}(t), t)}{dt} &= \frac{1}{2} \frac{d\text{Tr}(\bar{V}(t)^T \bar{V}(t))}{dt} \\ &= \frac{1}{2} \text{Tr}\left(\bar{V}(t)^T \frac{d\bar{V}(t)}{dt}\right) \\ &= -\frac{1}{2} \beta \text{Tr}\left\{\left((A + \lambda I)^{\lambda+1} \bar{V}(t)\right)^T F\left((A + \lambda I)^{\lambda+1} \bar{V}(t)\right)\right\} \quad (2.5.10) \end{aligned}$$

Από την στιγμή που η  $f(\bullet)$  είναι βαθμωτή συνάρτηση, περιττή αλλά και μονότονα αύξουσα, απευθείας προκύπτει ότι  $f(-x) = -f(x)$  και

$$f(x) = \begin{cases} > 0 & \text{αν } x > 0 \\ = 0 & \text{αν } x = 0 \\ < 0 & \text{αν } x < 0 \end{cases} \quad (2.5.11)$$

το οποίο σημαίνει πως

$$xf(x) = \begin{cases} > 0 & \text{αν } x \neq 0 \\ = 0 & \text{αν } x = 0 \end{cases} \quad (2.5.12)$$

Εξετάζοντας ξανά αν βάλουμε την εξίσωση  $W = (A + \lambda I)^{\lambda+1} \bar{V}(t) = (w_{ij})$ , προκύπτει ότι

$$\text{Tr}\{W^T F(W)\} = \sum_{i=1}^n w_{ij} f(w_{ij}) = \begin{cases} > 0 & \text{αν } w_{ij} \neq 0 \\ = 0 & \text{αν } w_{ij} = 0 \end{cases} \quad (2.5.13)$$

Η παράμετρος σχεδίασης ικανοποιεί  $\beta > 0$ , και βάσει της, προκύπτει το εξής

$$\frac{dL(\bar{V}(t), t)}{dt} = \begin{cases} < 0 & \text{αν } (A + \lambda I)^{\lambda+1} \bar{V}(t) \neq 0 \\ = 0 & \text{αν } (A + \lambda I)^{\lambda+1} \bar{V}(t) = 0 \end{cases} \quad (2.5.14)$$

Εφόσον ο πίνακας  $(A + \lambda I)^{\lambda+1}$  είναι μη ιδιάζων για κάθε  $\lambda > 0$ , μπορεί να διατυπωθεί στην ακόλουθη μορφή:

$$\frac{dL(\bar{V}(t), t)}{dt} = \begin{cases} < 0 & \text{if } \bar{V}(t) \neq 0 \\ = 0 & \text{if } \bar{V}(t) = 0 \end{cases} \quad (2.5.15)$$



αυτό σημαίνει ότι το  $\bar{V}_{v(0)}$  αποτελεί μια μορφή της γενικής λύσης της εξίσωσης  $AXA=A$ , οπότε  $V(0)$  κάνει τον ρόλο του αυθαίρετου πίνακα [49], [50].

**Θεώρημα 2.5.5.3[60]**

Θεωρούμε έναν αυθαίρετο πίνακα  $A \in R^{m \times n}$  ο οποίος αν χρησιμοποιηθεί με μια περιττή και μονότονα αύξουσα συνάρτηση  $f(\bullet)$  για την συνάρτηση ενεργοποίησης  $F(\bullet)$  και το  $\gamma > 0$  τότε ο πίνακας κατάστασης  $V(t) \in R^{n \times m}$  του GNN θα ικανοποιεί την σχέση  $AV(t)A \rightarrow A$  όταν το  $t \rightarrow \infty$  για μια αρχική κατάσταση  $V(0)$ .

**ΑΠΟΔΕΙΞΗ**

Έστω  $\bar{V}(t) = V(t) - A^\dagger$  να είναι η διαφορά αναμεσά στον πίνακα μεταβλητών της κατάστασης  $V(t)$  του νευρωνικού δικτύου και του αντίστροφου Moore-Penrose  $A^\dagger$ . Η αντίσταση της δυναμικής εξίσωσης συνεπάγεται ότι:

$$\begin{aligned} \frac{d\bar{V}(t)}{dt} &= \frac{dV(t)}{dt} = \gamma A^T F(A - AV(t)A) A^T \\ &= \gamma A^T F(A - A(\bar{V}(t) + A^\dagger)A) A^T \\ &= -\gamma A^T F(A\bar{V}(t)A) A^T \end{aligned} \quad (2.5.22)$$

έτσι η συνάρτηση Lyapunov θα οριστεί ως

$$L(\bar{V}(t), t) = \frac{1}{2} \|\bar{V}(t)\|_F^2 = \frac{1}{2} Tr(\bar{V}(t)^T \bar{V}(t)) \quad (2.5.23)$$

έτσι η ανισότητα  $L(\bar{V}(t), t) \geq 0$  θα ισχύει για  $\bar{V}(t) \neq 0$ . Σύμφωνα με το (2.5.24) αλλά και το (2.5.23) και την εξίσωση  $dTr(X^T X) = 2Tr(X^T dX)$  η χρονική παράγωγος  $L(\bar{V}(t), t)$  ικανοποιεί την παρακάτω σχέση:

$$\begin{aligned} \frac{dL(\bar{V}(t), t)}{dt} &= \frac{1}{2} \frac{dTr(\bar{V}(t)^T \bar{V}(t))}{dt} \\ &= Tr(\bar{V}(t)^T \frac{d\bar{V}(t)}{dt}) \\ &= -\gamma Tr[\bar{V}(t)^T A^T F(A\bar{V}(t)A) A^T] \\ &= -\gamma Tr[(A\bar{V}(t))^T F(A\bar{V}(t)A) A^T] \\ &= -\gamma Tr[A^T (A\bar{V}(t))^T F(A\bar{V}(t)A)] \\ &= -\gamma Tr[(A\bar{V}(t)A)^T F(A\bar{V}(t)A)] \end{aligned} \quad (2.5.24)$$

δεδομένου ότι η συνάρτηση  $f(\bullet)$  είναι περιττή και μονότονα αύξουσα, θα προκύψει άμεσα ότι:

$$f(x) \begin{cases} > 0 \text{ αν } x > 0 \\ = 0 \text{ αν } x = 0 \\ < 0 \text{ αν } x < 0 \end{cases} \quad (2.5.25)$$

και αυτό συνεπάγεται

$$xf(x) \begin{cases} > 0 \text{ αν } x \neq 0 \\ = 0 \text{ αν } x = 0 \end{cases} \quad (2.5.26)$$

σύμφωνα με προηγούμενες παρατηρήσεις αν χρησιμοποιηθεί το  $W(t) = A\bar{V}(t)A = (w_{ij})$  θα προκύψει ότι

$$w_{ij}f(w_{ij}) = \begin{cases} > 0 & \text{αν } w_{ij} \neq 0 \\ = 0 & \text{αν } w_{ij} = 0 \end{cases} \quad (2.5.27)$$

Λόγω του ότι η σχέση  $\gamma > 0$  ικανοποιείται από τις παραμέτρους και λαμβάνοντας υπόψιν την (2.5.25) θα προκύψει ότι

$$\frac{dL(\bar{V}(t),t)}{dt} = -\gamma \text{Tr}[(W^T F(W))] = -\gamma \sum_{i=1}^m \Sigma w_{ij}f(w_{ij}) = \begin{cases} < 0 & \text{αν } A\bar{V}(t)A \neq 0 \\ = 0 & \text{αν } A\bar{V}(t)A = 0 \end{cases} \quad (2.5.28)$$

και στην συνέχεια από την σχέση από την (2.5.29) προκύπτει

$$\frac{dL(\bar{V}(t),t)}{dt} = \begin{cases} < 0 & \text{αν } A\bar{V}(t)A - A \neq 0 \\ = 0 & \text{αν } A\bar{V}(t)A - A = 0 \end{cases} \quad (2.5.29)$$

έτσι αυτό θα συνεπάγεται:

- $\frac{dL(\bar{V}(t),t)}{dt} < 0$  για κάθε κατάσταση που το  $V(t)$  ικανοποιεί την σχέση  $W(t)=AV(t)A-A \neq 0$
- $\frac{dL(\bar{V}(t),t)}{dt} = 0$  στην κατάσταση ισορροπίας του  $V(t)$  που ικανοποιεί την σχέση  $W(t)=AV(t)A-A=0$  και με βάση την (2.5.30) το  $W(t)=AV(t)A-A$  θα συγκλίνει στον μηδενικό πίνακα από μια αυθαίρετη αρχική τιμή  $V(0)$ .

**Παράδειγμα 2.5.5.3 [60]** Περιλαμβάνοντας τον ακόλουθο πίνακα 4x4:

$$\begin{bmatrix} -0.3 & -0.4 & 0.1 & 0.1 \\ -0.4 & 0.3 & -0.2 & -0.2 \\ -1.9 & 3.8 & 5.1 & 5.1 \\ 2.1 & -4.2 & -4.9 & -4.9 \end{bmatrix}$$

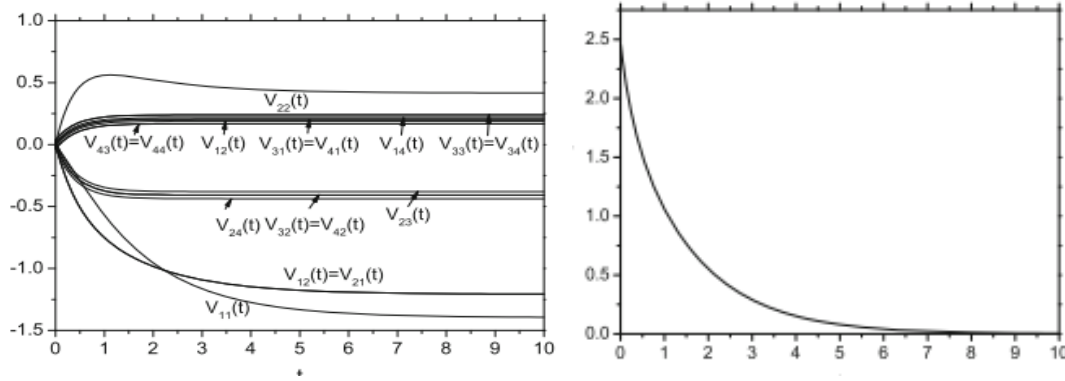
Η τάξη και ο δείκτης του πίνακα είναι αντίστοιχα  $\text{rank}(A)=2$  και  $k=\text{ind}(A)=2$ . Ο πίνακας αυτός χρησιμοποιείται για να παρουσιαστεί η συμπεριφορά των μοντέλων GNN-MP και GNN-D. Στο παρόν παράδειγμα, η συνάρτηση ενεργοποίησης S-P-S χρησιμοποιείται και στα δύο μοντέλα. Επιπλέον, η παράμετρος  $\gamma$  και ο χρόνος προσομοίωσης  $t_{\max}$  λαμβάνουν τις εξής τιμές:

για το μοντέλο GNN-MP,  $\gamma=10$  και  $t_{\max}=10$  ενώ για το μοντέλο ,GNN-D  $\gamma=10^4$  και  $t_{\max}=0.03$ .

