



ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ

ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ

"ΘΕΩΡΗΤΙΚΗ ΠΛΗΡΟΦΟΡΙΚΗ ΚΑΙ ΘΕΩΡΙΑ ΣΥΣΤΗΜΑΤΩΝ ΚΑΙ ΕΛΕΓΧΟΥ"

Υπολογισμός Αντιστρόφου Πίνακα με τη Χρήση Νευρωνικών Δικτύων

Μεταπτυχιακή διπλωματική εργασία

Δημήτριος Γεροντίτης

Επιβλέπων: Καραμπετάκης Νικόλαος

Καθηγητής Α.Π.Θ



ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ

ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ

"ΘΕΩΡΗΤΙΚΗ ΠΛΗΡΟΦΟΡΙΚΗ ΚΑΙ ΘΕΩΡΙΑ ΣΥΣΤΗΜΑΤΩΝ ΚΑΙ ΕΛΕΓΧΟΥ"

Υπολογισμός Αντιστρόφου Πίνακα με τη Χρήση Νευρωνικών Δικτύων

Μεταπτυχιακή διπλωματική εργασία

Δημήτριος Γεροντίτης

Επιβλέπων: Καραμπετάκης Νικόλαος

Καθηγητής Α.Π.Θ

Εγκρίθηκε από τη τριμελή εξεταστική επιτροπή τον Ιούνιο του 2016

.....

.....
Ν.Καραμπετάκης	Μ.Γουσίδου-Κουτίτα	Ε. Αντωνίου
Καθηγητής Α.Π.Θ.	Αν. Καθηγήτρια Α.Π.Θ.	Αν Καθηγητής ΑΤΕΙΘ

.....
Δημήτριος Κ. Γεροντίτης

Πτυχιούχος Μαθηματικός Α.Π.Θ.

Copyright©Δημήτριος Κ. Γεροντίτης, 2016

Απαγορεύεται η αντιγραφή αποθήκευση και διανομή της παρούσας εργασίας εξ' ολοκλήρου ή τμήματος αυτής Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα. Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευτεί ότι εκφράζουν τις επίσημες θέσεις του Α.Π.Θ.

Πρόλογος

Για την ολοκλήρωση της παρούσης εργασίας, που αποτελεί πόνημα πολύμηνης προσπάθειας, θα επιθυμούσα να ευχαριστήσω από καρδιάς τον επιβλέποντα καθηγητή Κ. Νικόλαο Καραμπετάκη, ο οποίος αποτέλεσε πολύτιμος καθοδηγητής καθ' όλη τη διάρκεια συγγραφής της παρούσας εργασίας καθώς επίσης και τα μέλη της τριμελούς επιτροπής για το χρόνο που αφιέρωσαν για τη μελέτη της. Συνεχίζοντας θα ήθελα να ευχαριστήσω τους Prof. Yunong Zhang και Prof. Predrag Stanimirovic, γιατί όποτε χρειάστηκε, μου έδωσαν σημαντικές παρατηρήσεις για τη μελέτη των δημοσιευμένων εργασιών τους.

Δε θα μπορούσα βέβαια να παραλείψω να ευχαριστήσω και τους γονείς μου οι οποίοι αποτελούν στυλοβάτες σε κάθε μου προσπάθεια καθώς και τους φίλους μου για τη πολύτιμη ψυχολογική στήριξη που μου παρείχαν τη δεδομένη χρονική περίοδο.

Περίληψη

Στη παρούσα εργασία εξετάζουμε την εφαρμογή επαναλαμβανόμενων νευρωνικών δικτύων για τη μελέτη χρονικά μεταβαλλόμενων συστημάτων. Στην ουσία κατασκευάζουμε συστήματα διαφορικών εξισώσεων, τα οποία μπορούν να υπολογίζουν τον αντίστροφο ενός χρονικά μεταβαλλόμενου τετραγωνικού πίνακα και να επιλύουν σταθερά γραμμικά συστήματα που έχουνε σημαντική εφαρμογή σε προβλήματα εύρεσης συνάρτησης μεταφοράς ηλεκτρονικών κυκλωμάτων. Θα ήταν σωστό να τονιστεί ότι για τη μελέτη της συγκεκριμένης κατηγορίας διαφορικών εξισώσεων χρησιμοποιούμε όρους και σχήματα που σχετίζονται με τα τεχνητά νευρωνικά δίκτυα, γι' αυτό το λόγο τις περισσότερες φορές αναφέρονται σε νευρωνικά δίκτυα. Στο πρώτο κεφάλαιο αναφέρουμε βασικούς ορισμούς και ιδιότητες που αφορούν τη θεωρητική μελέτη των τεχνητών νευρωνικών δικτύων, στο δεύτερο κεφάλαιο αναφέρουμε βασικές μαθηματικές έννοιες και στο τρίτο κεφάλαιο αναλύουμε τα ZNN (Zhang Neural Network) μοντέλα τα οποία συγκρίνουμε με τα GNN (Gradient Neural Network) μοντέλα. Στη συνέχεια κάνουμε αναφορά στην εργασία του κύριου Stanimirovic [17] που αφορά τη κατασκευή ενός επαναλαμβανόμενου νευρωνικού δικτύου για τον υπολογισμό του Drazin αντιστρόφου σταθερών τετραγωνικών πινάκων. Στο τέταρτο κεφάλαιο αναφερόμαστε στη μέθοδο του Osowski [25], που αφορά την εφαρμογή επαναλαμβανόμενων νευρωνικών δικτύων για τον υπολογισμό της συνάρτησης μεταφοράς ηλεκτρονικών κυκλωμάτων, και στα ZNN μοντέλα για την επίλυση γραμμικών συστημάτων. Στο τέλος του τέταρτου κεφαλαίου κάνουμε συγκριτική μελέτη των αποτελεσμάτων της LU παραγοντοποίησης με τα αντίστοιχα των ZNN μοντέλων (για επίλυση γραμμικών συστημάτων). Όλα τα παραδείγματα έχουν υλοποιηθεί στο περιβάλλον του Matlab, τα οποία μας οδήγησαν σε χρήσιμα συμπεράσματα όσον αφορά, τις προϋποθέσεις για την εφαρμογή των παραπάνω μεθόδων.

Λέξεις Κλειδιά

Χρονικά μεταβαλλόμενα συστήματα, επαναλαμβανόμενο νευρωνικό δίκτυο, ZNN μοντέλα, GNN μοντέλα, Μέθοδος μέγιστης κλίσης, Drazin αντίστροφος, Ευστάθεια Lyapunov

Abstract

In the present paper we examine the application of recurrent neural networks to study time-varying systems. In essence we construct systems of differential equations, which can calculate the inverse of a time-varying square matrix and solve linear constant systems with important applications to finding the transfer function of electronic circuits. Also it would be correct to note that the study of this class of differential equations we use terms and figures which related to artificial neural networks, so for this reason often mentioned as neural networks. In the first chapter we discuss basic definitions and properties related to the theoretical study of artificial neural networks, in the second chapter we discuss basic mathematical definitions and in the third chapter we analyze the ZNN (Zhang Neural Networks) models and we make the comparison with the GNN (Gradient Neural Networks) models. Then we analyze Stanimirovic's [17] paper which is concerning the construction of a recurrent neural network for calculating the Drazin inverse of square constant matrices. In the third chapter we refer to the main Osowski's method [25], on the application of recurrent neural networks for the calculation of electronic circuits transfer function, and ZNN models for solving linear systems. At the end of the fourth chapter we make a comparison study of the LU decomposition with ZNN models (for solving linear systems) results. All the examples have been implemented in the Matlab's environment, which led us to useful findings regarding the application of these methods.

Key Words

Time-Varying systems, recurrent neural network, ZNN models, GNN models, Gradient Descent, Drazin inverse, Lyapunov stability

Περιεχόμενα

Πρόλογος.....	4
Περίληψη.....	5
Abstract.....	6
Κεφάλαιο 1 Εισαγωγικές Έννοιες Τεχνητών Νευρωνικών Δικτύων.....	10
1.1 Εισαγωγή στα Νευρωνικά Δίκτυα.....	10
1.2 Εισαγωγή στους Τεχνητούς Νευρώνες.....	10
1.3 Αρχιτεκτονικές των ΤΝΔ.....	12
1.3.1 ΤΝΔ Πρόσθιας τροφοδότησης.....	13
1.3.2 ΤΝΔ με ανατροφοδότηση.....	13
1.3.3 Perceptrons.....	13
1.4 Πλεονεκτήματα των ΤΝΔ.....	14
1.5 Μέθοδοι εκπαίδευσης.....	15
1.6 Εισαγωγή στο Neural Network Toolbox του Matlab.....	16
1.6.1 Μοντέλο Νευρώνων.....	17
1.6.2 Συναρτήσεις Μεταφοράς.....	18
1.6.3 Νευρώνες με είσοδο διάνυσμα.....	20
1.6.4 Αρχιτεκτονικές δικτύου.....	21
Κεφάλαιο 2 Βασικές Μαθηματικές Έννοιες	25
2.1 Πίνακες.....	25
2.2 Πολυώνυμα.....	29
2.3 Πολυωνυμικοί πίνακες.....	30
2.4 Νόρμες πινάκων και διανυσμάτων.....	31

Κεφάλαιο 3 Εισαγωγή Στα ZNN και GNN Μοντέλα.....	38
3.1 Διατύπωση προβλήματος.....	38
3.2 Γενικό ZNN (Zhang Neural Network) μοντέλο.....	41
3.3 Σύγκλιση ZNN μοντέλου.....	42
3.4 Εφαρμογή ZNN μοντέλων στο Simulink του Matlab.....	48
3.4.1 Η οικογένεια ολοκληρωτών του Matlab.....	50
3.4.1.1 Εισαγωγικά.....	50
3.5 GNN (Gradient Neural Network) μοντέλα.....	62
3.5.1 GNN (Gradient Neural Network) μοντέλα για σταθερούς πίνακες.....	62
3.5.2 GNN (Gradient Neural Network) μοντέλα για χρονικά μεταβαλλόμενους πίνακες.....	66
3.5.3 Εφαρμογή GNN μοντέλου στο Simulink του Matlab.....	67
3.5.4 Συγκριτική μελέτη ZNN και GNN μοντέλων.....	71
3.6 Επαναλαμβανόμενο Νευρωνικό Δίκτυο για τον υπολογισμό Drazin αντιστρόφου.....	74
Κεφάλαιο 4 Εισαγωγή Στη Παρεμβολή και Επίλυση Αλγεβρικών Συστημάτων.....	98
4.1 Πολυωνυμική παρεμβολή συνάρτησης μίας μεταβλητής.....	98
4.2 Πολυώνυμο παρεμβολής Newton.....	99
4.3 Νευρωνική Προσεγγιση στη μέθοδο παρεμβολής.....	101
4.4 Βελτιωμένη Νευρωνική προσέγγιση στη μέθοδο παρεμβολής.....	105
4.5 Επαναλαμβανόμενο δίκτυο για επίλυση αλγεβρικών εξισώσεων.....	111
4.5.1 ZNN για γραμμικές χρονικά μεταβαλλόμενες εξισώσεις.....	111
4.5.2 ZNN για γραμμικές σταθερές εξισώσεις.....	112

4.6 Επίλυση γραμμικών συστημάτων.....	118
4.7 Μέθοδος LU παραγοντοποίησης.....	119
Γενικά Συμπεράσματα.....	134
Μελέτη για περαιτέρω έρευνα.....	135
Παράρτημα.....	136
Βιβλιογραφία.....	143

Κεφάλαιο 1

Εισαγωγικές Έννοιες Τεχνητών Νευρωνικών Δικτύων

1.1 Εισαγωγή στα Νευρωνικά Δίκτυα

Ένα τεχνητό νευρωνικό δίκτυο (ΤΝΔ) είναι ένα μαθηματικό μοντέλο επεξεργασίας πληροφορίας του οποίου η λειτουργία είναι εμπνευσμένη από τους βιολογικούς νευρώνες, όπως αυτοί του ανθρώπινου εγκεφάλου, για να επεξεργάζονται πληροφορία. Το βασικό στοιχείο ενός νευρωνικού δικτύου είναι ο νευρώνας. Ένα ΤΝΔ αποτελείται από ένα μεγάλο αριθμό διασυνδεδεμένων στοιχείων επεξεργασίας (νευρώνες) τα οποία εργάζονται από κοινού για την επίλυση συγκεκριμένων προβλημάτων. Τα ΤΝΔ, όπως και οι άνθρωποι μαθαίνουν από παραδείγματα και ρυθμίζονται προκειμένου να μπορούν να χρησιμοποιηθούν σε συγκεκριμένες εφαρμογές, όπως είναι η αναγνώριση προτύπων ή η ταξινόμηση δεδομένων, μέσα από μία διαδικασία εκπαίδευσης. Η εκπαίδευση στα βιολογικά συστήματα συνεπάγεται αναπροσαρμογές των συναπτικών συνδέσεων που υπάρχουν μεταξύ των νευρώνων. Το ίδιο ισχύει και για τα ΤΝΔ, όπου μέσω της διαδικασίας εκπαίδευσης ρυθμίζονται οι διασυνδέσεις των νευρώνων (βάρη).

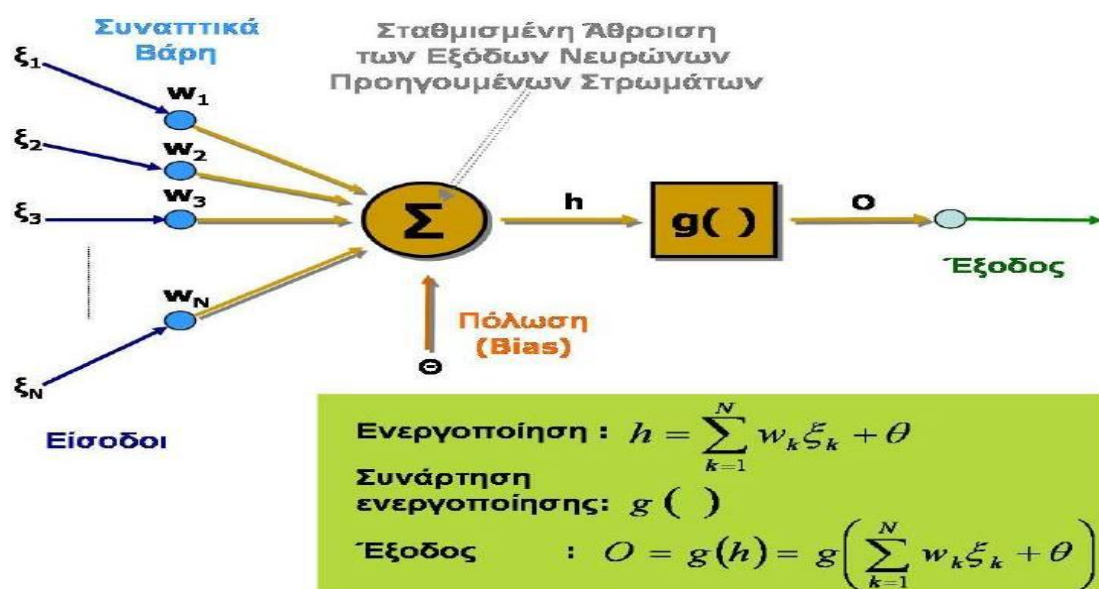
1.2 Εισαγωγή στους Τεχνητούς Νευρώνες

Στη προηγούμενη ενότητα αναφέρθηκε πως στα βιολογικά νευρωνικά δίκτυα οι νευρώνες διασυνδέονται μεταξύ τους και μέσα από την εκπαίδευση γίνεται τροποποίηση του κατά πόσο η έξοδος του ενός επηρεάζει τον άλλο νευρώνα. Κατά αναλογία, στις συνδέσεις μεταξύ των νευρώνων σε ένα ΤΝΔ αντιστοιχεί ένας πραγματικός αριθμός, που ονομάζεται συντελεστής βάρους ή απλά βάρος και χρησιμοποιείται (όπως και οι συνάψεις μεταξύ των ανθρώπινων νευρώνων) για τη τροποποίηση των εισόδων του νευρώνα.

Συνήθως, κάθε νευρώνας (εκτός από τους νευρώνες εισόδου) έχει ακόμα μία σύνδεση που ονομάζεται πόλωση (bias) και βοηθά το ΤΝΔ να έχει καλύτερη ικανότητα ταξινόμησης. Στα περισσότερα ΤΝΔ χρησιμοποιείται και η πόλωση και συνήθως όταν κανείς αναφέρεται στα βάρη υπονοεί και τη πόλωση.

Η εμπειρία και η γνώση που αποκτά το ΤΝΔ αποθηκεύεται στα βάρη του. Η εκπαίδευση του ΤΝΔ είναι η επαναληπτική διαδικασία κατά την οποία γίνεται συνεχής αναπροσαρμογή των τιμών των βαρών έτσι ώστε το ΤΝΔ να έχει την επιθυμητή απόκριση και γίνεται με τη βοήθεια διαφόρων αλγορίθμων εκπαίδευσης.

Στο παρακάτω σχήμα φαίνονται τα δομικά στοιχεία ενός νευρώνα. Τονίζεται ότι κάθε νευρώνας στέλνει μόνο ένα σήμα κάθε φορά στους γειτονικούς νευρώνες. Τα περισσότερα ΤΝΔ για να υπολογίσουν την ενεργοποίηση τους υπολογίζουν το γινόμενο κάθε εισόδου επί το αντίστοιχο βάρος και αθροίζουν όλα αυτά τα γινόμενα. 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 (ή συνάρτηση ενεργοποίησης). Η έξοδος της συνάρτησης είναι η έξοδος του νευρώνα.

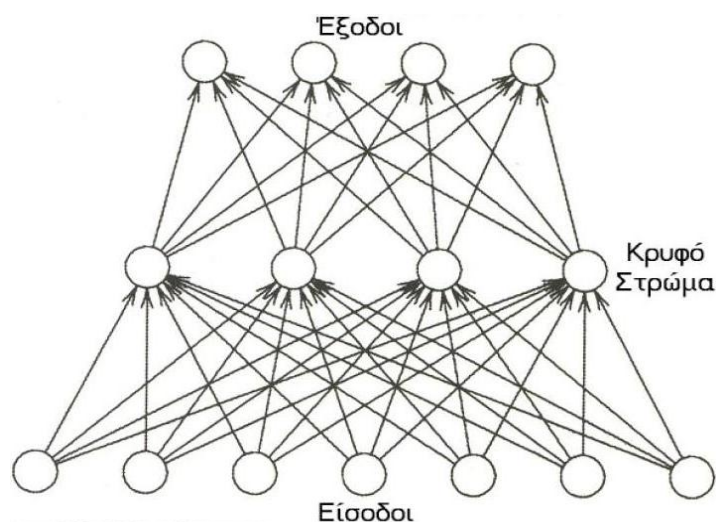


Δομικά στοιχεία ενός νευρώνα [4]

Ο τρόπος με τον οποίο είναι διασυνδεδεμένοι οι νευρώνες ενός δικτύου ονομάζεται τοπολογία ή αρχιτεκτονική του ΤΝΔ. Οι νευρώνες είναι λογικά χωρισμένοι σε επίπεδα, με τους νευρώνες του ίδιου επιπέδου να έχουν την ίδια συνάρτηση μεταφοράς και συνεπώς να έχουν παρόμοια συμπεριφορά. Σε κάθε ΤΝΔ αντιστοιχούν τουλάχιστον δύο στρώματα: το πρώτο στρώμα που ονομάζεται στρώμα εισόδου και έχει ως είσοδο την είσοδο του ΤΝΔ και το στρώμα εξόδου από το οποίο λαμβάνεται η έξοδος του δικτύου. Αν ένα ΤΝΔ δεν αποτελείται μόνο από το στρώμα εισόδου και το στρώμα εξόδου, τότε τα υπόλοιπα στρώματα του ονομάζονται κρυφά στρώματα και το ΤΝΔ λέγεται πολυστρωματικό. Όσο περισσότερα στρώματα έχει ένα νευρωνικό

δίκτυο τόσο πιο δύσκολο είναι να εκπαιδευτεί κατάλληλα, αλλά από την άλλη έχει και περισσότερες δυνατότητες προσέγγισης της επιθυμητής απόκρισης.

Η αρχιτεκτονική ενός ΤΝΔ εξαρτάται από το τρόπο διάδοσης των πληροφοριών μεταξύ των νευρώνων. Υπάρχουν διάφορα είδη δικτύων όπως τα ΤΝΔ πρόσθιας τροφοδότησης, τα ΤΝΔ οπίσθιας τροφοδότησης κ.τ.λ.π. Τα ΤΝΔ πρόσθιας τροφοδότησης για παράδειγμα είναι δίκτυα στα οποία το σήμα διαδίδεται έτσι ώστε να μην υπάρχει νευρώνας που η έξοδος του είναι είσοδος κάποιου νευρώνα του ίδιου ή προηγούμενου στρώματος. Επίσης, στα ΤΝΔ συνήθως κάθε νευρώνας ενός στρώματος, διασυνδέεται με όλους τους νευρώνες του επόμενου στρώματος. Στο επόμενο σχήμα παρουσιάζεται πολυστρωματικό πλήρως διασυνδεδεμένο πρόσθιας τροφοδοσίας ΤΝΔ.



Παράδειγμα Πολυστρωματικού Δικτύου [4]

Η συμπεριφορά ενός ΤΝΔ, δηλαδή η απόκριση σε συγκεκριμένες εισόδους εξαρτάται από: (α) την αρχιτεκτονική του, (β) τις συναρτήσεις ενεργοποίησης που χρησιμοποιεί, και (γ) τον αλγόριθμο εκπαίδευσης και τα δεδομένα εκπαίδευσης που χρησιμοποιήθηκαν για την εκπαίδευση του.

1.3 Αρχιτεκτονικές των ΤΝΔ

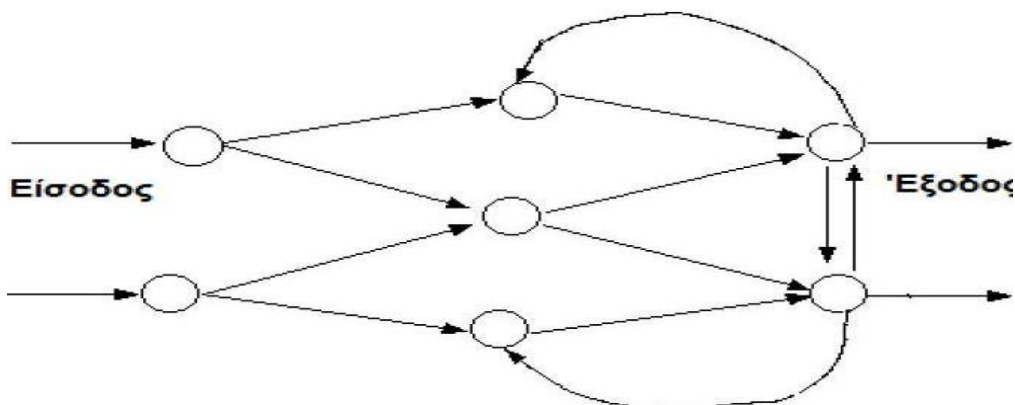
Στη συνέχεια θα γίνει παρουσίαση των διαφόρων αρχιτεκτονικών που χρησιμοποιούνται στα ΤΝΔ. Η αρχιτεκτονική ενός ΤΝΔ εξαρτάται και από τον τρόπο διάδοσης των πληροφοριών μεταξύ των νευρώνων.

1.3.1 ΤΝΔ πρόσθιας τροφοδότησης

ΤΝΔ πρόσθιας τροφοδότησης είναι δίκτυα στα οποία το σήμα διαδίδεται έτσι ώστε να μην υπάρχει νευρώνας που η έξοδος του είναι είσοδος κάποιου νευρώνα του ίδιου ή προηγούμενου στρώματος. Με άλλα λόγια, το σήμα μεταφέρεται μόνο προς μία κατεύθυνση, από την είσοδο του νευρωνικού προς την έξοδο. Δεν υπάρχουν βρόγχοι ανατροφοδότησης και η έξοδος ενός στρώματος επηρεάζει μονάχα τα επόμενα στρώματα.

1.3.2 ΤΝΔ με ανατροφοδότηση

Στα δίκτυα με ανατροφοδότηση τα σήματα ταξιδεύουν και προς τις δύο κατευθύνσεις καθώς υπάρχουν βρόγχοι στο δίκτυο. Τα δίκτυα αυτά είναι εν γένει πολύ ισχυρά και ταυτόχρονα ιδιαίτερα πολύπλοκα. Τα δίκτυα με ανατροφοδότηση είναι δυναμικά, η κατάσταση τους αλλάζει συνεχώς μέχρι να φτάσουν σε μια κατάσταση ισορροπίας, στην οποία παραμένουν μέχρι να αλλάξει η είσοδος και να βρεθεί καινούριο σημείο ισορροπίας.



Δίκτυο με ανατροφοδότηση [4]

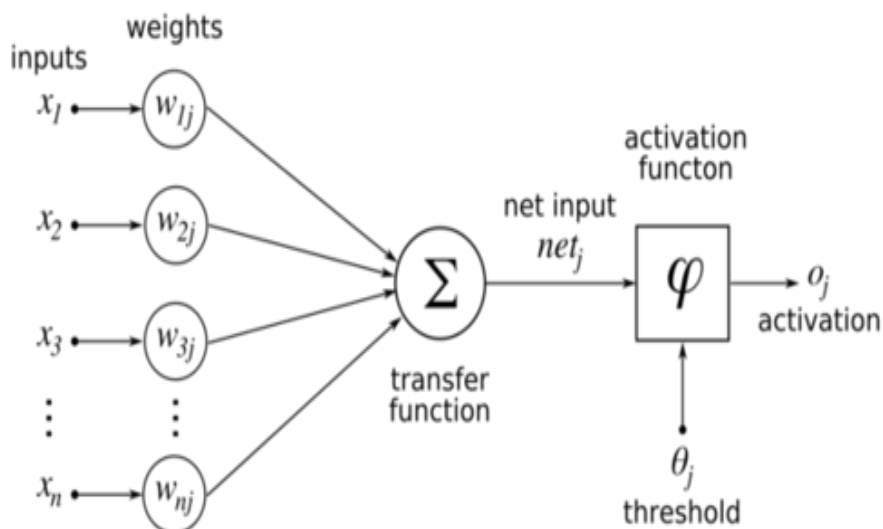
1.3.3 Perceptrons

Η έρευνα με τη μεγαλύτερη επιρροή στα νευρωνικά δίκτυα τη δεκαετία του 1960 συνδέθηκε με το τίτλο perceptrons ένας όρος που επινοήθηκε από τον Frank Rosenblatt. Το perceptron είναι ένα μοντέλο ΤΝΔ με νευρώνες με σταθμισμένες εισόδους και με κάποια πρόσθετη, σταθερή επεξεργασία. Τα perceptrons μιμούνται τη βασική ιδέα πίσω από το οπτικό σύστημα των θηλαστικών και χρησιμοποιήθηκαν κυρίως στην αναγνώριση προτύπων. Είναι δυαδικοί ταξινομητές οι οποίοι

αντιστοιχίζουν την είσοδο x (ένα διάνυσμα με πραγματικές τιμές) σε μια τιμή εξόδου $f(x)$ (μια δυαδική τιμή) τέτοια ώστε:

$$f(x) = \begin{cases} 1, & wx + b > 0 \\ 0, & wx + b \leq 0 \end{cases} \quad (1.1)$$

Το perceptron θεωρείται ως απλούστερη μορφή δικτύου πρόσθιας τροφοδότησης. Το 1969 οι Minsky και Papert [26] δημοσίευσαν τους περιορισμούς ενός perceptron με ένα στρώμα. Ο αντίκτυπος που είχε η δημοσίευση αυτή ήταν τεράστιος και προκάλεσε πολλούς ερευνητές να χάσουν το ενδιαφέρον τους όσον αφορά την εξερεύνηση των ΤΝΔ. Οι Minsky και Papert [26] απέδειξαν μαθηματικά ότι perceptrons με ένα στρώμα δε μπορούν να κάνουν κάποιες βασικές λειτουργίες αναγνώρισης προτύπων. Στη συνέχεια αποδείχθηκε ότι πολυστρωματικά perceptron μπορούν να κάνουν με επιτυχία κάθε τέτοιου είδους λειτουργία.



Μοντέλο νευρώνα Perceptron [4]

1.4 Πλεονεκτήματα των ΤΝΔ

Το επιστημονικό ενδιαφέρον για τα ΤΝΔ προκύπτει κυρίως από τη δυνατότητά τους να επιλύουν δύσκολα και ενδιαφέροντα υπολογιστικά προβλήματα του πραγματικού κόσμου. Η χρήση των ΤΝΔ προσφέρει τις ακόλουθες πολύ χρήσιμες ιδιότητες και δυνατότητες.

Αντικατάσταση μαθηματικού μοντέλου. Η αξία των ΤΝΔ έγκειται κυρίως στο γεγονός ότι η χρήση τους δε συνεπάγεται τη δημιουργία κάποιου μοντέλου για το περιβάλλον το οποίο καλούνται να προσομοιώσουν. Όταν η περιγραφή του περιβάλλοντος είναι περίπλοκη για να περιγραφεί με κάποιο μαθηματικό μοντέλο, τότε η χρήση των ΤΝΔ αποτελεί μία καλή λύση. Κατά την εκπαίδευση δίνονται στο ΤΝΔ πρότυπα εισόδου/εξόδου (δείγματα εισόδων και εξόδων της συνάρτησης που καλείται το ΤΝΔ να προσομοιώσει). Στόχος της διαδικασίας εκπαίδευσης είναι να φτάσει το ΤΝΔ σε μια τέτοια κατάσταση όπου για κάθε πρότυπο εκπαίδευσης, η έξοδος του να ταυτίζεται με την επιθυμητή έξοδο. Έτσι δημιουργείται μια συσχέτιση μεταξύ των δεδομένων εισόδου και εξόδου, χωρίς όμως τη χρήση κάποιου προκαθορισμένου στατιστικού ή άλλου μοντέλου.

Προσαρμογή. Τα ΤΝΔ έχουν την ικανότητα να μεταβάλλουν την απόκριση τους μεταβάλλοντας ελαφρώς τα βάρη των νευρώνων, ανάλογα με τα πρότυπα εισόδου. Συνεπώς τα περιβάλλοντα τα οποία αλλάζουν διαρκώς, τα ΤΝΔ μπορούν να εφαρμοστούν εφόσον υπόκεινται περιοδικά σε εκπαίδευση με τα επικαιροποιημένα πρότυπα εισόδου/εξόδου.

Δυνατότητα εύκολης υλοποίησης σε κατάλληλα λογισμικά. Η απλή δομή των νευρώνων και το γεγονός ότι ένα ΤΝΔ αποτελείται από πολλά ίδια στοιχεία (νευρώνες) καθιστά απλή την υλοποίηση τους σε διάφορα λογισμικά.

Ανεκτικότητα σε σφάλματα. Τα ΤΝΔ που έχουν υλοποιηθεί σε υλικό έχουν την ίδια ανεκτικότητα σε σφάλματα, γιατί η απόδοση του συστήματος μειώνεται ομαλά σε περίπτωση λάθους. Ακόμα και αν καταστραφεί ένας νευρώνας το ΤΝΔ θα συνεχίζει να λειτουργεί με κάπως μειωμένη απόδοση.

Ανοχή στο θόρυβο. Τα ΤΝΔ παρουσιάζουν ανοχή στο θόρυβο υπό την έννοια ότι αν εκπαιδευτούν για την αναγνώριση προτύπων τότε είναι σε γενικές γραμμές εφικτή η ταξινόμηση από το νευρωνικό ακόμα και όταν η είσοδος έχει υποστεί κάποια

1.5 Μέθοδοι εκπαίδευσης

Όλες οι μέθοδοι εκπαίδευσης που χρησιμοποιούνται στα νευρωνικά δίκτυα μπορούν να ταξινομηθούν σε δύο μεγάλες κατηγορίες:

•Εκπαίδευση με επίβλεψη η οποία απαιτεί την εμφάνιση ενός δασκάλου, έτσι ώστε το δίκτυο να μαθαίνει ποια είναι η επιθυμητή έξοδος για κάθε είσοδο. Κατά τη διαδικασία εκπαίδευσης μπορεί να απαιτείται γενικευμένη πληροφορία, παραδείγματα εκπαίδευσης με επίβλεψη αποτελούν η εκπαίδευση διόρθωσης σφάλματος, η εκπαίδευση με ενίσχυση και η στοχαστική εκπαίδευση. Ένα σημαντικό ζήτημα σχετικά με την εκπαίδευση με επίβλεψη είναι το πρόβλημα της σύγκλισης λάθους, δηλαδή η ελαχιστοποίηση του σφάλματος μεταξύ της επιθυμητής και της υπολογιζόμενης τιμής. Ο στόχος είναι να καθοριστεί ένα σύνολο βαρών το οποίο ελαχιστοποιεί το σφάλμα. Μία γνωστή μέθοδος, η οποία είναι κοινή για πολλά παραδείγματα εκπαίδευσης είναι η σύγκλιση του ελάχιστου μέσου τετραγώνων.

•Η Εκπαίδευση χωρίς επίβλεψη δε χρησιμοποιεί κάποιο δάσκαλο και βασίζεται μόνο σε τοπικές πληροφορίες. Αναφέρεται επίσης ως αυτό-οργάνωση. με την έννοια ότι αυτό-οργανώνει τα δεδομένα που παρουσιάζονται στο δίκτυο και εντοπίζει τις προκύπτουσες συλλογικές ιδιότητες τους. Ένα νευρωνικό δίκτυο εκπαιδεύεται off-line αν η φάση εκπαίδευσης και φάση λειτουργίας είναι διαφορετικές. Ένα νευρωνικό δίκτυο εκπαιδεύεται on-line, αν η εκπαίδευση και η λειτουργία του γίνονται ταυτόχρονα. Συνήθως, η εκπαίδευση με επίβλεψη γίνεται off-line, ενώ η εκπαίδευση χωρίς επίβλεψη γίνεται on-line.

1.6 Εισαγωγή στο Neural Network Toolbox του Matlab

Τα νευρωνικά δίκτυα αποτελούνται από απλά στοιχεία που λειτουργούν παράλληλα. Τα στοιχεία στα δίκτυα προσομοιώνουν σε ένα βαθμό τα βιολογικά νευρωνικά συστήματα. Όπως στη φύση, οι συνδέσεις μεταξύ των στοιχείων καθορίζουν κατά ένα μεγάλο μέρος τη συνάρτηση του δικτύου. Μπορεί κάποιος να εκπαιδεύσει ένα νευρωνικό δίκτυο έτσι ώστε αυτό να εκτελεί μια ιδιαίτερη συνάρτηση με τη ρύθμιση των τιμών των συνδέσεων (βάρη) μεταξύ των στοιχείων.

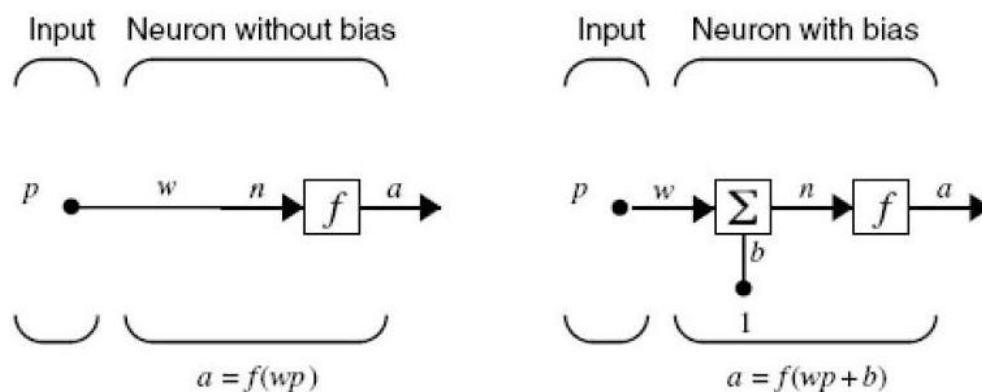
Χαρακτηριστικά, τα νευρωνικά δίκτυα εκπαιδεύονται έτσι ώστε μία συγκεκριμένη είσοδος να οδηγεί σε μια συγκεκριμένη έξοδο. Το δίκτυο ρυθμίζεται έτσι ώστε η έξοδος να ταιριάζει τελικά με το στόχο που έχει τεθεί. Αυτό επιτυγχάνεται με μία σύγκριση που γίνεται μεταξύ του στόχου και της επιθυμητής εξόδου. Τα νευρωνικά δίκτυα έχουν χρησιμοποιηθεί για την εκτέλεση σύνθετων λειτουργιών σε διάφορους

τομείς, όπως η αναγνώριση προτύπων, η επεξεργασία και αναγνώριση ομιλίας, εικόνας, και τα συστήματα αυτομάτου ελέγχου.

Τα νευρωνικά δίκτυα μπορούν να εκπαιδευθούν για να λύσουν τα προβλήματα που είναι δύσκολα για τους συμβατικούς υπολογιστές ή ανθρώπους. Οι επόμενες ενότητες περιγράφουν τα βασικά χαρακτηριστικά της εργαλειοθήκης του Matlab στα νευρωνικά δίκτυα. Πιο συγκεκριμένα παρουσιάζονται οι βασικές αρχές του προτύπου νευρώνων, οι αρχιτεκτονικές των νευρωνικών δικτύων, τα βασικά στοιχεία καθώς και η ονοματολογία γύρω από αυτά τα δίκτυα. Στη συνέχεια, γίνεται παρουσίαση της διαδικασίας δημιουργίας και εκπαίδευσης απλών δικτύων με τη χρήση της εργαλειοθήκης του Matlab όπως και του περιβάλλοντος του Simulink.