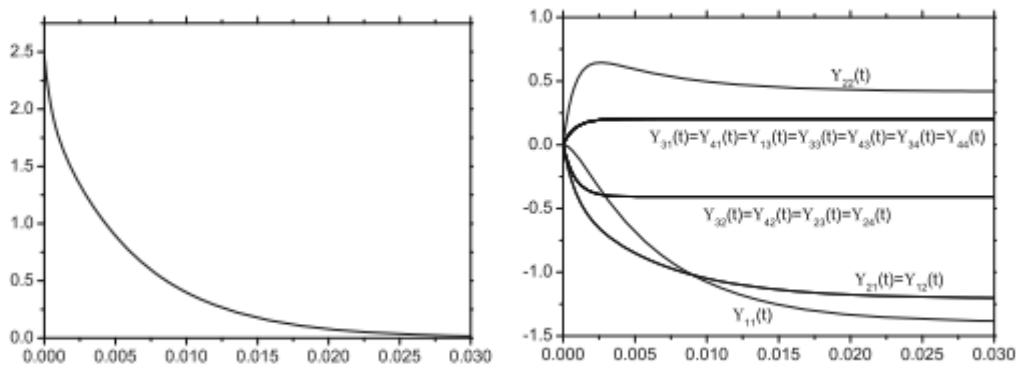
Και στις δύο περιπτώσεις, οι αρχικοί πίνακες κατάστασης (αντίστοιχα  $V(0)$  και  $Y(0)$ ) τίθενται ίσοι με τον μηδενικό πίνακα διαστάσεων 4x4. Οι τροχιές σύγκλισης όλων των στοιχείων του πίνακα εισόδου, καθώς και οι τροχιές σύγκλισης των αντίστοιχων υπολοίπων νορμών, παρουσιάζονται στα Σχήματα 5 και 6, αντίστοιχα. Το μοντέλο GNN-MP χρησιμοποιεί τη νόρμα υπολοίπου  $\|AV(t)A-A\|$ , ενώ το μοντέλο GNN-D βασίζεται στη νόρμα υπολοίπου  $\|A^2V(t)A-A\|$ .

Παρατηρείται ότι οι ακριβείς αντίστροφοι Moore-Penrose και Drazin είναι ίσοι με:

$$A^\dagger = \begin{bmatrix} -\frac{342}{245} & -\frac{296}{245} & \frac{93}{490} & \frac{107}{490} \\ \frac{296}{245} & \frac{102}{245} & -\frac{93}{245} & -\frac{107}{245} \\ \frac{10}{49} & -\frac{20}{49} & \frac{47}{196} & \frac{33}{196} \\ \frac{10}{49} & -\frac{20}{49} & \frac{47}{196} & \frac{33}{196} \end{bmatrix}, \quad A^{D=} = \begin{bmatrix} -\frac{342}{245} & -\frac{296}{245} & \frac{10}{49} & \frac{10}{49} \\ \frac{296}{245} & \frac{102}{245} & -\frac{20}{49} & -\frac{20}{49} \\ \frac{10}{49} & -\frac{20}{49} & \frac{10}{49} & \frac{10}{49} \\ \frac{10}{49} & -\frac{20}{49} & \frac{10}{49} & \frac{10}{49} \end{bmatrix}$$



Σχήμα 2.14: Τροχιές σύγκλισης του δικτύου GNN-MP(αριστερά σχήμα) και η τροχιά της νόρμας σφάλματος (δεξιά σχήμα) σε χρονικό διάστημα  $[0,10]$ s από παράδειγμα 2.5.5.3 [60].



Σχήμα 2.15 Τροχιές σύγκλισης του δικτύου GNN-MP(αριστερά σχήμα) και η τροχιά της νόρμας σφάλματος (δεξιά σχήμα) σε χρονικό διάστημα  $[0,0.03]$ s από παράδειγμα 2.5.5.3 [60].

### ΘΕΩΡΗΜΑ 2.5.5.4[61]

Ανάλογα με τον αρχικό πίνακα  $V(0)$ , ο οριακός πίνακας ικανοποιεί τα παρακάτω:

$$\bar{V}_{V(0)} \in A\{1, 2\} \text{ αν και μόνο αν } V(0)_{22} = V(0)_{21}A_{11}V(0)_{12};$$

$$\bar{V}_{V(0)} \in A\{1, 3\} \text{ αν και μόνο αν } V(0)_{12} = 0;$$

$$\bar{V}_{V(0)} \in A\{1, 4\} \text{ αν και μόνο αν } V(0)_{21} = 0;$$

$$\bar{V}_{V(0)} = A^\dagger \text{ αν και μόνο αν } V(0)_{12} = 0, V(0)_{21} = 0 \text{ και } V(0)_{22} = 0$$

### ΑΠΟΔΕΙΞΗ

Αποδεικνύοντας το γράφουμε τους πίνακες  $A$  και  $V(t)$  σε μορφή μπλοκ παρακάτω:

$$A = \begin{bmatrix} A_{11} & 0 \\ 0 & 0 \end{bmatrix}; \begin{bmatrix} R(A^T) \\ N(A) \end{bmatrix} \rightarrow \begin{bmatrix} R(A^T) \\ N(A^T) \end{bmatrix}$$

$$V(t) = \begin{bmatrix} V(t)_{11} & V(t)_{12} \\ V(t)_{21} & V(t)_{22} \end{bmatrix}; \begin{bmatrix} R(A) \\ N(A^T) \end{bmatrix} \rightarrow \begin{bmatrix} R(A^T) \\ N(A) \end{bmatrix} \quad (2.5.30)$$

Τώρα ο πίνακας  $A_{11}$  είναι κανονικός με διαστάσεις  $r \times r$  και  $r = \text{rank}(A)$ .

Επειδή όμως οι πίνακες  $AA^\dagger$  και  $A^\dagger A$  αποτελούν ορθογώνια προβολή στους χώρους  $R(A)$  και  $R(A^T)$  και προκύπτει από τον πίνακα  $\tilde{V}_{V(0)}$  και έχει την παρακάτω μορφή:

$$\tilde{V}_{V(0)} = \begin{bmatrix} A_{11}^{-1} & V(0)_{12} \\ V(0)_{21} & V(0)_{22} \end{bmatrix} : \begin{bmatrix} R(A) \\ N(A^T) \end{bmatrix} \rightarrow \begin{bmatrix} R(A^T) \\ N(A) \end{bmatrix} \quad (2.5.31)$$

### Θεώρημα 2.5.5.5[60]

Θεωρούμε έναν τετραγωνικό πίνακα  $A \in R^{n \times n}$ . Αν χρησιμοποιήσουμε μια περιττή και μονότονα αύξουσα συνάρτηση  $f(\bullet)$  για να ορίσουμε την συνάρτηση του πίνακα  $F(\bullet)$  τότε ο πίνακας κατάστασης  $Y(t) \in R^{n \times n}$  του GNN θα ικανοποιεί την σχέση  $A^k Y(t) A \rightarrow A^k$  όταν το  $t \rightarrow +\infty$  για ένα αρχικό πίνακα  $Y(0)$ .

### ΑΠΟΔΕΙΞΗ

Έστω ότι  $\bar{Y}(t) = Y(t) - A^D$  ότι είναι η διαφορά μεταξύ του πίνακα των μεταβλητών κατάστασης  $Y(t)$  του GNN και του Drazin αντίστροφου  $A^D$ . Η αντικατάσταση  $Y(t) = \bar{Y}(t) + A^D$  στην δυναμική εξίσωση θα συνεπάγεται ότι

$$\begin{aligned} \frac{d\bar{Y}(t)}{dt} &= \frac{dY(t)}{dt} = \gamma(A^k)^T F(A^k - A^k Y(t) A) A^T \\ &= \gamma(A^k)^T F(A^k - A^k(\bar{Y}(t) + A^D)A) A^T \quad (2.5.32) \\ &= -\gamma(A^k)^T F(A^k \bar{Y}(t) A) A^T \end{aligned}$$

Έτσι η υποψήφια συνάρτηση Lyapunov ορίζεται παρόμοια με την

$$L(\bar{Y}(t), t) = \frac{1}{2} \|\bar{Y}(t)\|_F^2 = \frac{1}{2} \text{Tr}(\bar{Y}(t)^T \bar{Y}(t)) \quad (2.5.33)$$

**Παράδειγμα 2.5.5.5[60]** Περιλαμβάνοντας τον πίνακα

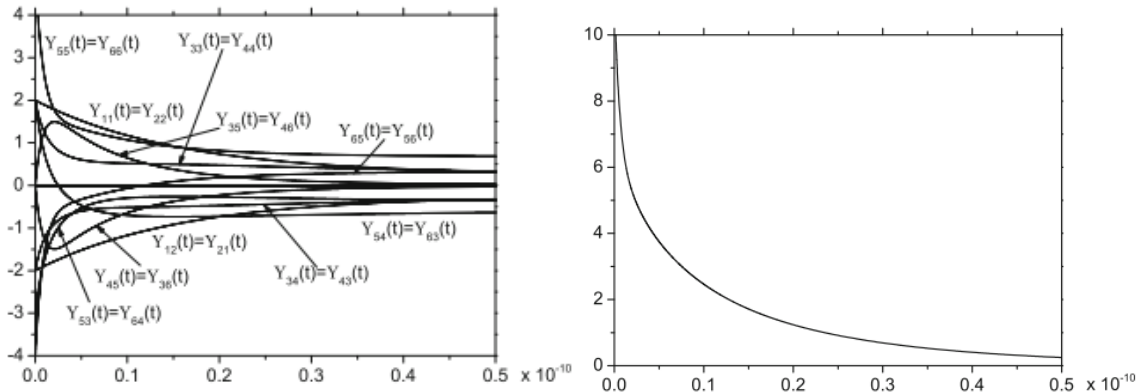
$$A = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ -1 & -1 & 1 & -1 & 0 & 0 \\ -1 & -1 & -1 & 1 & 0 & 0 \\ -1 & -1 & -1 & 0 & 2 & -1 \\ -1 & -1 & 0 & -1 & -1 & 2 \end{bmatrix}$$

με  $\text{ind}(A)=2$ . Χρησιμοποιώντας  $k=\text{ind}(A)=2$ , γραμμική συνάρτηση ενεργοποίησης, αρχική κατάσταση  $V(0) = A^2$ ,  $\gamma=10^9$  με τον επίλυτή ode15s στο χρονικό διάστημα  $[0, 10^{-9}]$ , το μοντέλο GNN-D παράγει την ακόλουθη προσέγγιση

$$A^d = \begin{bmatrix} 0.25 & -0.25 & 0 & 0 & 0 & 0 \\ -0.25 & 0.25 & 0 & -0.25 & 5.986e-14 & 5.587e-14 \\ 0 & 0 & -0.25 & 0.25 & 5.656e-14 & 5.277e-14 \\ 0 & 0 & -0.4166 & -0.5833 & 0.6666 & 0.3333 \\ 0 & 0 & -0.5833 & -0.4167 & 0.3333 & 0.6667 \end{bmatrix}$$

Από τον ακριβή αντίστροφο του Drazin

$$A^D = \begin{bmatrix} \frac{1}{4} & -\frac{1}{4} & 0 & 0 & 0 & 0 \\ -\frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} & -\frac{1}{4} & 0 & 0 \\ 0 & 0 & -\frac{1}{4} & \frac{1}{4} & 0 & 0 \\ 0 & 0 & -\frac{5}{12} & -\frac{7}{12} & \frac{2}{3} & \frac{1}{3} \\ 0 & 0 & -\frac{7}{12} & -\frac{5}{12} & \frac{1}{3} & \frac{2}{3} \end{bmatrix}$$



Σχήμα 2.16 Τροχίες σύγκλισης του δικτύου GNN-D (αριστερά σχήμα) και η τροχιά της νόρμας σφάλματος (δεξιά σχήμα) σε χρονικό διάστημα  $[0, 5 \times 10^{-11}]$ s από παράδειγμα 2.5.5.5[60]

### Παραπομπή 1

Με βάση τις προϋποθέσεις του θεωρήματος έχουμε τα εξής:

1.  $\tilde{V}_{V(0)} \in A\{1,2,3\}$  αν και μόνο αν  $N(V(0)) \supseteq N(A^T)$ .
2.  $\tilde{V}_{V(0)} \in A\{1,2,4\}$  αν και μόνο αν  $R(V(0)) \subseteq R(A^T)$ .
3.  $\tilde{V}_{V(0)} = A^\dagger$  αν και μόνο αν  $N(V(0)) \supseteq N(A^T)$  και  $R(V(0)) \subseteq R(A^T)$ .