1.6.1 Μοντέλο Νευρώνων

Ένας νευρώνας με μία μοναδική αριθμητική είσοδο και καμία πόλωση εμφανίζεται στο αριστερό μέρος του παρακάτω σχήματος.



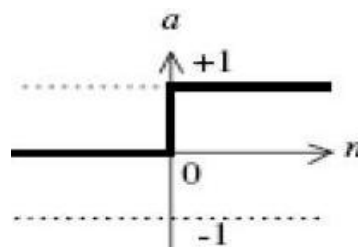
Μοντέλο Νευρώνων [4]

Η αριθμητική είσοδος P διαβιβάζεται μέσω μίας σύνδεσης που πολλαπλασιάζει τη δύναμή της με το αριθμητικό βάρος W για να διαμορφώσει το προϊόν wp , που είναι επίσης αριθμός. Εδώ η σταθμισμένη εισαγωγή wp είναι το μόνο όρισμα της συνάρτησης μεταφοράς f , η οποία παράγει την αριθμητική έξοδο a ο νευρώνας στα δεξιά του σχήματος έχει μια αριθμητική πόλωση (bias), b . Κανείς μπορεί να αντιληφθεί τη πόλωση σαν ένα προσθετέο που απλά προστίθεται στο προϊόν wp όπως παρουσιάζεται από το κόμβο άθροισμα Σ ή ως μετατόπιση της συνάρτησης f προς τα αριστερά κατά ένα ποσό b . Η πόλωση είναι σαν ένα βάρος με τη διαφορά ότι έχει σταθερή είσοδο 1.

Η συνάρτηση μεταφοράς n είναι επίσης αριθμός και ισοδυναμεί με το άθροισμα της σταθμισμένης εισαγωγής $w \cdot p$ με τη πόλωση b . Αυτό το ποσό είναι όρισμα της συνάρτησης μεταφοράς f . Στη περίπτωση αυτή η f είναι μία συνάρτηση μεταφοράς, για παράδειγμα μία συνάρτηση βημάτων ή μία σιγμοειδής συνάρτηση μεταφοράς, η οποία παίρνει το όρισμα n και παράγει την έξοδο a . Αξίζει να σημειωθεί ότι το w και το b είναι και οι δύο διευθετήσιμες αριθμητικές παράμετροι του νευρώνα. Η κεντρική ιδέα των νευρωνικών δικτύων είναι ότι τέτοιες παράμετροι μπορούν να ρυθμιστούν έτσι ώστε το δίκτυο να παρουσιάζει κάποια επιθυμητή ή ενδιαφέρουσα συμπεριφορά. Κατά συνέπεια, μπορεί κανείς να εκπαιδεύσει το δίκτυο να εκτελεί μία ιδιαίτερη εργασία με τη ρύθμιση του βάρους ή της πόλωσης, ή ίσως το ίδιο το δίκτυο θα ρυθμίσει αυτές τις παραμέτρους για να επιτύχει κάποιο επιθυμητό αποτέλεσμα.

1.6.2 Συναρτήσεις Μεταφοράς

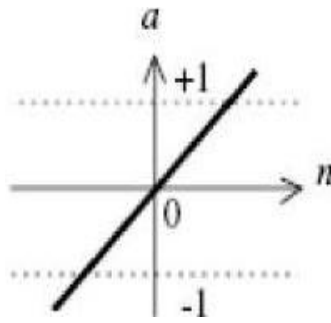
Πολλές συναρτήσεις μεταφοράς συμπεριλαμβάνονται στο Neural Network Toolbox του Matlab. Στη συνέχεια παρουσιάζονται τέσσερις από τις πιο συχνά χρησιμοποιούμενες συναρτήσεις.



Συνάρτηση μεταφοράς Hard-limit [4]

Η συνάρτηση μεταφοράς αυστηρού ορίου που εμφανίζεται στο παραπάνω σχήμα περιορίζει την έξοδο του νευρώνα είτε σε 0, εάν το όρισμα n είναι λιγότερο από 0, είτε 1, εάν το n είναι μεγαλύτερο ή ίσο με 0. Αυτή η λειτουργία χρησιμοποιείται στα perceptrons για να δημιουργήσει τους νευρώνες που λαμβάνουν αποφάσεις ταξινόμησης. Η υλοποίηση της παραπάνω μαθηματικής έννοιας λέγεται hardlim στο Neural Network Toolbox.

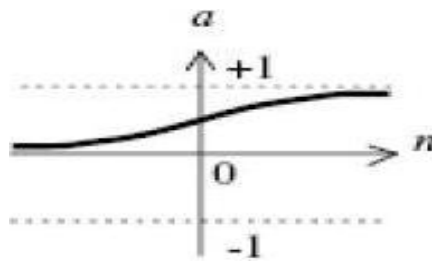
Μία ακόμα συνηθισμένη συνάρτηση μεταφοράς είναι η γραμμική συνάρτηση μεταφοράς. Η αντίστοιχη συνάρτηση του Neural Network Toolbox (purelin) και το σύμβολο της παρουσιάζονται παρακάτω.



Γραμμική Συνάρτηση μεταφοράς [4]

Μία ακόμα συνήθης συνάρτηση μεταφοράς στα νευρωνικά δίκτυα είναι η σιγμοειδής (sigmoid). Η συνάρτηση αυτή παίρνει σαν όρισμα μία τιμή ανάμεσα στο πλην άπειρο και το συν άπειρο και περιορίζει την έξοδο στο εύρος 0 με 1 και δίνεται από το τύπο:

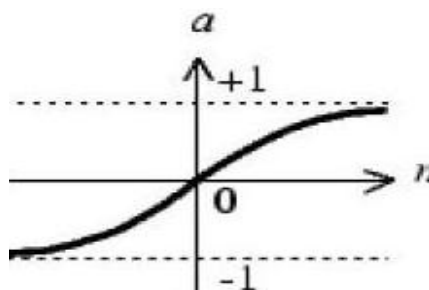
$$\log \text{sig}(n) = \frac{1}{1+e^{-n}} \quad (1.2)$$



Η Σιγμοειδής Συνάρτηση

Παραλλαγή της παραπάνω συνάρτηση μεταφοράς η tansig η οποία δίνεται από το τύπο:

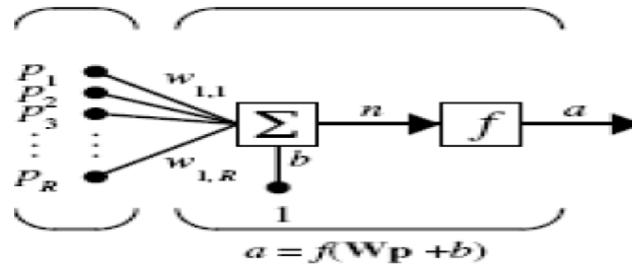
$$\tan \text{sig}(n) = \frac{2}{1-e^{-2n}} - 1 \quad (1.3)$$



Η Συνάρτηση Tansig(n)

1.6.3 Νευρώνες με είσοδο διάνυσμα

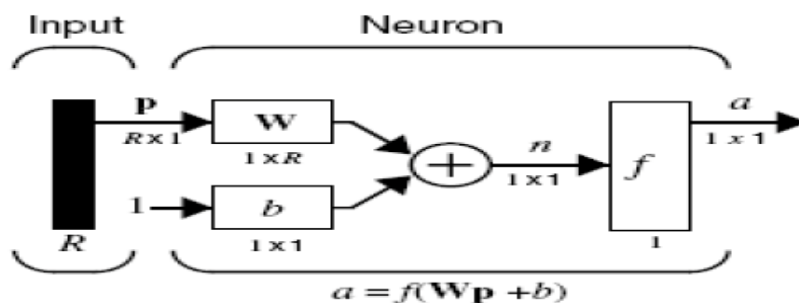
Ένας νευρώνας εκτός από αριθμητική τιμή μπορεί να έχει και ως είσοδο και ένα διάνυσμα



Είσοδος Διάνυσμα [4]

Αν η είσοδος είναι ένα διάνυσμα p με R ξεχωριστά στοιχεία $\{p_1, p_2, \dots, p_R\}$ τότε αυτό πολλαπλασιάζεται με το αντίστοιχο διάνυσμα βαρών $W = \{w_1, w_2, \dots, w_R\}$. Αν ο νευρώνας έχει και πόλωση b , τότε το όρισμα της συνάρτησης μεταφοράς του νευρώνα είναι: $n = w_1 p_1 + w_2 p_2 + \dots + w_R p_R + b$. Το διάγραμμα ενός μόνο νευρώνα που παρουσιάζεται στο προηγούμενο σχήμα περιέχει πολλή λεπτομέρεια. Όταν εξετάζει κανείς δίκτυα με πολλούς νευρώνες, και ίσως στρώματα πολλών νευρώνων χρειάζεται ένα πιο λιτό συμβολισμό. Για το λόγο αυτό, χρησιμοποιείται ένας πιο λιτός συμβολισμός ανά νευρώνα για τη σχεδίαση κυκλωμάτων πολλαπλών νευρώνων, ο οποίος παρουσιάζεται στο επόμενο σχήμα. Εδώ το διάνυσμα είσοδος p αντιπροσωπεύεται από τη κατακόρυφη μπάρα στα αριστερά.

Οι διαστάσεις του p παρουσιάζονται κάτω από το σύμβολο p στο σχήμα ως $R \times 1$. (Σημειώνετε ότι ένα κεφαλαίο γράμμα, όπως το R στη προηγούμενη πρόταση αναφέρεται στο μέγεθος ενός διανύσματος). Κατά συνέπεια, το p είναι ένα διάνυσμα R στοιχείων εισόδου.

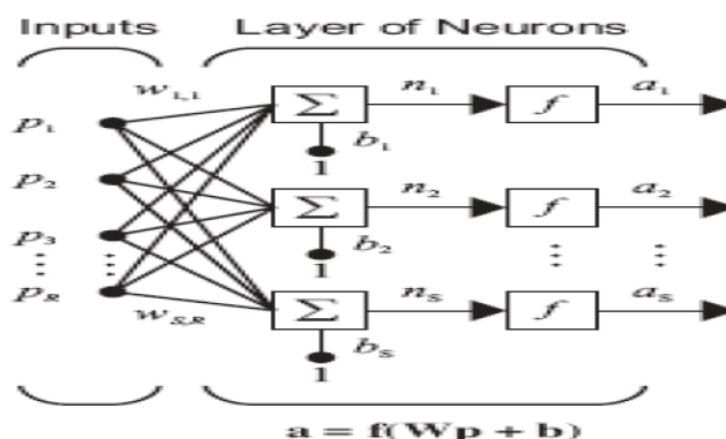


Σύντομος Συμβολισμός Νευρώνα [4]

Το παραπάνω σχήμα ορίζει ένα στρώμα δικτύου. Ένα στρώμα περιέχει το συνδυασμό βαρών και το πολλαπλασιασμό ($W \cdot p$), τη πρόσθεση της πόλωσης b και τη συνάρτηση μεταφοράς f . Το διάνυσμα εισόδων p δε συμπεριλαμβάνεται στο στρώμα. Κάθε φορά που χρησιμοποιείται αυτός ο σύντομος συμβολισμός, τα μεγέθη των πινάκων φαίνονται κάτω ακριβώς από τα ονόματα των μεταβλητών που τους αναπαριστούν. Στο παραπάνω σχήμα, όταν χρησιμοποιείται κάποια συγκεκριμένη συνάρτηση μεταφοράς, όπως αυτές που περιγράφηκαν προηγουμένως, ο ειδικός συμβολισμός τους αντικαθιστά το κουτί με το f .

1.6.4 Αρχιτεκτονικές Δικτύου

Δύο ή περισσότεροι από τους νευρώνες που παρουσιάστηκαν νωρίτερα μπορούν να συνδυαστούν σε ένα στρώμα, και ένα δίκτυο θα μπορούσε να περιέχει ένα ή περισσότερα τέτοια στρώματα.



Στρώμα Νευρώνων [4]

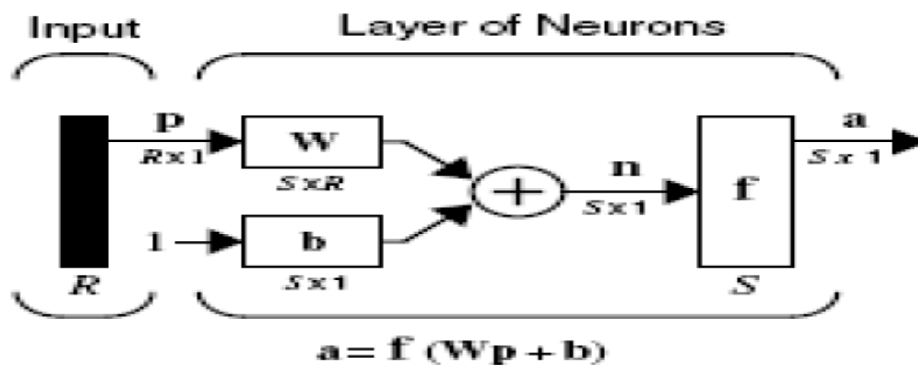
Στο παραπάνω σχήμα παρουσιάζεται ένα δίκτυο από S νευρώνες με R εισόδους. Σε αυτό το δίκτυο, κάθε στοιχείο του διανύσματος εισόδου p συνδέεται με κάθε νευρώνα που εισάγεται μέσω του πίνακα βάρους W . Ο i -στός νευρώνας έχει έναν αθροιστή που συλλέγει τις σταθμισμένες εισαγωγές και την πόλωση του για να διαμορφώσει την αριθμητική έξοδο $n(i)$. Τα διάφορα $n(i)$ όταν συλλέγονται δημιουργούν ένα δίκτυο αποτελούμενο από S στοιχεία. Τέλος, οι έξοδοι του στρώματος διαμορφώνουν μία στήλη διάνυσμα a . Η έκφραση για το a παρουσιάζεται στο κατώτατο σημείο του σχήματος. Στο παραπάνω σχήμα R είναι ο αριθμός στοιχείων στο διάνυσμα εισόδου, ενώ S είναι ο αριθμός νευρώνων στο στρώμα. Είναι σύνηθες ο αριθμός εισόδων να είναι διαφορετικός από τον αριθμό νευρώνων.

Μπορεί να οριστεί ένα σύνθετο στρώμα νευρώνων που να έχει διαφορετικές συναρτήσεις μεταφοράς βάζοντας δύο απλά δίκτυα παράλληλα. Και τα δύο δίκτυα θα πρέπει να έχουν τις ίδιες εισόδους, και κάθε δίκτυο θα παράγει κάποιες από τις εξόδους. Η είσοδος γίνεται με τη βοήθεια ενός πίνακα βαρών W :

$$W = \begin{bmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,R} \\ w_{2,1} & w_{2,2} & \ddots & w_{2,R} \\ \vdots & \vdots & \ddots & \vdots \\ w_{S,1} & w_{S,2} & \cdots & w_{S,R} \end{bmatrix} \quad (1.4)$$

Ο δείκτης γραμμών σε κάθε στοιχείο καθορίζει το νευρώνα για τον οποίο προορίζεται το συγκεκριμένο στοιχείο-βάρος ενώ ο δείκτης στήλης ποια είσοδος ταιριάζει με το συγκεκριμένο βάρος. Συνεπώς, οι δείκτες στο $w_{1,2}$ υποδηλώνουν ότι η ισχύς του σήματος από τη δεύτερη είσοδο στο πρώτο νευρώνα είναι $w_{1,2}$.

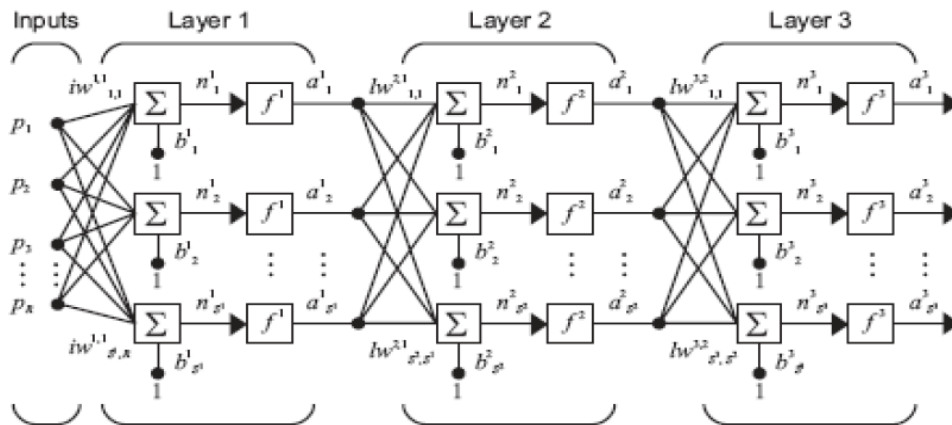
Το δίκτυο που αποτελείται από S νευρώνες και R εισόδους μπορεί επίσης να σχεδιαστεί με συντομία ως εξής:



Περιεκτική περιγραφή στρώματος [4]

Εδώ το p είναι ένα διάνυσμα εισόδου με μήκος R και W είναι ένας πίνακας $S \times R$, και τα a και b είναι διανύσματα μήκους S . Το δίκτυο αυτό περιλαμβάνει το πίνακα βαρών, τις πράξεις πολλαπλασιασμού, το διάνυσμα πόλωσης b , τον αθροιστή και τα κουτιά των συναρτήσεων μεταφοράς.

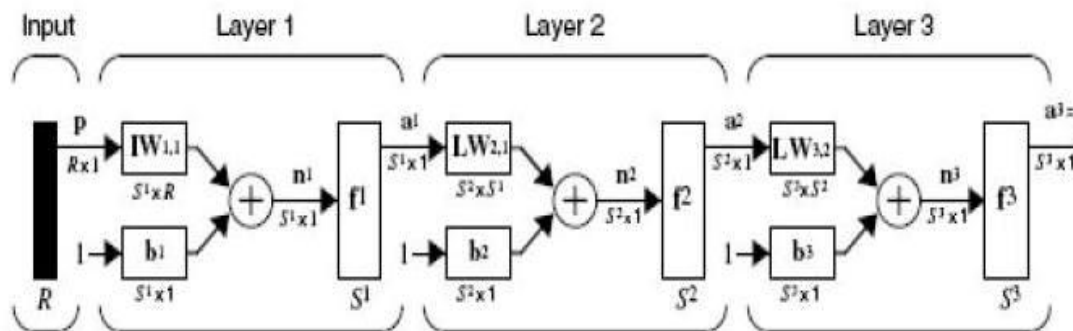
Ένα δίκτυο μπορεί να έχει πολλαπλά στρώματα νευρώνων, Κάθε στρώμα έχει ένα πίνακα βάρους W , το διάνυσμα πόλωσης b και το διάνυσμα εξόδου a . Για να ξεχωρίσει κανείς τους πίνακες βαρών, τα διανύσματα πόλωσης κτλ κάθε στρώματος προστίθενται ο αριθμός του στρώματος σαν άνω κείμενο σε κάθε μεταβλητή. Ο συμβολισμός αυτός φαίνεται στο παρακάτω σχήμα.



Δίκτυο Πολλαπλών Στρωμάτων [4]

Το παραπάνω δίκτυο έχει R^1 εισόδους, S^1 νευρώνες στο πρώτο στρώμα, S^2 νευρώνες στο δεύτερο στρώμα κτλ. Είναι σύνηθες κάθε στρώμα να έχει διαφορετικό αριθμό νευρώνων. Οι εξοδοί κάθε ενδιάμεσου στρώματος είναι οι εισοδοί του επομένου.

Τα στρώματα σε ένα δίκτυο νευρώνων παίζουν διαφορετικό ρόλο. Το στρώμα που παράγει την έξοδο του δικτύου ονομάζεται στρώμα εξόδου. Όλα τα υπόλοιπα στρώματα ονομάζονται κρυφά.



Περιεκτική περιγραφή Δικτύου Πολλαπλών στρωμάτων [4]

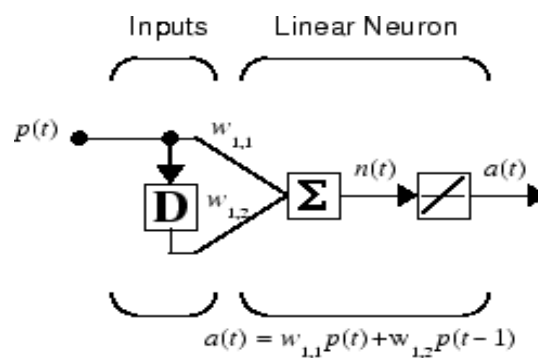
Τα δίκτυα που έχουν πολλαπλά στρώματα έχουν σημαντικές δυνατότητες. Για παράδειγμα, ένα δίκτυο με δύο στρώματα, όπου το πρώτο στρώμα έχει συνάρτηση μεταφοράς sigmoid και το δεύτερο γραμμική, μπορεί να εκπαιδευτεί ώστε να προσεγγίζει οποιαδήποτε συνάρτηση (με πεπερασμένο πλήθος ασυνεχειών) αρκετά καλά. Στο Neural Network Toolbox η έξοδος του δικτύου συμβολίζεται με y .

Συνεχίζοντας θα ήταν σωστό να παραθέσουμε στη συνέχεια τους παρακάτω ορισμούς οι οποίοι θα μας βοηθήσουν να κατανοήσουμε καλύτερα τα επόμενα παραδείγματα.

Ορισμός 1.1 [4] Μια *εποχή* εκπαίδευσης ονομάζεται η διαδικασία εισαγωγής όλων των ζευγαριών εισόδων-εξόδων στο δίκτυο μια φορά.

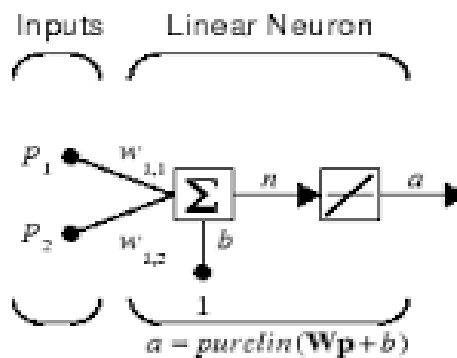
Ορισμός 1.2 [4] *Ρυθμός εκπαίδευσης* είναι η παράμετρος που ελέγχει το μέγεθος των βαρών και των αλλαγών των πολώσεων στην εκμάθηση του αλγόριθμου εκπαίδευσης (κυμαίνεται μεταξύ [0,1]).

Ορισμός 1.3 [4] *Δυναμικά* τεχνητά νευρωνικά δίκτυα είναι τα ΤΝΔ στα οποία τα διανύσματα εξόδου περιγράφονται από πολύπλοκες εξισώσεις διαφορών ή διαφορικές εξισώσεις.



Παράδειγμα Δυναμικού Δικτύου [4]

Ορισμός 1.4 [4] *Στατικά* τεχνητά νευρωνικά δίκτυα είναι τα ΤΝΔ στα οποία η έξοδος κάθε νευρώνα αποτελεί μία συγκεκριμένη, σχετικά απλή, συνάρτηση μόνο του διανύσματος εισόδου.



Παράδειγμα Στατικού Δικτύου [4]

Κεφάλαιο 2

Βασικές Μαθηματικές Έννοιες

Στο κεφάλαιο αυτό θα αναφερθούμε σε κάποιες βασικές έννοιες, αναγκαίες για τη κατανόηση των επόμενων ενοτήτων. Θα δώσουμε τους ορισμούς των πινάκων, των πολυωνύμων, πολυωνυμικών πινάκων, του αντίστροφου ενός πολυωνυμικού πίνακα καθώς και εννοιών που αφορούν την εμπέδωση των μεθόδων που πρόκειται να αναλύσουμε.

2.1 Πίνακες

Έστω F ένα σώμα. Μία ορθογώνια διάταξη στοιχείων a_{ij} ($i=1,2,3,\dots,n$, $j=1,2,\dots,m$) ενός σώματος F .

$$A = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1m} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ \alpha_{n1} & \alpha_{n2} & \cdots & \alpha_{nm} \end{bmatrix} \quad (2.1)$$

θα λέγεται πίνακας. Ο παραπάνω πίνακας θα λέμε ότι είναι *διάστασης* $n \times m$, ότι έχει δηλαδή n γραμμές και m στήλες. Αν έχουμε $m=n$, τότε ο πίνακας είναι *τετραγωνικός*. Συμβολίζουμε με $M_{n \times m}(F)$ το σύνολο των $n \times m$ πινάκων με στοιχεία από το σώμα F . Συγκεκριμένα αν ο πίνακας είναι τετραγωνικός θα γράφουμε $M_n(F)$.

Ορισμός 2.1.1 [6]

Ένας τετραγωνικός πίνακας λέγεται *μοναδιαίος* αν όλα τα στοιχεία στην κύρια διαγώνιο είναι 1 και τα υπόλοιπα 0. Τον πίνακα αυτόν τον συμβολίζουμε με I_n .

Ορισμός 2.1.2 [6]

Έστω $A \in M_{n \times m}(F)$. Ονομάζουμε βαθμίδα του πίνακα A το μεγαλύτερο αριθμό από γραμμικά ανεξάρτητες στήλες (γραμμές) του A . Συμβολίζεται με $\text{rank}(A)$.

Συμβολίζουμε την ορίζουσα ενός πίνακα $A \in M_n(F)$ με $\det(A)$ ή με $|A|$. Η ορίζουσα είναι πάντα αριθμός και περιέχει κάποιες πληροφορίες για το πίνακα. Υπολογίζεται από το τύπο:

$$\det(A) = \sum_j (-1)^{i+j} \det(A_{ij})$$

όπου A_{ij} προκύπτει αν αφαιρέσουμε από τον A τη γραμμή i και τη στήλη j και $\det[a_{ij}] = a_{ij}$.

Ορισμός 2.1.3 [6]

Ένας πίνακας $A \in M_n(F)$ λέγεται ομαλός (non-singular) αν ισχύει $|A| \neq 0$ και ιδιόμορφος (singular) αν ισχύει $|A| = 0$.

Ορισμός 2.1.4 [6]

Έστω ένας πίνακας $A \in M_n(F)$. Ο *ανάστροφος* πίνακας του A είναι ένας νέος πίνακας που προκύπτει αν τις γραμμές του A τις κάνουμε στήλες του καινούριου πίνακα. Συμβολίζεται με A^T .

Ορισμός 2.1.5 [6]

Ένας πίνακας $A \in M_n(F)$ θα λέγεται *συμμετρικός* αν ισχύει $A = A^T$ και *αντί-συμμετρικός* αν ισχύει $A = -A^T$.

Ορισμός 2.1.6 [8]

Ένας πίνακας λέγεται *κάτω τριγωνικός* αν για κάθε $i < j$ έχουμε $a_{ij} = 0$. Δηλαδή ένας τετραγωνικός πίνακας λέγεται *κάτω τριγωνικός* αν τα στοιχεία που βρίσκονται πάνω από τη κύρια διαγώνιο είναι ίσα με 0.

Ορισμός 2.1.7 [8]

Ένας πίνακας λέγεται *άνω τριγωνικός* αν για κάθε $i > j$ έχουμε $a_{ij} = 0$. Δηλαδή ένας τετραγωνικός πίνακας λέγεται *άνω τριγωνικός* αν τα στοιχεία που βρίσκονται κάτω από τη κύρια διαγώνιο είναι ίσα με 0.

Ορισμός 2.1.8 [8]

Αν για το πίνακα $A \neq 0$ του $M_n(F)$ υπάρχει ένας πίνακας $B \in M_n(F)$ ώστε να ισχύει η σχέση $AB = I_n = BA$, τότε ο πίνακας B λέγεται *αντίστροφος* πίνακας του A και συμβολίζεται με A^{-1} .

Ένας τρόπος υπολογισμού του αντίστροφου πίνακα είναι με το τύπο

$$A^{-1} = \frac{1}{|A|} \text{adj}(A)$$

όπου $\text{adj}(A)$ είναι ο πίνακας που περιέχει στοιχεία της μορφής $\text{adj}(A) = ((-1)^{i+j}|A_{ij}|)^T$ και ο οποίος λέγεται *προσαρτημένος πίνακας*.

Ορισμός 2.1.9 [6]

Έστω $A \in M_{n \times m}(F)$. Το πολυώνυμο $|A - \lambda I_n|$ λέγεται χαρακτηριστικό πολυώνυμο του πίνακα A . Θα το συμβολίζουμε με $c(\lambda)$. Μας βοηθάει να υπολογίσουμε τις *ιδιοτιμές* και τα *ιδιοδιανύσματα* του πίνακα A .

Ορισμός 2.1.10 [8]

Έστω A πραγματικός συμμετρικός πίνακας.

- εάν $x^T A x > 0 \forall x \neq 0$ τότε ο A ονομάζεται θετικά ορισμένος.
- εάν $x^T A x \geq 0 \forall x \neq 0$ τότε ο A ονομάζεται θετικά ημί-ορισμένος.
- εάν $x^T A x < 0 \forall x \neq 0$ τότε ο A ονομάζεται αρνητικά ορισμένος.
- εάν $x^T A x \leq 0 \forall x \neq 0$ τότε ο A ονομάζεται αρνητικά ημί-ορισμένος.

Ορισμός 2.1.11 [8]

Το άθροισμα των στοιχείων της κυρία διαγωνίου ενός τετραγωνικού πίνακα A ονομάζεται ίχνος (ή **bold** ή tr) του πίνακα και συμβολίζεται με $\text{Trace}(A)$.

Ορισμός 2.1.12 [8]

Δύο τετράγωνοι πίνακες $A, B \in R^{n \times n}$ ονομάζονται όμοιοι, αν και μόνο αν, υπάρχει αντιστρέψιμος πίνακας $P \in R^{n \times n}$ τέτοιος ώστε:

$$B = P^{-1} A P$$

Κάθε τετράγωνος πίνακας A αποδεικνύεται ότι είναι όμοιος με ένα πίνακα της μορφής:

$$J = \begin{bmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & J_n \end{bmatrix}$$

όπου τα J_i , $i=1,2,\dots,n$ αντιστοιχούν στις ιδιοτιμές λ_i του A και έχουν τη παρακάτω μορφή:

$$J_i = \begin{bmatrix} J_{i1} & 0 & \cdots & 0 \\ 0 & J_{i2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & J_{im_i} \end{bmatrix} \in F^{\mu_i \times \mu_i}, i = 1, 2, \dots, n$$

όπου μ_i η αλγεβρική πολλαπλότητα της ιδιοτιμής και m_i η γεωμετρική της πολλαπλότητα. Οι πίνακες J_{ij} ονομάζονται Jordan Blocks και είναι της μορφής:

$$J_{ij} = \begin{bmatrix} \lambda_i & 1 & \cdots & 0 \\ 0 & \lambda_i & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 0 & \cdots & 0 & \lambda_i \end{bmatrix} \in F^{p_{ij} \times p_{ij}}$$

όπου $\mu_i = \sum_{j=1}^{m_i} p_{ij}$. Ο πίνακας J , ονομάζεται κανονική μορφή Jordan, είναι μοναδική

μέχρι αναδιάταξης των Jordan Blocks για κάθε κλάση ισοδυναμίας που ορίζει η ομοιότητα πινάκων και χαρακτηρίζει πλήρως τη δομή των πινάκων. Η κανονικά μορφή Jordan αποκαλύπτει πλήρως τη φασματική δομή του πίνακα A .

Ορισμός 2.1.13 [17]

Ορίζω ως Drazin γενικευμένο αντίστροφο ενός τετράγωνου πίνακα $A \in R^{m \times m}$, το μοναδικό πίνακα που θα συμβολίζεται $A^D \in R^{m \times m}$ που ικανοποιεί τις παρακάτω συνθήκες:

$$A^{k+1} A^D = A^k \text{ για } k := \text{ind}(A) := \min(k \in N^+ : \text{rank}(A^k) = \text{rank}(A^{k+1})),$$

$$A^D A A^D = A^D$$

$$A A^D = A^D A$$

Στην ειδική περίπτωση όπου ο πίνακας A είναι αντιστρέψιμος, ο Drazin γενικευμένος αντίστροφος συμπίπτει με το κλασσικό αντίστροφο του A δηλαδή $A^D=A^{-1}$.

2.2 Πολυώνυμο

Ένα πολυώνυμο με συντελεστές από το σώμα F θα γράφεται στη μορφή:

$$f(s) = \alpha_m s^m + \alpha_{m-1} s^{m-1} + \dots + \alpha_1 s + \alpha_0 \quad (2.2)$$

με $\alpha_0, \alpha_1, \dots, \alpha_m \in F$

Ο βαθμός του πολυωνύμου είναι ο μεγαλύτερος εκθέτης του πολυωνύμου. Συμβολίζεται με $\deg f = m$. Θα λέμε ότι το πολυώνυμο είναι *σταθερό ή μηδενικού βαθμού* αν ισχύει $f(s) = \alpha_0$. Το μηδενικό πολυώνυμο δεν έχει βαθμό και είναι το $f(s) = 0$.

Ορισμός 2.2.1 [6]

Ένα πολυώνυμο θα λέγεται *κανονικό* αν ο συντελεστής της μεγαλύτερης δύναμης είναι 1.

Ορισμός 2.2.2 [6]

Θα λέμε ότι το πολυώνυμο f λέγεται πολυώνυμο μηδενιστής του πίνακα $A \in M_n(F)$ αν ισχύει η σχέση $f(A) = 0$.

Θεώρημα Cayley-Hamilton [8]

Κάθε τετραγωνικός πίνακας $A \in M_n(F)$ ικανοποιεί τη χαρακτηριστική εξίσωση. Δηλαδή $c(A) = 0$ όπου $c(\lambda) = |A - \lambda I_n|$ το χαρακτηριστικό πολυώνυμο του A .

Ορισμός 2.2.3 [8]

Ένα ιδιοδιάνυσμα ενός τετραγωνικού πίνακα A είναι ένα μη μηδενικό διάνυσμα u που όταν πολλαπλασιαστεί με τον A , ισούται με το αρχικό διάνυσμα, πολλαπλασιασμένο με έναν αριθμό λ έτσι ώστε $Au = \lambda u$.

Ορισμός 2.2.4 [8]

Εάν $A \in M_n(\mathbb{R})$ τότε, ονομάζεται *ιδιοτιμή (eigenvalue)* του πίνακα A κάθε ρίζα του χαρακτηριστικού πολυωνύμου, δηλαδή κάθε πραγματικός ή μιγαδικός αριθμός λ_i για τον οποίο ισχύει $c(\lambda_i) = 0$. Επειδή το χαρακτηριστικό πολυώνυμο ενός πίνακα είναι n -

οστού βαθμού, θα έχει n ρίζες στο σώμα των μιγαδικών αριθμών. Τις ρίζες αυτές τις συμβολίζουμε με $\lambda_1, \lambda_2, \dots, \lambda_n$.

2.3 Πολυωνυμικοί Πίνακες

Ένας πίνακας του οποίου τα στοιχεία είναι πολυώνυμα με συντελεστές πάνω στο σώμα F θα λέγεται *πολυωνυμικός πίνακας* στο σώμα F .

$$A(s) = \left[\alpha_{ij}(s) \right]_{\substack{i=1,2,\dots,n \\ j=1,2,\dots,m}} = \begin{bmatrix} \alpha_{11}(s) & \cdots & \alpha_{1m}(s) \\ \vdots & \ddots & \vdots \\ \alpha_{n1}(s) & \cdots & \alpha_{nm}(s) \end{bmatrix}$$

Στα πλαίσια αυτής της εργασίας θα ασχοληθούμε με χρονικά μεταβαλλόμενους πίνακες όπου η μεταβλητή s αντικαθίσταται από τη μεταβλητή $t \geq 0$. Επίσης θα ασχοληθούμε και με πίνακες που το κάθε στοιχείο μπορεί να είναι γενικά οποιαδήποτε συνάρτηση (χρονικά μεταβαλλόμενη).

Η διάσταση του πολυωνυμικού πίνακα θα είναι $n \times m$. Το σύνολο των πολυωνυμικών πινάκων με διάσταση $n \times m$ πάνω στο σώμα F θα συμβολίζεται με $F[s]^{n \times m}$.

Ένας πολυωνυμικός πίνακας $A(s)$ μπορεί να γραφεί και ως

$$A(s) = A_n s^n + A_{n-1} s^{n-1} + \cdots + A_1 s + A_0$$

όπου, ο αριθμός n , θα είναι ο *βαθμός* του $A(s)$ και θα συμβολίζεται με $\deg A(s) = n$ δεδομένου ότι $A_0 \neq 0$.

Ορισμός 2.3.1 [6]

Ένας τετραγωνικός πολυωνυμικός πίνακας $A(s) \in R[s]^{p \times p}$ θα λέγεται *ομαλός (regular)* αν ισχύει $\text{rank}_R A_0 = p$ ή $\det A(s) \neq 0$.

Ορισμός 2.3.2 [6]

Ένας πολυωνυμικός πίνακας $U(s) \in R[s]^{p \times p}$ θα λέγεται *αντιστρέψιμος (unimodular)* (η $R[s]$ αντιστρέψιμος) αν υπάρχει ένας $\hat{U}(s) \in R[s]^{p \times p}$ τέτοιος ώστε $U(s)\hat{U}(s) = I_p$.

2.4 Νόρμες πινάκων και διανυσμάτων

Ορισμός 2.4 [8]

Έστω V ο διανυσματικός χώρος των πραγματικών συναρτήσεων, πάνω στο σώμα K . Ορίζουμε στον V την απεικόνιση $\|\cdot\|:V\rightarrow\mathbb{R}^+$ η οποία ονομάζεται νόρμα (norm) επί του V , με τις εξής ιδιότητες:

Για κάθε $\lambda\in K$ και $x(t),y(t)\in V$

$$\bullet \|x(t)\|=0 \text{ αν και μόνο αν } x(t)=0 \text{ για κάθε } t\in[a,b]$$

$$\bullet \|\lambda x(t)\|=\lambda\|x(t)\|$$

$$\bullet \|x(t)+y(t)\|\leq\|x(t)+y(t)\|$$

Σε έναν νορμικό γραμμικό χώρο $(V,\|\cdot\|)$ ορίζουμε ως απόσταση δύο στοιχείων του x_1 και x_2 τη νόρμα $\|x_1-x_2\|$. Έτσι ο νορμικός χώρος γίνεται μετρικός χώρος, δηλαδή ένα χώρος εφοδιασμένος με μία μετρική $d:V\times V\rightarrow\mathbb{R}^+$ που ικανοποιεί τις παρακάτω ιδιότητες.

$$\bullet d(x,y)=0\iff x=y$$

$$\bullet d(x,y)=d(y,x) \forall x,y\in V$$

$$\bullet d(x,z)\leq d(x,y)+d(y,z) \forall x,y,z\in V$$

Η συνάρτηση $\|\cdot\|_\infty:C[a,b]\rightarrow\mathbb{R}$, $\|x(t)\|_\infty=\sup\{|x(t)|\}$, $t\in[a,b]$ αποτελεί νόρμα του διανυσματικού χώρου $C[a,b]$.

Άλλες γνωστές νόρμες στο χώρο $C[a,b]$ είναι οι εξής:

$$\|\cdot\|_1:C[a,b]\rightarrow\mathbb{R}^+, \|x(t)\|_1 = \int_a^b |x(t)| dt$$

$$\|\cdot\|_2:C[a,b]\rightarrow\mathbb{R}^+, \|x(t)\|_2 = \sqrt{\int_a^b x^2(t) dt} \text{ (L}_2 \text{ νόρμα)}$$

$$\|\cdot\|_p:C[a,b]\rightarrow\mathbb{R}^+, \|x(t)\|_p = \left(\int_a^b |x(t)|^p dt \right)^{1/p}$$

Ένας επίσης σημαντικός χώρος είναι ο χώρος $C^1[a,b]$ των συνεχών συναρτήσεων στο κλειστό διάστημα $[a,b]$, που επιπλέον έχουν και συνεχείς πρώτες παραγώγους στο ίδιο διάστημα. Ο χώρος αυτός είναι εφοδιασμένος με τη γνωστή πρόσθεση και το πολλαπλασιασμό με σταθερά, αλλά η νόρμα με την οποία είναι εφοδιασμένος είναι της μορφής:

$$\|\bullet\|: C^1[a,b] \rightarrow \mathbb{R}, \|x(t)\| = \sup_{t \in [a,b]} \{|x(t)|\} + \sup_{t \in [a,b]} \{|x'(t)|\}$$

Ορισμός 2.4.1 [8]

Έστω V ένας διανυσματικός χώρος των πραγματικών συναρτήσεων, πάνω στο σώμα K (\mathbb{R} ή \mathbb{C}). Ορίζουμε στον V την απεικόνιση $\|\bullet\|: V \rightarrow \mathbb{R}$ η οποία ονομάζεται διανυσματική νόρμα με τις εξής ιδιότητες:

Για κάθε x, y ισχύει:

$$\|x\| \geq 0$$

$$\|ax\| = |a| \|x\| \text{ για όλα τα } a \in K$$

$$\|x+y\| \leq \|x\| + \|y\|$$

$$\|x\| = 0 \text{ αν και μόνο αν } x=0$$

Εάν μία συνάρτηση ικανοποιεί τις τρεις πρώτες ιδιότητες αλλά κατ' ανάγκη και τη τέταρτη, τότε αυτή ονομάζεται διανυσματική ημι-norm.

Οι πιο γνωστές νόρμες διανύσματος είναι οι εξής:

1) Η L_2 norm στον C^n (Ευκλείδεια)

Αν $x \in C^n$ με (x_1, x_2, \dots, x_n) η L_2 norm ορίζεται ως εξής

$$\|x\|_2 = \sqrt{(|x_1|^2 + |x_2|^2 + \dots + |x_n|^2)}$$

2) Η L_1 norm ως εξής

$$\|x\| = |x_1| + |x_2| + \dots + |x_n|$$

3) Η L_∞ (max norm)

$$\|x\|_{\infty} = \max\{|x_1|, |x_2|, \dots, |x_n|\}$$

Ορισμός 2.4.2 [8]

Έστω ένας $C^{n \times m}$ διανυσματικός χώρος. Ορίζουμε στον $C^{n \times m}$ τη συνάρτηση $\|\bullet\|: C^{n \times m} \rightarrow \mathbb{R}$ η οποία ονομάζεται norm πίνακα με τις εξής ιδιότητες:

Για κάθε πίνακα $A, B \in C^{n \times m}$ ισχύει:

- 1) $\|A\| \geq 0$
- 2) $\|\alpha A\| = |\alpha| \|A\|$ για όλα τα $\alpha \in \mathbb{C}$
- 3) $\|A+B\| \leq \|A\| + \|B\|$
- 4) $\|AB\| \leq \|A\| \|B\|$
- 5) $\|A\| = 0$ αν και μόνο αν $A=0$

Εάν μία συνάρτηση ικανοποιεί τις ιδιότητες (1), (2), (3) και (5) τότε αυτή ονομάζεται γενικευμένη norm πίνακα.

Οι πιο γνωστές νόρμες πινάκων είναι:

- 1) Η L_1 norm στον $C^{n \times m}$ που ορίζεται με το παρακάτω τύπο:

$$\|A\|_1 = \sum_{i=1}^n \sum_{j=1}^m |\alpha_{ij}|$$

- 2) Η L_2 norm στον $C^{n \times m}$ που ορίζεται με το παρακάτω τύπο:

$$\|A\|_2 = \left(\sum_{i=1}^m \sum_{j=1}^n |\alpha_{ij}|^2 \right)^{\frac{1}{2}}$$

- 3) Η L_{∞} norm στον $C^{n \times m}$ που ορίζεται με το παρακάτω τύπο:

$$\|A\|_{\infty} = \max_{i,j} |\alpha_{ij}|$$

- 4) Η Frobenius norm με τύπο:

$$\|A\|_F^2 = \sum_{i=1}^m \sum_{j=1}^n |\alpha_{ij}|^2 = \text{Tr}(A^T A)$$

Ορισμός 2.4.3 [10]

Το γινόμενο $\|A\| \cdot \|A^{-1}\| = \|I\|$ ορίζεται σαν *αριθμός κατάστασης* και συμβολίζεται με $\text{Cond}(A) = K(A)$ του ομαλού πίνακα A ως προς τη νόρμα $\|\cdot\|$. Είναι γνωστό ότι για οποιονδήποτε ομαλό πίνακα A ισχύει $A \cdot A^{-1} = I_n$ και $\|A\| \cdot \|A^{-1}\| = \|I\| = 1$. Επίσης είναι γνωστό ότι $\|A \cdot A^{-1}\| \leq \|A\| \cdot \|A^{-1}\|$.

Ο αριθμός κατάστασης εξαρτάται από τη νόρμα στην οποία εργαζόμαστε.

Ορισμός 2.4.3.1 [10]

Εάν ο αριθμός κατάστασης είναι πολύ μεγάλος τότε λέμε ότι το σύστημα η ο πίνακας είναι σε *αρρωστημένη κατάσταση*. Εάν ο πίνακας είμαι μη αντιστρέψιμος τότε ο αριθμός κατάστασης είναι το άπειρο.

Όταν ο πίνακας A είναι σε *αρρωστημένη κατάσταση*, τότε μικρές αλλαγές (διαταραχές) στα στοιχεία A και b (σε ένα γραμμικός σύστημα) μπορούν να παράγουν μεγάλες αλλαγές στη προσέγγιση της λύσης x του συστήματος. Διαφορετικά, όταν ο A είναι σε *καλή κατάσταση* συμβαίνει το αντίθετο. Εάν εργαζόμαστε με την L_2 norm τότε:

$$K(A) = \|A\|_2 \|A^{-1}\|_2 = \frac{\max |\lambda_i(A)|}{\min |\lambda_i(A)|}$$

Ορισμός 2.4.3.2 [21]

Ένα σύστημα διαφορικών $\dot{x} = f(x(t))$, $t > 0$ είναι *ασυμπτωτικά ευσταθές* εάν για κάθε φραγμένη είσοδο ή έξοδος είναι επίσης φραγμένη.

Ορισμός 2.4.3.2.1 [21]

Ένα σύστημα $\dot{x} = f(x(t))$, $t > 0$ είναι *οριακά ευσταθές* αν και μόνο αν η φυσική του απόκριση παραμένει φραγμένη για $t \rightarrow \infty$.

Ορισμός 2.4.3.2.2 [21]

Η ιδιότητα που μετριέται από το σχετικό πραγματικό μέρος κάθε ρίζας ή ζεύγους ριζών της χαρακτηριστική εξίσωσης καλείται *σχετική ευστάθεια*.

Ορισμός 2.4.3.3 [21]

Μία κατάσταση ισορροπίας \underline{x}_0 ενός αυτόνομου δυναμικού συστήματος καλείται ευσταθής κατά Lyapunov εάν για κάθε $\varepsilon > 0$ υπάρχει $\delta(\varepsilon) > 0$ τέτοιο ώστε $\|\underline{x}(0) - \underline{x}_0\| < \delta$ να συνεπάγεται $\|\underline{x}(t) - \underline{x}_0\| < \varepsilon$ για κάθε $t > 0$.

Από φυσική άποψη ο ορισμός (2.4.3.3) δηλώνει ότι όταν η κατάσταση ισορροπίας του συστήματος διαταραχθεί τη μηδενική χρονική στιγμή, η κίνηση την οποία θα εκτελέσει το διάνυσμα κατάστασης στο χώρο καταστάσεων θα παραμείνει μέσα σε μία περιοχή της κατάστασης ισορροπίας, το δε μέγεθος της περιοχής εξαρτάται από το μέγεθος της διαταραχής.

Ορισμός 2.4.4 [21]

Έστω $U \subset \mathbb{R}^n$ μία περιοχή γύρω από την αρχή των αξόνων. Μία συνάρτηση $V: U \rightarrow \mathbb{R}$ θα ονομάζεται συνάρτηση Lyapunov (για ένα σύστημα $\dot{x} = f(x(t))$, $t > 0$) εάν για κάθε $x \in U$ ικανοποιούνται οι επόμενες προϋποθέσεις:

- i. Η V είναι C^1 , δηλαδή η V είναι συνεχής και οι μερικές παράγωγοι πρώτης τάξης ορίζονται και είναι συνεχείς.
- ii. $V(0) = 0$
- iii. $V(x) > 0$ για κάθε $x \neq 0$
- iv. $\dot{V}(x(t)) = \frac{dV}{dt} = \frac{dV}{dx} \cdot \frac{dx}{dt} < 0$, $\forall x \neq 0$

Θεώρημα 2.4.5 (Θεώρημα Ευστάθειας Lyapunov) [21]

Έστω το σύστημα διαφορικών εξισώσεων $\dot{x} = f(x(t))$, $t > 0$. Εάν για το σύστημα αυτό μπορεί να βρεθεί μία συνάρτηση Lyapunov $V(x(t))$, τότε το σύστημα είναι *ασυμπτωτικά ευσταθές*.

Θεώρημα 2.4.6 [21]

Έστω $\Omega \subset \mathbb{R}^n$ ανοιχτό, $\vec{a} \in \Omega$ και $f: \Omega \rightarrow \mathbb{R}$ είναι διαφορίσιμη στο \vec{a} . Υποθέτουμε ότι το \vec{a} είναι σημείο τοπικού ακροτάτου για την f τότε $\nabla f(\vec{a}) = \vec{0}$.

Στη περίπτωση της μίας μεταβλητής όταν παρουσιάζει ακρότατο τότε:

$$\frac{df(a)}{dt} = 0$$

Πρόταση 2.4.7 [2]

Έστω ο ομαλός πίνακας A . Εάν ο πίνακας $A^T A$ είναι θετικά ορισμένος τότε ο αντίστροφος $(A^T A)^{-1}$ υπάρχει.

Μέθοδος Μέγιστης Κλίσης [21]

Μία από τις πιο παλιές και γνωστές μεθόδους για εύρεση ελαχίστου μίας συνάρτησης πολλών μεταβλητών, είναι η μέθοδος της μέγιστης κλίσης.

Έστω f με συνεχείς πρώτες παραγώγους στο $\Omega \subset \mathbb{R}^n$. Έστω $\nabla f(x)$ το διάνυσμα κλίσης της f (ένα διάνυσμα-γραμμή n διαστάσεων). Για λόγους ευκολίας ορίζουμε το διάνυσμα στήλη n διαστάσεων $g(x) = \nabla f(x)^T$. Για να μην υπάρχει ασάφεια η προηγούμενη σχέση κάποιες φορές αντί για x τοποθετούμε στη θέση του x_k , δηλαδή $g(x_k) = \nabla f(x_k)^T$.

Ο αλγόριθμος ορίζεται σύμφωνα με τη παρακάτω επαναληπτική σχέση:

$$x_{k+1} = x_k - \alpha_k g_k$$

όπου α_k είναι μία μη-αρνητική σταθερά. Με άλλα λόγια ξεκινώντας από το σημείο x_k ακολουθώντας σα <<διεύθυνση>> την αρνητική τιμή $-g_k$ σύμφωνα με τη παραπάνω επαναληπτική σχέση καταλήγουμε στο σημείο x_{k+1} που αποτελεί θέση ελαχίστου (μετά από κάποιο αριθμό επαναλήψεων).

Η συγκεκριμένη μέθοδος χρησιμοποιείται για την επίλυση γραμμικών συστημάτων της μορφής $Ax=b$. Αναδιατυπώνεται σα τετραγωνικό πρόβλημα ελαχιστοποίησης (όπως για παράδειγμα χρησιμοποιώντας τη μέθοδο των ελαχίστων τετραγώνων) της

συνάρτησης $F(x)=\|Ax-b\|^2$ με $\nabla F(x)=2A^T(Ax-b)$ η ταχύτητα σύγκλισης της μεθόδου εξαρτάται από τη μέγιστη και ελάχιστη ιδιοτιμή του πίνακα A .

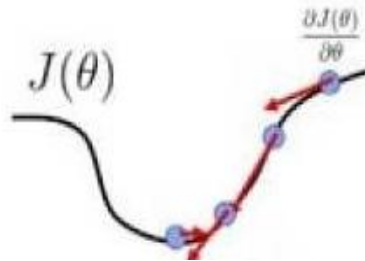
Θα πρέπει να τονιστεί ότι η συγκεκριμένη μέθοδος αποτελεί τη βάση για τη δημιουργία των GNN μοντέλων και για τη μέθοδο Osowski που θα αναλύσουμε στη συνέχεια. Παρακάτω βλέπουμε την αντιστοιχία της μεθόδου για τη δημιουργία των GNN δικτύων:

$$\left\{ \begin{array}{l} x_{k+1} - x_k \rightarrow \dot{X}(t) \\ a_k \rightarrow \gamma \\ \nabla f(x_k)^T \rightarrow \frac{dE(X(t), t)}{dX} \\ x_{k+1} = x_k - a_k \nabla f(x_k)^T \rightarrow \dot{X}(t) = -\gamma \frac{dE}{dX} \end{array} \right.$$

Αντιστοιχία μεθόδου μέγιστης κλίσης με μοντέλο GNN δικτύου.

Η παραπάνω μέθοδος δημιούργησε τον τρόπο σκέψης για τα GNN δίκτυα που θα δούμε στο δεύτερο κεφάλαιο.

• Initialization	Initialize θ
• Step size	Do {
- Can change as a function of iteration	$\theta \leftarrow \theta - \alpha \nabla_{\theta} J(\theta)$
• Gradient direction	} while ($\alpha \ \nabla J\ > \epsilon$)
• Stopping condition	



Απεικόνιση της μεθόδου [21]

Κεφάλαιο 3

Εισαγωγή Στα ZNN και GNN Μοντέλα.

Εισαγωγικά

Η αντιστροφή πινάκων χρησιμοποιείται ευρέως σε ποικιλία εφαρμογών στη ρομποτική, θεωρία ελέγχου και ανάλυση σημάτων. Σε πολλά προβλήματα μεγάλης κλίμακας συχνά χρειάζεται ο υπολογισμός του αντιστρόφου ενός πίνακα για την παρακολούθηση και των έλεγχου δυναμικών συστημάτων.

3.1 Διατύπωση Προβλήματος

Έστω ένα χρονικά μεταβαλλόμενος $n \times n$ πίνακας $A(t)$. Ψάχνουμε να βρούμε έναν $n \times n$ πίνακα $X(t)$ ο οποίος να ικανοποιεί τη παρακάτω εξίσωση πινάκων:

$$A(t)X(t) - I = 0, \forall t \in [0, \infty) \quad (3.1)$$

όπου I είναι ο $n \times n$ μοναδιαίος πίνακας. Χωρίς περιορισμό της γενικότητας γνωρίζουμε το πίνακα $A(t)$ καθώς και τη παράγωγο $\dot{A}(t)$. Στα πλαίσια αυτού του κεφαλαίου θα ασχοληθούμε με τη μέθοδο του Yunong Zhang και θα αναλύσουμε τα αποτελέσματα και τη σύγκλιση της συγκεκριμένης μεθόδου.

Πρώτα θα πρέπει να ορίσουμε τη παρακάτω συνθήκη αντιστρεψιμότητας.

Συνθήκη Αντιστρεψιμότητας (ZNN μοντέλο)

Υπάρχει θετικός πραγματικός αριθμός a τέτοιος ώστε:

$$\min_{\forall i \in \{1, 2, \dots, n\}} |\lambda_i(A(t))| \geq a, \forall t \geq 0 \quad (3.1.1)$$

όπου $\lambda_i(\cdot)$ υποδηλώνει την ιδιοτιμή της i του πίνακα $A \in R^{n \times n}$. Εάν ισχύει η παραπάνω συνθήκη αντιστρεψιμότητας τότε υπάρχει μοναδική λύση της εξίσωσης πινάκων (3.1).

Έστω $\|A\|_F$ η Frobenius νόρμα με:

$$\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2} \quad (3.1.2)$$

Τότε η συνθήκη αντιστρεψιμότητας οδηγεί στο ακόλουθο λήμμα που αφορά κατά πόσο φράζονται οι νόρμες $\|A(t)\|_F$ και $\|A^{-1}(t)\|_F$.

Λήμμα 3.1 [13]

Εάν ο πίνακας $A(t)$ ικανοποιεί τη συνθήκη αντιστρεψιμότητας και η νόρμα του είναι ομοιόμορφα άνω φραγμένη από έναν αριθμό β (δηλαδή $\|A(t)\|_F \leq \beta, \forall t \geq 0$) τότε:

$$\|A^{-1}(t)\|_F \leq \phi(\alpha, \beta, n) \quad (3.1.3)$$

όπου

$$\phi(\alpha, \beta, n) = \sum_{i=0}^{n-2} C_n^i \frac{\beta^{n-i-1}}{\alpha^{n-i}} + \frac{n^{3/2}}{\alpha} \quad \text{και} \quad C_n^i = \frac{n!}{i!(n-i)!} \quad (3.1.4)$$

Απόδειξη

Από το θεώρημα Cayley-Hamilton ο πίνακας A ικανοποιεί το χαρακτηριστικό του πολυώνυμο οπότε:

$$A^n + \omega_{n-1}A^{n-1} + \dots + \omega_{n-i}A^{n-i} + \dots + \omega_0I = 0 \quad (3.1.5)$$

όπου οι συντελεστές ω_i ορίζονται σε σχέση με τις ιδιοτιμές λ_i του πίνακα σύμφωνα με τις παρακάτω σχέσεις:

$$\left\{ \begin{array}{l} \omega_0 = (-\lambda_1)(-\lambda_2)\dots(-\lambda_n) \\ \dots \\ \omega_{n-i} = \sum_{j_1 > \dots > j_i}^n (-\lambda_{j_1})(-\lambda_{j_2})\dots(-\lambda_{j_i}) \\ \dots \\ \omega_{n-1} = \sum_{j_1=1}^n (-\lambda_{j_1}) \end{array} \right.$$

Ειδικά $\omega_0 \neq 0$ και $|\omega_0| \geq \alpha^n$ (λογικό εφόσον ισχύει η συνθήκη αντιστρεψιμότητας θα έχουμε $(|\omega_0| = |-\lambda_1| |-\lambda_2| \dots |-\lambda_n| \geq \alpha \cdot \alpha \cdot \dots \cdot \alpha = \alpha^n)$). Στη συνέχεια έχουμε ότι

$$A^{-1} = -(\nu_n A^{n-1} + \nu_{n-1} A^{n-2} + \dots + \nu_{n-i} A^{n-i-1} + \dots + \nu_1 I) \quad (3.1.6)$$

όπου οι συντελεστές ν_i δίνονται από τους παρακάτω τύπους.

$$\left\{ \begin{array}{l} v_n = \frac{1}{\omega_0} = \frac{1}{(-\lambda_1)(-\lambda_2)\cdots(-\lambda_n)} \\ v_{n-1} = \frac{\omega_{n-1}}{\omega_0} = \sum_{j_1=1}^n \frac{1}{(-\lambda_1)\cdots(-\lambda_{j_1-1})(-\lambda_{j_1+1})\cdots(-\lambda_n)} \\ \dots \\ v_{n-i} = \frac{\omega_{n-i}}{\omega_0} = \sum_{j_1 > \dots > j_i > j_{i-1}=1}^n \frac{(-\lambda_{j_1})(-\lambda_{j_2})\cdots(-\lambda_{j_i})}{(-\lambda_1)(-\lambda_2)\cdots(-\lambda_n)} \\ \dots \\ v_1 = \frac{\omega_1}{\omega_0} = \sum_{j_1=1}^n \frac{1}{(-\lambda_{j_1})} \end{array} \right.$$

Από τη συνθήκη αντιστρεψιμότητας και από τη σχέση (3.1.4) για κάθε $i \in \{1, 2, \dots, n\}$ έχουμε:

$$\left\{ \begin{array}{l} |v_n| \leq \frac{1}{\alpha^n} \\ |v_{n-1}| \leq \frac{n}{\alpha^{n-1}} \\ \dots \\ |v_{n-i}| \leq \frac{C_n^i}{\alpha^{n-i}} \\ \dots \\ |v_1| \leq \frac{n}{\alpha^n} \end{array} \right.$$

Από τις παραπάνω ανισότητες και τη σχέση (3.1.6) έχουμε το παρακάτω αποτέλεσμα:

$$\begin{aligned} \|A^{-1}\|_F &\leq |v_n| \|A\|_F^{n-1} + |v_{n-1}| \|A\|_F^{n-2} + \dots + |v_{n-i}| \|A\|_F^{n-i-1} + \dots + |v_1| \|I\|_F \\ &\leq \frac{\beta^{n-1}}{\alpha^n} + \frac{n\beta^{n-2}}{\alpha^{n-1}} + \dots + \frac{C_n^i \beta^{n-i-1}}{\alpha^{n-i}} + \dots + \frac{n\sqrt{n}}{a} = \frac{\sum_{i=0}^{n-2} C_n^i \beta^{n-i-1}}{\alpha^{n-i}} + \frac{n^{3/2}}{\alpha} \end{aligned}$$

Το οποίο μας οδηγεί στην απόδειξη του παραπάνω λήμματος.

Παρατήρηση

Η συνθήκη αντιστρεψιμότητας μεταφράζεται στο γεγονός ότι όλες οι ιδιοτιμές του πίνακα $A(t)$ θα πρέπει να είναι διάφορες από τη τιμή 0 για κάθε $t \geq 0$ και όπως θα

δούμε στη συνέχεια παίζει σημαντικό ρόλο, ιδιαίτερα όσον αφορά την αντιστροφή χρονικά μεταβαλλόμενων πολυωνυμικών πινάκων.

3.2 Γενικό ZNN (Zhang Neural Network) μοντέλο

Αξίζει να σημειωθεί ότι η συνθήκη αντιστρεψιμότητας, η παράμετρος α και το λήμμα 1 χρησιμοποιούνται για αναλυτικό σκοπό και δεν υπάρχει ανάγκη να γνωρίζουμε συγκεκριμένη τιμή για τη παράμετρο α . Αντ' αυτού, σε πρακτική εφαρμογή, για να ελέγξουμε τη συνθήκη αντιστρεψιμότητας μπορούμε απλά να παρατηρήσουμε εάν η τιμή $\|AX-I\|_F$ γίνεται πολύ μικρή (δηλαδή συγκλίνει στη τιμή 0) από την έναρξη λειτουργίας του δικτύου. Στη βιβλιογραφία πολλά νευρωνικά δίκτυα έχουν κατασκευαστεί για την αντιστροφή ενός χρονικά μεταβαλλόμενου πίνακα, στα πλαίσια αυτού του κεφαλαίου θα ασχοληθούμε, όπως προαναφέραμε, με τη μέθοδο του *Yunong Zhang (ZNN Neural Network)*.

Όσον αφορά τη διαδικασία αντιστροφής του πίνακα, η ακόλουθη συνάρτηση πίνακα σφάλματος ορίζεται σύμφωνα με το παρακάτω τύπο:

$$E(X(t), t) := A(t)X(t) - I \quad (3.2)$$

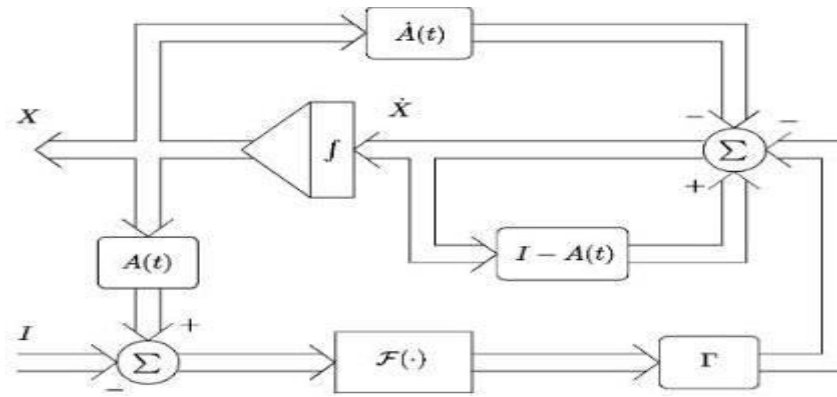
Η παράγωγος της συνάρτησης σφάλματος $\dot{E}(X(t), t)$ πρέπει να κατασκευαστεί έτσι ώστε κάθε είσοδος $e_{ij}(t)$, $i, j=1, 2, \dots, n$ της $E(X(t), t)$ να συγκλίνει στο 0. Ειδικά, η $\dot{E}(X(t), t)$ μπορεί να περιγραφεί με τη παρακάτω γενική μορφή.

$$\frac{dE(X(t), t)}{dt} = -\Gamma F(E(X(t), t)) \quad (3.2.1)$$

Όπου Γ είναι ένας θετικά ορισμένος πίνακας που χρησιμοποιείται για τη σύγκλιση της μεθόδου και $F(\cdot): R^{n \times n} \rightarrow R^{n \times n}$ είναι μία πολυμεταβλητή συνάρτηση ενεργοποίησης $n \times n$ διαστάσεων. Συνδυάζοντας τις σχέσεις (3.2) και (3.2.1) οδηγούμαστε στο παρακάτω νευρωνικό μοντέλο:

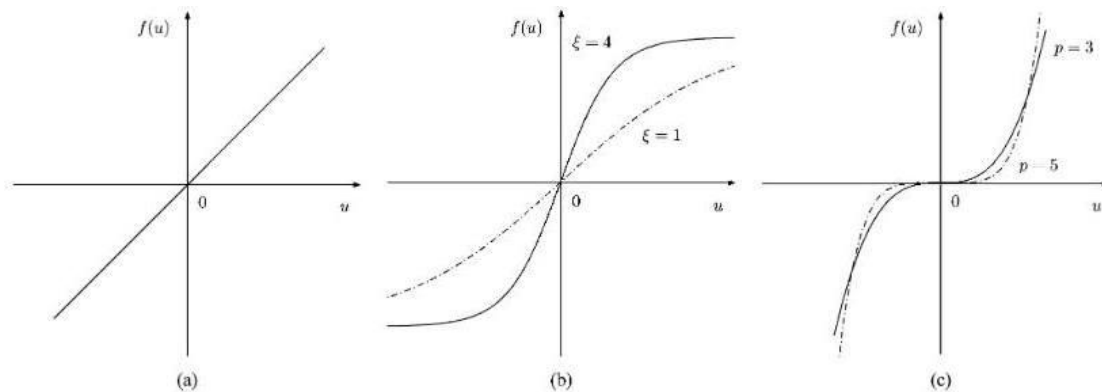
$$A(t)\dot{X}(t) = -\dot{A}(t)X(t) - \Gamma F(A(t)X(t) - I) \quad (3.2.2)$$

όπου $X(t)$, ξεκινά από μία αρχική κατάσταση $X(0)=X_0$, και συγκλίνει στη θεωρητική λύση $X^*(t)$ της σχέσης (3.1). Στο παρακάτω διάγραμμα μπορούμε να παρατηρήσουμε τη λειτουργία της σχέσης (3.2.2) στη γενική της μορφή.



Διάγραμμα για τη μέθοδο αντιστροφής του χρονικά μεταβαλλόμενου πίνακα [13]

Διαφορετικές επιλογές για το πίνακα Γ και τη συνάρτηση F θα μας οδηγήσουν σε διαφορετικά αποτελέσματα. Γενικά οποιαδήποτε αύξουσα συνάρτηση $f(\cdot)$, που αποτελεί την ij -στη θέση του πίνακα F , μπορεί να χρησιμοποιηθεί για τη κατασκευή του νευρωνικού κυκλώματος. Στα πλαίσια αυτού του κεφαλαίου χρησιμοποιούμε τη γραμμική, τη συνάρτηση δύναμης και τη διπολική-σιγμοειδή των οποίων οι γραφικές παραστάσεις φαίνονται παρακάτω.



Συναρτήσεις ενεργοποίησης $f(\cdot)$ για το νευρωνικό μοντέλο. (a) γραμμική συνάρτηση. (b) διπολική σιγμοειδής συνάρτηση. (c) συνάρτηση δύναμης. [13]

3.3 Σύγκλιση ZNN μοντέλου

Για να διατηρήσουμε τη διαφορική εξίσωση (3.2.2) καλά ορισμένη θα πρέπει να είναι ευσταθή και ο πίνακας Γ να είναι θετικά ορισμένος. Για αυτό το λόγο και για δική μας ευκολία επιλέγουμε $\Gamma = \gamma \cdot I$ όπου I είναι ο μοναδιαίος $n \times n$ πίνακας και γ ένας θετικός αριθμός (μεγάλος όσο μας το επιτρέπει το λογισμικό που εργαζόμαστε). Τα ακόλουθα θεωρητικά αποτελέσματα που αφορούν τη σύγκλιση της μεθόδου παρουσιάζονται παρακάτω.

Θεώρημα 3.3 [13]

Έστω ένας χρονικά μεταβαλλόμενος πίνακας $A(t)$ ο οποίος ικανοποιεί τη συνθήκη αντιστρεψιμότητας, εάν χρησιμοποιούμε ένα $n \times n$ πίνακα $F(\cdot)$, που να περιέχει στη κάθε θέση της μία περιττή και αύξουσα συνάρτηση $f(\cdot)$ (μίας μεταβλητής), τότε ο πίνακας κατάστασης $X(t)$ του νευρωνικού συστήματος (3.3.2) ξεκινά από μία αρχική κατάσταση X_0 και συγκλίνει στη θεωρητική λύση της εξίσωσης πινάκων (3.1). Επιπλέον για το νευρωνικό σύστημα (3.2.2) ισχύουν οι παρακάτω ιδιότητες:

- i) εκθετική σύγκλιση με ποσοστό γ εάν χρησιμοποιήσουμε τη γραμμική συνάρτηση ενεργοποίησης.
- ii) εκθετική σύγκλιση με ποσοστό $\xi\gamma/2$, για εύρος σφάλματος $e_{ij}(t) \in (-\ln 2/\xi, \infty)$, εάν χρησιμοποιήσουμε τη διπολική-σιγμοειδή.
- iii) γρηγορότερη και καλύτερη σύγκλιση σε σχέση με τις περιπτώσεις (i), (ii) για εύρος σφάλματος $|e_{ij}(t)| > 1$ εάν χρησιμοποιήσουμε τη συνάρτηση δύναμης.

Απόδειξη

Έστω

$$\tilde{X}(t) := X(t) - X^*(t)$$

η διαφορά μεταξύ της λύσης $X(t)$ που παράγεται από το νευρωνικό δίκτυο και της θεωρητικής λύσης $X^*(t)$ της (2.5). Εφόσον $X^*(t)$ αποτελεί τη θεωρητική λύση της σχέσης (3.2) έχουμε ότι $E(X^*(t), t) = 0$ οπότε κάνοντας παραγωγή στη καταλήγουμε στη παρακάτω ταυτότητα:

$$\dot{A}(t)X^*(t) + A(t)\dot{X}^*(t) = 0$$

Χρησιμοποιώντας τη παραπάνω ταυτότητα έχουμε ότι ο πίνακας $\tilde{X}(t)$ αποτελεί λύση του νευρωνικού συστήματος με αρχική κατάσταση $\tilde{X}(0) := X_0 - X^*(0)$. Οπότε καταλήγουμε στη σχέση:

$$A(t)\dot{\tilde{X}}(t) + \dot{A}(t)\tilde{X}(t) = -\gamma F(A(t)\tilde{X}(t)) \quad (3.3)$$

από τη σχέση $E(t) = A(t)\tilde{X}(t)$, η (3.3) μπορεί να ξαναγραφεί ως $\dot{E}(t) = -\gamma F(E(t))$ η οποία αποτελεί μία συμπαγή μορφή πίνακα του παρακάτω συστήματος n^2 εξισώσεων.

$$\dot{e}_{ij}(t) = -\gamma f(e_{ij}(t)), \forall i, j \in \{1, 2, \dots, n\} \quad (3.3.1)$$

Σαφώς, μπορούμε να ορίσουμε μία υποψήφια Lyapunov συνάρτηση, για το κάθε ij-υποσύστημα σύμφωνα με τη παρακάτω σχέση:

$$v_{ij} = \frac{e_{ij}^2}{2} \geq 0$$

Παραγωγίζοντας καταλήγουμε ότι

$$\frac{dv_{ij}}{dt} = e_{ij}\dot{e}_{ij} = -\gamma e_{ij}f(e_{ij}) \quad (3.3.2)$$

Συνεχίζοντας εφόσον έχουμε γνήσιες περιττές αύξουσες συναρτήσεις, σαν συνάρτηση ενεργοποίησης, θα ισχύει ότι $f(-u) = -f(u)$ οπότε για τη συνάρτηση f της σχέσης (3.3.2) θα ισχύουν οι παρακάτω περιπτώσεις:

$$f(u) = \begin{cases} > 0 & , \alpha \nu \quad u > 0 \\ = 0 & , \alpha \nu \quad u = 0 \\ < 0 & , \alpha \nu \quad u < 0 \end{cases}$$

το οποίο μας εγγυάται ότι η παράγωγος της συνάρτησης v_{ij} είναι αρνητική, σε κάθε περίπτωση (για παράδειγμα εάν $\dot{v}_{ij} < 0$ για κάθε $e_{ij} \neq 0$ και $\dot{v}_{ij} = 0$ για $e_{ij} = 0$). Από τον ορισμός (2.4.4) και το θεώρημα ευστάθειας (2.4.5) κατά Lyapunov [1] το μοντέλο (3.2.2) είναι ευσταθές και ο κάθε όρος $e_{ij}(t)$ συγκλίνει στο 0 για κάθε $i, j \in \{1, 2, \dots, n\}$ (όπως φαίνεται στην απόδειξη των ιδιοτήτων (i), (ii), (iii)). Έτσι από τη σχέση $E(t) = A(t)\tilde{X}(t)$ και τη (3.1.1) έχουμε ότι $\tilde{X}(t) \rightarrow 0 \in R^{n \times n}$ καθώς $t \rightarrow \infty$ οπότε ο πίνακας κατάστασης $X(t)$ συγκλίνει στο θεωρητικό αντίστροφο $X^*(t)$.

Παρατηρώντας το τύπο $E = A\tilde{X}$, τη σχέση (3.1.1) και το λήμμα (3.1) καταλήγουμε στη παρακάτω ανισότητα:

$$\|X(t) - X^*(t)\|_F \leq \|A^{-1}\|_F \|E\|_F \leq \phi \left(\sum_{i=1}^n \sum_{j=1}^n e_{ij}^2 \right)^{1/2} \leq n\phi \max_{1 \leq i, j \leq n} |e_{ij}|$$

πράγμα που συνεπάγεται ότι το σφάλμα υπολογισμού και η σύγκλιση του δικτύου μπορούν να εκτιμηθούν αρκεί να ξέρουμε το μέγιστο σφάλμα εισόδου.

Τώρα πρόκειται να αποδείξουμε τις επιπλέον ιδιότητες που αφορούν την εφαρμογή της κάθε συνάρτησης ενεργοποίησης $f(\cdot)$ χωριστά.

(α) Στην απλή περίπτωση που έχουμε τη γραμμική συνάρτηση ενεργοποίησης η (3.3.2) απλοποιείται στη :

$$\dot{e}_{ij}(t) = -\gamma e_{ij}(t)$$

Από θεωρία, η παραπάνω διαφορική εξίσωση έχει σαν γενική λύση τη:

$$e_{ij}(t) = e^{-\gamma t} e_{ij}(0) = \exp(-\gamma t) e_{ij}(0)$$

Έτσι παρατηρούμε ότι υπάρχει μία σταθερά $\zeta(\alpha, \beta, n, e_{ij}(0)) > 0$ τέτοια ώστε να έχουμε $\|X(t) - X^*(t)\|_F \leq \zeta \exp(-\gamma t)$. Αυτό σημαίνει ότι το νευρωνικό κύκλωμα (3.3.2) διαθέτει την εκθετική σύγκλιση με ποσοστό γ όταν έχουμε τη γραμμική συνάρτηση ενεργοποίησης.