## 2.6 Μη Γραμμικά Ενεργοποιημένο Αναδρομικό Νευρωνικό Δίκτυο για τον Υπολογισμό του Αντίστροφου Drazin

Ο αντίστροφος Drazin έχει εφαρμοστεί με επιτυχία σε διάφορα πεδία της επιστήμης. Οι κύριοι τομείς εφαρμογής περιλαμβάνουν την εύρεση λύσεων κλειστής μορφής για ιδιαίζουσες διαφορικές εξισώσεις με συντελεστές πινάκων, την επίλυση εξισώσεων, τη δυναμική συστημάτων πολλαπλών σωμάτων, καθώς και την εύρεση λύσεων σε διάφορες επαναληπτικές μεθόδους [69-70, 71, 72-73].

Με βάση το [74] χρησιμοποιείται στη θεωρία του Drazin αντίστροφου για να γίνει ανάλυση των ιδιαίζοντων συστημάτων πολλαπλών πρακτόρων. Επίσης, η λύση του αντίστροφου Drazin  $A^D b$  επιλύει το ιδιαίζον γραμμικό σύστημα  $Ax = b$  αν και μόνο αν  $b \in R(A^k)$  [75].

Η προσέγγιση των νευρωνικών δικτύων στον παράλληλο υπολογισμό και την επεξεργασία σημάτων έχει επιδειχθεί με επιτυχία μέσω μιας ποικιλίας νευροδυναμικών μοντέλων με δυνατότητες μάθησης

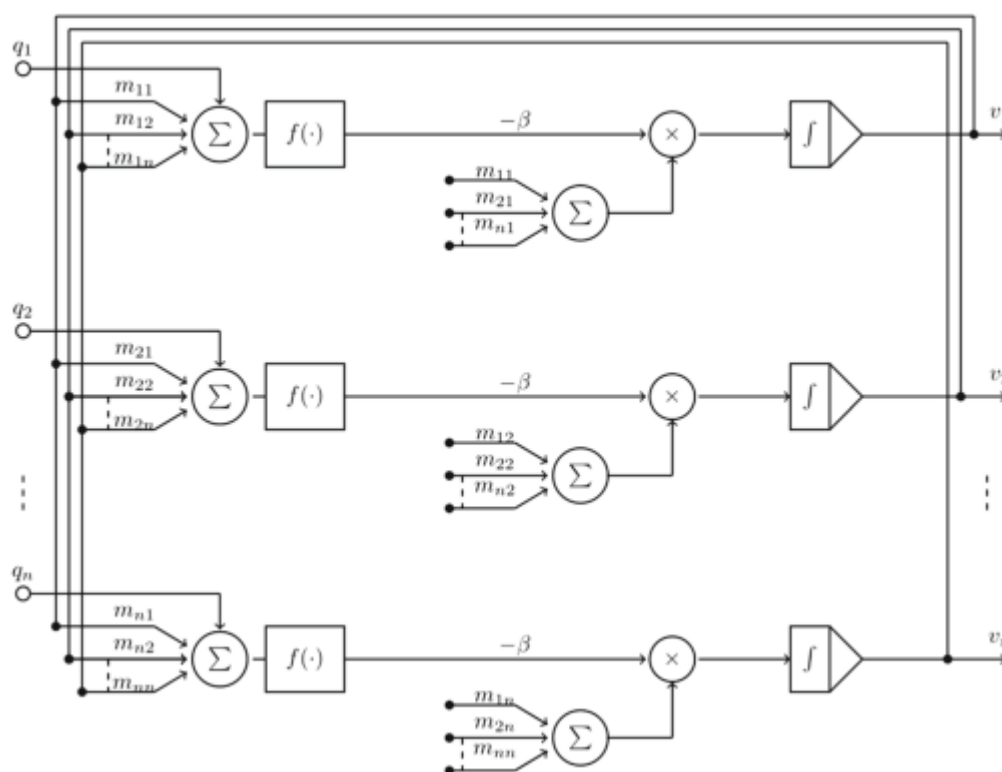
[76,77]. Πρόσφατα, έχουν αναπτυχθεί διάφορα μοντέλα γραμμικών νευρωνικών δικτύων (RNN) με σκοπό την αριθμητική αξιολόγηση του αντίστροφου πίνακα και του γενικευμένου αντίστροφου για ορθογώνιους πίνακες πλήρους βαθμού γραμμών ή στηλών [78,79,80,81].

Ένας νέος τύπος RNN με μιγαδικές τιμές, που ονομάζεται νευρωνικό δίκτυο Zhang (ZNN), προτάθηκε το 2001 και έχει αξιοποιηθεί εκτενώς στην επίλυση διαφόρων χρονικά μεταβαλλόμενων μιγαδικών προβλημάτων γενικευμένης αντιστροφής.

Ο σχεδιασμός των μιγαδικών μοντέλων ZNN προκύπτει από την επιλογή μιας μιγαδικής συνάρτησης παρακολούθησης σφάλματος με τιμές σε μορφή πίνακα, η οποία ονομάζεται συνάρτηση Zhang. Διαφορετικά μοντέλα ZNN για τον υπολογισμό του αντίστροφου Moore-Penrose ενός χρονικά μεταβαλλόμενου πίνακα πλήρους βαθμού γενικεύτηκαν, διερευνήθηκαν και αναλύθηκαν στο [82].

Οι Liao και Zhang στο [83] πρότειναν πέντε διαφορετικές μιγαδικές συναρτήσεις και, αντίστοιχα ανέπτυξαν και διερεύνησαν πέντε διαφορετικά μιγαδικά μοντέλα ZNN για τον υπολογισμό του χρονικά μεταβαλλόμενου μιγαδικού ψευδοαντίστροφου. Μια αρχιτεκτονική νευρωνικού δικτύου εμπρόσθιας τροφοδότησης για τον υπολογισμό του αντίστροφου Drazin  $A^D$  προτάθηκε στο [65].

Ένα RNN για τον υπολογισμό του αντίστροφου Drazin ενός πραγματικού τετραγωνικού πίνακα σε πραγματικό χρόνο παρουσιάστηκε στο [63]. Η δυναμική εξίσωση και το σχετικό τεχνητό RNN για τον υπολογισμό του αντίστροφου Drazin ενός αυθαίρετου τετραγωνικού πραγματικού πίνακα, χωρίς κανέναν περιορισμό στις ιδιοτιμές του, προτάθηκε στο [66].



Σχήμα 2.6.1: Αρχιτεκτονική του RNN που υλοποιεί το μοντέλο Ia [62].

**Θεώρημα 2.6.1[62]**

Θεωρούμε έναν πραγματικό πίνακα  $A \in R^{n \times n}$  με δείκτη  $\text{Ind}(A) = k$ . Αν γίνει χρήση μιας περιττής και μονότονα αύξουσας συνάρτησης  $f(\bullet)$  για να οριστεί η συνάρτηση ενεργοποίησης  $f(\bullet)$ , τότε ο πίνακας κατάστασης  $V(t) \in R^{n \times n}$  συγκλίνει καθολικά στο Drazin αντίστροφο  $A^D$ , ξεκινώντας από μια αυθαίρετη αρχική κατάσταση  $V(0)$ .

**Απόδειξη**

Έστω  $\bar{V}(t) = V(t) - A^D$  η διαφορά της λύσης μεταξύ του  $V(t)$  που παράγεται από το μοντέλο νευρωνικού δικτύου αλλά και του Drazin αντίστροφου  $A^D$ . Με βάση το εξής:

$$(A + \lambda I)^{l+1} A^D - A^l \approx 0 \quad (2.6.1.1)$$

Η αντικατάσταση  $\bar{V}(t) = V(t) - A^D$  στην δυναμική εξίσωση συνεπαγεται ότι :

$$\frac{d\bar{V}(t)}{dt} = -\beta((A + \lambda I)^{l+1})^T F((A + \lambda I)^{l+1} \bar{V}(t)) \quad (2.6.1.2)$$

Η συνάρτηση Lyapunov θα οριστεί ως:

$$L(\bar{V}(t), t) = \frac{1}{2} \|\bar{V}(t)\|_F^2 = \frac{1}{2} \text{Tr}(\bar{V}(t)^T \bar{V}(t)) \quad (2.6.1.3)$$

Η ανισότητα  $L(\bar{V}(t), t) \geq 0$  ισχύει για  $\bar{V}(t) \neq 0$ . Οσον αφορά την εξίσωση (2.6.1.3) λαμβάνοντας υπόψιν την (2.6.1.2) η χρονική παράγωγος  $L(\bar{V}(t), t) \geq 0$  ικανοποιεί :

$$\begin{aligned} \frac{dL(\bar{V}(t), t)}{dt} &= \frac{1}{2} \frac{d\text{Tr}(\bar{V}(t)^T \bar{V}(t))}{dt} \\ &= \frac{1}{2} \text{Tr}(\bar{V}(t)^T \frac{d\bar{V}(t)}{dt}) \\ &= -\frac{1}{2} \beta \text{Tr}\{((A + \lambda I)^{l+1} \bar{V}(t))^T F((A + \lambda I)^{l+1} \bar{V}(t))\} \quad (2.6.1.4) \end{aligned}$$

Επειδή η βαθμωτη συνάρτηση  $f(\bullet)$  είναι περιττή και μονότονα αύξουσα θα προκύψει ότι  $f(-x) = -f(x)$

$$f(x) = \begin{cases} > 0 & \text{αν } x > 0 \\ = 0 & \text{αν } x = 0 \\ < 0 & \text{αν } x < 0 \end{cases} \quad (2.6.1.5)$$

Το οποίο συνεπάγεται

$$xf(x) = \begin{cases} > 0 & \text{αν } x \neq 0 \\ = 0 & \text{αν } x = 0 \end{cases} \quad (2.6.1.6)$$

Θεωρούμε ξανά την εξίσωση (2.6.1.4). Αν γίνει χρήση του συμβολισμού  $W = (A + \lambda I)^{l+1} \bar{V}(t) = (w_{ij})$  προκύπτει ότι

$$\text{Tr}\{W^T F(W)\} = \sum_{i=1}^n w_{ij} f(w_{ij}) = \begin{cases} > 0 & \text{αν } w_{ij} \neq 0 \\ = 0 & \text{αν } w_{ij} = 0 \end{cases} \quad (2.6.1.7)$$

Λόγω αυτού η παράμετρος ικανοποιεί την συνθήκη  $\beta > 0$ , λαμβάνοντας υπόψη την (2.6.1.4) θα προκύψει το παρακάτω

$$\frac{dL(\bar{V}(t), t)}{dt} = \begin{cases} < 0 & \text{αν } (A + \lambda I)^{l+1} \bar{V}(t) \neq 0 \\ = 0 & \text{αν } (A + \lambda I)^{l+1} \bar{V}(t) = 0 \end{cases} \quad (2.6.1.8)$$

Επειδή ο πίνακας είναι μη ιδιάζων για οποιοδήποτε  $\lambda > 0$  η (2.6.1.8) μπορεί να γραφεί στην ακόλουθη μορφή

$$\frac{dL(\bar{V}(t), t)}{dt} = \begin{cases} < 0 \text{ αν } \bar{V}(t) \neq 0 \\ = 0 \text{ αν } \bar{V}(t) = 0 \end{cases} \quad (2.6.1.9)$$

Σύμφωνα με την (2.6.1.9), με βάση την θεωρία ευστάθειας του Lyapunov, το  $\bar{V}(t) = V(t) - A^D$  συγκλίνει καθολικά στον μηδενικό πίνακα, από οποιαδήποτε αρχική τιμή. Λαμβάνοντας υπόψη τη μοναδικότητα του Drazin αντίστροφου, προκύπτει ότι  $V(t) \rightarrow A^D$  καθώς  $\lambda \rightarrow 0$ . Επομένως, ο πίνακας κατάστασης  $V(t)$  συγκλίνει καθολικά στο θεωρητικό αντίστροφο Drazin  $A^D$  και η απόδειξη της καθολικής σύγκλισης του RNN, το οποίο ορίζεται από τη δυναμική Εξίσωση, έχει ολοκληρωθεί.