(β) Όσον αφορά τη διπολική σιγμοειδή συνάρτηση έχει τύπο:

$$f(u) = \frac{(1 - e^{-\xi u})}{(1 + e^{-\xi u})}$$

Ορίζουμε τη παρακάτω σταθερά:

$$\eta = \frac{(e^{-\xi e_{ij}(0)} - 1)^2}{e^{-\xi e_{ij}(0)}}$$

Στη περίπτωση που έχουμε τη συγκεκριμένη συνάρτηση ενεργοποίησης η διαφορική εξίσωση (3.3.2) μας δίνει τη παρακάτω λύση:

$$z(t) = \frac{\eta e^{-\xi \gamma t}}{2} - \operatorname{sgn}(e_{ij}(0)) \sqrt{\left(1 + \frac{\eta e^{-\xi \gamma t}}{2}\right)^2 - 1} \quad (3.3.3)$$

Χρησιμοποιώντας τα αναπτύγματα Taylor

$$\ln(1+z) = z - \frac{z^2}{2} + \frac{z^3}{3} - + \dots$$

και

$$\frac{z}{(1-z)} = z + z^2 + z^3 + \dots \quad \text{για } |z| < 1$$

έχουμε

$$|e_{ij}(t)| \leq \frac{z}{\xi} + \frac{z^2}{(2\xi)} + \frac{z^3}{(3\xi)} + \dots \leq \frac{(|z| + |z|^2 + \dots)}{\xi} = \frac{|z|}{(\xi(1-|z|))}$$

Εφόσον $|z| < 1$, για το εύρος τιμών $e_{ij}(t)$ ορίζεται για $e_{ij}(t) > -\ln 2/\xi$. Από τη σχέση (3.3.3) έχουμε:

$$\begin{aligned} |z(t)| &\leq \frac{\eta e^{-\xi\gamma t}}{2} + \sqrt{\frac{\eta^2 e^{-2\xi\gamma t} + 4\eta e^{-\xi\gamma t}}{4}} \\ &\leq \frac{\eta}{2} e^{-\xi\gamma t} + \frac{\sqrt{\eta^2 + 4\eta}}{2} e^{-\frac{\xi\gamma t}{2}} \\ &\leq \frac{\eta + \sqrt{\eta^2 + 4\eta}}{2} e^{-\frac{\xi\gamma t}{2}} \end{aligned}$$

Όσον αφορά τη σύγκλιση του σφάλματος $e_{ij}(t)$ και του όρου $z(t)$ στο 0 υπάρχει $z_0 = |z(t)| < 1$ τέτοιο ώστε:

$$|e_{ij}(t)| \leq \frac{1}{\xi(1-z_0)} |z| \leq \frac{\eta + \sqrt{\eta^2 + 4\eta}}{2\xi(1-z_0)} e^{-\frac{\xi\gamma t}{2}}$$

Συνεπώς το σύστημα (3.3.2) συγκλίνει εκθετικά με ποσοστό $\xi\gamma/2$ στη περίπτωση της σιγμοειδής συνάρτησης ενεργοποίησης.

(γ) Στη περίπτωση που χρησιμοποιούμε τη συνάρτηση ενεργοποίησης $f(u) = u^p$ η (3.3.2) απλοποιείται και παίρνει τη παρακάτω μορφή:

$$\dot{e}_{ij} = -\gamma e_{ij}^p$$

Η οποία έχει σα γενική λύση:

$$e_{ij}(t) = e_{ij}(0) \{(p-1)e_{ij}^{p-1}(0)\gamma t + 1\}^{-\frac{1}{p-1}}$$

Ειδικά για $p=3$ η παραπάνω σχέση μπορεί να πάρει τη μορφή:

$$e_{ij}(t) = \frac{e_{ij}(0)}{\sqrt{2e_{ij}^2(0)\gamma t + 1}}$$

Καθώς $t \rightarrow \infty$ το σφάλμα $e_{ij}(t) \rightarrow 0$ (χρησιμοποιώντας της συνάρτηση Lyapunov v_{ij} και τη παράγωγο της όπως στη (3.3.2)). Για εύρος $|e_{ij}| \gg 1$, έχουμε $e_{ij}^{p+1} \gg e_{ij}^2 \gg |e_{ij}|$. Αυτό σημαίνει ότι χρησιμοποιώντας τη συνάρτηση δύναμης πετυχαίνουμε πολύ γρηγορότερη σύγκλιση για $|e_{ij}| > 1$ (σε σχέση με τις προηγούμενες περιπτώσεις).

Παρατήρηση 1

Από το παραπάνω θεώρημα προκύπτει ότι για να πετύχουμε καλύτερη σύγκλιση, ένα νευρωνικό δίκτυο υψηλής απόδοσης μπορεί να κατασκευαστεί αλλάζοντας τη συνάρτηση δύναμης σε σιγμοειδή ή γραμμική στα σημεία αλλαγής $|e_{ij}(t)|=1$, για $i,j \in \{1,2,\dots,n\}$. Για παράδειγμα μία τέτοια συνάρτηση είναι η δυναμική-σιγμοειδή με τύπο:

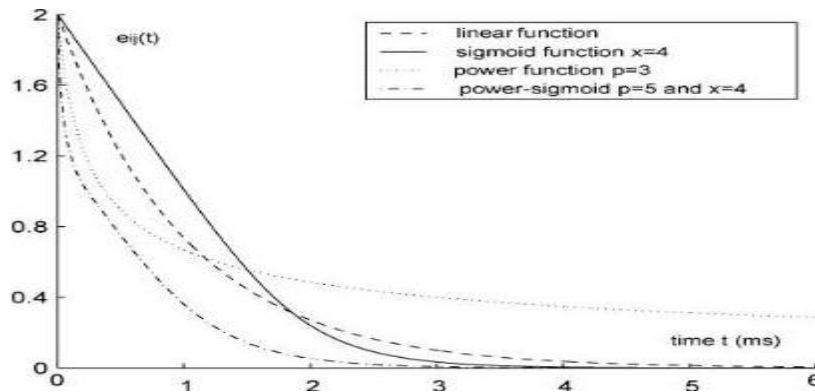
$$f(u) = \begin{cases} u^p, & \text{εάν } |u| \geq 1 \\ \frac{1+e^{-\xi}}{1-e^{-\xi}} \cdot \frac{1-e^{-\xi u}}{1+e^{-\xi u}}, & \text{αλλιώς} \end{cases} \quad (3.3.3)$$

Επιλέγοντας κατάλληλα της παραμέτρους ξ και p (p περιττός). Ένα επιπλέον πλεονέκτημα της σιγμοειδής σε σχέση με τη γραμμική (για το δεύτερο κλάδο) έχει να κάνει με την επιλογή της παραμέτρου ξ , η οποία λειτουργεί σαν πολλαπλασιαστής του εκθετικού ρυθμού σύγκλισης. Όταν υπάρχει ένα ανώτατο όριο για τη τιμή γ , η παράμετρος ξ θα είναι ένας άλλος παράγοντας επιτάχυνσης της σύγκλισης του δικτύου (για παράδειγμα η σύγκλιση για μη γραμμικές συναρτήσεις ενεργοποίησης μπορεί να είναι πιο γρήγορη από εκείνη για γραμμικές). Γι' αυτό το λόγο στο δεύτερο κλάδο επιλέγουμε να τοποθετήσουμε τη διπολική σιγμοειδή συνάρτηση παρά τη γραμμική.

Παρατήρηση 2

Αυξάνοντας τη τιμή της παραμέτρου γ , έχουμε γρηγορότερη σύγκλιση του δικτύου (3.3.2) στη θεωρητική λύση.

Στο παρακάτω γράφημα μπορούμε να παρατηρήσουμε τη συμπεριφορά των σφαλμάτων εισόδου για τις διάφορες συναρτήσεις ενεργοποίησης (θέτουμε $\gamma=10^3$).



Χαρακτηριστικά σύγκλισης για τα σφάλματα εισόδων για τις διάφορες συναρτήσεις ενεργοποίησης. [13]

3.4 Εφαρμογή ZNN μοντέλου στο Simulink του Matlab

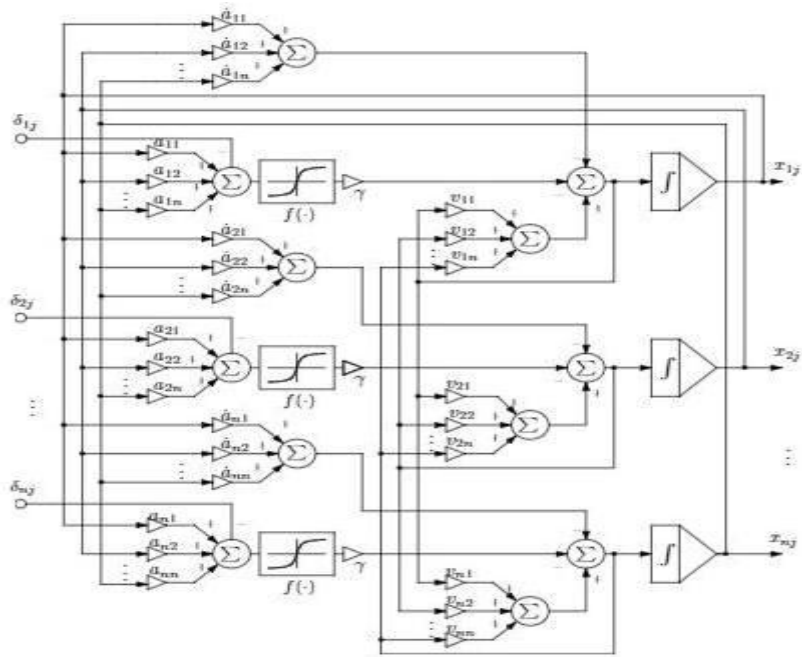
Στα πλαίσια αυτής της υπό-ενότητας θα εξετάσουμε και θα προγραμματίσουμε τα παραπάνω θεωρητικά αποτελέσματα στο περιβάλλον του Matlab. Κατ' αρχήν θα πρέπει να παρατηρήσουμε ότι η σχέση (3.3.2) μπορεί με τις κατάλληλες πράξεις να μετασχηματιστεί στη παρακάτω εξίσωση πινάκων.

$$\dot{X}(t) = -\dot{A}(t)X(t) + (I - A(t))\dot{X}(t) - \gamma F(A(t)X(t) - I). \quad (3.4)$$

Η οποία αναλυτικά μπορεί να περιγραφεί σύμφωνα με το παρακάτω τύπο:

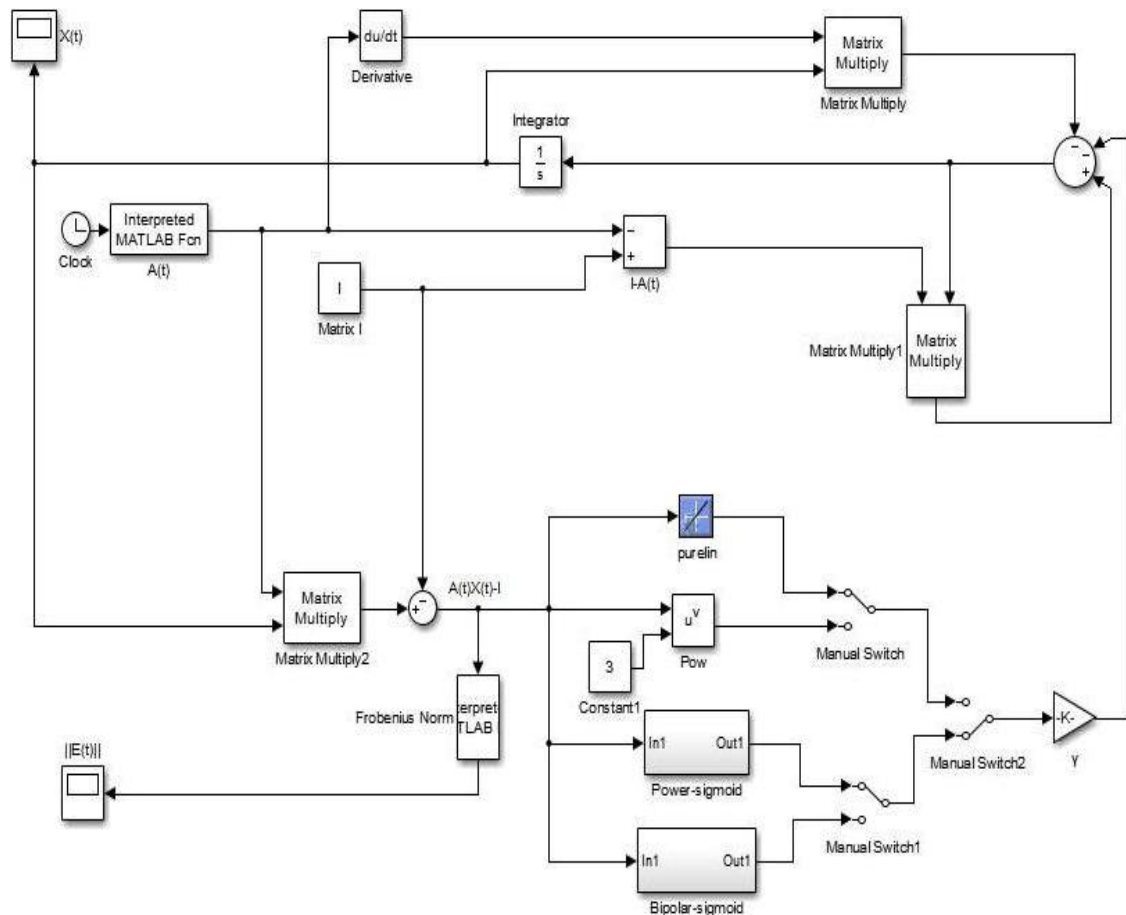
$$\dot{x}_{ij} = -\sum_{k=1}^n \dot{a}_{ik} x_{kj} - \gamma f \left(\sum_{k=1}^n (a_{ik} x_{kj}) - \delta_{ij} \right) + \sum_{k=1}^n v_{ik} \dot{x}_{kj} \quad (3.4.1)$$

- όπου x_{ij} είναι το ij -ό στοιχείο του πίνακα κατάστασης $X(t)$ (σε πολλές εργασίας ο όρος x_{ij} παρομοιάζεται σαν νευρώνας του πίνακα $X(t)$).
- οι χρονικά μεταβαλλόμενοι όροι a_{ij} και \dot{a}_{ij} είναι τα ij -στα στοιχεία του πίνακα $A(t)$, $\dot{A}(t)$ αντίστοιχα.
- δ_{ij} είναι το δέλτα του Kronecker
- $v_{ij} = \delta_{ij} - a_{ij}$.



Κύκλωμα που αφορά τη j-στήλη του πίνακα X(t) σύμφωνα με τη (2.8.1) [14]

Παρακάτω έχουμε το νευρωνικό κύκλωμα που επιλύει το πρόβλημα της αντιστροφής του πίνακα για ένα τυχαίο χρονικά μεταβαλλόμενο πίνακα (στο simulink του Matlab).



ZNN μοντέλο στο Matlab

Παρατήρηση

Προφανώς αν ο πίνακας A είναι σταθερός, τότε η παράγωγος του είναι 0, το παραπάνω κύκλωμα απλοποιείται και έτσι μπορούμε να έχουμε τον υπολογισμό σταθερού (ομαλού). Επίσης όταν πρόκειται να εφαρμόσουμε το παραπάνω κύκλωμα για σταθερό πίνακα τότε σύμφωνα με το κύριο Zhang [16] ενδείκνυται η χρησιμοποίηση της γραμμικής συνάρτησης ενεργοποίησης.

3.4.1 Η Οικογένεια ολοκληρωτών του Matlab

3.4.1.1 Εισαγωγικά [12]

Η οικογένεια αριθμητικής επίλυσης διαφορικών εξισώσεων του Matlab είναι εφοδιασμένη με πληθώρα ολοκληρωτών για την επίλυση προβλημάτων αρχικών συνθηκών της μορφής:

$$\dot{x} = f(t, x), \quad x(t_0) = x_0 \in R^n$$

ή γενικότερα

$$M(t, x)\dot{x} = f(t, x), \quad x(t_0) = x_0 \in R^n$$

όπου ο πίνακας $M(t, x)$ ενδεχομένως να είναι ιδιάζων. Καθένας από αυτούς υλοποιεί και μία διαφορετική αριθμητική μέθοδο και η χρήση του ενδείκνυται σε διαφορετικά προβλήματα και περιστάσεις σε σχέση με τους υπόλοιπους. Ο αναγνώστης που ενδιαφέρεται για θεωρητικά ζητήματα καθώς και ζητήματα υλοποίησης των ολοκληρωτών αυτών μπορεί να ανατρέξει στα αρχεία των ολοκληρωτών στο search του matlab πχ edit ode45.m, τα οποία περιέχουν χρήσιμα σχόλια.

Στα πλαίσια αυτής της εργασίας [14] η μέθοδος που χρησιμοποιείται είναι η υπό-ρουτίνα ode45 (Dormand-Price method) [28] η οποία αποτελεί μια προσεγγιστική μέθοδο επίλυσης διαφορικών εξισώσεων κατηγορίας των τύπων Runge-Kutta. Το αντίστοιχο Butcher Tableau είναι σημειωμένο στο παράθεμα της εργασίας.

Παράδειγμα 2.8.1

Έστω ότι θέλουμε να υπολογίσουμε τον αντίστροφο του χρονικά μεταβαλλόμενου πίνακα $A(t)$ με τύπο:

$$A(t) = \begin{bmatrix} \sin(t) & \cos(t) \\ -\cos(t) & \sin(t) \end{bmatrix}$$

με ορίζουσα:

$$\det(A(t)) = 1 \neq 0$$

ιδιοτιμές:

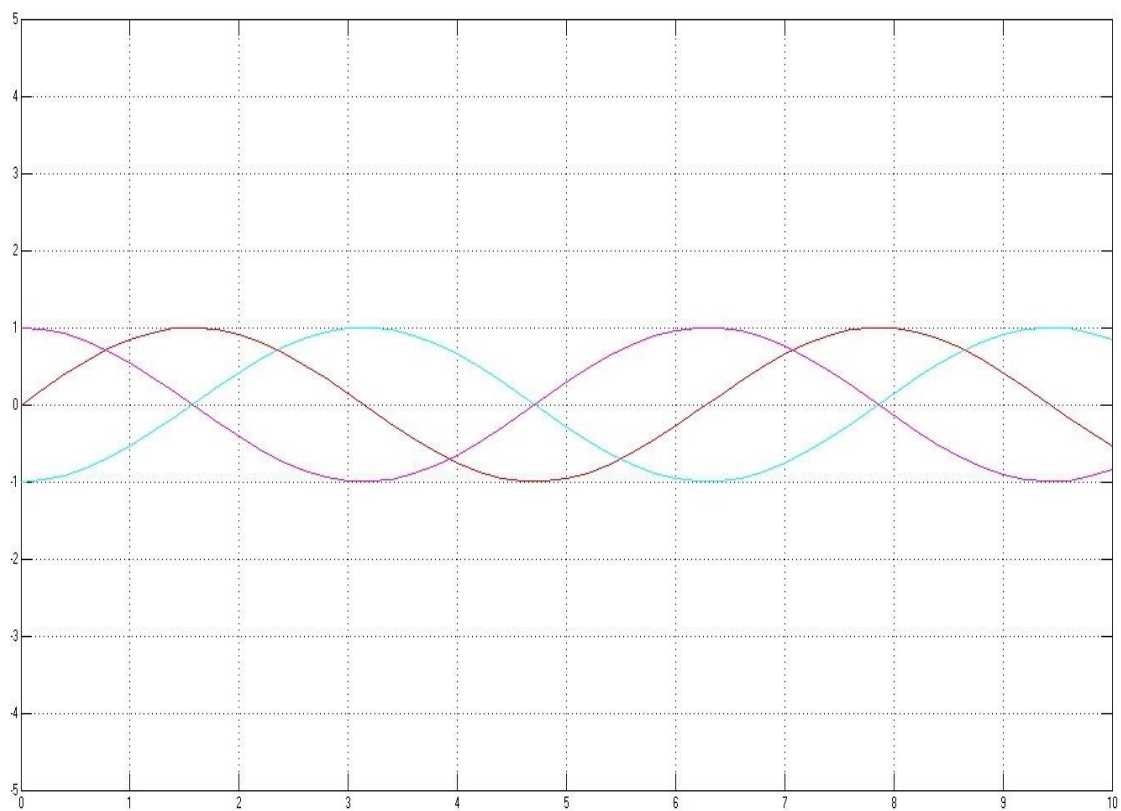
$$\varepsilon_1 = \sin(t) - \cos(t)i, \quad \varepsilon_2 = \sin(t) + \cos(t)i$$

και παράγωγο:

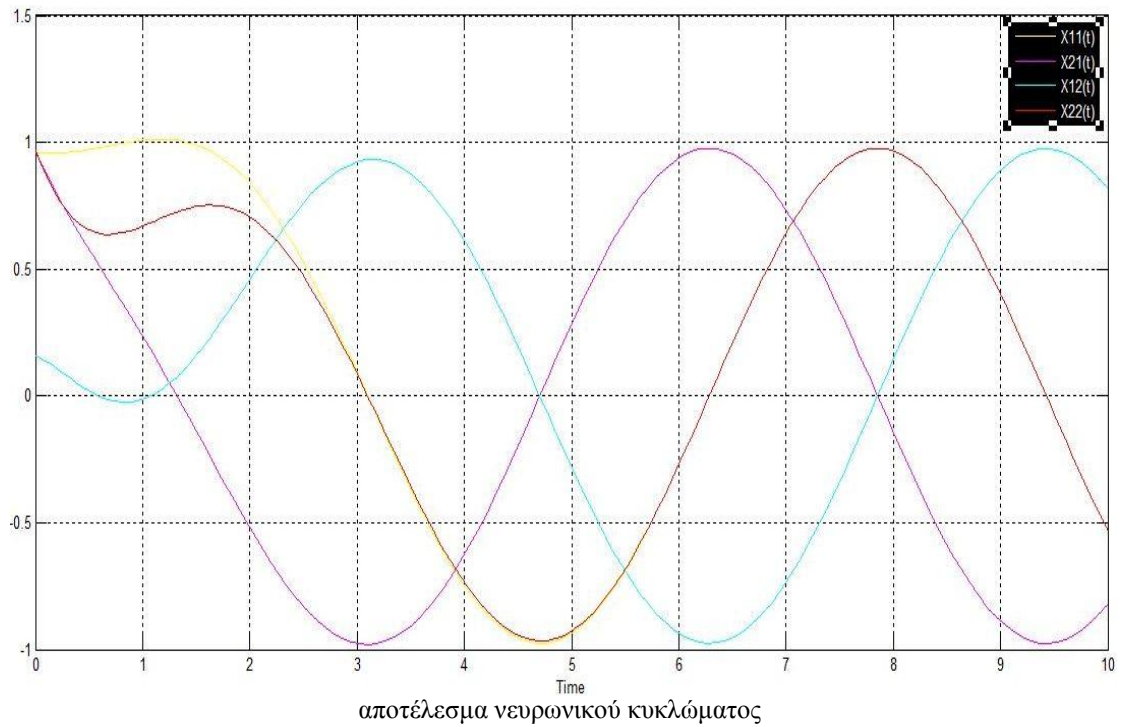
$$\dot{A}(t) = \begin{bmatrix} \cos(t) & -\sin(t) \\ \sin(t) & \cos(t) \end{bmatrix}$$

Ο παραπάνω πίνακας έχει αντίστροφο τον:

$$X^*(t) = \begin{bmatrix} \sin(t) & -\cos(t) \\ \cos(t) & \sin(t) \end{bmatrix}$$



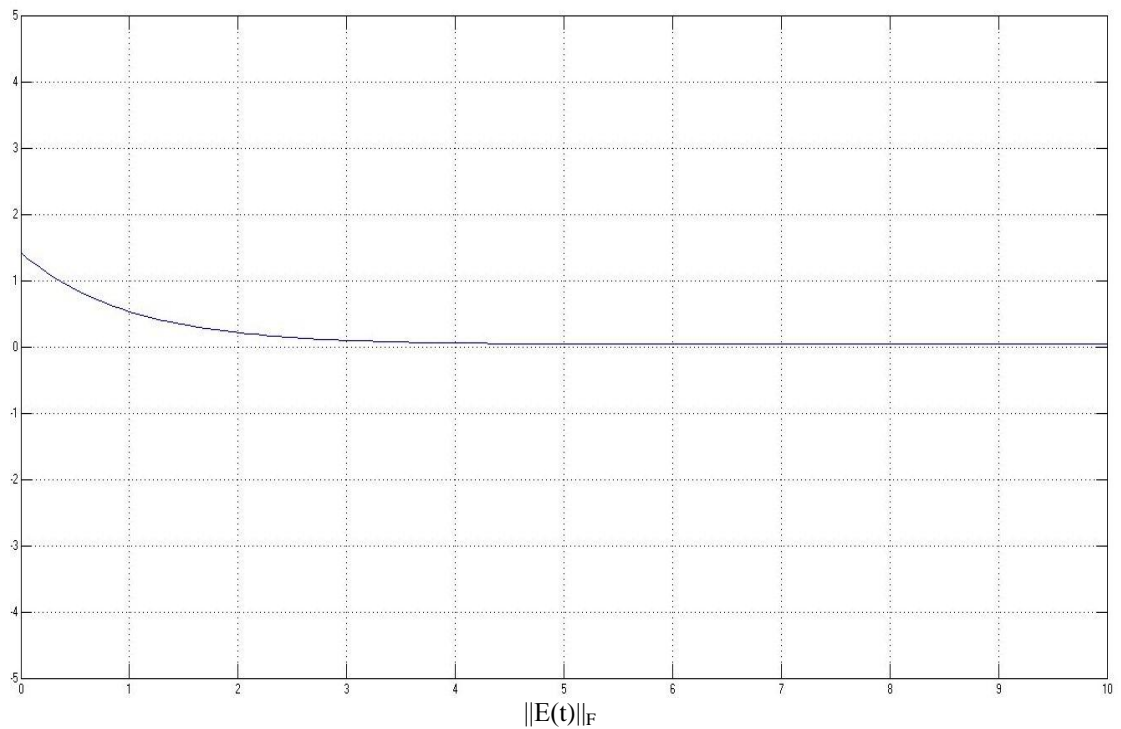
Εφαρμόζοντας το παραπάνω κύκλωμα αντιστροφής για $\gamma=1$ και επιλέγοντας τη γραμμική συνάρτηση ενεργοποίησης έχουμε τα παρακάτω αποτελέσματα:



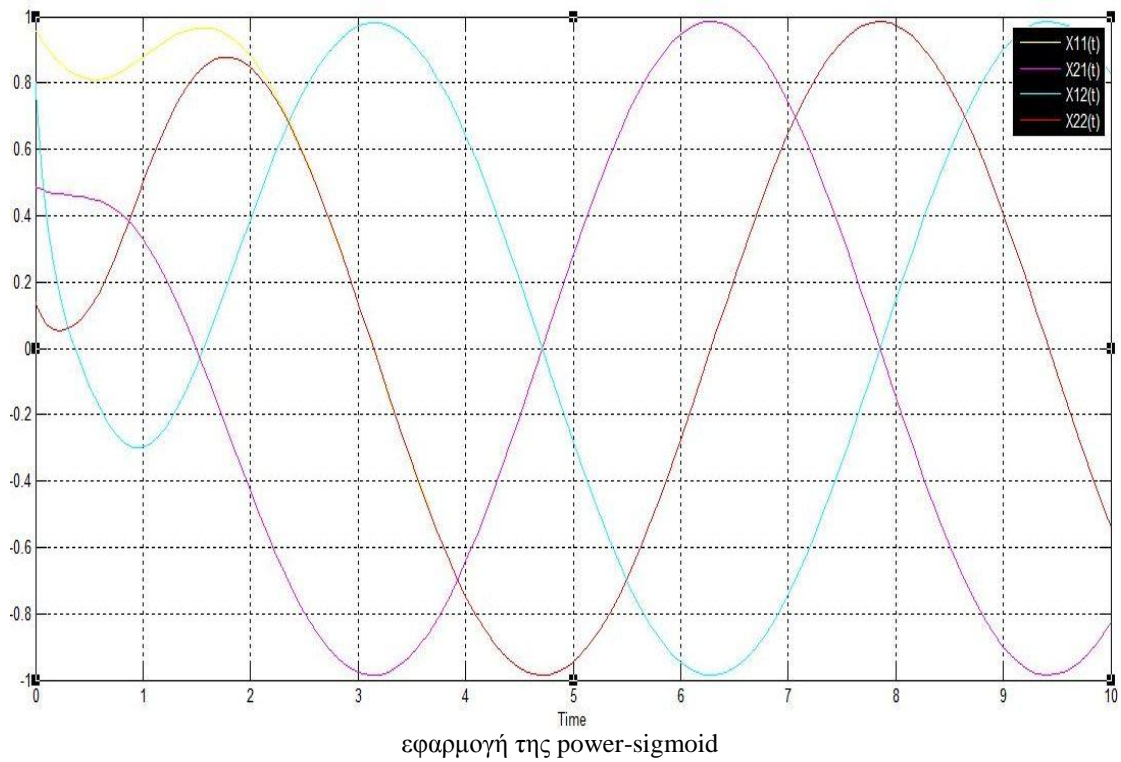
όπου

$X_{11}(t) \rightarrow$, $X_{21}(t) \rightarrow$, $X_{12}(t) \rightarrow$, $X_{22}(t) \rightarrow$

Για το σφάλμα $\|E(t)\|_F$ γραφικά έχουμε:

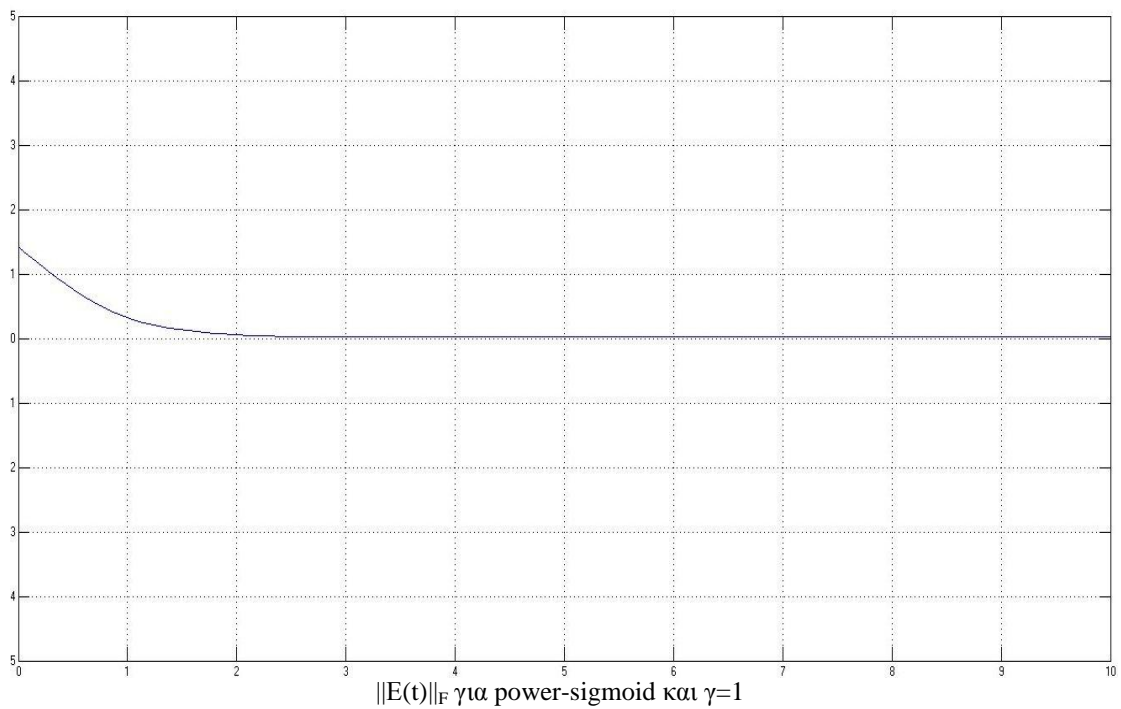


Αν τώρα επιλέξουμε διαφορετική συνάρτηση ενεργοποίησης (τη power-sigmoid) για την ίδια τιμή της τιμής γ ($\gamma=1$) θα πάρουμε τα παρακάτω αποτελέσματα:

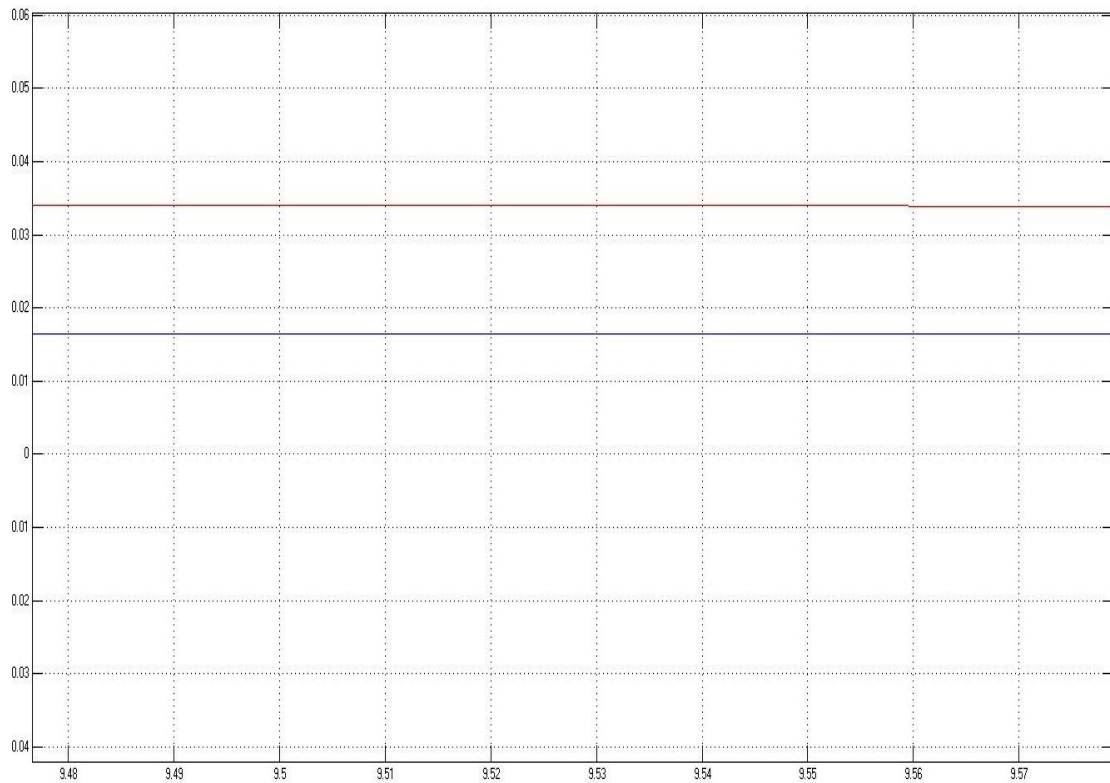


όπου

$X_{11}(t) \rightarrow$ ---, $X_{21}(t) \rightarrow$ ---, $X_{12}(t) \rightarrow$ ---, $X_{22}(t) \rightarrow$ ---



Όσον αφορά τη σύγκλιση για να βγάλουμε πιο ασφαλή συμπέρασμα κάνουμε μεγέθυνση στο κοινό γράφημα σφάλματος και παίρνουμε τα παρακάτω γραφικά αποτελέσματα:



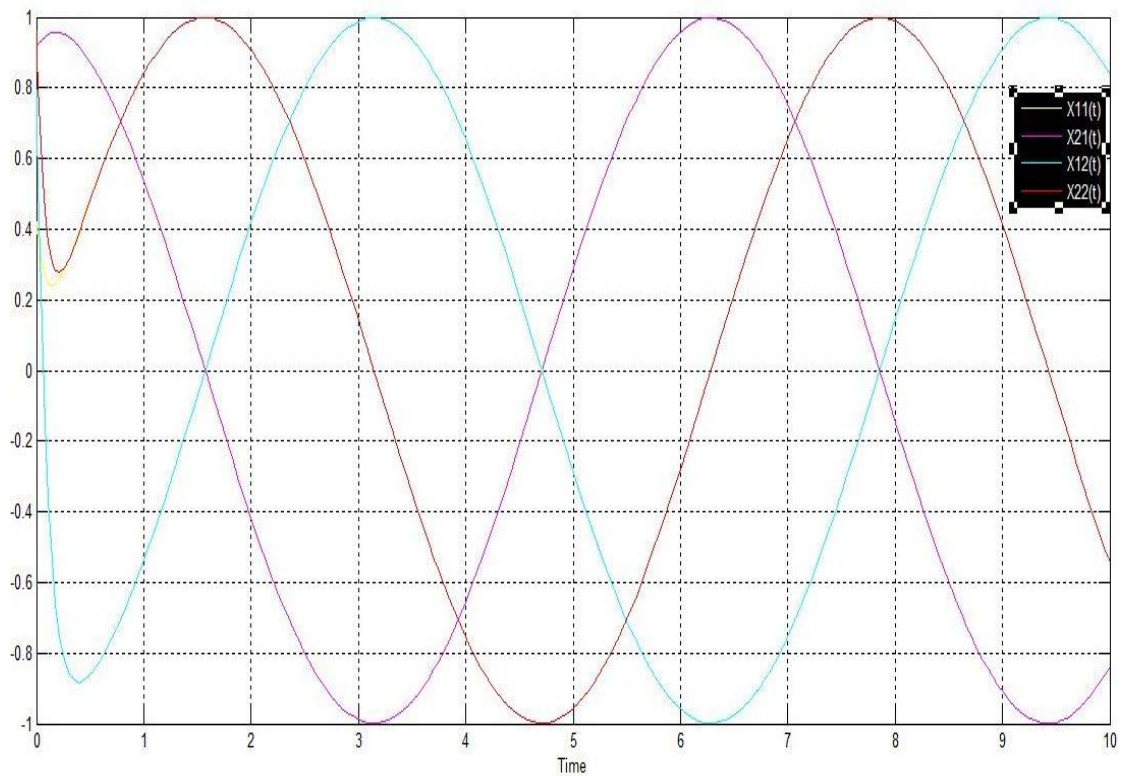
όπου με **μπλε γραμμή** είναι το σφάλμα για τη power-sigmoid και με **κόκκινη γραμμή** για τη γραμμική συνάρτησης ενεργοποίησης.