### Παράδειγμα 2.6.1[62]

Ο ακριβής αντίστροφος Drazin του παραμετρικού πίνακα

$$A(a) = \begin{bmatrix} 1+a & a & 1+a \\ a^2 & -1+a & a \\ 1+a & a & 1+a \end{bmatrix}, A \in \mathbb{R}$$

Είναι ίσος με :

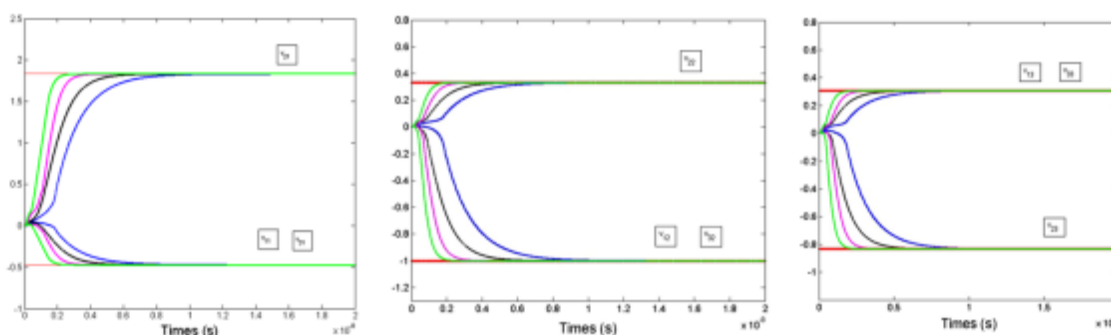
$$A(a)^D = \begin{bmatrix} \frac{1-a+2a^3-2a^4}{(2-a^2+a^3)^2} & \frac{a}{2-a^2+a^3} & \frac{1-a+a^2+a^4}{(2-a^2+a^3)^2} \\ \frac{a(-1+a^2+a^3)}{(1+a)(2-2a+a^2)^2} & \frac{2}{2-2a+a^2} & \frac{3a-2a^3}{(1+a)(2-2s+s^2)^2} \\ \frac{1-a+2a^3-2a^4}{(2-a^2+a^3)^2} & \frac{a}{2-a^2+a^3} & \frac{1-a+a^2+a^4}{(2-a^2+a^3)^2} \end{bmatrix}$$

Επειδή  $k = \text{ind}(A) = 1$ , είναι δυνατό να επιλέξουμε  $k = 1$ . Δοκιμάζουμε την τροχιά των μεταβλητών και τα υπολειπόμενα σφάλματα χρησιμοποιώντας τις εξισώσεις:

$$\frac{dv_j(t)}{dt} = -\beta \sum_{k=1}^n m_{jk} F(\sum_{k=1}^n m_{kj} v_j(t) - q_j) \text{ και}$$

$$\frac{dv_j(t)}{dt} = -\beta F(\sum_{k=1}^n m_{kj} v_j(t) - q_j) \text{ με } l = 10^{-6} \text{ και } \beta = 10^9.$$

Το ακριβές αντίστροφο Drazin του πίνακα  $A(1)$  είναι:



Σχήμα 2.6.1: Συμπεριφορά σύγκλισης των μεταβλητών κατάστασης του μοντέλου Ia στο Παράδειγμα 2.6.1 1ο σχημα, 2ο σχήμα και 3ο σχήμα[62]

Matrix	$\lambda$	$\beta$	t (s)	$\ Ia(A) - A^D\ $	$\ I(A) - A^D\ $	$\ II(A) - A^D\ $	$\ FF(A) - A^D\ $
A(1)	$10^{-8}$	$10^8$	$10^{-6}$	$1.1286 \times 10^{-6}$	$3.5423 \times 10^{-4}$	$1.8954 \times 10^{-4}$	$4.3451 \times 10^{-2}$
A(2)	$10^{-8}$	$10^8$	$10^{-8}$	$4.8952 \times 10^{-6}$	$2.1203 \times 10^{-4}$	$6.2546 \times 10^{-4}$	$2.8645 \times 10^{-2}$

Πίνακας 2:Αριθμητική συγκριτική δοκιμή με το FF και τα μοντέλα του Stanimirović [66, 63] [65].

Matrix	$\lambda$	$\beta$	t (s)	$\ I Ib(A) - A^D\ $	$\ I(A) - A^D\ $	$\ II(A) - A^D\ $	$\ FF(A) - A^D\ $
A(1)	$10^{-8}$	$10^8$	$10^{-6}$	$4.0256 \times 10^{-6}$	$1.2536 \times 10^{-4}$	$2.7845 \times 10^{-4}$	$1.3215 \times 10^{-2}$
A(2)	$10^{-8}$	$10^8$	$10^{-8}$	$2.7723 \times 10^{-6}$	$3.0021 \times 10^{-4}$	$4.8562 \times 10^{-4}$	$3.2221 \times 10^{-2}$

Πίνακας 3:Αριθμητική συγκριτική δοκιμή με το FF και τα μοντέλα του Stanimirović [66, 63] [65].

$$A(1)^D = \begin{bmatrix} 0 & 0.5 & 0 \\ 0.5 & -2 & 0.5 \\ 0 & 0.5 & 0 \end{bmatrix}$$

Οι τροχιές των μεταβλητών κατάστασης οι οποίες αντιστοιχούν στις μηδενικές αρχικές συνθήκες του μοντέλου Ia σε  $10^{-7}$  φαίνονται στο σχήμα 2.6.1. Αποτελέσματα σχετικά με την ακρίβεια από τις προσεγγίσεις από τις μεθόδους, το μοντέλο RNN από τα [ 66,63 ] και από την προσέγγιση του νευρωνικού δικτύου FF από [65] επισημαίνονται στους πίνακες 2 και 3. Οι συμβολισμοί  $Ia(\bullet)$ ,  $I Ib(\bullet)$ ,  $I(\bullet)$ ,  $II(\bullet)$ ,  $FF(\bullet)$  δείχνουν τις αριθμητικές τιμές που παράγονται από το μοντέλο Ia.

### Θεώρημα 2.6.2[63]

Θεωρήστε έναν πραγματικό πίνακα  $A \in \mathbb{R}^{n \times n}$  με δείκτη  $\text{Ind}(A) = k$ , αν χρησιμοποιηθεί μια περιττή και μονότονα αύξουσα συνάρτηση  $f(\bullet)$  στο εσωτερικό της συνάρτησης ενεργοποίησης  $f(\bullet)$  τότε ο πίνακας κατάστασης  $V(t) \in \mathbb{R}^{n \times n}$  του μοντέλου RNN συγκλίνει καθολικά στο ακριβές αντίστροφο Drazin  $A^D$  του A, ξεκινώντας από οποιαδήποτε αρχική κατάσταση  $V(0)$  και μοντέλο RNN να είναι το παρακάτω:

$$\frac{dV_0(t)}{dt} = -\beta(G_0 A + \lambda I)^T F((G_0 A + \lambda I) V_0(t) - G_0)$$

$$G_0 = A^l (A^{2l+1})^T A^l, l \geq k, V(0) = V_0(2.6.2.1)$$

### Απόδειξη

Έστω  $\bar{V}_0(t) = V_0(t) + A^D$  η διαφορά μεταξύ της λύσης  $V_0(t)$  που παράγεται από το μοντέλο νευρωνικού δικτύου (2.6.2.1) και του Drazin αντίστροφου  $A^D$ . Η αντικατάσταση  $V_0(t) = \bar{V}_0(t) + A^D$  στο μοντέλο (2.6.2.1) οδηγεί σε:

$$\frac{d\bar{V}_0(t)}{dt} = -\beta((G_0 A + \lambda I)^T F((G_0 A + \lambda I) \bar{V}_0(t))) (2.6.2.2)$$

Η συνάρτηση Lyapunov μπορεί να οριστεί παρόμοια με την (2.6.1.3), ως εξής:

$$L(\bar{V}_0(t), t) = \frac{1}{2} \|\bar{V}_0(t)\|_F^2 = \frac{1}{2} \text{Tr}(\bar{V}_0(t)^T \bar{V}_0(t)) (2.6.2.3)$$

Προφανώς,  $L(\bar{V}_0(t), t) \geq 0$  για  $\bar{V}_0(t) \neq 0$ . Η χρονική παράγωγος της συνάρτησης  $L(\bar{V}_0(t), t)$  ικανοποιεί:

$$\begin{aligned} \frac{dL(\bar{V}_0(t), t)}{dt} &= \frac{1}{2} \frac{dTr(\bar{V}_0(t)^T \bar{V}_0(t))}{dt} \\ &= Tr(\bar{V}_0(t)^T \frac{d\bar{V}_0(t)}{dt}) \\ &= -\beta Tr\{((G_0 A + \lambda I) \bar{V}_0(t))^T F((G_0 A + \lambda I) \bar{V}_0(t))\} \end{aligned} \quad (2.6.2.4)$$

Επειδή η βαθμωτη συνάρτηση  $f(\bullet)$  είναι περιττή και μονότονα αύξουσα και η παράμετρος σχεδιασμού ικανοποιεί τη συνθήκη  $\beta > 0$ , παρόμοια με την απόδειξη του Θεωρήματος (2.6.1), μπορεί να επαληθευτεί ότι:

$$\frac{dL(\bar{V}_0(t), t)}{dt} = \begin{cases} < 0 & \text{αν } (G_0 A + \lambda I) \bar{V}_0(t) \neq 0 \\ = 0 & \text{αν } (G_0 A + \lambda I) \bar{V}_0(t) = 0 \end{cases} \quad (2.6.2.5)$$

Λαμβάνουμε υπόψη ότι  $s(G_0 A) \geq 0$ . Επειδή ο πίνακας  $G_0 A + \lambda I$  είναι μη ιδιάζων για οποιοδήποτε  $\lambda > 0$ , ισχύει το εξής

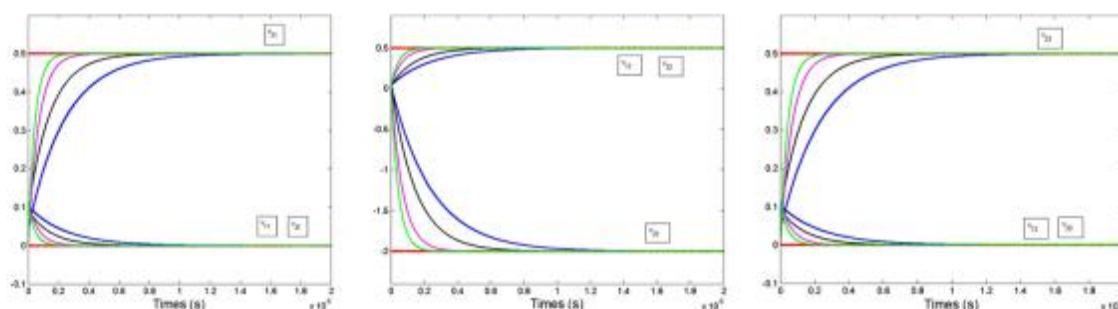
$$\frac{dL(\bar{V}_0(t), t)}{dt} = \begin{cases} < 0 & \text{αν } \bar{V}_0(t) \neq 0 \\ = 0 & \text{αν } \bar{V}_0(t) = 0 \end{cases} \quad (2.6.2.6)$$

Βάσει της θεωρίας ευστάθειας του Lyapunov, το  $\bar{V}_0(t) = V_0(t) - A^D$  συγκλίνει καθολικά στον μηδενικό πίνακα, από μια αρχική τιμή. Λαμβάνοντας υπόψη τη μοναδικότητα του Drazin αντίστροφου, προκύπτει ότι  $V_0(t) \rightarrow A^D$  καθώς  $\lambda \rightarrow 0$  δηλαδή η κατάσταση  $V_0(t)$  συγκλίνει καθολικά στο ακριβές αντίστροφο Drazin  $A^D$ . Η καθολική σύγκλιση του μοντέλου RNN (2.6.21) επαληθεύεται με αυτόν τον τρόπο.

### Παράδειγμα 2.6.2[63]

Θεωρούμε τον παρακάτω πίνακα παραμέτρων  $A(a)$  από το [ 67 ]:

$$A(a) = \begin{bmatrix} a^2 + a & 1 & a^2 + a \\ 1 & a^2 - a & 1 \\ a^2 + a & 1 & a^2 + a \end{bmatrix}$$



Σχήμα 2.6.2(α): Συμπεριφορά σύγκλισης των μεταβλητών κατάστασης του μοντέλου Πα στο Παράδειγμα 2.6.2 1ο σχήμα, 2ο σχήμα και 3ο σχήμα [63]

Ο ακριβής Drazin αντίστροφος  $A(a)^D$  του  $A(a)$  είναι ίσος με:

$$A(a)^D = \begin{bmatrix} \frac{a(a-1)}{4(a^4-a^2-1)} & -\frac{1}{2(a^4-a^2-1)} & -\frac{a(a-1)}{4(a^4-a^2-1)} \\ -\frac{1}{2(a^4-a^2-1)} & \frac{a(a+1)}{(a^4-a^2-1)} & -\frac{1}{2(a^4-a^2-1)} \\ \frac{a(a-1)}{4(a^4-a^2-1)} & -\frac{1}{2(a^4-a^2-1)} & \frac{a(a-1)}{4(a^4-a^2-1)} \end{bmatrix}$$

Στην συνέχεια γίνεται μια δοκιμή με  $a = 1$ . Το ακριβές αντίστροφο Drazin του  $A(1)$  θα είναι:

$$A(1)^D = \begin{bmatrix} 0 & 0.5 & 0 \\ 0.5 & -2 & 0.5 \\ 0 & 0.5 & 0 \end{bmatrix}$$

Χρησιμοποιούμε τις Εξισώσεις

$$\frac{d(V_0)_j(t)}{dt} = -\beta \sum_{k=1}^n h_{jk} F(\sum_{k=1}^n m_{kj} (V_0)_k(t) - (G_0)_j), G_0 = A^l (A^{2l+1})^T A^l \text{ και}$$

$$\frac{d(V_0)_j(t)}{dt} = -\beta F(\sum_{k=1}^n m_{kj} (V_0)_k(t) - (G_0)_j), G_0 = A^l (A^{2l+1})^T A^l$$

για να δοκιμάσουμε το μοντέλο. Επειδή ο  $A(a)$  είναι συμμετρικός και με  $k = 1$  μια επιλογή και στις δύο παραπάνω εξισώσεις και μπορεί να είναι η  $l = 1$ . Επίσης επιλέγουμε  $10^8$  για τα  $v_1, v_2$  και  $v_3$ . Η τροχιά της συμπεριφοράς σύγκλισης σε  $10^{-6}$  s κάτω από μηδενικές αρχικές συνθήκες φαίνονται στο Σχήμα 2.6.2.(α). Τα απόλυτα υπόλοιπα στις εκτιμώμενες τιμές που αντιστοιχούν και στις πραγματικές τιμές των στηλών

$v_j, j = 1, 2, 3$  φαίνονται στο Σχήμα 2.6.2.(β). Οι κόκκινες γραμμές στα σχήματα 2.6.2.(α) και 2.6.2.(β) υποδηλώνουν τον ακριβή αντίστροφο Drazin, οι μπλε τελείες δείχνουν την  $f_{in}(x)$ , οι ροζ σταυροί αντιστοιχούν στη συνάρτηση ενεργοποίησης δυναμική - σιγμοειδής  $f_{ps}(x, 3, 7)$ , τα μαύρα αστέρια υποδηλώνουν τη συνάρτηση διπολική - σιγμοειδής  $f_{bs}(x, 3)$  και τα πράσινα τρίγωνα παράγονται χρησιμοποιώντας τη ομαλή δυναμο-σιγμοειδής συνάρτηση  $f_{sps}(x, 3, 11)$ .

### Θεώρημα 2.6.4[64]

Έστω  $A \in R^{n \times n}$  ένας δεδομένος πίνακας. Τότε το μοντέλο (2.6.4.1) στη γραμμική περίπτωση συγκλίνει καθολικά ασυμπτωτικά στο αντίστροφο Drazin  $A^D$ , δηλαδή:

$$\bar{V} = A^D \quad (2.6.4.3)$$

ανεξάρτητα από την επιλογή του  $V(0)$ .

### Απόδειξη

Σύμφωνα με τη θεωρία γραμμικών συστημάτων [64], η λύση κλειστής μορφής του πίνακα κατάστασης του μοντέλου (2.6.4.1) στη γραμμική περίπτωση ισούται με:

$$V(t) = \exp(-\beta(A + \lambda I)^{l+1}t)V(0) + \beta \exp(-\beta(A + \lambda I)^{l+1}t) \int_0^t \exp(\beta(A + \lambda I)^{l+1}\tau) A^l d\tau \quad (2.6.4.4)$$

Μια εφαρμογή του [63] συνεπάγεται

$$\lim_{t \rightarrow \infty} \exp(-\beta(A + \lambda I)^{l+1}t) = 0 \quad (2.6.4.5)$$

Τώρα, οι Εξισώσεις (2.6.4.4) και (2.6.4.5) συνεπάγονται την ακόλουθη αναπαράσταση για το  $\lim_{t \rightarrow \infty} V(t) = \bar{V}$

$$\bar{V} = \lim_{t \rightarrow \infty} \beta \exp(-\beta(A + \lambda I)^{l+1}t) \int_0^t \exp(\beta(A + \lambda I)^{l+1}\tau) A^l d\tau \quad (2.6.4.6)$$

Ξεκινώντας από την (2.6.4.6), χρησιμοποιώντας

$$\lim_{\lambda \rightarrow 0} (A + \lambda I)^{l+1} A^D = A^l \quad (2.6.4.7)$$

μετά από διάφορους μετασχηματισμούς της άλγεβρας πινάκων, μπορεί κανείς να επαληθεύσει ότι

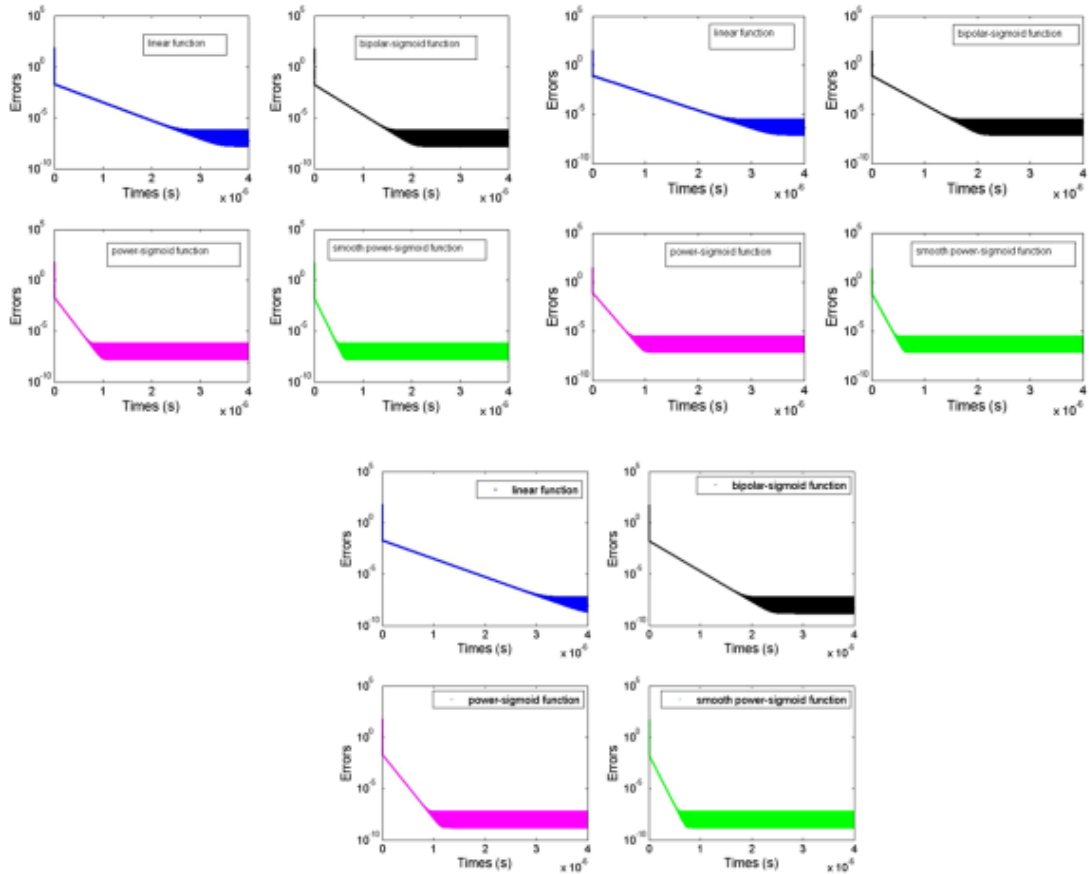
$$\begin{aligned} V &= [\lim_{t \rightarrow \infty} \exp(-\beta(A + \lambda I)^{l+1}t) \int_0^t \exp(\beta(A + \lambda I)^{l+1}\tau) (\beta(A + \lambda I)^{l+1}) d\tau] A^D \\ &= [\lim_{t \rightarrow \infty} \exp(-\beta(A + \lambda I)^{l+1}t) \int_0^t d(\exp(\beta(A + \lambda I)^{l+1}\tau))] A^D \\ &= [\lim_{t \rightarrow \infty} \exp(-\beta(A + \lambda I)^{l+1}t) \exp(\beta(A + \lambda I)^{l+1}\tau) \Big|_{\tau=0}^{\tau=t}] A^D \\ &= \{\lim_{t \rightarrow \infty} \exp(-\beta(A + \lambda I)^{l+1}t) [\exp(\beta(A + \lambda I)^{l+1}t) - I]\} A^D \\ &= [I - \lim_{t \rightarrow \infty} \exp(-\beta(A + \lambda I)^{l+1}t)] A^D \quad (2.6.4.8) \end{aligned}$$