Παρατήρηση 1

Παρατηρούμε ότι σε αυτή τη περίπτωση έχουμε γρηγορότερη σύγκλιση σε σχέση με τη γραμμική συνάρτηση ενεργοποίησης, γεγονός που συμφωνεί με τα θεωρητικά αποτελέσματα τα οποία προαναφέραμε.

Στη συνέχεια εάν αυξήσουμε τη τιμή γ από 1 σε 10 θα πάρουμε τα παρακάτω γραφικά αποτελέσματα:

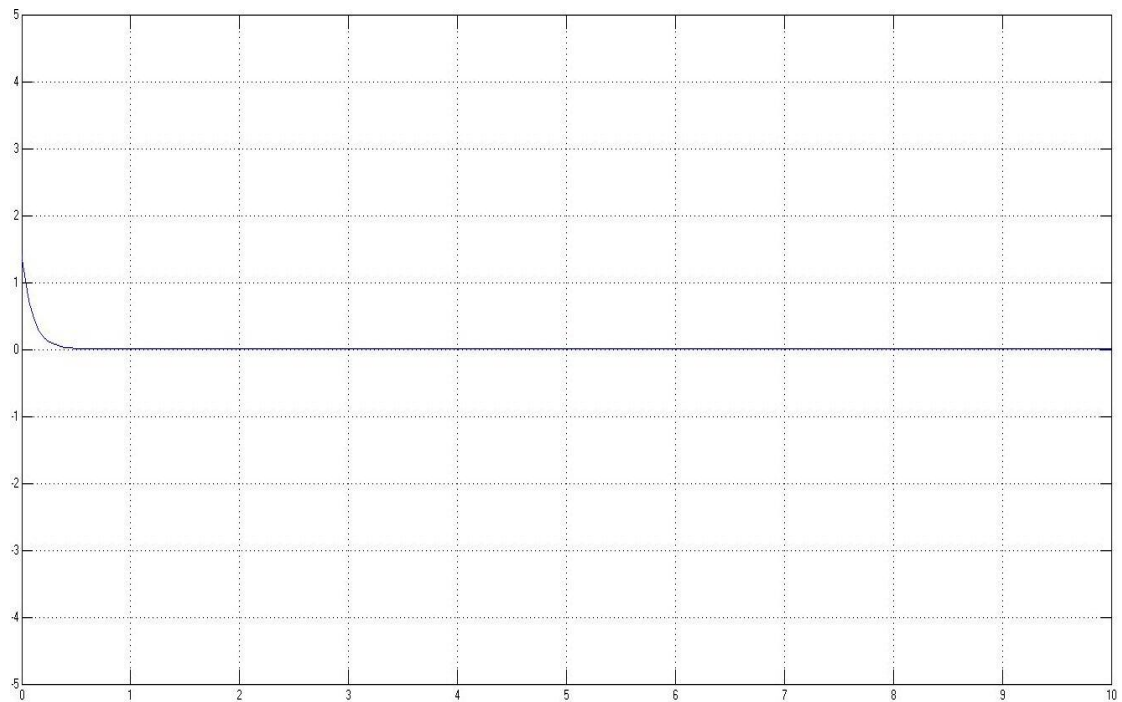
A] Στη γραμμική περίπτωση για $\gamma=10$ το κύκλωμα θα μας δώσει τα παρακάτω αποτελέσματα:



$\gamma=10$ και εφαρμογή γραμμικής συνάρτησης ενεργοποίησης

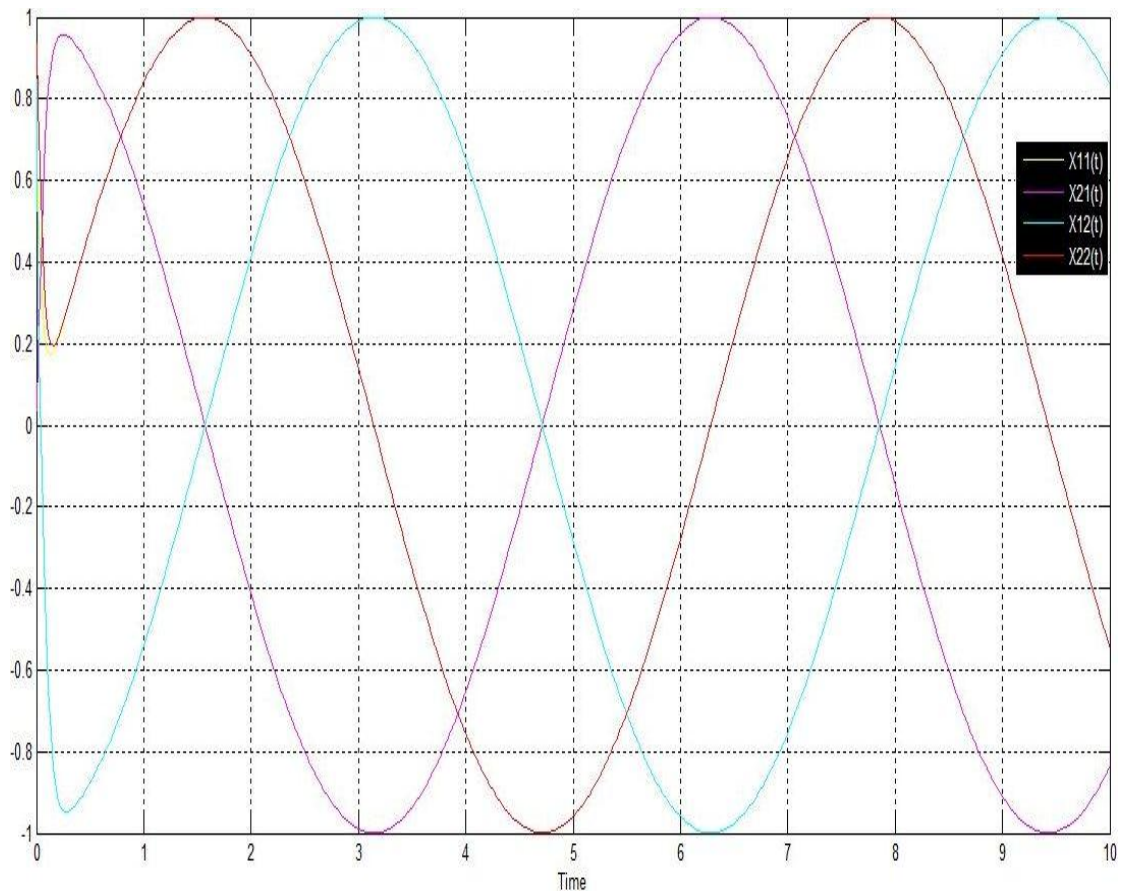
όπου

$X_{11}(t) \rightarrow$ -----, $X_{21}(t) \rightarrow$ - - - - -, $X_{12}(t) \rightarrow$ - - - - -, $X_{22}(t) \rightarrow$ -----



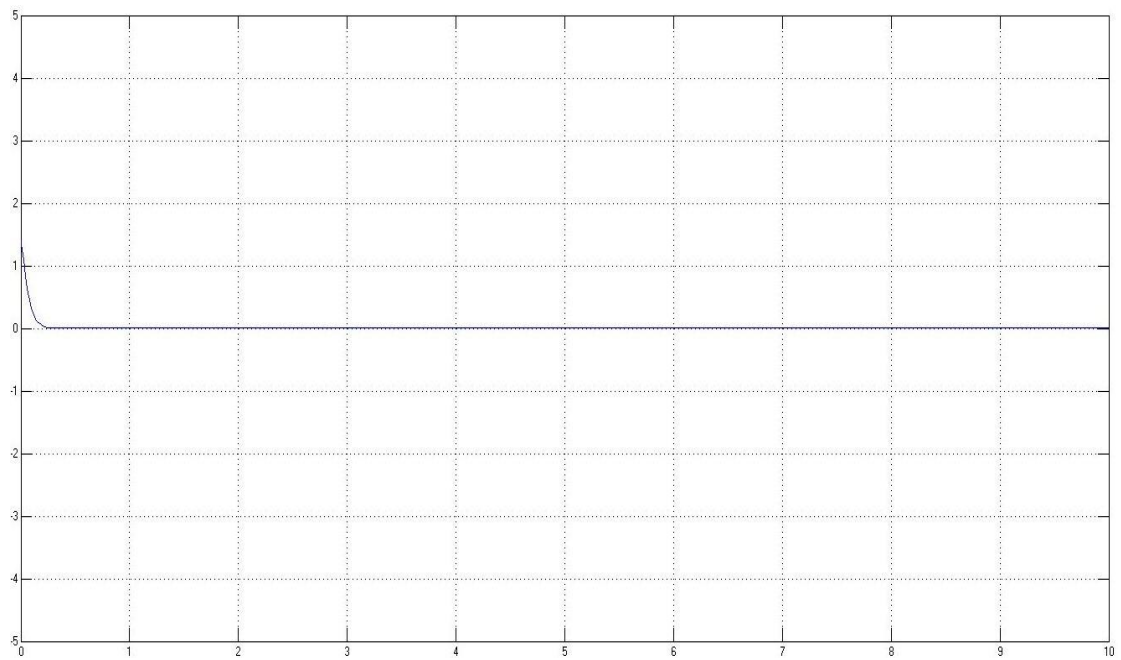
$\|E(t)\|$ για $\gamma=10$ και εφαρμογή γραμμικής συνάρτησης ενεργοποίησης

Β] Στη περίπτωση της power-sigmoid και $\gamma=10$ έχουμε:



$\gamma=10$ και εφαρμογή της power-sigmoid συνάρτησης ενεργοποίησης

Με σφάλμα:



$\|E(t)\|_F$ για $\gamma=10$ και εφαρμογή power-sigmoid συνάρτησης ενεργοποίησης

Παρατήρηση 2

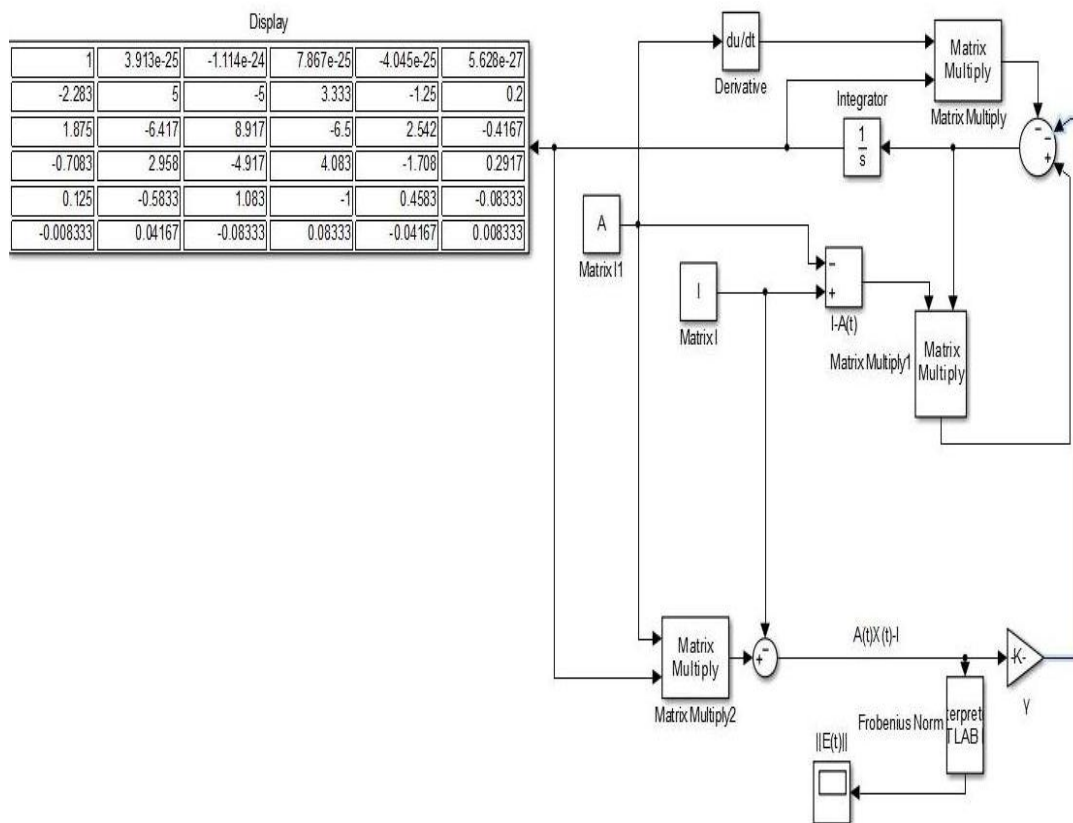
Καθώς η τιμή γ μεγαλώνει τότε η σύγκλιση γίνεται γρηγορότερη (και στις δύο περιπτώσεις).

Παράδειγμα 2.3.3

Θέλουμε να ελέγξουμε πως δουλεύει το ZNN μοντέλο για σταθερό πίνακα με μεγάλο αριθμό κατάστασης. Έστω ο πίνακας A:

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 & 16 & 32 \\ 1 & 3 & 9 & 27 & 81 & 243 \\ 1 & 4 & 16 & 64 & 256 & 1024 \\ 1 & 5 & 25 & 125 & 625 & 3125 \end{bmatrix}$$

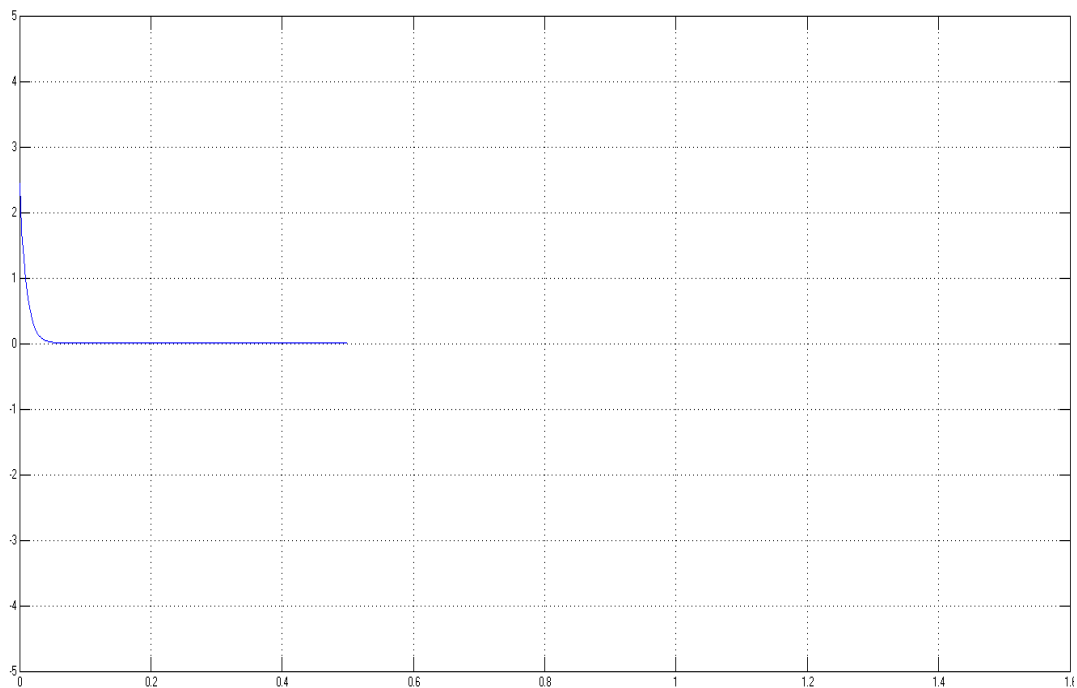
Ο πίνακας A έχει μεγάλο $\text{cond}(A)$ (συγκεκριμένα $\text{cond}(A) = 5.7689 \cdot 10^4$), επιλέγοντας τη τιμή $\gamma = 100$ παίρνουμε τα παρακάτω αποτελέσματα:



Ενώ ο θεωρητικός αντίστροφος είναι:

$$A^{-1} = \begin{bmatrix} 1.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ -2.2833 & 5.0000 & -5.0000 & 3.3333 & -1.2500 & 0.2000 \\ 1.8750 & -6.4167 & 8.9167 & -6.5000 & 2.5417 & -0.4167 \\ -0.7083 & 2.9583 & -4.9167 & 4.0833 & -1.7083 & 0.2917 \\ 0.1250 & -0.5833 & 1.0833 & -1.0000 & 0.4583 & -0.0833 \\ -0.0083 & 0.0417 & -0.0833 & 0.0833 & -0.0417 & 0.0083 \end{bmatrix}$$

Παρατηρούμε ότι η λύση του δικτύου είναι πολύ κοντά στη θεωρητική λύση. Όσον αφορά το χρόνο σύγκλισης μπορούμε να ελέγξουμε από το παρακάτω γράφημα του σφάλματος:



Στη συνέχεια θα εφαρμόσουμε το παραπάνω κύκλωμα για χρονικά μεταβαλλόμενους πολυωνυμικούς πίνακες οι οποίοι έχουν μεγάλη εφαρμογή σε πολλά προβλήματα της θεωρίας ελέγχου.

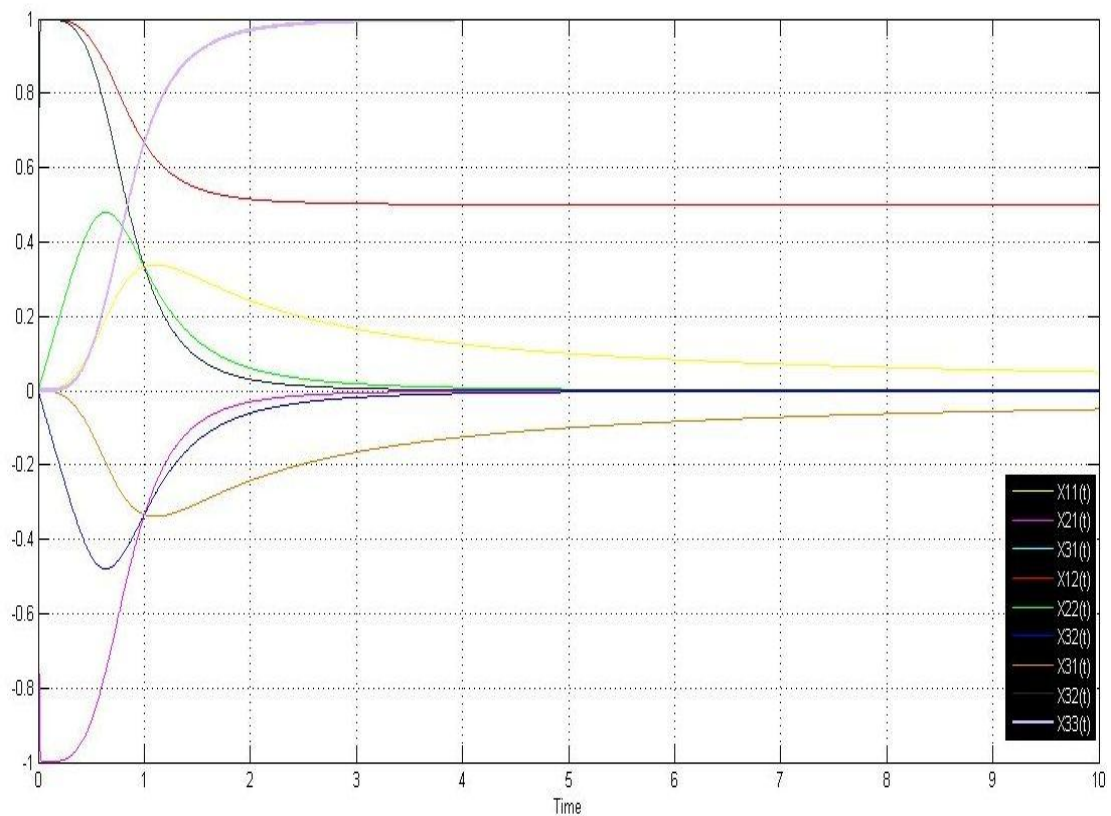
Παράδειγμα 2.3.4

Έστω ο χρονικά μεταβαλλόμενος πολυωνυμικός πίνακας:

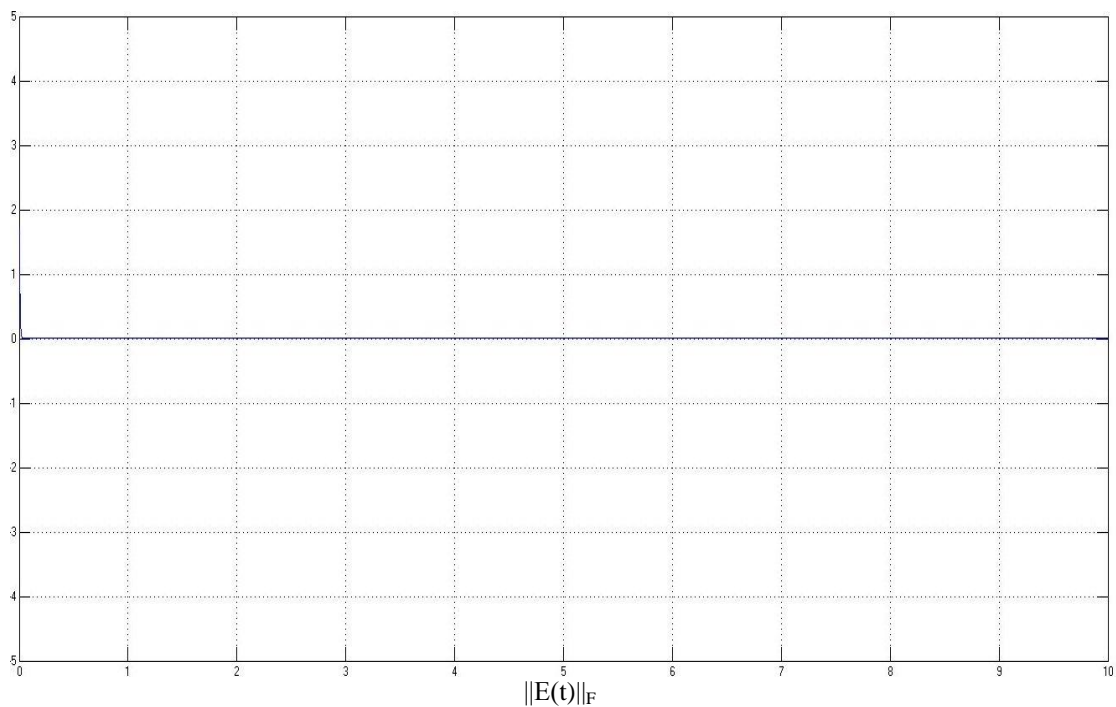
$$A(t) = \begin{bmatrix} t & -t^4 & 1 \\ 1 & t^3 & 0 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} t^4 + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} t^3 + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} t^2 + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} t + \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} I$$

με ορίζουσα: $\det A(t) = 2t^4 + 1$

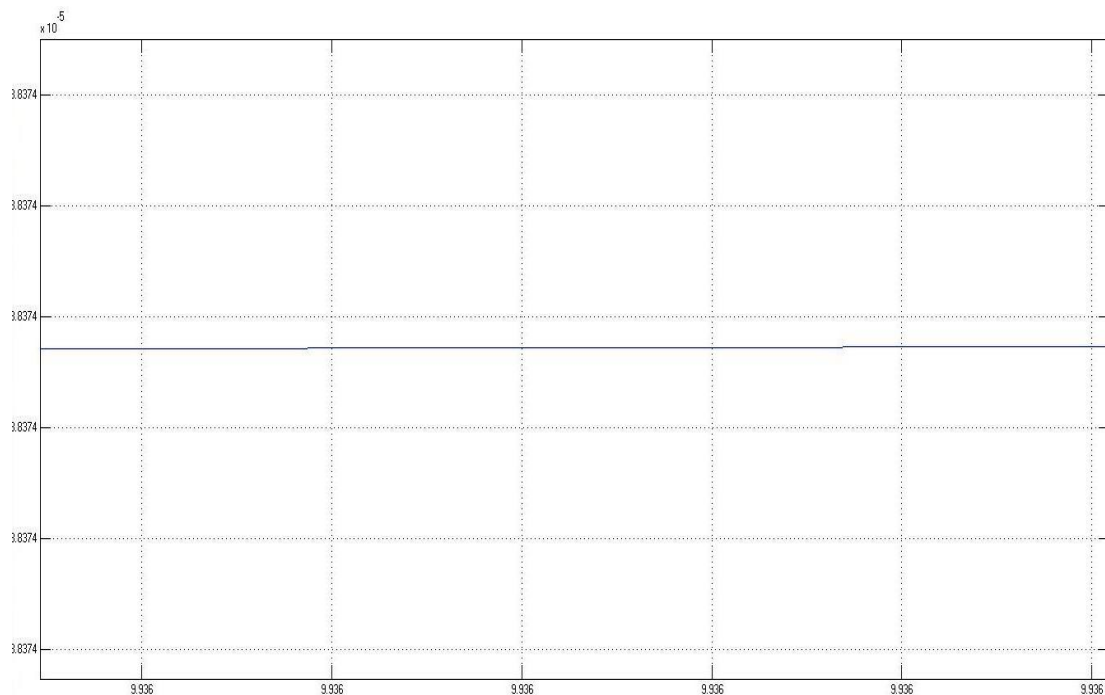
Εφαρμόζοντας το κύκλωμα μας για power sigmoid συνάρτηση ενεργοποίησης και $\gamma=100$ έχουμε τα παρακάτω γραφικά αποτελέσματα:



με σφάλμα:



Κάνοντας μεγέθυνση παίρνουμε:



Παρατηρώντας το παραπάνω παράδειγμα μας δημιουργήθηκε η απορία εάν τα παραπάνω μοντέλα μπορούν να εφαρμοστούν για οποιοδήποτε πολυωνυμικό πίνακα $n \times n$ διαστάσεων. Η απάντηση που θα δώσουμε είναι αρνητική και αυτό οφείλεται στη συνθήκη αντιστρεψιμότητας (3.1.1) που κάποιες φορές παραβιάζεται όπως θα δούμε στη παρακάτω εφαρμογή.

Εφαρμογή

Έστω ότι θέλουμε να υπολογίσουμε τον αντίστροφο του πολυωνυμικού πίνακα $A(t)$ με τύπο:

$$A(t) = \begin{bmatrix} t-1 & 0 \\ 1 & t-2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} t + \begin{bmatrix} -1 & 0 \\ 1 & -2 \end{bmatrix} I$$

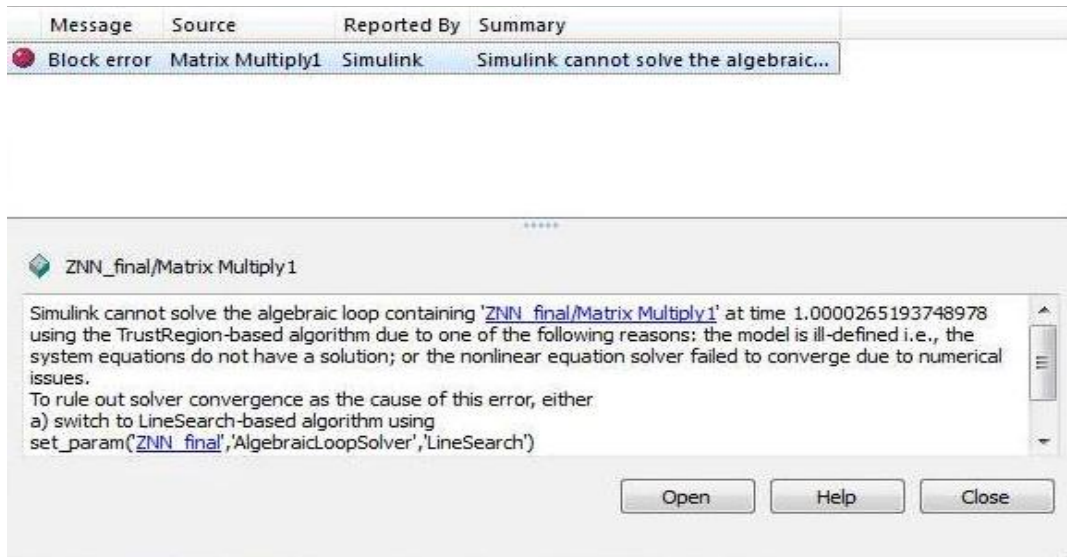
έχει ορίζουσα

$$\det(A(t)) = (t-1)(t-2)$$

και ιδιοτιμές:

$$\varepsilon_1 = t-1, \quad \varepsilon_2 = t-2$$

Το Matlab μας βγάζει το παρακάτω πίνακα



Δεν σημαίνει ότι έχουμε κάνει εμείς κάποιο λάθος απλώς **δεν ισχύει η συνθήκη αντιστρεψιμότητας** και αυτό οφείλεται στο γεγονός ότι οι ιδιοτιμές του πίνακα $A(t)$ μηδενίζονται για $t=1\text{sec}$ και $t=2\text{sec}$. Συνεπώς η ZNN μέθοδος δε μπορεί να εφαρμοστεί στο συγκεκριμένο παράδειγμα και δε μπορεί να γενικευτεί για οποιοδήποτε χρονικά μεταβαλλόμενο πολυωνυμικό πίνακα.

Συμπεράσματα

Πλεονεκτήματα ZNN μοντέλων

- Δίνουν γρήγορα αποτελέσματα για μικρή τιμή του γ
- Τα προτιμούμε για αντιστροφή χρονικά μεταβαλλόμενων συστημάτων που ικανοποιούν τη συνθήκη αντιστρεψιμότητας, εφόσον έχουμε καλή σύγκλιση μεταξύ της θεωρητικής λύσης και της προσεγγιστικής λύσης που παίρνουμε από το ZNN μοντέλο.
- Μπορούν να εφαρμοστούν και για σταθερό πίνακα με μεγάλο αριθμό κατάστασης.

Μειονέκτημα ZNN μοντέλων

- Λόγω της συνθήκης αντιστρεψιμότητας (3.1.1) δεν μπορούν να εφαρμοστούν για την αντιστροφή οποιοδήποτε χρονικά μεταβαλλόμενου πολυωνυμικό πίνακα.

Στη συνέχεια θα αναφερθούμε στα GNN δίκτυα και θα κάνουμε σύγκριση με τα ZNN μοντέλα.

3.5 GNN (Gradient Neural Networks) μοντέλα

3.5.1 GNN (Gradient Neural Network) για σταθερούς πίνακες

Τα GNN μοντέλα, είναι από τις πρώτες κατηγορίες τεχνητών νευρωνικών δικτύων που κατασκευάστηκαν για να επιλύουν προβλήματα αντιστροφής πινάκων και η ιδέα τους στηρίζεται στη μέθοδο μέγιστης κλίσης. Στα GNN μοντέλα ορίζουμε τη παρακάτω συνάρτηση σφάλματος:

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

Παρατηρούμε ότι λόγω της νόρμας ισχύει $E(X(t), t) \geq 0$, συνεπώς το σφάλμα E παρουσιάζει ελάχιστο στη τιμή $\bar{X}(t)$ (δηλαδή $E(X(t), t) \geq E(\bar{X}(t), t) = 0$). Οπότε χρησιμοποιώντας τη λογική της μεθόδου μέγιστης κλίσης καταλήγουμε στο μοντέλο (2.9). Πρώτα όμως θα αποδείξουμε τη παρακάτω σχέση.

Λήμμα 3.5.1 [20]

Η παράγωγος της E ως προς X είναι:

$$\frac{\partial E}{\partial X} = A^T (AX(t) - I) (*)$$

Ακολουθώντας το παρακάτω τρόπο σχεδιασμού (στηρίζεται στη μέθοδο μέγιστης κλίσης):

$$\dot{X}(t) = -\gamma \frac{\partial E}{\partial X}$$

καταλήγουμε στο παρακάτω μοντέλο:

$$\dot{X}(t) = -\gamma A^T AX(t) + \gamma A^T \quad (3.5)$$

όπου γ είναι μία θετική παράμετρος και A^T είναι ο ανάστροφος του πίνακα A . Την ποσότητα $-\gamma A^T A$ την ορίζουμε σαν πίνακα βάρους και συμβολίζεται με W (δηλαδή $W = -\gamma A^T A$). Το νευρωνικό δίκτυο (3.5) μπορεί να διαιρεθεί σε n ανεξάρτητα υπό-

δίκτυα, όπου n είναι η τάξη του πίνακα A . Κάθε υπό-δίκτυο αποτελείται ουσιαστικά από ένα διάνυσμα στήλης του πίνακα $X(t)$. Ας ορίζουμε $x_i(t)$ την i -στήλη του πίνακα κατάστασης $X(t)$ για $i=1,2,\dots,n$. Δηλαδή έχουμε:

$$\frac{dx_i}{dt} = -\gamma A^T A x_i(t) + \gamma A_i^T \quad (3.5.1)$$

Εφόσον έχουμε ορίσει W τον πίνακα βάρους από τη παραπάνω σχέση παρατηρούμε ότι ο πίνακας W είναι ίδιος για κάθε υπό-δίκτυο της (3.5.1) και ο όρος A_i^T συμπεριφέρεται σαν όρος πόλωσης (bias) για το δίκτυο.

Συνεχίζοντας, θα αναφέρουμε τις παρακάτω 2 προτάσεις που αφορούν την ευστάθεια και τη δυνατότητα σύγκλισης του νευρωνικού δικτύου (3.5) στο θεωρητικό αντίστροφο του πίνακα A

Πρόταση 3.5.1.1 [18]

Αν ο σταθερός πίνακας A είναι ομαλός τότε το επαναλαμβανόμενο νευρωνικό δίκτυο (2.9) είναι ασυμπτωτικά ευσταθές.

Απόδειξη

Πρώτα θα αποδείξουμε ότι ο πίνακας $A^T A$ είναι θετικά ορισμένος. Εάν ο πίνακας A είναι ομαλός τότε $Ax=0$ αν και μόνο αν $x=0$ για ένα διάνυσμα x . Για κάθε $x \neq 0$, $x^T A^T A x = (Ax)^T Ax = \|Ax\|^2 > 0$. Έτσι, ο πίνακας $A^T A$ είναι ένας θετικά ορισμένος πίνακας. Επίσης ο πίνακας $-\gamma A^T A$ είναι συμμετρικός, οπότε όλες οι ιδιοτιμές του είναι πραγματικοί αριθμοί. Επιπλέον αφού $\gamma > 0$ και $A^T A$ θετικά ορισμένος, όλες οι ιδιοτιμές του πίνακα $-\gamma A^T A$ είναι αρνητικές. Συνεπώς το σύστημα (2.9) είναι ασυμπτωτικά ευσταθές.

Πρόταση 3.5.1.2 [18]

Αν ο πίνακας A είναι ομαλός, τότε ο πίνακας κατάστασης της (2.9) συγκλίνει στο θεωρητικό αντίστροφο του πίνακα A .