### Παράδειγμα 2.6.3

Θεωρήστε τους συμμετρικούς πίνακες  $n \times n$  πίνακες  $S_n$  βαθμού  $n-1$  [ 68 ].

$$\begin{bmatrix} \alpha + 1 & \alpha & \dots & \dots & \alpha & \alpha + 1 \\ \alpha & \alpha - 1 & \alpha & \dots & \alpha & \alpha \\ \alpha & \alpha & \alpha + 1 & \ddots & \alpha & \alpha \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \alpha & \alpha & \ddots & \ddots & \alpha - 1 & \alpha \\ \alpha + 1 & \alpha & \dots & \dots & \alpha & \alpha + 1 \end{bmatrix}$$

Οι παρακάτω πίνακες είναι δοκιμαστικοί και ορίζονται για περιττές τάξεις  $n = 3, 5, 7, \dots$ . Ο Drazin αντίστροφος του  $S_n$  συμπίπτει με το αντίστροφο του Moore-Penrose και θα ισούται με:



Σχήμα 2.6.2(β): Τροχιές των υπολειπόμενων σφαλμάτων του μοντέλου Ib στο Παράδειγμα 2.6.2 για τις διάφορες συναρτήσεις ενεργοποίησης.

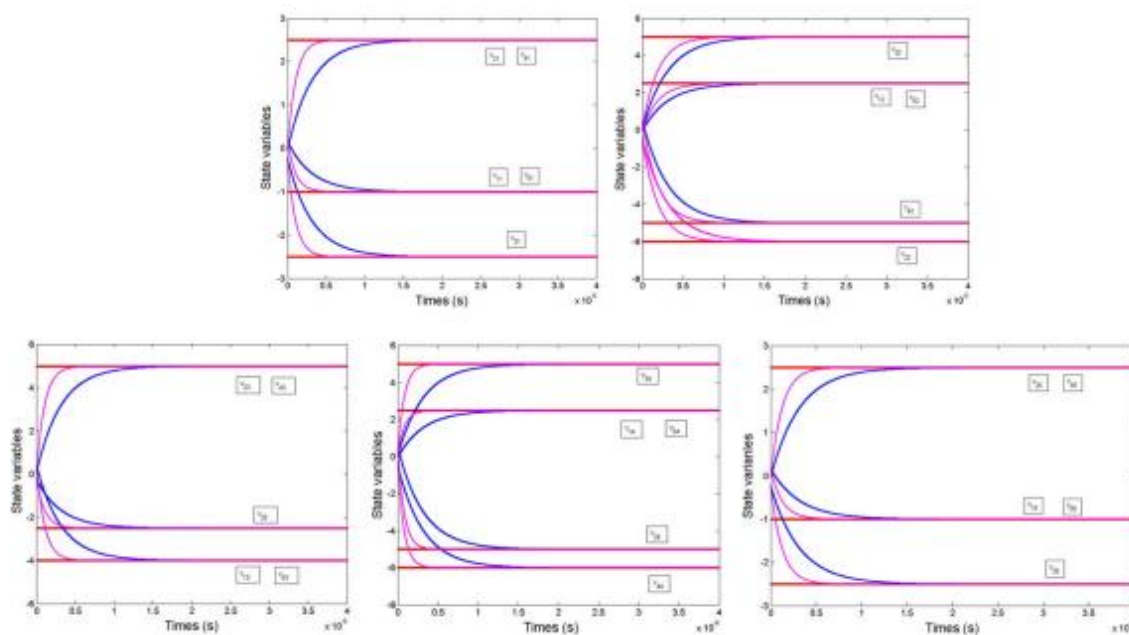
$$\begin{bmatrix} -\alpha + 1 & 2\alpha & -2\alpha & \dots & 2\alpha & -\alpha + 1 \\ 2\alpha & -4\alpha - 4 & 4\alpha & \ddots & -4\alpha & 2\alpha \\ -2\alpha & 4\alpha & -4\alpha + 4 & \ddots & 4\alpha & -2\alpha \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 2\alpha & -4\alpha & 4\alpha & \ddots & -4\alpha - 4 & 2\alpha \\ -\alpha + 1 & 2\alpha & -2\alpha & \dots & 2\alpha & -\alpha + 1 \end{bmatrix}$$

Έστω  $n = 5$  και το αρχικό διάνυσμα  $v(0) = (0, 0, 0, 0, 0)^T$ . Επειδή ο βαθμός του  $S_5(t)$  είναι  $n-1 = 4$ , λαμβάνοντας υπόψη ότι ο  $S_5(t)$  είναι συμμετρικός, συμπεραίνουμε ότι  $(S_5(t)) = 1$ , και στη συνέχεια επιλέγουμε  $l = 1$ . Οι τροχιές των μεταβλητών κατάστασης του μοντέλου Ib που αντιστοιχούν σε  $a = 5$ ,  $\beta = 10^9$  και  $\lambda = 10^{-6}$  και παρουσιάζονται στο Σχήμα 2.6.3, όπου οι κόκκινες καμπύλες υποδηλώνουν το θεωρητικό αντίστροφο Drazin, ενώ οι μπλε καμπύλες και οι ροζ καμπύλες υποδηλώνουν τη λύση που υπολογίζεται με την  $f_{in}\{x, 3, 11\}$  και την  $f_{ps}\{x, 3, 6\}$ .

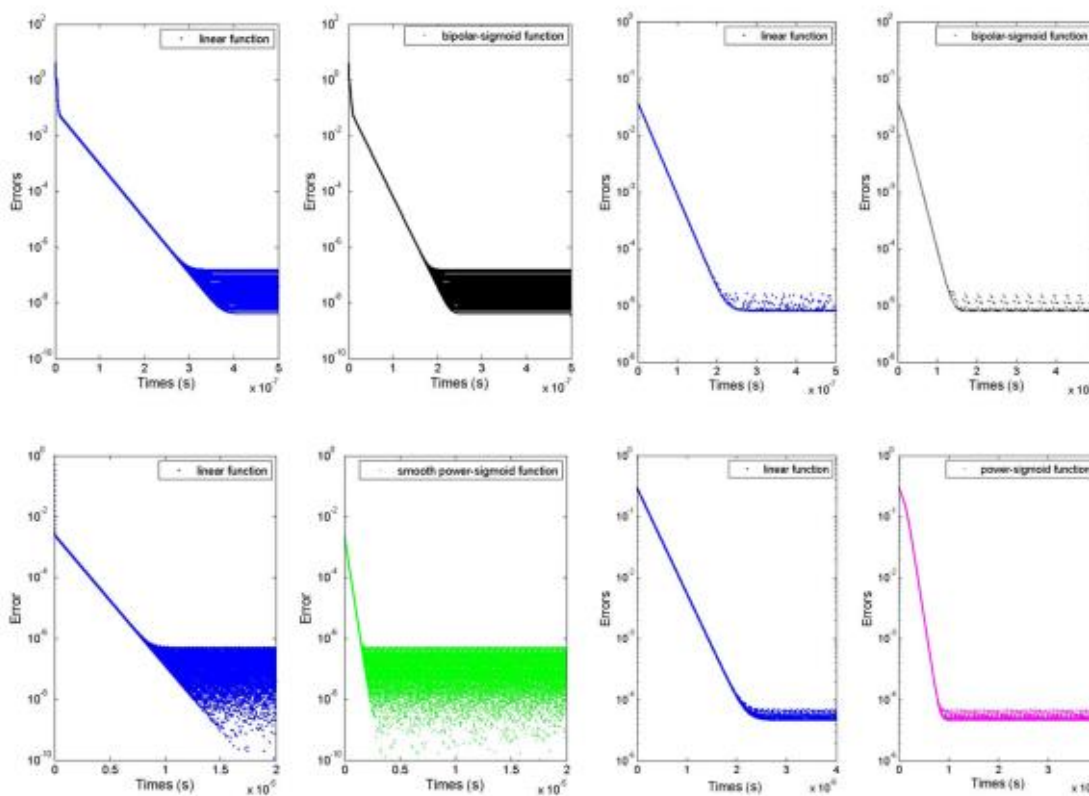
Επιπλέον, οι τάξεις  $n = 10, 51, 101, 501$ . Οι τροχιές των υπολειπόμενων σφαλμάτων

$$\|(A + \lambda I)^{\lambda+1} V(t) - Q(t)\|_F$$

οι οποίες προκύπτουν εφαρμόζοντας διαφορετικές συναρτήσεις ενεργοποίησης,  $\lambda = 10^{-6}$  και  $\beta = 10^8$  και παρουσιάζονται στο Σχήμα 2.6.4. Η διάσταση  $n = 10$  σε συνδυασμό με το  $a = 5$ , η γραμμική συνάρτηση ενεργοποίησης.



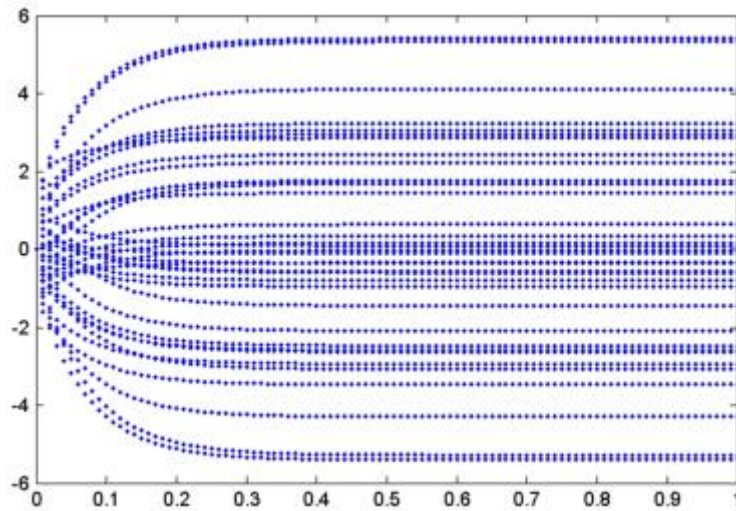
σχήμα 2.6.3: Συμπεριφορά σύγκλισης των μεταβλητών κατάστασης στο Παράδειγμα 2.6.4 [64].



Σχήμα 2.6.4: Τροχιές των υπολειπόμενων σφαλμάτων του μοντέλου I b στο Παράδειγμα 2.6.4. **a**  $n = 10$ ,  $a = 5$ , **b**  $n = 51$ ,  $a = 1$ , **c**  $n = 101$ ,  $a = 0.1$ , και **d**  $n = 501$ ,  $a = 0.2$

$F_{lin}(x)$  και η ομαλή δυναμο-σιγμοειδής συνάρτηση  $f(x, 3, 6)$  εξετάζονται στο Σχήμα 2.6.4(6a).