Απόδειξη

Στη κατάσταση $\bar{X}(t)$ (έχω το ελάχιστο του σφάλματος άρα $dE/d\bar{X}=0$), ισχύει $\dot{\bar{X}}(t)=0$. Έτσι, $A^T A \bar{X}=A$. Επίσης ο πίνακας $A^T A$ είναι θετικά ορισμένος, άρα ο αντίστροφος

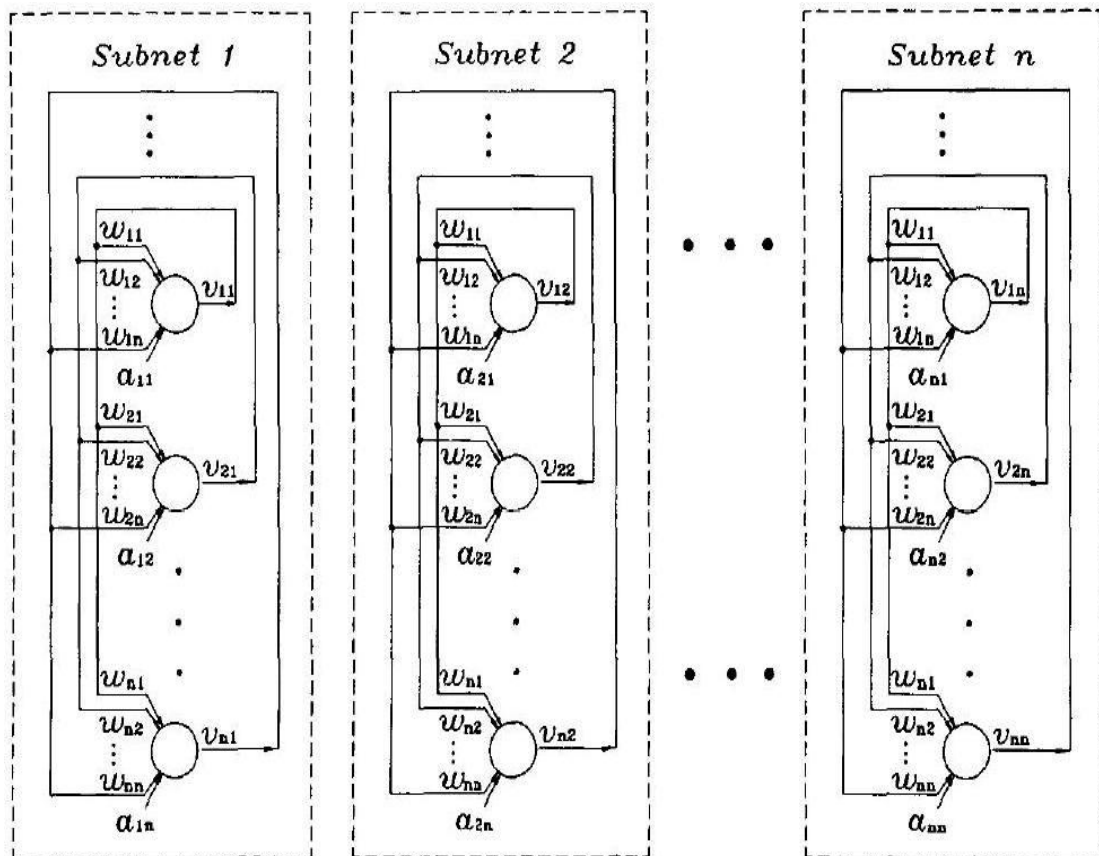
του πίνακα $(A^T A)^{-1}$ υπάρχει. Επιπλέον $\bar{X} = (A^T A)^{-1} A^T = A^{-1} (A^T)^{-1} A^T = A^{-1} I = A^{-1}$. Άρα αποδείξαμε τη πρόταση (2.9.1.2).

Παρατήρηση

1] Η σύγκλιση του δικτύου (2.9) εξαρτάται από τη ελάχιστη ιδιοτιμή, του πίνακα $\gamma A^T A$, $\lambda_{\min}(\gamma A^T A) = \min\{\lambda_i: i=1,2,\dots,n\}$ όπου λ_i είναι i -στη ιδιοτιμή του πίνακα $\gamma A^T A$. Όσον αφορά το χρόνο σύγκλισης του επαναλαμβανόμενου νευρωνικού δικτύου, συγκλίνει στη θεωρητική λύση περίπου μετά από χρόνο:

$$\frac{5}{\lambda_{\min}(\gamma A^T A)} \text{ sec}$$

2] Αυξάνοντας τη τιμή της παραμέτρου γ αυξάνουμε τη ταχύτητα σύγκλισης του δικτύου.



Αρχιτεκτονική δικτύου (2.9) [18]

Παράδειγμα

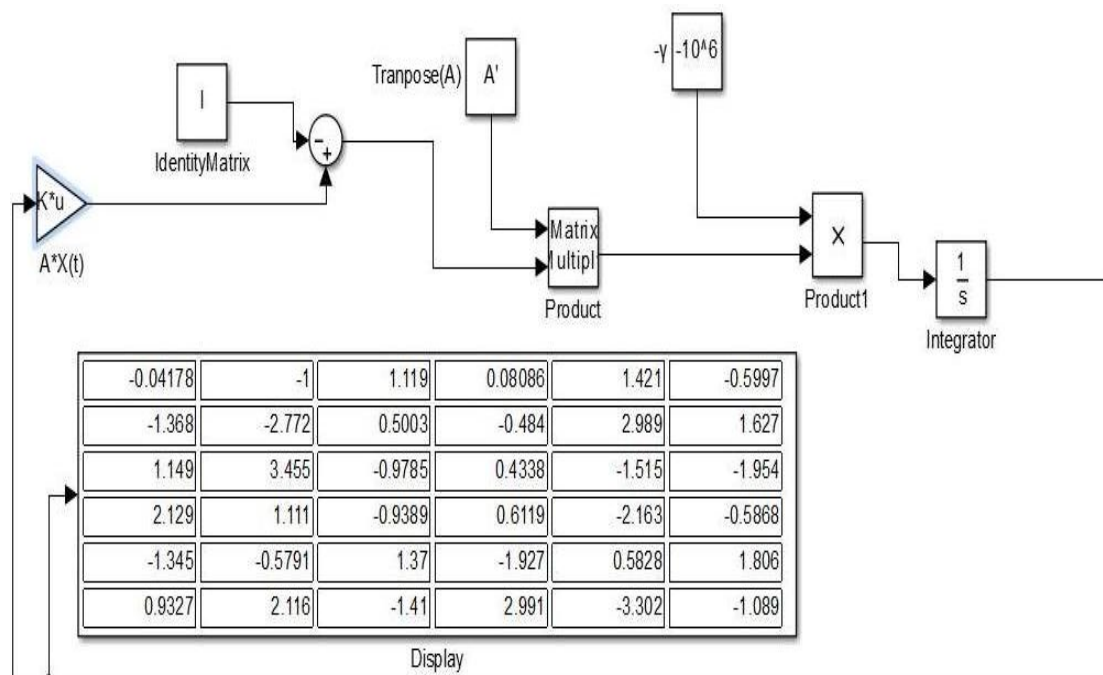
Να βρεθεί ο αντίστροφος του πίνακα A:

$$A = \begin{bmatrix} 0.43 & 0.55 & 0.32 & 0.82 & 0.29 & 0.05 \\ 0.24 & 0.61 & 0.69 & 0.23 & 0.52 & 0.28 \\ 0.83 & 0.09 & 0.14 & 0.25 & 0.62 & 0.32 \\ 0.44 & 0.58 & 0.26 & 0.19 & 0.28 & 0.52 \\ 0.27 & 0.73 & 0.56 & 0.24 & 0.17 & 0.09 \\ 0.15 & 0.92 & 0.45 & 0.62 & 0.71 & 0.41 \end{bmatrix}$$

Ο θεωρητικός αντίστροφος του A είναι:

$$A^{-1} = \begin{bmatrix} -0.0418 & -1.0001 & 1.1190 & 0.0808 & 1.4208 & -0.5997 \\ 1.3680 & -2.7717 & 0.5004 & -0.4841 & 2.9891 & 1.6270 \\ 1.1492 & 3.4549 & -0.9786 & 0.4339 & -1.5148 & -1.9536 \\ 2.1293 & 1.1111 & -0.9389 & 0.6120 & -2.1636 & -0.5868 \\ -1.3450 & -0.5792 & 1.3696 & -1.9267 & 0.5829 & 1.8063 \\ 0.9328 & 2.1163 & -1.4101 & 2.9915 & -3.3021 & -1.0887 \end{bmatrix}$$

Επιλέγοντας $\gamma=10^6$ για το πίνακα $\gamma A^T A$, οι ιδιοτιμές είναι το σύνολο $\{128800, 188000, 12300, 364500, 730800, 6622300\}$ οπότε έχουμε ότι $\lambda_{\min}=12300$. Άρα το δίκτυο θα συγκλίνει στη θεωρητική λύση μετά από περίπου $5/\lambda_{\min}=407\mu s$. Στο Matlab παίρνουμε τα παρακάτω αποτελέσματα:



Παρατηρούμε ότι η λύση του δικτύου (2.9) είναι πολύ κοντά στη θεωρητική λύση, αφού έχουμε:

$$\bar{X}(t) = \begin{bmatrix} -0.04178 & -1.0000 & 1.1190 & 0.08086 & 1.4210 & -0.5997 \\ 1.36800 & -2.7720 & 0.5003 & -0.48400 & 2.9890 & 1.6270 \\ 1.14900 & 3.4550 & -0.9785 & 0.43380 & -1.5150 & -1.9540 \\ 2.12930 & 1.1110 & -0.9389 & 0.61190 & -2.1630 & -0.5868 \\ -1.34500 & -0.5791 & 1.3700 & -1.92700 & 0.5828 & 1.8060 \\ 0.93270 & 2.1160 & -1.4100 & 2.99100 & -3.3020 & -1.0890 \end{bmatrix}$$

Παρατήρηση

Το παραπάνω μοντέλο δίνει αποτελέσματα σε μικρό χρόνο μόνο για πίνακες που δεν βρίσκονται σε αρρωστημένη κατάσταση (δηλαδή παρουσιάζουν μικρό $\text{cond}(A)$).

3.5.2 GNN (Gradient Neural Network) για χρονικά μεταβαλλόμενους πίνακες

Τα GNN μοντέλα, είναι από τις πρώτες κατηγορίες τεχνητών νευρωνικών δικτύων που κατασκευάστηκαν για να επιλύουν προβλήματα αντιστροφής τετραγωνικών σταθερών όπως είπαμε στη προηγούμενη υπό-ενότητα.

Στη συνέχεια και με τη πάροδο των χρόνων έγιναν προσπάθειες κατασκευής, για ένα προσεγγιστικό μοντέλο, για την αντιστροφή χρονικά μεταβαλλόμενων πινάκων. Στα GNN μοντέλα ορίζουμε την παρακάτω συνάρτηση σφάλματος

$$E(X(t), t) = \frac{\|A(t)X(t) - I\|_F^2}{2} \quad (2.9.2)$$

Στη συνέχεια έχουμε ότι το παρακάτω μοντέλο διαφορικών εξισώσεων:

$$\dot{X}(t) = -\gamma \frac{dE}{dX} \quad (2.9.2.1)$$

είναι ευσταθές και προσεγγίζει τη λύση του προβλήματος της αντιστροφής του πίνακα $A(t)$. Στα πλαίσια αυτής της εργασίας ασχοληθήκαμε με τα ZNN μοντέλα γι' αυτό δε θα επεκταθούμε περαιτέρω στη θεωρητική απόδειξη της μεθόδου. Ο αναγνώστης που ενδιαφέρεται για την απόδειξη της μεθόδου που αφορά τη σύγκλιση και το σφάλμα της μεθόδου, μπορεί να ανατρέξει στις εργασίες [19], [20],

Συνεχίζοντας συνδυάζοντας τις σχέσεις (3.5.2), (3.5.2.1) καταλήγουμε στο παρακάτω μοντέλο διαφορικών εξισώσεων:

$$\dot{X}(t) = -\gamma A^T(t)(A(t)X(t) - I) \quad (3.5.2.3)$$

Η σχέση (3.5.2.3) προσεγγίζει το θεωρητικό αντίστροφο του πίνακα $A(t)$. Στα παραδείγματα που ακολουθούν μπορούμε να διαπιστώσουμε κατά πόσο είναι καλή η σύγκλιση (θα ελέγξουμε κατά πόσο το σφάλμα $E(X(t),t)$ τείνει στο 0) και θα τα συγκρίνουμε με τα ZNN μοντέλα. Επίσης θα πρέπει να τονιστεί ότι τα GNN μοντέλα στηρίζονται στη παρακάτω συνθήκη αντιστρεψιμότητας:

Συνθήκη Αντιστρεψιμότητας (GNN μοντέλα)

Υπάρχει θετικός πραγματικός αριθμός a τέτοιος ώστε:

$$\min_{\forall i \in \{1,2,\dots,n\}} |\lambda_i(A^T(t)A(t))| \geq a, \forall t \geq 0 \quad (3.5.2.4)$$

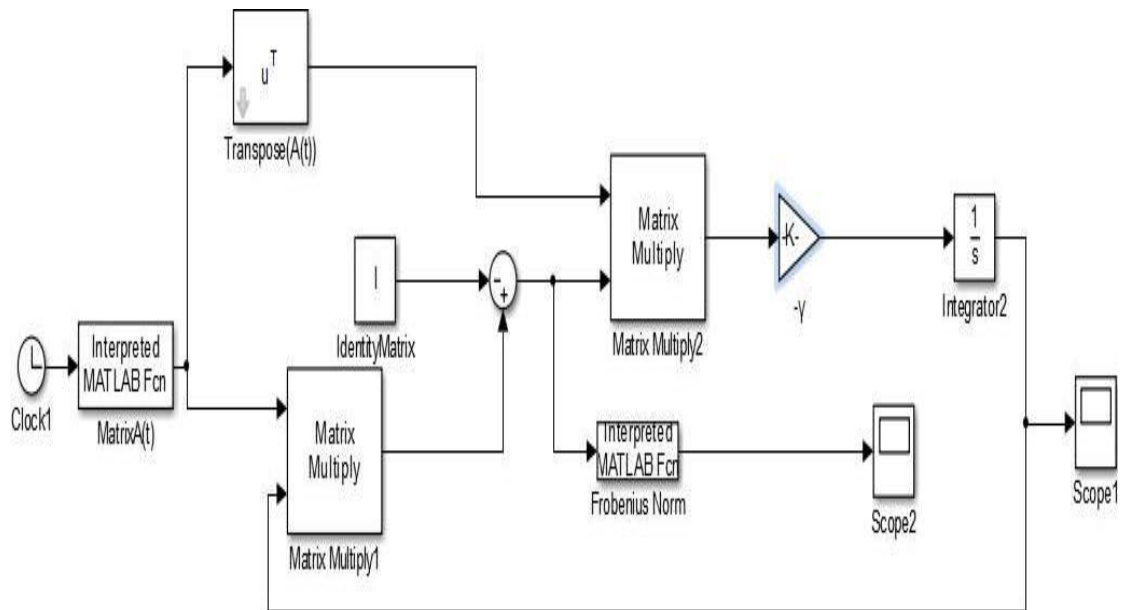
όπου $\lambda_i(\cdot)$ υποδηλώνει την ιδιοτιμή της i του πίνακα $A^T A \in R^{n \times n}$. Εάν ισχύει η παραπάνω συνθήκη αντιστρεψιμότητας τότε υπάρχει μοναδική λύση της εξίσωσης πινάκων (3.1).

Παρατήρηση

Όπως και στα ZNN μοντέλα, έτσι και στα GNN αυξάνοντας τη τιμή γ έχουμε γρηγορότερη σύγκλιση του δικτύου (3.5.2).

3.5.3 Εφαρμογή GNN μοντέλου στο Simulink του Matlab

Στα πλαίσια αυτής της υπό-ενότητας θα εξετάσουμε και θα προγραμματίσουμε τα παραπάνω θεωρητικά αποτελέσματα στο περιβάλλον του Matlab. Παρακάτω έχουμε το νευρωνικό κύκλωμα της σχέσης (3.5.2) που επιλύει το πρόβλημα προσέγγισης της αντιστροφής του πίνακα $A(t)$. Θα πρέπει να τονιστεί ότι το σύστημα διαφορικών εξισώσεων λύνεται με τη μέθοδο των ολοκληρωτών (ode 45) που αναλύσαμε προηγουμένως στα πλαίσια αυτού του κεφαλαίου.



GNN μοντέλο για χρονικά μεταβαλλόμενο πίνακα στο Matlab

Παράδειγμα 3.5.1

Έστω ότι θέλουμε να υπολογίσουμε τον αντίστροφο του χρονικά μεταβαλλόμενου πίνακα $A(t)$ του παραδείγματος (3.4.1) με τύπο:

με ορίζουσα:

$$\det(A(t))=1 \neq 0$$

ιδιοτιμές:

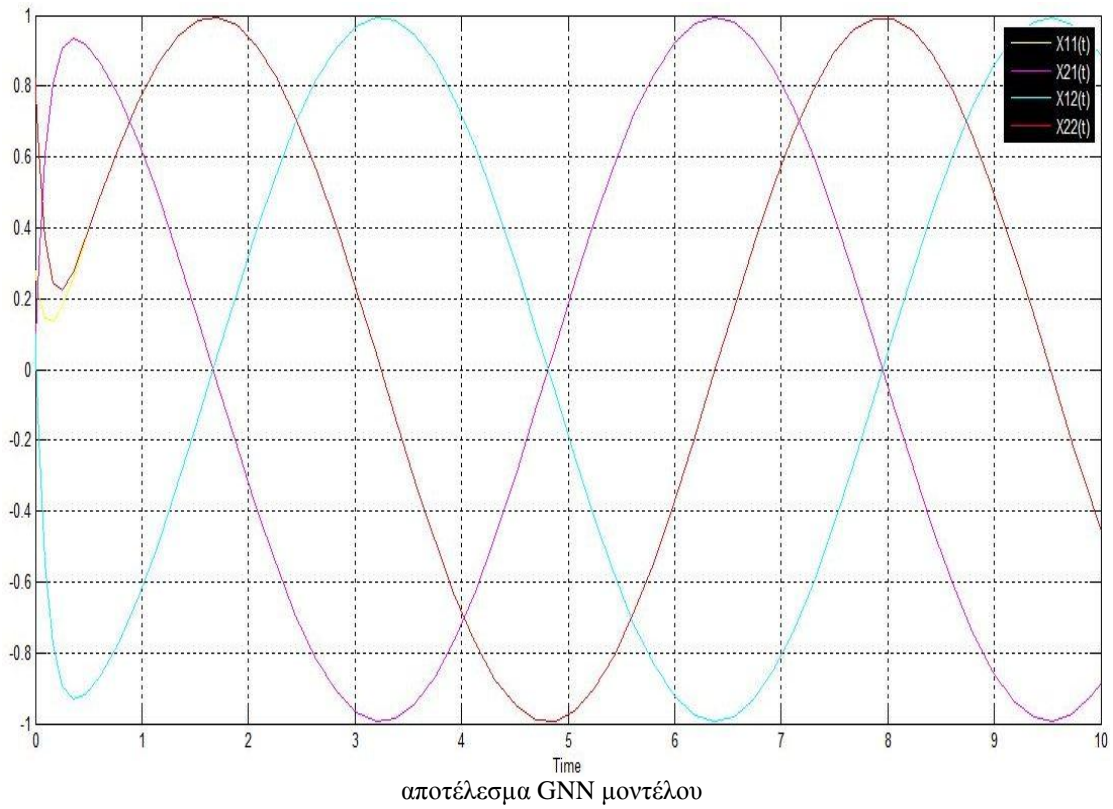
$$\varepsilon_1 = \sin(t) - \cos(t) \cdot i, \quad \varepsilon_2 = \sin(t) + \cos(t) \cdot i$$

Ο πίνακας $A^T(t)A(t)$ έχει ιδιοτιμές:

$$k_1 = \sin^2(t) + \cos^2(t) = 1 \neq 0, \quad k_2 = \sin^2(t) + \cos^2(t) = 1 \neq 0$$

Επομένως ο πίνακας $A(t)$ ικανοποιεί τη συνθήκη αντιστρεψιμότητας (3.5.2.4) και μπορούμε να εφαρμόσουμε το μοντέλο (3.5.2.3).

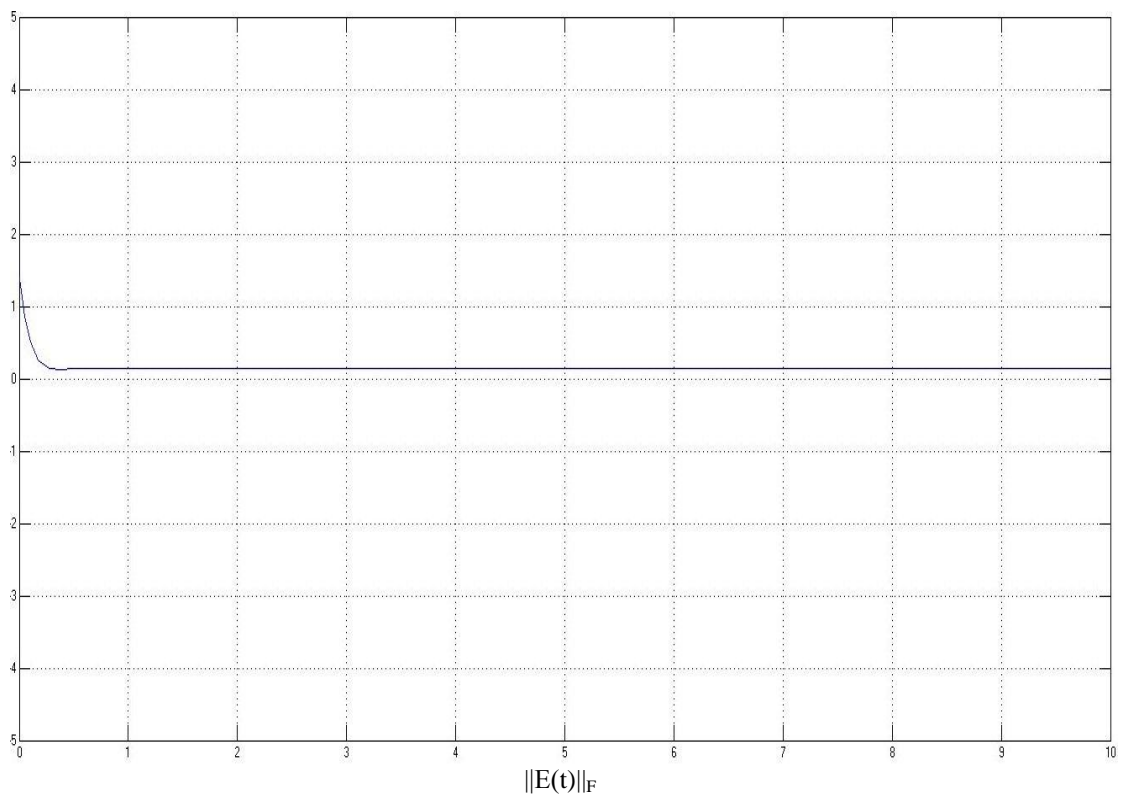
Εφαρμόζοντας το παραπάνω κύκλωμα αντιστροφής για $\underline{\gamma=10}$ έχουμε τα παρακάτω αποτελέσματα:



όπου

$X_{11}(t) \rightarrow$ -----, $X_{21}(t) \rightarrow$ -----, $X_{12}(t) \rightarrow$ -----, $X_{22}(t) \rightarrow$ -----

Όσον αφορά το σφάλμα έχουμε:



Ο παραπάνω πίνακας έχει αντίστροφο τον:

$$X^*(t) = \begin{bmatrix} \sin(t) & -\cos(t) \\ \cos(t) & \sin(t) \end{bmatrix}$$

Παρατήρηση

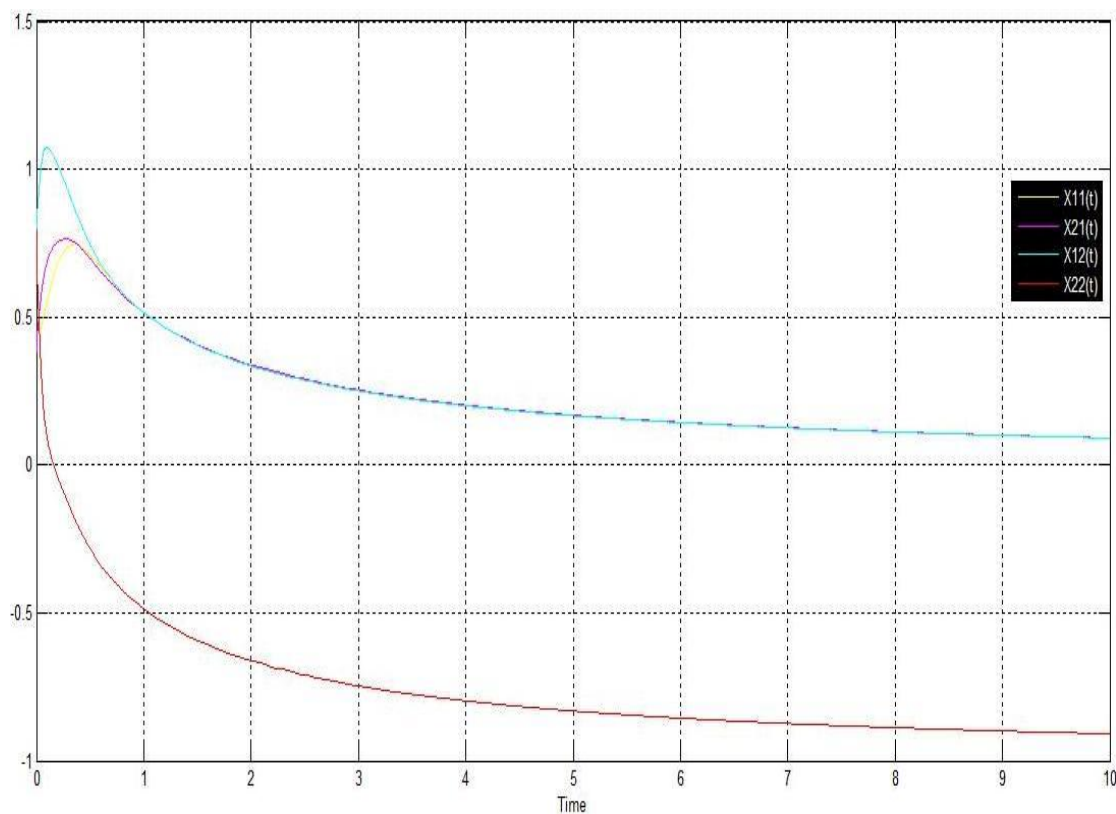
Παρατηρούμε ότι το σφάλμα $E(t)$ τείνει σε τιμή που βρίσκεται λίγο πιο πάνω από τι τιμή 0, εν αντιθέσει με τα ZNN μοντέλα, όπου είχαμε γρηγορότερη σύγκλιση στη τιμή 0. Στη προκειμένη περίπτωση δε κάνουμε μεγέθυνση εφόσον είναι πασιφανές ότι έχουμε μεγάλο σφάλμα.

Εφαρμογή (για πολυωνυμικούς πίνακες)





Έστω ότι θέλουμε να υπολογίσουμε τον αντίστροφο του πολυωνυμικού πίνακα $A(t)$ με τύπο (για $\gamma=10$):

$$A(t) = \begin{bmatrix} t & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} t + \begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix} I$$

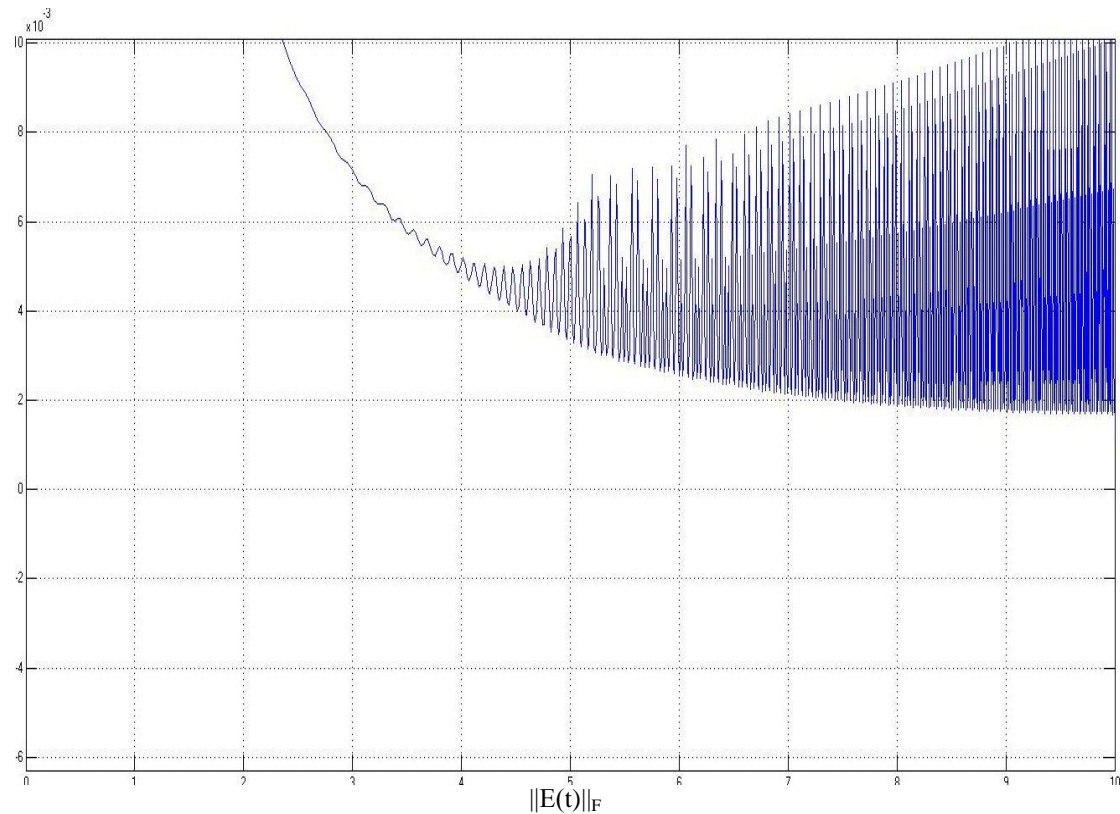
Παίρνουμε τα παρακάτω αποτελέσματα:



όπου

$X_{11}(t) \rightarrow$ , $X_{21}(t) \rightarrow$ , $X_{12}(t) \rightarrow$ , $X_{22}(t) \rightarrow$ 

με σφάλμα (υπό μεγέθυνση):



Παρατηρούμε ότι δεν έχουμε τόσο καλή σύγκλιση όσο θα θέλαμε. Γι' αυτό το λόγο αποτελεί ένα ακόμα κίνητρο αυτή της εργασίας, να συγκρίνουμε τα ZNN με τα GNN μοντέλα (επίσης το λογισμικό μας δίνει τα αποτελέσματα σε αργό ρυθμό).

3.5.4 Συγκριτική μελέτη GNN και ZNN μοντέλων

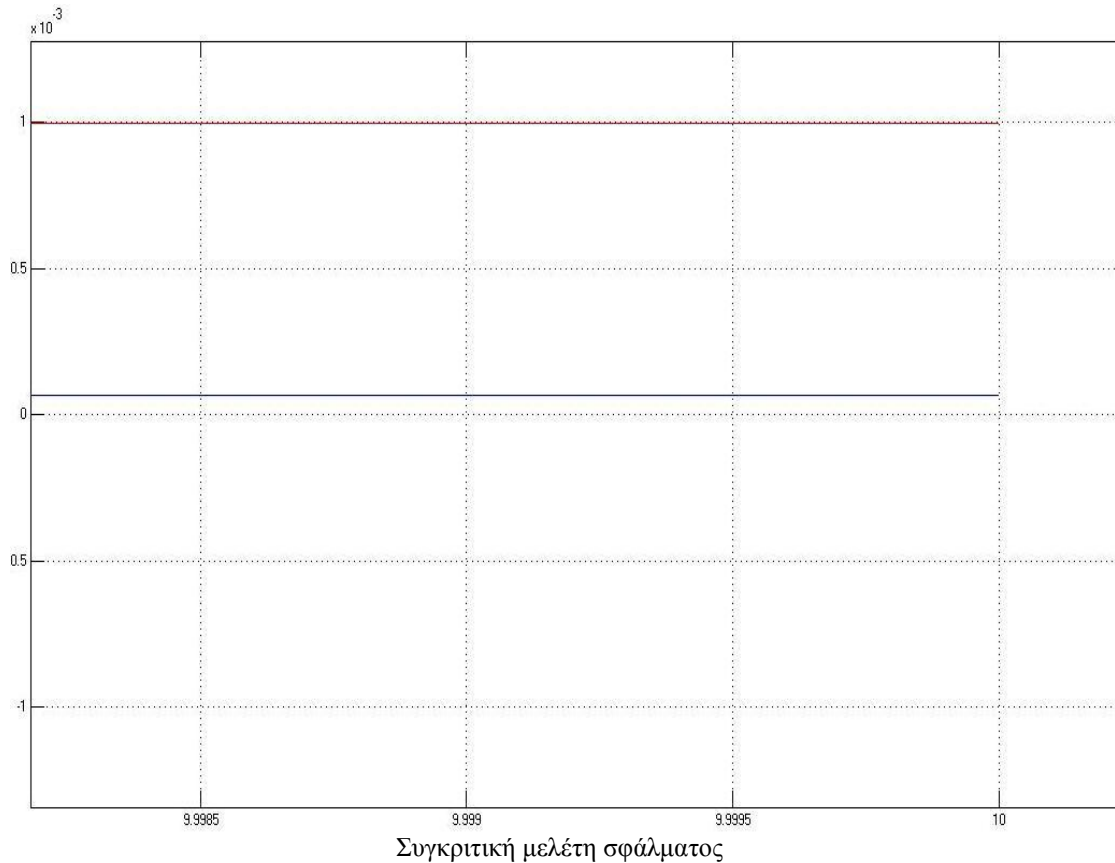
Παρακάτω θα εφαρμόσουμε τις δύο μεθόδους για την ίδια τιμή της παραμέτρου γ .

Εφαρμογή 1

Έστω ότι θέλουμε να συγκρίνουμε τις δύο μεθόδους για τον υπολογισμό του αντιστρόφου του χρονικά μεταβαλλόμενου πίνακα $A(t)$ με τύπο:

$$A(t) = \begin{bmatrix} \sin(t) & \cos(t) \\ -\cos(t) & \sin(t) \end{bmatrix}$$

Παίρνουμε τα παρακάτω συγκριτικά αποτελέσματα



όπου με **κόκκινη γραμμή** αφορά το σφάλμα που δίνει το GNN μοντέλο ενώ με **μπλε γραμμή** το σφάλμα που έχουμε από το ZNN μοντέλο.

Παρατήρηση

Παρατηρούμε ότι τα ZNN μοντέλα δίνουνε καλύτερα αποτελέσματα εφόσον το σφάλμα τείνει πολύ κοντά στη τιμή 0.

Εφαρμογή 2

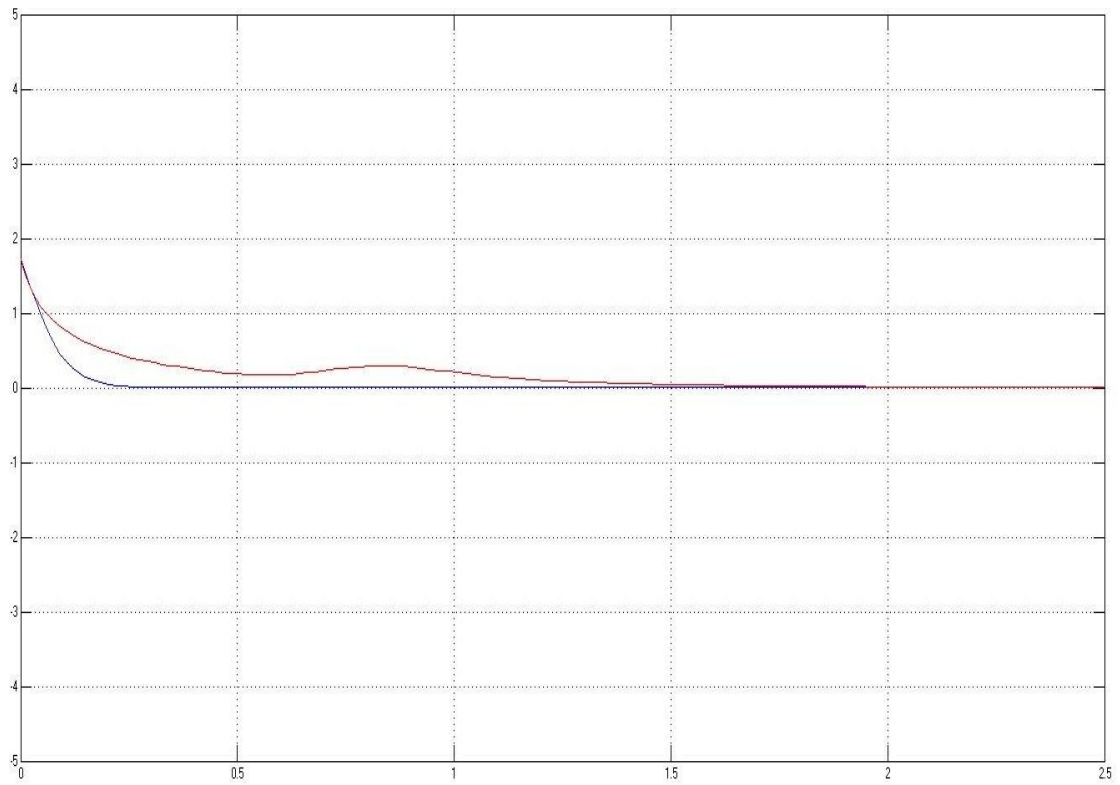
Έστω ότι θέλουμε να συγκρίνουμε τις δύο μεθόδους για τον υπολογισμό του αντιστρόφου του πολυωνυμικού πίνακα $A(t)$ με τύπο:

$$A(t) = \begin{bmatrix} t & -t^4 & 1 \\ 1 & t^3 & 0 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} t^4 + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} t^3 + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} t^2 + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} t + \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} I$$

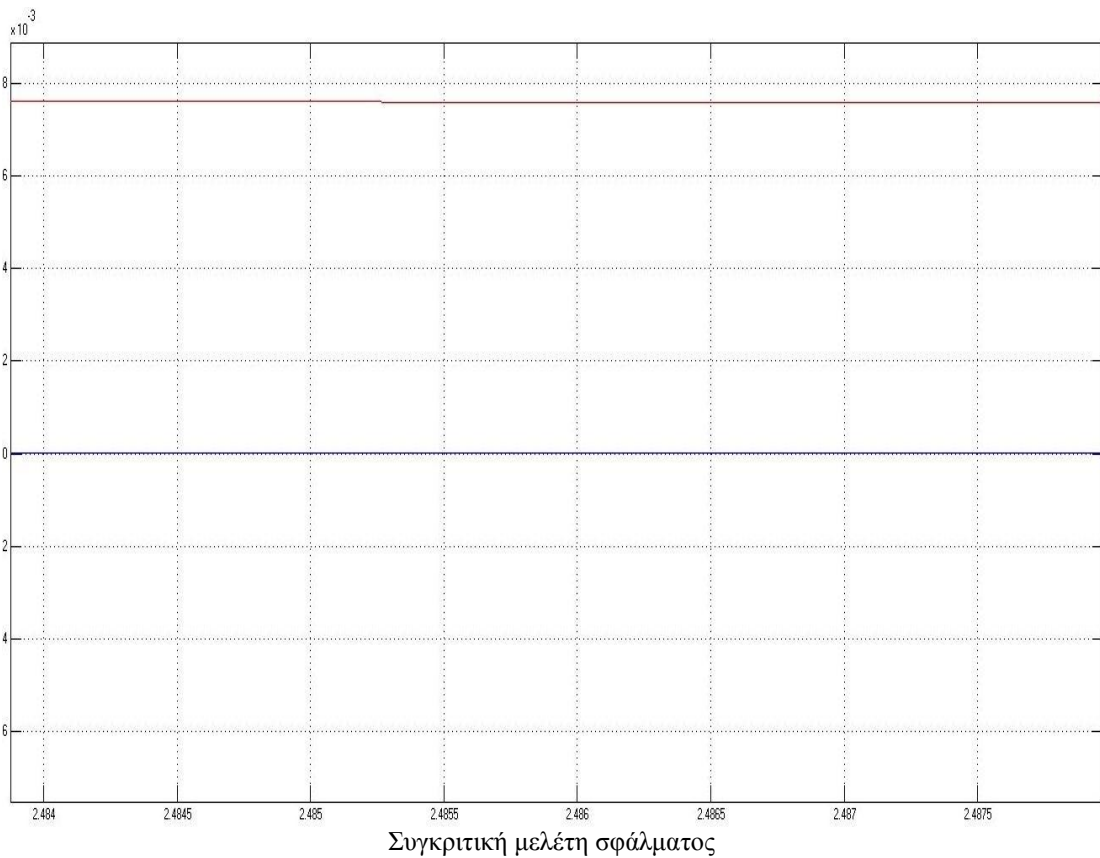
με ορίζουσα

$$\det(A(t)) = 2t^4 + 1 \neq 0$$

Παίρνουμε τα παρακάτω συγκριτικά αποτελέσματα:



Πληκτρολογώντας τη μεγέθυνση:



όπου με **κόκκινη γραμμή** αφορά το σφάλμα που δίνει το GNN μοντέλο ενώ με **μπλέ γραμμή** το σφάλμα που έχουμε από το ZNN μοντέλο.

Παρατήρηση

Τα ZNN παρουσιάζουν γρηγορότερη σύγκλιση από τα GNN.

Συμπεράσματα

Από τα παραπάνω παραδείγματα εξάγουμε τα εξής συμπεράσματα όσον αφορά τη σύγκριση των δύο μοντέλων:

1] Ο σχεδιασμός των ZNN μοντέλων βασίζεται στην εξάλειψη του σφάλματος $E(X(t),t)=A(t)X(t)-I$ ενώ στα GNN στο μηδενισμό της νόρμας του σφάλματος $E(X(t),t)=\|A(t)X(t)-I\|_F^2 / 2$.

2] Τα ZNN δίνουν καλύτερα αποτελέσματα από τα GNN μοντέλα αφού το σφάλμα μεταξύ της θεωρητικής λύσης και της λύσης του ZNN τείνει πολύ κοντά στο 0.

3] Διαφορετικά είδη γραμμικών και μη γραμμικών συναρτήσεων ενεργοποίησης χρησιμοποιούνται στα ZNN μοντέλα, εν αντιθέσει με τα GNN που χρησιμοποιούν μόνο τη γραμμική. Αυτό έχει ως αποτέλεσμα τα ZNN μοντέλα να εμφανίζουν γρηγορότερη σύγκλιση από τα GNN.

4] Στα ZNN έχουμε καλά αποτελέσματα για μικρή τιμή του γ σε σχέση με τα GNN δίκτυα.

Στη συνέχεια θα ασχοληθούμε με το νευρωνικό δίκτυο του κυρίου Stanimorovic [17] που αφορά τον υπολογισμό Drazin αντιστρόφου ενός σταθερού $n \times n$ πίνακα.

3.6 Επαναλαμβανόμενο Νευρωνικό Δίκτυο για τον υπολογισμό του Drazin-Inverse

Έστω $A \in C^{n \times n}$ ένας πίνακας. Ο ελάχιστος μη αρνητικός ακέραιος k για τον οποίο ισχύει:

$$\text{rank}(A^{k+1}) = \text{rank}(A^k) \quad (3.6)$$

καλείται δείκτης του πίνακα A και συμβολίζεται με $\text{ind}(A)=k$. Για λόγους πληρότητας επαναλαμβάνουμε τον ορισμό του Drazin αντιστρόφου. Έστω $A \in C^{n \times n}$ και $\text{ind}(A)=k$. Τότε ο πίνακας $X \in C^{n \times n}$ που ικανοποιεί τις εξής ιδιότητες:

$$A^{l+1}A = A^l, \quad l \geq \text{ind}(A) \quad (3.6.1)$$

$$XAX = X \quad (3.6.2)$$

$$AX = XA \quad (3.6.3)$$

καλείται Drazin αντιστρόφου και συμβολίζεται με $X=A^D$.

Εάν ο πίνακας A είναι ομαλός τότε $\text{ind}(A)=0$ και $A^{-1}=A^D$ αλλιώς, εάν ο πίνακας A είναι μη-ομαλός τότε $\text{ind}(A) \geq 1$. Στόχος μας σε αυτό το κεφάλαιο είναι να ολοκληρωθεί το αποτέλεσμα που δεν έχει διερευνηθεί: Να αναπτύξουμε ένα επαναλαμβανόμενο νευρωνικό δίκτυο που να υπολογίζει τον Drazin αντίστροφο, συνεχίζοντας τη προσέγγιση που έχει γίνει σε παλαιότερες εργασίες.

Αρχιτεκτονική Νευρωνικού δικτύου

Πολλά νευρωνικά δίκτυα έχουν κατασκευαστεί για την αντιστροφή πινάκων και ψευδο-αντιστροφή πινάκων. Σε αυτό το κεφάλαιο θα αποδείξουμε ότι ο Drazin αντίστροφος ενός πίνακα μπορεί να καθοριστεί χρησιμοποιώντας ένα επαναλαμβανόμενο νευρωνικό δίκτυο.

Θα εφαρμόσουμε τα θεωρητικά συμπεράσματα μας σε πραγματικούς τετραγωνικούς πίνακες $A \in R^{n \times n}$. Μπορούμε να ξαναγράψουμε τη σχέση (3.6.1) όπως βλέπουμε παρακάτω:

$$A^{m+1}X - A^m = 0 \quad (3.6.4)$$

όπου $m \geq \text{ind}(A)$, και $X \in R^{n \times n}$ δηλώνει έναν άγνωστο πίνακα που προσεγγίζει το πίνακα A^D . Για να λύσουμε τη παραπάνω εξίσωση πινάκων μπορούμε να ορίσουμε τη διανυσματική συνάρτηση σφάλματος (βασισμένη σε νόρμα) με τύπο:

$$E(t) = \frac{\|A^{m+1}X(t) - A^m\|_F^2}{2}, \quad m \geq \text{ind}(A)$$

Παρατηρούμε ότι $E(t) \geq 0$ άρα η ελάχιστη τιμή $E(\bar{t}) = 0$ της συνάρτησης σφάλματος $E(t)$ επιτυγχάνεται για τη τιμή $\bar{X}(t) = X(\bar{t})$ αν και μόνο αν η $X(\bar{t})$ αποτελεί ακριβή λύση της (3.6.4). Το διάνυσμα κλίσης της συνάρτησης $E(t)$ (ως προς το πίνακα X) είναι:

$$\frac{\partial E(t)}{\partial X} = (A^{m+1})^T (A^{m+1}X(t) - A^m) \quad (3.6.5)$$

Ο όρος $(A^{m+1})^T$ μπορεί να θεωρηθεί σταθερός παράγοντας. Επομένως το διάνυσμα κλίσης εξαρτάται από τη παράσταση $(A^{m+1}X(t) - A^m)$ και ο όρος $(A^{m+1})^T$ να παραληφθεί.

Σύμφωνα με το παρακάτω τύπο:

$$\frac{dX(t)}{dt} = -\gamma \frac{dE(t)}{dX}$$

Καταλήγουμε στο παρακάτω δυναμικό μοντέλο:

$$\begin{aligned} \frac{dX(t)}{dt} &= -\gamma (A^{m+1}X(t) - A^m) \\ m &\geq \text{ind}(A), \quad X(0) = X_0 \end{aligned} \quad (3.6.6)$$

Ο λόγος για τον ορισμό αυτό, έχει να κάνει με το γεγονός ότι εάν υπολογίσουμε το σφάλμα στο πίνακα $\bar{X}(t) = X(\bar{t})$ για το δυναμικό σύστημα (3.6.6) τότε έχουμε το ελάχιστο της συνάρτησης σφάλματος $E(t)$, οπότε συνεπάγεται η παρακάτω εξίσωση:

$$\frac{d\bar{X}}{dt} = 0 \quad (3.6.7)$$

και

$$-\gamma (A^{m+1}\bar{X} - A^m) = 0 \quad (3.6.8)$$

Επίσης η τιμή γ είναι ένας σταθερός αριθμός, τόσο μεγάλος όσο μας το επιτρέπει το λογισμικό πάνω στο οποίο εργαζόμαστε. Από τη (3.6.8) παρατηρούμε ότι ο πίνακας $\bar{X}(t)$ ικανοποιεί την εξίσωση πινάκων (3.6.4).

Το επαναλαμβανόμενο νευρωνικό δίκτυο που ορίζεται παραπάνω είναι ένα δυναμικό σύστημα σε μορφή πινάκων. Από τη θεωρία ελέγχου για γραμμικά συστήματα η λύση την εξίσωσης (πινάκων) δίνεται από το παρακάτω τύπο:

$$X(t) = e^{-\gamma A^{m+1}t} X(0) + \gamma e^{-\gamma A^{m+1}t} \int_0^t e^{\gamma A^{m+1}u} A^m du \quad (3.6.9)$$

Για να αναλύσουμε τη σύγκλιση και ευστάθεια του νευρωνικού δικτύου, είναι απαραίτητο να γνωρίζουμε πληροφορία για τις ιδιοτιμές του πίνακα A^m . Πρώτα θα αναφέρουμε, όμως τη πρόταση (3.6.1) και λήμμα (3.6.1).

Πρόταση 3.6.1 [17]

Εάν x είναι ένα ιδιοδιάνυσμα της αντίστοιχης ιδιοτιμής $\lambda \in C$ του πίνακα $A \in R^{n \times n}$, τότε x είναι ένα ιδιοδιάνυσμα του πίνακα A^m της αντίστοιχης ιδιοτιμής λ^m για κάθε ακέραιο $m \geq 0$.

Χρησιμοποιώντας την πρόταση 3.6.1, καταλήγουμε στο παρακάτω χρήσιμο συμπέρασμα.

Λήμμα 3.6.1 [17]

Έστω ο μη-ομαλός πίνακας $A \in R^{n \times n}$ με φάσμα ιδιοτιμών $\sigma(A) = \{\lambda_1, \dots, \lambda_n\}$ και $m \geq 0$. Υποθέτουμε ότι

$$\operatorname{Re}(\lambda_j^{m+1}) \geq 0, \quad j = 1, 2, \dots, n \quad (2.10.10)$$

Τότε ισχύει:

$$\lim_{t \rightarrow \infty} e^{-\gamma A^{m+1}t} = 0 \quad (2.10.11)$$

Απόδειξη

Η κανονική Jordan μορφή του πίνακα A^{m+1} ορίζεται όπως βλέπουμε παρακάτω:

$$A^{m+1} = P \operatorname{diag}[J_1, J_2, \dots, J_p] P^{-1}$$

όπου ο P είναι ένας αντιστρέψιμος πίνακας και ο όρος $\text{diag}[J_1, J_2, \dots, J_p]$ ένας διαγώνιος πίνακας του οποίου τα Jordan blocks J_k $k=1, 2, \dots, p$ (που βρίσκονται στη κύρια διαγώνιο) είναι της μορφής:

$$J_k = J_k(\lambda_k^{m+1}) = \begin{bmatrix} \lambda_k^{m+1} & 1 & & \\ & \lambda_k^{m+1} & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_k^{m+1} \end{bmatrix} \in \mathbb{R}^{n_k \times n_k}$$

ο εκθετικός πίνακας του $-\gamma A^{m+1} t$ ορίζεται όπως βλέπουμε παρακάτω:

$$e^{-\gamma A^{m+1} t} = P \text{diag}[e^{-\gamma J_1}, e^{-\gamma J_2}, \dots, e^{-\gamma J_p}] P^{-1}$$

Συνεχίζοντας από [24] δεν είναι δύσκολο να παρατηρήσουμε ότι:

$$e^{-\gamma J_k} = e^{-\gamma t \lambda_k^{m+1}} \begin{bmatrix} 1 & -\gamma t & \dots & \frac{(-\gamma t)^{n_k-1}}{(n_k-1)!} \\ & 1 & \ddots & \vdots \\ & & \ddots & -\gamma t \\ & & & 1 \\ & & & & 1 \end{bmatrix} \quad (3.6.12)$$

όπου $k=1, 2, \dots, p$.

Η ιδιοτιμή λ_k^{m+1} της ιδιοτιμής λ_k είναι ένας μιγαδικός αριθμός για τον οποίο ξέρουμε ότι:

$$\lambda_k^{m+1} = u_k + i v_k$$

όπου i είναι η φανταστική μονάδα, και u_k, v_k το πραγματικό και φανταστικό μέρος της ιδιοτιμής λ_k^{m+1} .

Από τη (3.6.10) έχουμε ότι η παρακάτω ανισότητα:

$$-\gamma \text{Re}(\lambda_k^{m+1}) = -\gamma u_k \leq 0$$

ισχύει για κάθε $\lambda_k \in \sigma(A)$. Επειδή τα πραγματικά μέρη όλων των ιδιοτιμών που βρίσκονται στο $\sigma(A^{m+1})$ είναι μη αρνητικά έχουμε ότι:

$$\lim_{t \rightarrow \infty} e^{-\gamma t \lambda_k^{m+1}} \cdot (-\gamma t)^l = \lim_{t \rightarrow \infty} e^{(t(-\gamma u_k))(\cos(-\gamma u_k t) + i \sin(-\gamma u_k t))} \cdot (-\gamma t)^l = 0$$

για κάθε l , $0 \leq l \leq n_k - 1$. Από τη σχέση (3.6.12) το όριο $\lim_{t \rightarrow \infty} e^{-\gamma t J_k}$ είναι ο μηδενικός πίνακας για κάθε $k=1, 2, \dots, p$. Από αυτό συνεπάγεται περαιτέρω ότι το όριο $\lim_{t \rightarrow \infty} e^{-\gamma A^{m+1} t}$ είναι ο μηδενικός πίνακας.

Τώρα, από τη σχέση (3.6.9) και το λήμμα (3.6.1) συνεπάγεται η παρακάτω παρουσίαση για το όριο $\lim_{t \rightarrow \infty} X(t) = \bar{X}(t)$

$$\bar{X} = \lim_{t \rightarrow \infty} \gamma e^{-\gamma A^{m+1} t} \int_0^t e^{\gamma A^{m+1} u} A^m du \quad (3.6.13)$$

Μπορεί να επαληθευτεί με βάση τον ορισμό και τις ιδιότητες του εκθετικού πίνακα, ότι η λύση \bar{X} της (3.6.13) αναπαριστά τον Drazin αντίστροφο και εξαρτάται από τη τιμή γ . Συνεχίζοντας έχουμε το παρακάτω θεώρημα.

Θεώρημα 3.6.1 [17]

Έστω ο πίνακας $A \in R^{n \times n}$ με φάσμα ιδιοτιμών $\sigma(A) = \{\lambda_1, \dots, \lambda_n\}$ και $m \geq \text{ind}(A)$. Υποθέτουμε ότι ικανοποιείται η συνθήκη (3.6.10), τότε η έκφραση (3.6.13) παράγει τον Drazin αντίστροφο A^D δηλαδή:

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

Απόδειξη

Εφαρμόζοντας αρκετούς μετασχηματισμούς και ιδιότητες που αφορούν τον Drazin αντίστροφο έχουμε:

$$\begin{aligned} \bar{X} &= \lim_{t \rightarrow \infty} \gamma e^{-\gamma A^{m+1} t} \int_0^t e^{\gamma A^{m+1} u} A^m du \\ &= \lim_{t \rightarrow \infty} \left[e^{-\gamma A^{m+1} t} \int_0^t (e^{\gamma A^{m+1} u}) (\gamma A^{m+1}) du \right] A^D \\ &= \left[\lim_{t \rightarrow \infty} e^{-\gamma A^{m+1} t} e^{\gamma A^{m+1} u} \Big|_{u=0}^{u=t} \right] A^D \end{aligned}$$

$$\begin{aligned}
&= \left\{ \lim_{t \rightarrow \infty} e^{-\gamma A^{m+1} t} \left[e^{\gamma A^{m+1} t} - I \right] \right\} A^D \\
&= \left[I - \lim_{t \rightarrow \infty} e^{-\gamma A^{m+1} t} \right] A^D
\end{aligned}$$

Από το λήμμα (3.6.1) έχουμε ότι

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

και η απόδειξη τελείωσε.

Η εξίσωση (3.6.9) μπορεί να απλοποιηθεί, εάν θέσουμε τον πίνακα κατάστασης να έχει για αρχική κατάσταση το μηδενικό πίνακα ($X(0)=0$) οπότε ο όρος:

$$e^{-\gamma A^{m+1} t} X(0)$$

θα μηδενιστεί. Επομένως μπορεί να χρησιμοποιηθεί το παρακάτω δυναμικό μοντέλο για να υπολογιστεί ο Drazin αντίστροφος ενός πίνακα.

$$\begin{aligned}
\frac{dX(t)}{dt} &= -\gamma(A^{m+1}X(t) - A^m), \\
m &\geq \text{ind}(A), \quad X(0) = X_0 = 0
\end{aligned} \tag{3.6.14}$$

Συνεχίζοντας θα ασχοληθούμε με τη ευστάθεια και τη σύγκλιση του δικτύου (3.6.14) χρησιμοποιώντας την θεωρία ευστάθειας Lyapunov. Πρώτα ας αναφέρουμε τα παρακάτω θεωρητικά αποτελέσματα.

Πρόταση 3.6.2 [17]

Εάν C και D είναι $2 \times n$ πραγματικοί συμμετρικοί, θετικά ημί-ορισμένοι πίνακες, τότε

$$\lambda_{\min}(C)Tr(D) \leq Tr(CD) \leq \lambda_{\max}(C)Tr(D) \tag{3.6.15}$$

όπου $\lambda_{\min}(A)$ είναι η ελάχιστη ιδιοτιμή του πίνακα A και $\lambda_{\max}(A)$ η μέγιστη, αντίστοιχα.

Λήμμα 3.6.2 [17]

Έστω M και N 2 θετικά ημί-ορισμένοι πίνακες. Τότε ισχύουν τα ακόλουθα συμπεράσματα.

$$\bullet \text{Tr}(MN) \geq 0$$

$$\bullet \text{Tr}(M) = 0 \Leftrightarrow M = 0.$$

Στο λήμμα 3.6.3 θα αποδείξουμε ότι το ερμιτιανό μέρος ενός πραγματικού τετραγωνικού πίνακα M που ορίζεται σύμφωνα με το παρακάτω τύπο:

$$H(M) = \frac{1}{2}(M + M^T)$$

κατέχει μία χρήσιμη ιδιότητα.

Λήμμα 3.6.3 [17]

Έστω M και N δύο $n \times n$ πραγματικοί τετραγωνικοί πίνακες. Εάν ο πίνακας N είναι συμμετρικός, τότε ισχύει η παρακάτω ισότητα:

$$\text{Tr}(MN) = \text{Tr}(H(M)N) \quad (3.6.16)$$

Απόδειξη

Ορίζουμε την (ij) -θέση του πίνακα M (ανάλογα για το πίνακα N) με το σύμβολο m_{ij} (ανάλογα για το πίνακα N n_{ij}). εφαρμόζοντας τον ορισμό της συνάρτησης Trace και αρκετούς αλγεβρικούς μετασχηματισμούς έχουμε ότι:

$$\begin{aligned} \text{Tr}(H(M)N) &= \sum_{i=1}^n (H(M)N)_{ii} \\ &= \sum_{i=1}^n \sum_{j=1}^n H(M)_{ij} n_{ji} \\ &= \frac{1}{2} \left(\sum_{i=1}^n \sum_{j=1}^n m_{ij} n_{ji} + \sum_{i=1}^n \sum_{j=1}^n m_{ji} n_{ji} \right) \end{aligned}$$

Αφού ο πίνακας N είναι συμμετρικός, επιπλέον έχουμε:

$$\begin{aligned} \text{Tr}(H(M)N) &= \frac{1}{2} \left(\sum_{i=1}^n \sum_{j=1}^n m_{ij} n_{ji} + \sum_{j=1}^n \sum_{i=1}^n m_{ji} n_{ji} \right) \\ &= \frac{1}{2} \left(\sum_{i=1}^n (MN)_{ii} + \sum_{j=1}^n (MN)_{jj} \right) \end{aligned}$$

$$= Tr(MN)$$

που ήταν ο αρχικός μας στόχος.

Θεώρημα 3.6.2 [17]

Έστω $A \in R^{n \times n}$ ένας μη-ομαλός πίνακας με φάσμα $\sigma(A) = \{\lambda_1, \dots, \lambda_n\}$. Υποθέτουμε ότι ισχύει η σχέση (3.6.10). Τότε το νευρωνικό δίκτυο (3.6.14) είναι ευσταθές λόγω θεωρίας Lyapunov.

Πόρισμα 3.6.1 [17]

Έστω $A \in R^{n \times n}$ ένας ομαλός πίνακας και έστω ότι ισχύει η υπόθεση (3.6.10). Τότε το νευρωνικό δίκτυο (3.6.14) είναι ασυμπτωτικά ευσταθές λόγω της θεωρίας Lyapunov.

Από το θεώρημα (3.6.1) καταλήγουμε ότι στόχος μας είναι να βρούμε εκείνον τον ακέραιο m για τον οποίο ο πίνακας A^{m+1} έχει ιδιοτιμές με μη-αρνητικό πραγματικό μέρος, έτσι ώστε το νευρωνικό δίκτυο (3.6.14) να παράγει τον Drazin αντίστροφο δηλαδή:

$$\sigma(A^{m+1}) \subset \{z : \operatorname{Re}(z) \geq 0\} \quad (3.6.17)$$

Υπάρχουν πολλές προϋποθέσεις σύμφωνα με τους οποίους μπορούμε να επιλέξουμε τη τιμή m , που να μας εγγυάται τη σχέση (3.6.17). Αυτές τις προϋποθέσεις, θα τις συζητήσουμε στο θεώρημα (3.6.3). Πρώτα θα αναφέρουμε το λήμμα (3.6.4) για να καταλάβουμε τις προϋποθέσεις του θεωρήματος (3.6.3).

Λήμμα 3.6.4 [17]

Έστω $A \in R^{n \times n}$ με $\lambda_j \in \sigma(A)$ και k ένας θετικός ακέραιος. Τότε η συνθήκη $\operatorname{Re}(\lambda_j^k) \geq 0$ επιτυγχάνεται στις παρακάτω περιπτώσεις:

$$\Sigma 1: \lambda_j \in R^+$$

$$\Sigma 2: \lambda_j \in R^-, k \text{ άρτιος}$$

$$\Sigma 3: \lambda_j \text{ και ο } k \text{ ικανοποιεί:}$$

$$\phi_j = \operatorname{Arg} \lambda_j = \pm \frac{\pi}{2}, \quad k \in \{4l : l \in \mathbb{N}^+\} \quad (3.6.18)$$

$\Sigma_4: \lambda_j$ ικανοποιεί:

$$\lambda_j \in \mathbb{C} \cap \{z = r_j e^{i\phi_j} : 0 < |\phi_j| < \pi\} \quad (3.6.19)$$

και η παράμετρος k ικανοποιεί τη:

$$\frac{4s-1}{2} \frac{\pi}{|\phi_j|} < k < \frac{4s+1}{2} \frac{\pi}{|\phi_j|}, \quad s \in N_0 \quad (3.6.20)$$

Απόδειξη

Η απόδειξη για τις συνθήκες Σ_1 - Σ_2 - Σ_3 είναι προφανή. Όσον αφορά τη συνθήκη Σ_4 έχουμε:

$$\lambda_j = a_j + ib_j = r_j e^{i\phi_j} = r_j (\cos(\phi_j) + i \sin(\phi_j))$$

όπου

$$r_j = |\lambda_j| = \sqrt{a_j^2 + b_j^2}$$

και

$$\phi_j = \begin{cases} -1 & , & a_j = 0 & , & b_j = 0 \\ \frac{\pi}{2} & , & a_j = 0 & , & b_j > 0 \\ -\frac{\pi}{2} & , & a_j = 0 & , & b_j < 0 \\ \arctan\left(\frac{b_j}{a_j}\right) & , & & & a_j \neq 0 \end{cases}$$

Το πραγματικό μέρος της παρακάτω ιδιοτιμής :

$$\lambda_j^k = r_j^k (\cos(k\phi_j) + i \sin(k\phi_j))$$

ισούται με:

$$\operatorname{Re}(\lambda_j^k) = r_j^k \cos(k\phi_j)$$

Επομένως, η συνθήκη $\operatorname{Re}(\lambda_j^k) > 0$ ικανοποιείται στη περίπτωση που $\cos(k|\phi_j|) > 0$, ή ισοδύναμα:

$$-\frac{\pi}{2} + 2s\pi < k|\phi_j| < \frac{\pi}{2} + 2s\pi, \quad s \in \mathbb{N}_0$$

Επομένως, από τη συνθήκη (3.6.20) και για $\gamma > 0$ έχουμε ότι $\gamma \operatorname{Re}(\lambda_j^k) \geq 0$. Το μήκος της σχέσης (3.6.20) που περιορίζει τις τιμές που μπορεί να πάρει το k ισούται με $\pi/|\phi_j|$. Επομένως

$$\frac{\pi}{|\phi_j|} > 1 \quad (3.6.21)$$

Ως εκ τούτου η αναγκαιότητα των (3.6.19), (3.6.20) επαληθεύεται.

Θεώρημα 3.6.3 [17]

Έστω $A \in \mathbb{R}^{n \times n}$ ένα τετράγωνος πραγματικός πίνακας με δείκτη $k = \operatorname{ind}(A)$ και φάσμα ιδιοτιμών $\sigma(A) = \{\lambda_j, j=1,2,\dots,n\}$. Για κάθε ιδιοτιμή λ_j , θεωρούμε το σύνολο των πιθανών τιμών για το οποίο οι ιδιοτιμές του πίνακα A^{m+1} έχουν μη αρνητικό πραγματικό μέρος (δηλαδή $\operatorname{Re}(\lambda_j^{m+1}) \geq 0$). Το σύνολο M_1^j ορίζεται:

$$M_1^j = [l, +\infty], \quad l \geq \operatorname{ind}(A)$$

για κάθε $j=1,2,\dots,n$

Συνεχίζοντας, έστω M_2^j το σύνολο των τιμών της παραμέτρου m που ορίζεται, από την εφαρμογή των περιορισμών που επιβάλλονται στη συνθήκη $\Sigma 2$ από το λήμμα 3.6.4 στην ιδιοτιμή $\lambda_j \in \sigma(A)$. Παρομοίως με M_3^j, M_4^j ορίζουμε τα σύνολα των αποδεκτών τιμών της τιμής m εφαρμόζοντας τις συνθήκες $\Sigma 3, \Sigma 4$ από το λήμμα 3.6.4 αντίστοιχα.

Επομένως έχουμε:

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

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

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

Αργότερα, το σύνολο των θετικών ακεραίων που αντιστοιχεί στην ιδιοτιμή λ_j το συμβολίζουμε με U_j . Είναι εύκολο να καταλήξουμε ότι $U_j = M_2^j$ ή $U_j = M_3^j$ ή $U_j = M_4^j$. Το επαναλαμβανόμενο νευρωνικό δίκτυο που ορίζεται στη σχέση (3.6.14) είναι ευσταθές και ο πίνακας κατάστασης $X(t)$ συγκλίνει στον Drazin αντίστροφο. Δηλαδή $\lim_{t \rightarrow +\infty} X(t) = A^D$ στις παρακάτω περιπτώσεις:

Συνθήκη 1:

$$\sigma(A) \subset \mathbb{R}^+$$

και

$$m+1 \geq \text{ind}(A)$$

Συνθήκη 2:

$$\sigma(A) \subset \mathbb{R}, \lambda_j < 0 \text{ για τουλάχιστον ένα δείκτη } j$$

και

$$m+1 \geq \text{ind}(A), \text{ με } m+1 \text{ άρτιο}$$

Συνθήκη 3:

$$\sigma(A) \subset \mathbb{C} \cap \{z = re^{i\theta} : 0 < |\theta| < \pi\} \quad (3.6.22)$$

και

$$m+1 = \min \left(\bigcap_{i=1}^n (U_j \cap M_i^j) \right) \quad (3.6.23)$$

όπου

$$U_j = M_1^j \text{ στη περίπτωση που } \lambda_j \geq 0.$$

Απόδειξη

Είναι απαραίτητο να εξασφαλίσουμε την ύπαρξη της συνθήκης (3.6.17). Επιπλέον, για να εξασφαλίσουμε την ύπαρξη του Drazin-Inverse, πρέπει να ισχύει $m \geq \text{ind}(A)$ σε όλες τις περιπτώσεις. Η απόδειξη στη Συνθήκη 1 και Συνθήκη 2 είναι προφανές από

το λήμμα 3. Από πρόταση 1 έχουμε $\sigma(A)=\{\lambda_j^{m+1}, j=1,2,\dots,n\}$. Στη Συνθήκη 1 και Συνθήκη 2 η (3.6.19) ικανοποιείται. Σύμφωνα με το λήμμα (3.6.4), η σχέση (3.6.22) εξασφαλίζει ότι $\text{Re}(\lambda_j^m) \geq 0$. Τελικά (3.6.22), (3.6.23) εξασφαλίζουν την (3.6.17)

Η επιλογή της παραμέτρου m σύμφωνα με την (3.6.23) σε συνδυασμό με το γεγονός ότι $\gamma > 0$ εξασφαλίζει ότι το φάσμα του πίνακα A^{m+1} από τη σχέση (3.6.6) ανήκει στο ανοικτό δεξιά μέρος του μιγαδικού επιπέδου.

Σύνθεση Δικτύου

Παρόμοια με νευρωνικά δίκτυα που έχουν αναπτυχθεί για αντιστροφή πινάκων [18], [13] το νευρωνικό δίκτυο που αναλύσαμε διαιρείται σε έναν αριθμό ανεξάρτητων υπό-δικτύων. Κάθε υπό-δίκτυο αποτελείται από ένα διάνυσμα στήλης του πίνακα κατάστασης $X(t)=V(t)$. Ας συμβολίσουμε $v_j(t)$ (παρόμοια, a_j^m) το j -στό διάνυσμα στήλης του πίνακα $V(t)$ (παρόμοια, A^m). Η δυναμική του j -στού υπό-δικτύου μπορεί να εκφραστεί με τη παρακάτω διαφορική εξίσωση:

$$\frac{dv_j(t)}{dt} = -\gamma(A^{m+1}v_j(t) - a_j^m) \quad (3.6.24)$$

Έστω W ο πίνακας βαρών (σύνδεσης) με τύπο $W = -\gamma A^{m+1}$ και :

$$\gamma a_j^m = \{\gamma a_{1j}^m, \dots, \gamma a_{nj}^m\}$$

Τα στοιχεία του πίνακα W (αντίστοιχα του V) τα συμβολίζουμε με w_{ij} (αντίστοιχα v_{ij}). Οπότε τα στοιχεία του άγνωστου πίνακα V υπολογίζονται από τη παρακάτω σχέση

$$\dot{v}_{ij} = \frac{dv_{ij}}{dt} = \sum_{k=1}^n v_{kj} w_{ik} + \gamma a_{ji}^m, \quad i, j = 1, 2, \dots, n \quad (3.6.25)$$

Παρατήρηση 1 [17]

Ο χρόνος σύγκλισης του επαναλαμβανόμενου νευρωνικού δικτύου είναι περίπου ίσος με:

$$\frac{5}{|\lambda_{\min}(W)|}$$

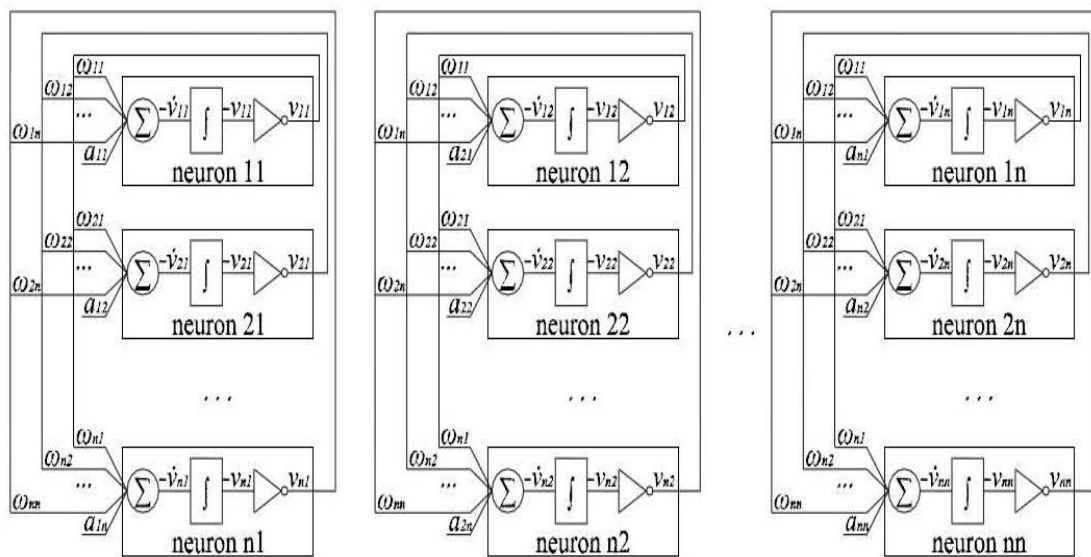
όπου $\lambda_{\min}(W)$ είναι η ελάχιστη μη μηδενική ιδιοτιμή του πίνακα W .

Παρατήρηση 2 [17]

Η παράμετρος γ πρέπει να επιλέγεται τόσο μεγάλη, όσο μας επιτρέπει το λογισμικό πάνω στο οποίο εργαζόμαστε έτσι ώστε να επιταχυνθεί η διαδικασία του υπολογισμού Drazin αντιστρόφου, διότι τότε ο όρος:

$$\lim_{t \rightarrow +\infty} e^{-\gamma A^{m+1}t}$$

τείνει γρηγορότερα στο μηδενικό πίνακα.