Οι επιλογές στο Σχήμα 2.6.4(6b) είναι  $n = 51$ ,  $a = 1$ , η  $f_{lin}(x)$  και η  $f_{bs}(x, 5)$ . Στο Σχήμα 2.6.4(6c), λαμβάνουμε  $n = 101$ ,  $a = 0.1$  και οι συναρτήσεις ενεργοποίησης είναι η  $f_{lin}(x)$  και η  $f_{sps}(x, 3, 6)$ . Το Σχήμα 2.6.4(6d) παράγεται



Σχήμα 2.6.5 Συμπεριφορά σύγκλισης των μεταβλητών κατάστασης του μοντέλου Ib στο Παράδειγμα 2.6.4.

χρησιμοποιώντας  $a = 0.2$  τη διάσταση  $n = 501$  και τις συναρτήσεις ενεργοποίησης  $f_{lin}(x)$  και  $f_{ps}(x, 3, 7)$ . Στο Σχήμα 2.6.4, οι μπλε καμπύλες υποδεικνύουν την  $f_{lin}(x)$ , οι ροζ καμπύλες υποδηλώνουν την  $f_{ps}(x, 3, 7)$ , οι μαύρες καμπύλες υποδηλώνουν την  $f_{bps}(x, 3)$  και οι πράσινες καμπύλες αναπαριστούν τιμές που προκύπτουν επιλέγοντας την  $f_{ps}(x, 3, 11)$ , αντίστοιχα.

#### Παράδειγμα 2.6.4

Ο παρακάτω μη ιδιάζων πίνακας:

$$A = \begin{bmatrix} 0.4447 & 0.4057 & 0.3529 & 0.6038 & 0.9318 & 0.6721 \\ 0.6154 & 0.9355 & 0.8132 & 0.2722 & 0.4660 & 0.8381 \\ 0.7919 & 0.9169 & 0.0099 & 0.1988 & 0.4186 & 0.0196 \\ 0.9218 & 0.4103 & 0.1389 & 0.0153 & 0.8462 & 0.6813 \\ 0.7382 & 0.8936 & 0.2028 & 0.7468 & 0.5252 & 0.3795 \\ 0.1763 & 0.0579 & 0.1987 & 0.4451 & 0.2026 & 0.8318 \end{bmatrix}$$

από [21]. Επομένως,  $\text{Ind}(A) = 0$  και  $A^{-1} = A^D$ . Σύμφωνα με τα αποτελέσματα από το [21], η επιλογή  $l = 1$  είναι κατάλληλη για τον υπολογισμό του  $A^{-1}$ . Έστω ότι το αρχικό διάνυσμα είναι το μηδενικό διάνυσμα  $v(0) = (0, 0, 0, 0, 0, 0)^T$ . Το μοντέλο νευρωνικού δικτύου Ib με τη γραμμική συνάρτηση ενεργοποίησης  $f_{lin}(x)$  δίνει την ακόλουθη προσέγγιση του  $A^D$  μετά από  $10^{-7}$  s, για  $\beta = 10^{10}$  και  $\lambda = 10^{-8}$ :

$$A^D = \begin{bmatrix} -3.0746 & 0.1593 & -5.2799 & 3.2284 & 5.3347 & -2.6300 \\ 1.7626 & 0.0864 & 5.4113 & -2.5822 & -4.2963 & 2.4363 \\ -1.4431 & 1.4461 & -5.4049 & 1.7459 & 4.1102 & -3.4690 \\ -0.5694 & -0.3461 & -2.4723 & 0.3365 & 3.0550 & -0.8025 \\ 2.2305 & -0.3552 & 1.6884 & -0.6080 & -2.0921 & -0.0317 \\ 0.6351 & -0.1135 & 2.9452 & -0.9536 & -2.9387 & 2.8558 \end{bmatrix}$$

## Κεφάλαιο 3<sup>ο</sup> :

### Επίλογος και προτάσεις για περαιτέρω έρευνα

Η παρούσα διπλωματική εργασία ανέδειξε τη δυναμική σύνδεση μεταξύ της θεωρίας πινάκων και των σύγχρονων υπολογιστικών μεθόδων που βασίζονται στα επαναλαμβανόμενα νευρωνικά δίκτυα. Μέσα από τη μελέτη και ανάλυση μοντέλων όπως τα Gradient Neural Networks, κατέστη σαφές ότι η επίλυση προβλημάτων που σχετίζονται με τον υπολογισμό γενικευμένων αντιστρόφων μπορεί να υλοποιηθεί αποδοτικά μέσω δυναμικών συστημάτων.

Η εστίαση σε σταθερούς πίνακες επέτρεψε την εις βάθος κατανόηση της συμπεριφοράς των προτεινόμενων μοντέλων, αναδεικνύοντας τόσο τα πλεονεκτήματα της ταχείας σύγκλισης όσο και τη σημασία της επιλογής κατάλληλων συναρτήσεων ενεργοποίησης και παραμέτρων. Ιδιαίτερα σημαντική είναι η παρατήρηση ότι τα νευρωνικά δίκτυα δεν λειτουργούν απλώς ως αριθμητικά εργαλεία, αλλά ως ένα εναλλακτικό πλαίσιο επίλυσης προβλημάτων, όπου η έννοια της εξέλιξης στο χρόνο αντικαθιστά τις κλασικές επαναληπτικές διαδικασίες.

Παράλληλα, η ανάλυση του Drazin και του Moore–Penrose γενικευμένου αντιστρόφου μέσα από το πρίσμα των νευρωνικών μοντέλων αναδεικνύει νέες προοπτικές σε εφαρμογές μεγάλης κλίμακας, όπου η παραλληλοποίηση και η υπολογιστική αποδοτικότητα αποτελούν κρίσιμους παράγοντες. Η δυνατότητα υλοποίησης τέτοιων μοντέλων σε πραγματικό χρόνο ενισχύει περαιτέρω τη σημασία τους σε σύγχρονες τεχνολογικές εφαρμογές.

Ωστόσο, η εργασία αυτή δεν εξαντλεί το πεδίο μελέτης. Αντιθέτως, ανοίγει τον δρόμο για περαιτέρω διερεύνηση, όπως η βελτίωση της ευστάθειας των μοντέλων, η μελέτη της συμπεριφοράς τους υπό αριθμητικό θόρυβο, καθώς και η επέκτασή τους σε πιο σύνθετες κατηγορίες προβλημάτων. Επιπλέον, η διερεύνηση εναλλακτικών συναρτήσεων ενεργοποίησης και υβριδικών προσεγγίσεων θα μπορούσε να οδηγήσει σε ακόμα πιο αποδοτικά και ευέλικτα μοντέλα.

Συνολικά, η παρούσα εργασία επιβεβαιώνει ότι η σύγκλιση της γραμμικής άλγεβρας με τη θεωρία νευρωνικών δικτύων δεν αποτελεί απλώς μια σύγχρονη τάση, αλλά μια ουσιαστική κατεύθυνση εξέλιξης των μαθηματικών μεθόδων επίλυσης προβλημάτων, με σημαντικές προοπτικές για το μέλλον της επιστήμης και της τεχνολογίας.

Κατά την εκπόνηση της συγκεκριμένης διπλωματικής εργασίας δημιουργήθηκαν οι παρακάτω προβληματισμοί για **μελλοντική έρευνα**:

- Η εφαρμογή των διαφόρων Gradient Neural Networks για την επίλυση των εξισώσεων Riccati ή Yang-Baxter με σταθερούς συντελεστές.
- Εφαρμογή των GNN μοντέλων για τη μελέτη και την επίλυση τανυστικών εξισώσεων με το Einstein γινόμενο ή το T-product.
- Ανάπτυξη και ανάλυση νέων αρχιτεκτονικών RNN που στοχεύουν στη βελτίωση της ταχύτητας σύγκλισης και της ακρίβειας των λύσεων, υπερτερώντας των υφιστάμενων προσεγγίσεων.
- Η επέκταση της εφαρμογής των υφιστάμενων μοντέλων στον υπολογισμό λύσεων και στον προσδιορισμό ευρύτερων κατηγοριών γενικευμένων αντιστρόφων, όπως ο σταθμισμένος (με βάρη) Drazin αντίστροφος.

- Εφαρμογή και άλλων προσεγγιστικών μεθόδων επίλυσης διαφορικών εξισώσεων εκτός της ode15s στα GNN μοντέλα.
- Η αξιοποίηση των υφιστάμενων μοντέλων GNN για την επίλυση προβλημάτων που περιλαμβάνουν τον υπολογισμό γενικευμένων αντιστρόφων και τη λύση συστημάτων, με εφαρμογές σε τομείς όπως η ρομποτική, η στατιστική και συναφή επιστημονικά πεδία.

## ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] Adi Ben-Israel, Thomas N.E. Greville, "Generalized Inverses Theory and Applications, Second Edition", pp. 14-67, 2002.
- [2] Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. "Deep Learning." (MIT Press, 2016)
- [3] Rumelhart, David E., Georey E. Hinton, and Ronald J. Williams. "Learning representations by backpropagating errors." *Nature* 323.6088 (1986): 533-536.
- [4] Graves, Alex. "Supervised Sequence Labelling with Recurrent Neural Networks." (Springer, 2012)
- [5] Vapnik, Vladimir. "The Nature of Statistical Learning Theory." (Springer, 1995)
- [6] Σ. Δ.Βολογιαννίδης, «Αλγεβρο-Πολυωνυμικές υπολογιστικές μέθοδοι στη θεωρία ελέγχου» Διδακτορική διατριβή, Θεσσαλονίκη, 2005.
- [7] L. Jin, Y. Zhang, and S. Li, Integration-enhanced Zhang neural network for real-time-varying matrix inversion in the presence of various kinds of noises, *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 12, pp. 26152627, Dec. 2016.
- [8] D. Guo, Z. Nie, and L. Yan, The application of noise-tolerant ZD design formula to robots' kinematic control via time-varying nonlinear equations solving, *IEEE Trans.Syst. Man Cybern. Syst.*, vol. 48, no. 12, pp. 21882197, Dec. 2018.
- [9] X. Hu, C. Sun, and B. Zhang, "Design of recurrent neural networks for solving constrained least absolute deviation problems," *IEEE Trans.Neural Netw.*, vol. 21, no. 7, pp. 1073–1086, Jul. 2010

- [10] Q. Liu and J. Wang, "Finite-time convergent recurrent neural network with a hard-limiting activation function for constrained optimization with piecewise-linear objective functions," *IEEE Trans. Neural Netw.*, vol. 22, no. 4, pp. 601–613, Apr. 2011.
- [11] A. Cichocki and R. Unbehauen, "Neural networks for solving systems of linear equations and related problems," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 39, no. 2, pp. 124–138, Feb. 1992.
- [12] J. Wang and H. Li, "Solving simultaneous linear equations using recurrent neural networks," *Inf. Sci.*, vol. 76, nos. 3–4, pp. 255–277, 1994.
- [13] J. Wang, "Electronic realisation of recurrent neural network for solving simultaneous linear equations," *Electron. Lett.*, vol. 28, no. 5, pp. 493–495, Feb. 1992.
- [14] J. Wang, "Recurrent neural networks for solving linear matrix equations," *Comput. Math. Appl.*, vol. 26, no. 9, pp. 23–34, 1993.
- [15] L.-X. Wang and J. M. Mendel, "Three-dimensional structured networks for matrix equation solving," *IEEE Trans. Comput.*, vol. 40, no. 12, pp. 1337–1346, Dec. 1991.
- [16] J.-S. Jang, S.-Y. Lee, and S.-Y. Shin, "An optimization network for matrix inversion," in *Neural Information Processing Systems*. College Park, MD, USA: AIP, 1988, pp. 397–401.
- [17] J. Wang, "A recurrent neural network for real-time matrix inversion", *Appl. Math. Comput.*, vol. 55, no. 1, pp. 89–100, 1993.
- [18] B. Liao and Y. Zhang, "Different complex ZFs leading to different complex ZNN models for time-varying complex generalized inverse matrices," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 9, pp. 1621–1631, Sep. 2014
- [19] Y. Zhang, Y. Yang, N. Tan, and B. Cai, "Zhang neural network solving for time-varying full-rank matrix Moore–Penrose inverse," *Computing*, vol. 92, no. 2, pp. 97–121, 2011.
- [20] J. Zhong, X. Liu, G. Zhou, and Y. Yu, "A new iterative method for computing the Drazin inverse," *Filomat*, vol. 26, no. 3, pp. 597–606, 2012.

- [21] A. Cichocki, T. Kaczorek, and A. Stajniak, "Computation of the Drazin inverse of a singular matrix making use of neural networks," *Bull. Polish Acad. Sci., Tech. Sci.*, vol. 40, no. 4, pp. 387–394, 1992.
- [24] Ji J (1994) An alternative limit expression of Drazin inverse and its application. *Appl Math Comput*, 61:151–156
- [25] Chen YL (1997) Representation and approximation for the Drazin inverse AD. *Appl Math Comput*, 48:83–92
- [26] Stanimirovic PS, Živkovic IS, Wei Y (2015) Recurrent neural network for computing the Drazin inverse. *IEEE Trans Neural Netw Learn Syst* 26:2830–2843
- [27] Stanimirovic PS, Živkovic IS, Wei Y (2015) Recurrent neural network approach based on the integral representation of the Drazin inverse. *Neural Comput* 27:2107–2131
- [28]. Zhang Y, Danchi J, Jun W (2002) A recurrent neural network for solving Sylvester equation with timevarying coefficients. *IEEE Trans Neural Netw* 13:1053–1063
- [29]. Li S, Chen S, Liu B (2013) Accelerating a recurrent neural network to finite-time convergence for solving time-varying Sylvester Equation by using a sign-bi-power activation function. *Neural Process Lett* 37:189–205
- [30]. Li Z, Zhang Y (2010) Improved Zhang neural network model and its solution of time-varying generalized linear matrix equations. *Expert Syst Appl* 37:7213–7218
- [31]. Zhang Y, Ge SS (2003) A general recurrent neural network model for time-varying matrix inversion. In: *Proceedings of 42nd IEEE conference on decision and control, San Diego*, vol 6, pp 6169–6174
- [32]. Zhang Y, Shi Y, Chen K, Wang C (2009) Global exponential convergence and stability of gradient-based neural network for online matrix inversion. *Appl Math Comput* 215:1301–1306
- [33]. Zhang Y, Ma W, Cai B (2009) From Zhang neural network to Newton iteration for matrix inversion. *IEEE Trans Circuits Syst I* 56(7):1405–1415
- [34]. Sheng X, Wang T (2013) An iterative method to compute Moore–Penrose inverse based on gradient maximal convergence rate. *Filomat* 27(7):1269–1276

- [35]. Higham NJ (2002) Accuracy and stability of numerical algorithms, 2nd edn. SIAM, Philadelphia
- [36]. Soleymani F, Stanimirović PS (2013) A higher order iterative method for computing the Drazin inverse. *Sci World J* 2013, Article ID 708647. doi:10.1155/2013/708647
- [37]. Wang J (1993) A recurrent neural network for real-time matrix inversion. *Appl Math Comput* 55:89–100
- [38]. Wang J (1997) Recurrent neural networks for computing pseudoinverses of rank-deficient matrices. *SIAM J Sci Comput* 18:1479–1493
- [39]. Stanimirović PS, Živković IS, Wei Y (2015) Recurrent neural network for computing the Drazin inverse. *IEEE Trans Neural Netw Learn Syst* 26:2830–2843
- [40]. Stanimirović PS, Živković I, Wei Y (2015) Recurrent neural network approach based on the integral representation of the Drazin inverse. *Neural Comput* 27(10):2107–2131
- [41]. Zhang Y (2005) Design and analysis of a general recurrent neural network model for time-varying matrix inversion. *IEEE Trans Neural Netw* 16(6):1477–1490
- [42]. Zhang Y, Shi Y, Chen K, Wang C (2009) Global exponential convergence and stability of gradient-based neural network for online matrix inversion. *Appl Math Comput* 215:1301–1306
- [43] Campbell, S. L., Meyer, C. D., Jr., & Rose, N. J. (1976). Applications of the Drazin inverse to linear systems of differential equations with singular constant coefficients. *SIAM Journal on Applied Mathematics*, 31(3), 411–425.
- [44] Cichocki, A., Kaczorek, T., & Stajniak, A. (1992). Computation of the Drazin inverse of a singular matrix making use of neural networks. *Bulletin of the Polish Academy of Sciences: Technical Sciences*, 40(4), 387–394.
- [45] D. Guo, C. Yi, Y. Zhang, Zhang neural network versus gradient-based neural network for Time varying linear matrix equation solving, *Neurocomputing* 74 (2011), 3708–3712.

- [46] P.S. Stanimirović, M.D. Petković, D. Gerontitis, Gradient neural network with nonlinear activation for computing inner inverses and the Drazin inverse, *Neural Processing Letters*, DOI:10.1007/s11063-017-9705-4
- [47] Y. Zhang, Design and analysis of a general recurrent neural network model for time-varying matrix inversion, *IEEE Trans. Neural Netw.* 16(6) (2005), 1477–1490.
- [48] Ben-Israel A, Greville TNE (2003) *Generalized inverses: theory and applications*, 2nd edn. Springer, New York
- [49] Wang G, Wei Y, Qiao S (2003) *Generalized inverses: theory and computations*. Science Press, New York
- [50] Wang X-Z, Ma H, Stanimirović PS (2017) Nonlinearly activated recurrent neural network for computing the Drazin inverse. *Neural Process Lett.* doi:10.1007/s11063-017-9581-y
- [51] J. Zhong, X. Liu, G. Zhou, and Y. Yu, "A new iterative method for computing the Drazin inverse," *Filomat*, vol. 26, no. 3, pp. 597–606, 2012
- [52] P. S. Stanimirović, I. Živković, and Y. Wei, "Recurrent Neural Network Approach Based on the Integral Representation of the Drazin Inverse," *Neural Computation*, vol. 27, pp. 1-25, 2015
- [53] J. Zhong, X. Liu, G. Zhou, and Y. Yu, "A new iterative method for computing the Drazin inverse," *Filomat*, vol. 26, no. 3, pp. 597–606, 2012.
- [54] Y. Wei and H. Wu, "The representation and approximation for Drazin inverse," *Journal of Computational and Applied Mathematics*, vol. 126, no. 1, pp. 417-432, 2000.
- [55] P. S. Stanimirović, I. Živković, and Y. Wei, "Recurrent Neural Network Approach Based on the Integral Representation of the Drazin Inverse," *Neural Computation*, vol. 27, pp. 1-25, 2015
- [57] P. S. Stanimirović and M. D. Petković, "Gradient neural dynamics for solving matrix equations and their applications," *Neurocomputing*, 2018, doi: 10.1016/j.neucom.2018.03.058.
- [60] P. S. Stanimirović, M. D. Petković, and D. Gerontitis, "Gradient Neural Network with Nonlinear Activation for Computing Inner Inverses and the Drazin Inverse," *Neural Process. Lett.*, 2017, doi: 10.1007/s11063-017-9705-4.

- [61] Y. Wei and H. Wu, “The representation and approximation for Drazin inverse,” *J. Comput. Appl. Math.*, vol. 126, nos. 1–2, pp. 417–432, 2000.
- [62] Zhang Y, Shi Y, Chen K, Wang C (2009) Global exponential convergence and stability of gradient-based neural network for online matrix inversion. *Appl Math Comput* 215:1301–1306
- [63] Stanimirovic PS, Zivkovic IS, Wei Y (2015) Recurrent neural network for computing the Drazin inverse. *IEEE Trans Neural Netw Learn Syst* 26:2830–2843
- [64] Kailath T (1980) *Linear systems*. Prentice-Hall, Englewood Cliffs
- [65] Cichocki A, Kaczorek T, Stajniak A (1992) Computation of the Drazin inverse of a singular matrix making use of neural networks. *Bull Pol Acad Sci Tech Sci* 40:387–394
- [66] Stanimirovic PS, Zivkovic IS, Wei Y (2015) Recurrent neural network approach based on the integral representation of the Drazin inverse. *Neural Comput* 27:2107–2131
- [67] Ji J (2002) A finite algorithm for the Drazin inverse of a polynomial matrix. *Appl Math Comput* 130:243–251
- [68] Zielke G (1986) Report on test matrices for generalized inverses. *Computing* 36:105–162
- [69] Campbell SL, Meyer CD Jr (1991) *Generalized inverses of linear transformations*. Dover Publications Inc, New York Corrected reprint of the 1979 original
- [70] Castro González N, Koliha JJ, Wei Y (2000) Perturbation of the Drazin inverse for matrices with equal eigenprojections at zero. *Linear Algebra Appl* 312:181–189
- [71] González NC, Koliha J, Wei Y (2002) Error bounds for perturbation of the Drazin inverse of closed operators with equal spectral projections. *Appl Anal* 81:915–928
- [72] Wei Y (2002) The Drazin inverse of a modified matrix. *Appl Math Comput* 125:295–301
- [73] Wei Y, Wu H (2001) Challenging problems on the perturbation of Drazin inverse. *Ann Oper Res* 103:371–378
- [74] Fu M, Wang S, Cong Y (2013) Swarm stability analysis of high-order linear time-invariant singular multiagent systems. *Math Probl Eng*. Article ID 469747, p 11
- [75] Campbell SL, Meyer CD Jr (1991) *Generalized inverses of linear transformations*. Dover Publications Inc, New York Corrected reprint of the 1979 original

- [76] Hopfield JJ (1984) Neurons with graded response have collective computational properties like those of two-state neurons. *Proc Natl Acad Sci* 81:3088–3092
- [77] Lippmann RP (1987) An introduction to computing with neural nets. *IEEE ASSP Mag* 4:4–22
- [78] Jang J, Lee S, Shin S (1988) An optimization network for matrix inversion. In: Anderson D (ed) *Neural information processing systems*. American Institute of Physics, New York, pp 397–401
- [79] Luo FL, Bao Z (1992) Neural network approach to computing matrix inversion. *Appl Math Comput* 47:109–120
- [80] Wang J (1993) A recurrent neural network for real-time matrix inversion. *Appl Math Comput* 55:89–100
- [81] Wang J (1993) Recurrent neural networks for solving linear matrix equations. *Comput Math Appl* 26:23–34
- [82] Zhang Y, Yang Y, Tan N, Cai B (2011) Zhang neural network solving for time-varying full-rank matrix Moore–Penrose inverse. *Computing* 92:97–121
- [83] Liao B, Zhang Y (2014) Different complex ZFs leading to different complex ZNN models for time-varying complex generalized inverse matrices. *IEEE Trans Neural Netw Learn Syst* 25:1621–1631
- [84] <https://dione.lib.unipi.gr/xmlui/bitstream/handle/unipi/4822/Vidali.pdf?sequence=2&isAllowed=y>
- [85] The representation and approximation for Drazin inverse. *Journal of Computational and Applied Mathematics*, 126(1), 417–432