Αρχιτεκτονική δικτύου για τον υπολογισμό του Drazin αντιστρόφου [17]

Παραδείγματα στο Matlab

Παράδειγμα 3.6.1

Έστω ο πίνακας:

$$A_1 = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & -1 & 0 \end{bmatrix}$$

Παρατηρούμε ότι $\text{rank}(A_1) = \text{rank}(A_1^2) = 2$, οπότε $\text{ind}(A_1) = 1$. Ο πίνακας A_1 έχει φάσμα ιδιοτιμών το σύνολο:

$$\sigma(A_1) = \{0, 1, 1\}$$

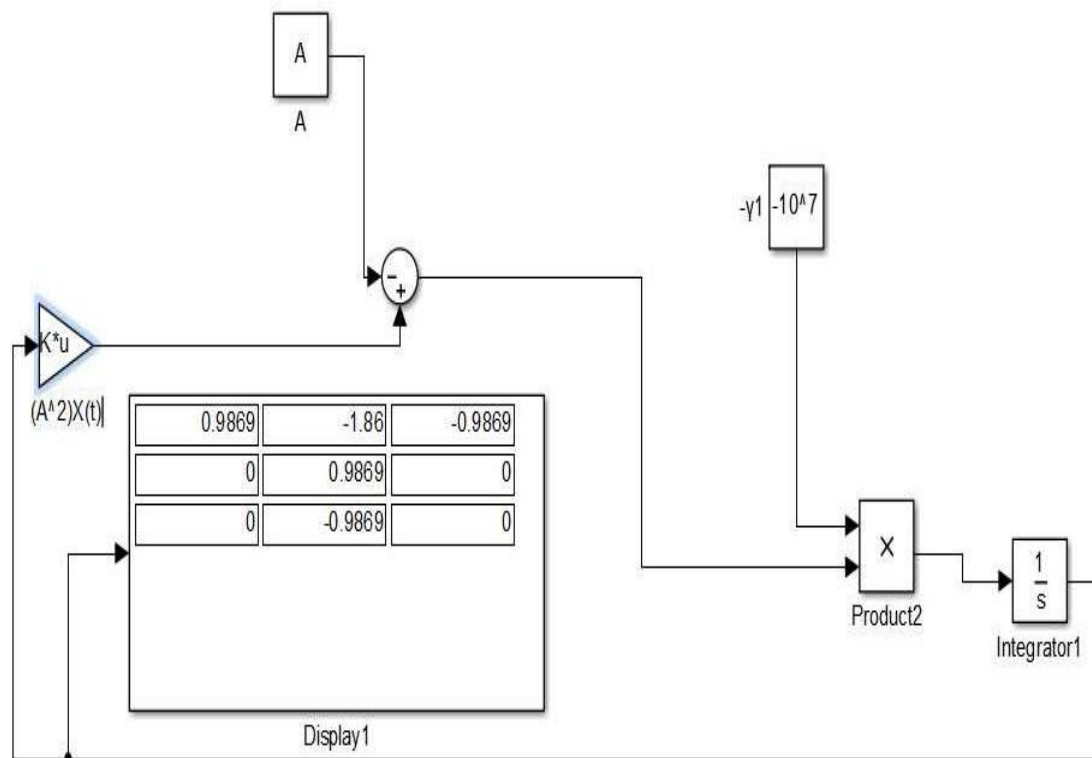
συνεπώς όλες οι ιδιοτιμές είναι μη αρνητικές άρα ισχύει η συνθήκη (2.10.10) και $\sigma(A^{m+1}) \subset \{z: \text{Re}(z) \geq 0\}$ οπότε επιλέγουμε $m = \text{ind}(A_1) = 1$ (παρατηρούμε ότι ο A^2 έχει ιδιοτιμές το σύνολο $\{0, 1, 1\}$). Ο πίνακας $W = -\gamma A^{m+1} = -\gamma A^2$ έχει φάσμα ιδιοτιμών το σύνολο:

$$\sigma(W) = \{-\gamma, -\gamma, 0\}$$

Επιλέγοντας $\gamma = 10^7$ έχουμε $\{-10^7, -10^7, 0\}$ έχουμε ότι $|\lambda_{\min}(W)| = 10^7$ και σύγκλιση μετά από:

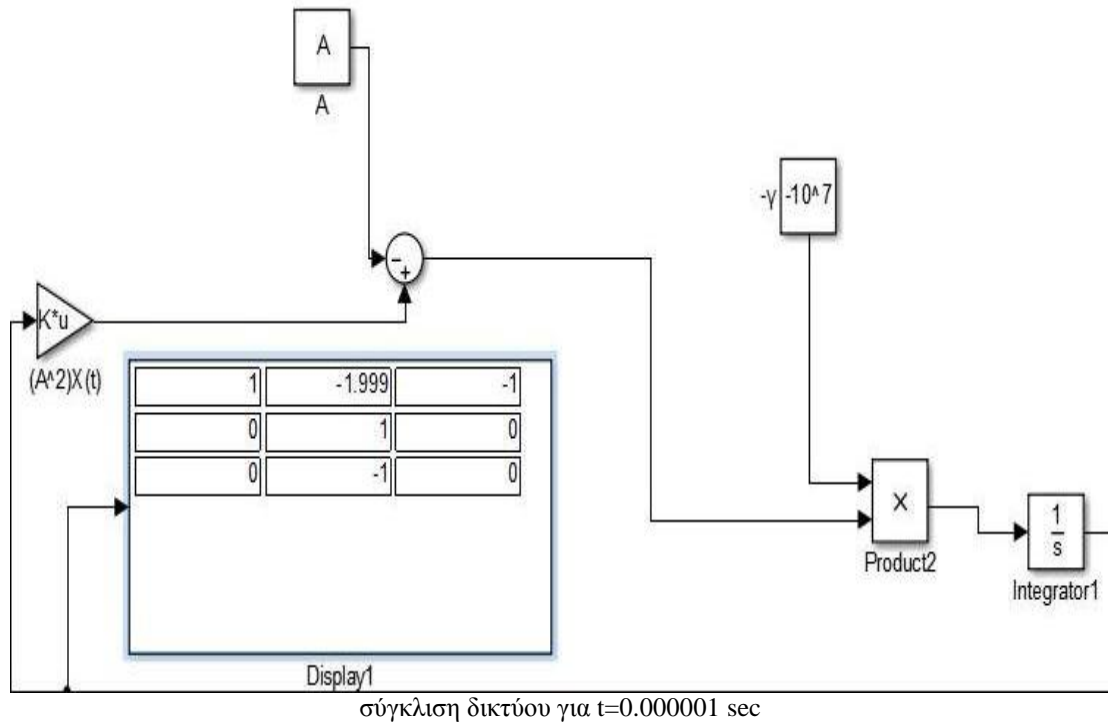
$$\frac{5}{|\lambda_{\min}(W)|} = 0.0000005 \text{ sec}$$

Τέλος μας μένει να υπολογίσουμε τον Drazin αντίστροφο του πίνακα A_1 σύμφωνα με τη σχέση (2.10.14). Στο Matlab έχουμε τα παρακάτω αποτελέσματα:



σύγκλιση δικτύου για $t = 0.0000005 \text{ sec}$

Μετά από χρόνο $t = 0.000001 \text{ sec}$:



Ενώ το θεωρητικό αποτέλεσμα είναι:

$$A_1^D = \begin{bmatrix} 1 & -2 & -1 \\ 0 & 1 & 0 \\ 0 & -1 & 0 \end{bmatrix}$$

Παράδειγμα 3.6.2

Παρατηρώντας το παραπάνω παράδειγμα αποφασίζουμε να αλλάξουμε το στοιχείο που βρίσκεται στην τρίτη γραμμή και τρίτη στήλη του πίνακα A_1 από 0 σε 10^{-4} . Δηλαδή:

$$A = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & -1 & 10^{-4} \end{bmatrix}$$

στη προκειμένη περίπτωση ο πίνακας έχει ορίζουσα $\det(A) = 10^{-4} \neq 0$ (αλλά πολύ κοντά 0). Οπότε ο πίνακας A είναι αντιστρέψιμος $A^{-1} = A^D$ και $\text{ind}(A) = 0$. Επίσης ο πίνακας A έχει φάσμα ιδιοτιμών το σύνολο:

$$\sigma(A) = \{1, 1, 10^{-4}\}$$

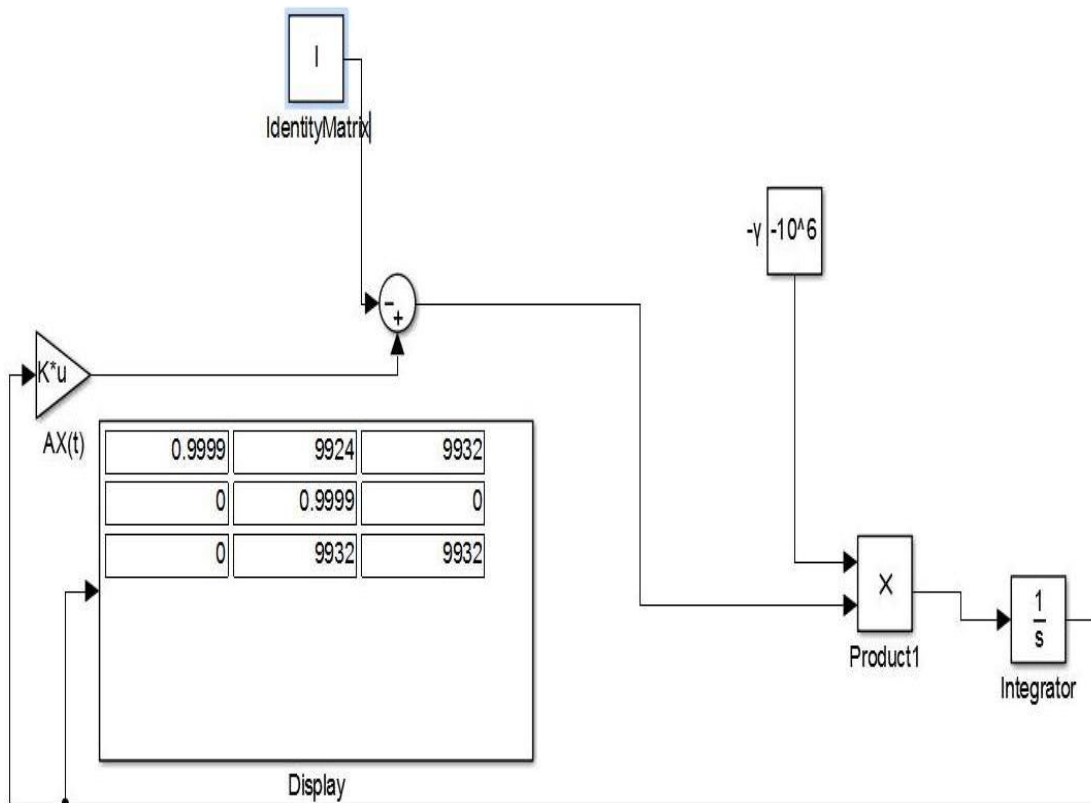
Οπότε επιλέγουμε $m=0$ (εφόσον ο A έχει ιδιοτιμές που ανήκουν μόνο στη πρώτη περίπτωση) για $\gamma=10^6$ ο πίνακας $W=-\gamma A^{m+1}=-10^6 \cdot A$ έχει φάσμα ιδιοτιμών το σύνολο:

$$\sigma(W)=\{10^6, 10^6, 100\}$$

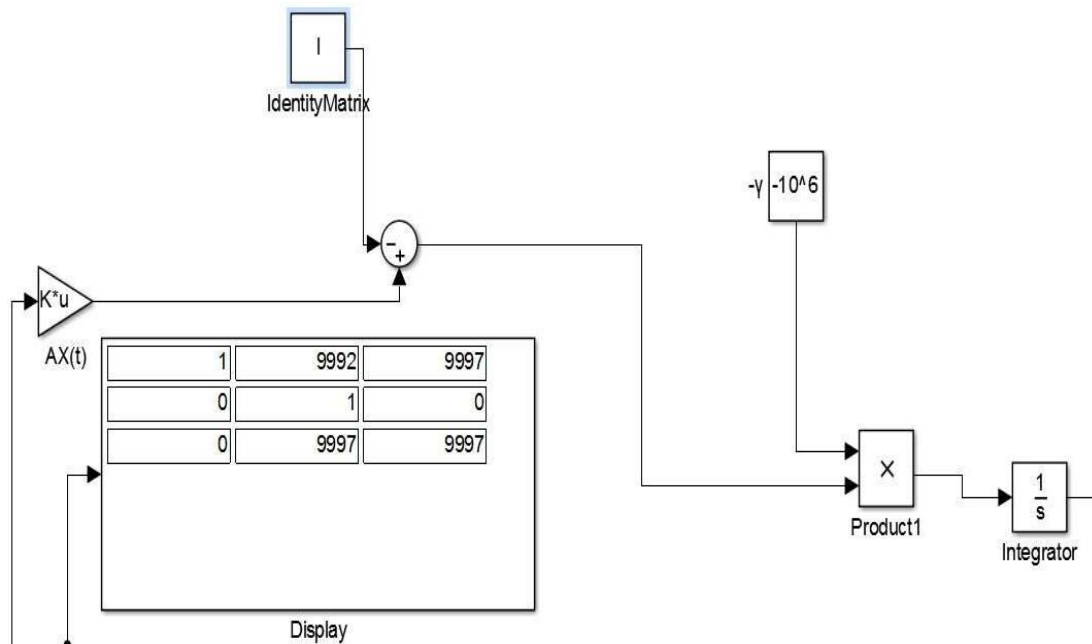
Οπότε $|\lambda_{\min}(W)|=100$, συνεπώς η σύγκλιση επιτυγχάνεται περίπου μετά από:

$$\frac{5}{|\lambda_{\min}(W)|} = 0.05 \text{ sec}$$

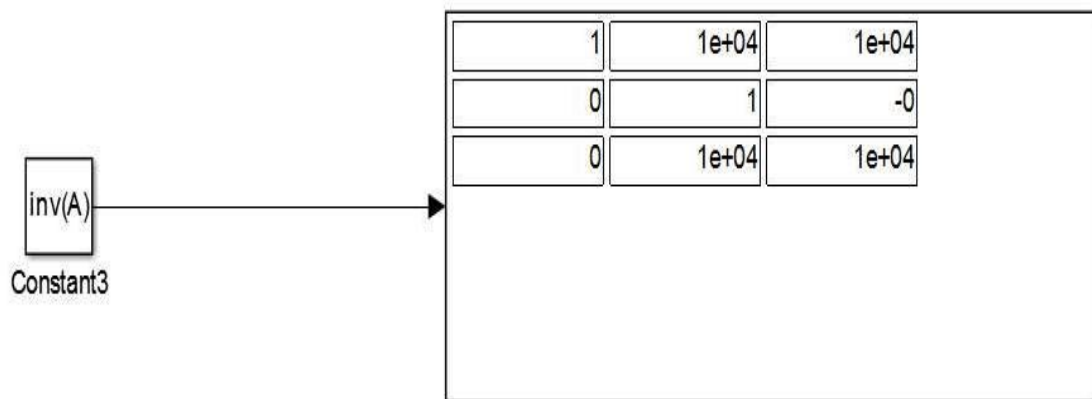
Στο περιβάλλον του Matlab παίρνουμε τα παρακάτω αποτελέσματα και όπως θα παρατηρήσουμε μετά από 0.05 sec έχουμε



Μετά από χρόνο $t=0.08 \text{ sec}$ παρατηρούμε ότι το αποτέλεσμα του δικτύου συγκλίνει καλύτερα στη λύση όπως βλέπουμε στη παρακάτω εικόνα.



σύγκλιση δικτύου για t=0.08 sec



Θεωρητική λύση

Παράδειγμα 3.6.3

Να υπολογιστεί ο Drazin αντίστροφος του πίνακα A:

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

Ο πίνακας A έχει δείκτη $\text{ind}(A)=2$. Φάσμα ιδιοτιμών το σύνολο:

$$\sigma(A) = \{3, 2, 2, 1, 0, 0\}$$

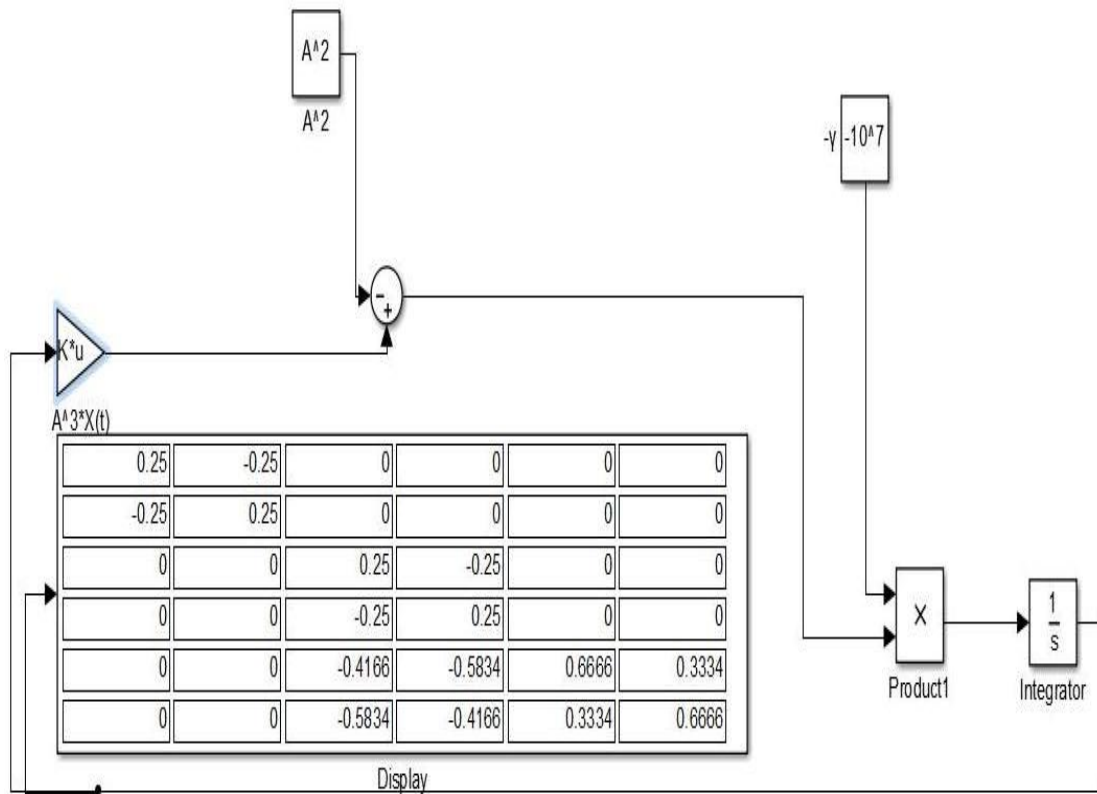
Παρατηρούμε ότι ο πίνακας A δεν έχει καμία ιδιοτιμή αρνητική επομένως επιλέγουμε $m=2$. Επιλέγοντας $\gamma=10^7$, ο πίνακας $W=-10^7A^3$ έχει φάσμα ιδιοτιμών το σύνολο:

$$\sigma(W)=10^7 \cdot \{0, -0.8, -2.7, -0.8, 0, -0.1\}$$

Οπότε έχουμε σύγκλιση περίπου μετά από χρόνο:

$$\frac{5}{|\lambda_{\min}(W)|} = 0.000005 \text{ sec}$$

Στο Matlab παίρνουμε τα παρακάτω αποτελέσματα:



Ενώ

$$A^D = \begin{bmatrix} 0.25 & -0.25 & 0 & 0 & 0 & 0 \\ -0.25 & 0.25 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.25 & -0.25 & 0 & 0 \\ 0 & 0 & -0.25 & 0.25 & 0 & 0 \\ 0 & 0 & -0.4167 & -0.5833 & 0.6667 & 0.3333 \\ 0 & 0 & -0.5833 & -0.4167 & 0.3333 & 0.6667 \end{bmatrix}$$

Παράδειγμα 3.6.4

$$\text{Έστω ο πίνακας: } A = \begin{bmatrix} 0.447 & 0.4057 & 0.3529 & 0.6038 & 0.9318 & 0.6721 \\ 0.6154 & 0.9355 & 0.8132 & 0.2722 & 0.4660 & 0.8381 \\ 0.7919 & 0.9169 & 0.0099 & 0.1988 & 0.4186 & 0.0196 \\ 0.9218 & 0.4103 & 0.1389 & 0.0153 & 0.8462 & 0.6813 \\ 0.7382 & 0.8936 & 0.2028 & 0.7468 & 0.5252 & 0.3795 \\ 0.1763 & 0.0579 & 0.1987 & 0.4451 & 0.2026 & 0.8318 \end{bmatrix}$$

Ο πίνακας A είναι αντιστρέψιμος οπότε $\text{ind}(A)=0$ και $A^{-1}=A^D$. Ο πίνακας A έχει φάσμα ιδιοτιμών:

$$\sigma(A)=\{3.0601, 0.6701, -0.6206+0.1081 \cdot i, 0.6206-0.1081 \cdot i, 0.3905, -1170\}.$$

Εν αντιθέσει με τα προηγούμενα παραδείγματα εδώ το φάσμα του πίνακα A έχει τρεις ιδιοτιμές θετικές, μία αρνητική και δύο ιδιοτιμές συζυγείς μιγαδικές. Οπότε στόχος μας είναι να βρούμε τον αριθμό m έτσι ώστε ο πίνακας A^{m+1} να έχει ιδιοτιμές στο δεξιά μιγαδικό ημί-επίπεδο. Οπότε παρατηρώντας τις ιδιοτιμές του πίνακα A^{m+1} μπορούμε να συμπεράνουμε ότι η παράμετρος $m+1$ πρέπει να πληροί τις παρακάτω προϋποθέσεις :

1. Πρέπει $m+1 \geq \text{ind}(A)=0$
2. Η ιδιοτιμή -1170 είναι αρνητική άρα η παράμετρος $m+1$ πρέπει να είναι άρτια. Δηλαδή $U_6=\{2,4,6,\dots\}$.
3. Έχουμε 2 μιγαδικές ιδιοτιμές, οπότε θα προσπαθήσουμε να κατασκευάσουμε σύνολα πιθανών τιμών για τη παράμετρο $m+1$ σύμφωνα με τη παρακάτω σχέση:

$$\frac{4s-1}{2} \frac{\pi}{|\phi|} < m+1 < \frac{4s+1}{2} \frac{\pi}{|\phi|} : s \in N_0$$

Επίσης θα πρέπει να επισημανθεί ότι $\text{Re}(z^{m+1}) = \text{Re}(\bar{z}^{m+1})$ οπότε θα χρειαστεί να εφαρμόσουμε το παραπάνω τύπο για τη μία από τις δύο συζυγείς ιδιοτιμές. Συνεχίζοντας πρώτα βρίσκουμε το όρισμα ϕ , άρα έχουμε:

$$\phi = \arctan\left(\frac{b}{a}\right) = -9.88100^\circ = 180^\circ - 9.88100^\circ = 170.119^\circ$$

Για $s=0$ έχουμε ότι:

$$-\frac{1}{2} \cdot 1.05805 < m+1 < \frac{1}{2} \cdot 1.05805$$

όμως η τιμή $m+1$ δεν παίρνει επιθυμητή τιμή.

Για $s=1$ καταλήγουμε ότι:

$$\frac{3}{2} \cdot 1.05805 < m+1 < \frac{5}{2} \cdot 1.05805$$

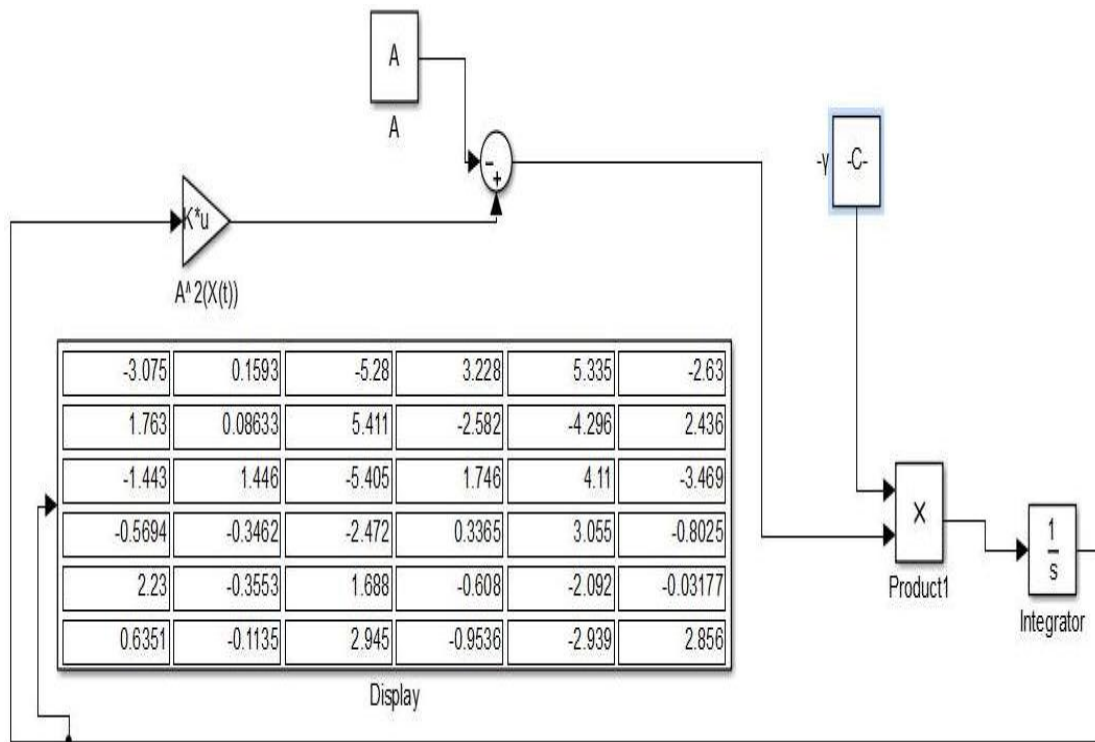
Άρα

$$1.5870 < m+1 < 2.645125$$

οπότε πρέπει $m+1=2$ (που είναι και άρτιος) επομένως επιλέγουμε $m=1$. Ο πίνακας A^2 έχει ιδιοτιμές

$$\{9.3643, 0.4490, 0.3735+0.1341 \cdot i, 0.3735-0.1341 \cdot i, 0.1525, 0.0137\}$$

και όλες έχουν θετικό πραγματικό μέρος. Επιλέγουμε $\gamma=10^{10}$ και μετά από χρόνο $t=10^{-7}$ sec παίρνουμε τα παρακάτω αποτελέσματα:



σύγκλιση δικτύου για $t=10^{-7}$ sec

Ενώ ο A έχει αντίστροφο τον:

$$A^{-1} = \begin{bmatrix} -3.0746 & 0.1594 & -5.2799 & 3.2284 & 5.3347 & -2.6300 \\ 1.7626 & 0.0864 & 5.4113 & -2.5822 & -4.2963 & 2.4364 \\ -1.4430 & 1.4461 & -5.4049 & 1.7459 & 4.1102 & -3.4690 \\ -0.5694 & -0.3461 & -2.4723 & 0.3365 & 3.0550 & -0.8024 \\ 2.2305 & -0.3552 & 1.6884 & -0.6080 & -2.0921 & -0.0317 \\ 0.6351 & -0.1135 & 2.9452 & -0.9536 & -2.9387 & 2.8558 \end{bmatrix}$$

Παράδειγμα 3.6.5

Τώρα θα επιλέξουμε ένα πίνακα με μεγάλο αριθμό κατάστασης. Έστω ο πίνακας A :

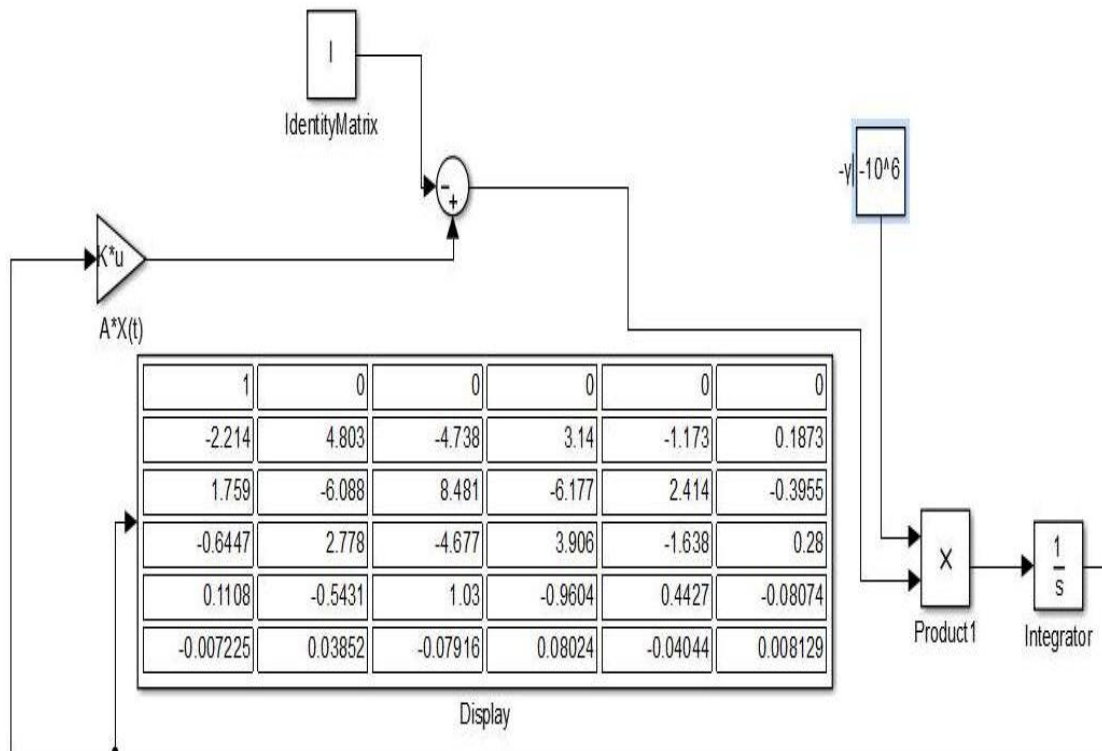
$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 & 16 & 32 \\ 1 & 3 & 9 & 27 & 81 & 243 \\ 1 & 4 & 16 & 64 & 256 & 1024 \\ 1 & 5 & 25 & 125 & 625 & 3125 \end{bmatrix}$$

με $\text{cond}(A) = 5.7689 \cdot 10^4$. Ο πίνακας A είναι αντιστρέψιμος οπότε $A^{-1} = A^D$ και $\text{ind}(A) = 0$. Επίσης ο πίνακας A έχει φάσμα ιδιοτιμών το σύνολο:

$$\sigma(A) = \{3.3438 \cdot 10^3, 0.0643 \cdot 10^3, 0.0042 \cdot 10^3, 0.006 \cdot 10^3, 0.001 \cdot 10^3, 0.0010 \cdot 10^3\}$$

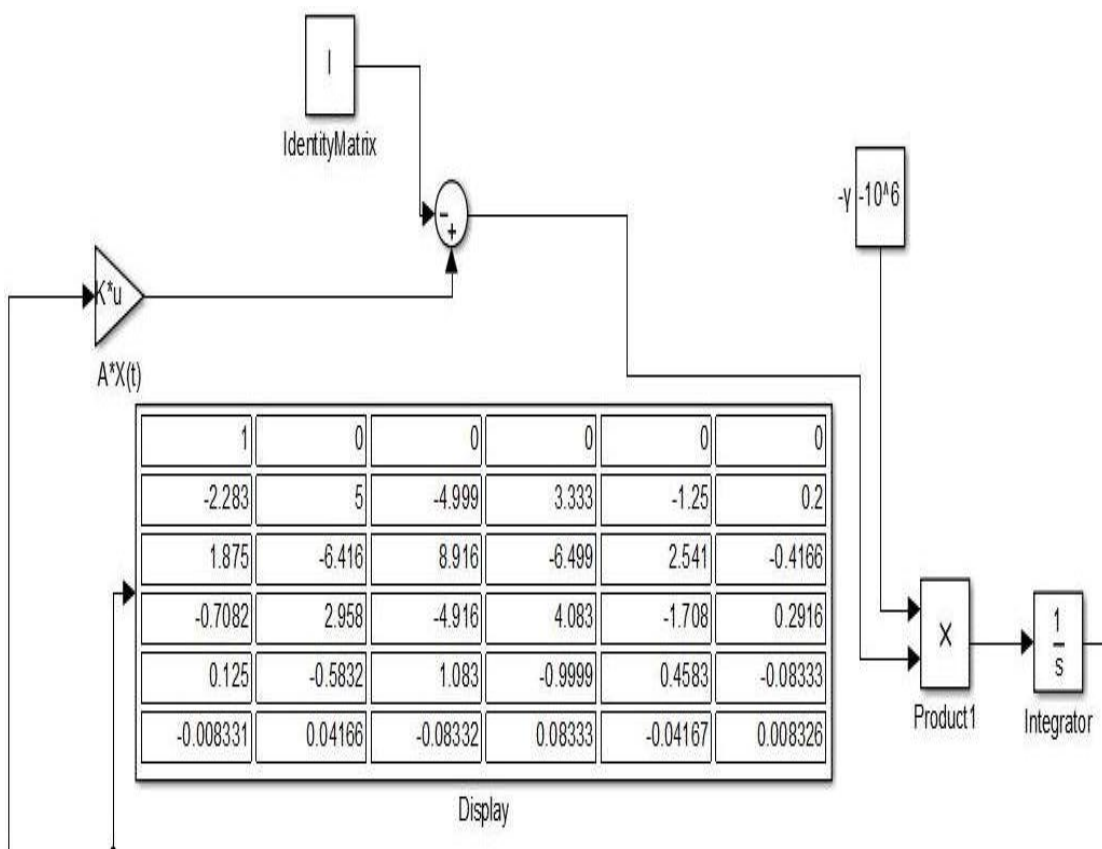
Όλες οι ιδιοτιμές είναι θετικές, άρα ισχύει η συνθήκη (2.10.10) οπότε διαλέγουμε $m=0$. Επιλέγοντας $\gamma=10^6$ βλέπουμε ότι ο πίνακας $W = -\gamma A^{m+1}$ έχει φάσμα ιδιοτιμών το σύνολο: $\sigma(W) = 10^9 \cdot \{-3.3438, -0.0643, -0.0042, -0.0006, -0.0001, -0.0010\}$

Επομένως μετά από περίπου $5/|\lambda_{\min}(W)| = 0.00005$ sec θα πάρουμε τα παρακάτω αποτελέσματα:



σύγκλιση δικτύου για $t=0.00005 \text{ sec}$

Μετά από περίπου $t=0.000015 \text{ sec}$:



σύγκλιση δικτύου για $t=0.00015 \text{ sec}$

1	0	0	0	0	-0
-2.283	5	-5	3.333	-1.25	0.2
1.875	-6.417	8.917	-6.5	2.542	-0.4167
-0.7083	2.958	-4.917	4.083	-1.708	0.2917
0.125	-0.5833	1.083	-1	0.4583	-0.08333
-0.008333	0.04167	-0.08333	0.08333	-0.04167	0.008333

Θεωρητικός αντίστροφος του πίνακα A

Παρατηρούμε ότι αυξάνοντας το χρόνο, το συγκεκριμένο μοντέλο έδωσε καλά αποτελέσματα και για πίνακα με μεγάλο αριθμό κατάστασης.

Οπότε από τα παραπάνω παραδείγματα καταλήγουμε στα παρακάτω συμπεράσματα.

Συμπεράσματα

1] Στη συγκεκριμένη υπό-ενότητα αναπτύξαμε ένα επαναλαμβανόμενο νευρωνικό δίκτυο για τον υπολογισμό του Drazin αντίστροφο σταθερών τετραγωνικών πινάκων. Το συγκεκριμένο νευρωνικό δίκτυο στηρίζεται στη σχέση (3.6.1), με τη προϋπόθεση ότι ο πίνακας A^{m+1} έχει όλες του τις ιδιοτιμές με πραγματικό μέρος μεγαλύτερο ή ίσο με το 0, εν αντιθέση με παλαιότερα νευρωνικά δίκτυα που είχαν κατασκευαστεί και είχαν ως βάση και τις 3 σχέσεις ($\{(3.6.10), (3.6.11), (3.6.12)\}$, όπως για παράδειγμα το νευρωνικό δίκτυο του Chickoski), με αποτέλεσμα να έχουμε ένα πολύπλοκο σύστημα.

2] Το συγκεκριμένο μοντέλο (Stanimirovic) έδωσε καλά αποτελέσματα και για πίνακα που βρίσκεται σε αρρωστημένη κατάσταση (μεγάλο $\text{cond}(A)$)

3] Οι δυσκολίες για τη γενίκευση του συγκεκριμένου μοντέλου για χρονικά μεταβαλλόμενους πίνακες οφείλεται στη μη απλοποίηση της σχέσης (3.6.5) (ο χρονικά μεταβαλλόμενος όρος $(A^{m+1}(t))^T$ επηρεάζει το σύστημα διαφορικών εξισώσεων) και στο πρόσημο των ιδιοτιμών του πίνακα $A^{m+1}(t)$ (δε διατηρούν σταθερό πρόσημο οι ιδιοτιμές του συγκεκριμένου πίνακα).

Κεφάλαιο 4

Εισαγωγή Στη Παρεμβολή και Επίλυση Αλγεβρικών Συστημάτων

4.1 Πολυωνυμική παρεμβολή συνάρτησης μίας μεταβλητής

Παρεμβολή είναι το πρόβλημα προσαρμογής μίας κατάλληλης και εύχρηστης συνάρτησης f σε ένα σύνολο τιμών ή στις τιμές μίας συνάρτησης g . έτσι ώστε η συνάρτηση f να έχει τις ίδιες τιμές με τις αντίστοιχες τιμές των δεδομένων ή με τις αντίστοιχες τιμές της συνάρτησης g , δηλαδή για τα σημεία (x_i, y_i) με $i=0, 1, \dots, n$ να ισχύει $f(x_i)=y_i$ για κάθε $i=0, 1, \dots, n$ ή $f(x_i)=g(x_i)$ για κάθε $i=0, 1, \dots, n$. Όταν η συνάρτηση f είναι ένα πολυώνυμο τότε η παρεμβολή ονομάζεται πολυωνυμική.

Σκοπός της πολυωνυμικής παρεμβολής σε συναρτήσεις μίας μεταβλητής είναι να βρεθεί ένα πολυώνυμο βαθμού n που να παρεμβάλλει τα σημεία (x_i, y_i) με $i=0, 1, \dots, n$ το οποίο δίνεται από το τύπο

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0 = \sum_{i=0}^n a_i x^i \quad (4.1)$$

Ο υπολογισμός των συντελεστών του πολυωνύμου, με βάση ένα σύνολο σημείων, οδηγεί στη λύση ενός γραμμικού συστήματος του οποίου, ο πίνακας των συντελεστών είναι ο γνωστός πίνακας Vandermonde. Η ορίζουσα του πίνακα Vandermonde είναι διάφορη του μηδενός, όταν τα σημεία παρεμβολής είναι διακεκριμένα. Η επίλυση του γραμμικού συστήματος, δηλαδή η εύρεση των συντελεστών του πολυωνύμου παρεμβολής, δεν αποτελεί τη καλύτερη αντιμετώπιση του προβλήματος. Αρκετές μέθοδοι έχουν αναπτυχθεί για την εύρεση του πολυωνύμου παρεμβολής όπως η παρεμβολή με το πολυώνυμο Taylor, η παρεμβολή με το πολυώνυμο Lagrange, η παρεμβολή του πολυώνυμο Newton με διαιρεμένες διαφορές, η παρεμβολή του Hermite κ.τ.λ.π.

Ας είναι $n+1$ διακεκριμένα σημεία x_0, x_1, \dots, x_n στα οποία οι τιμές $f(x_0), f(x_1), \dots, f(x_n)$ της συνάρτησης είναι γνωστές. Συνήθως ζητάμε να βρεθεί πολυώνυμο $p_n(x)$, βαθμού n , το οποίο να εμφανίζει τις ίδιες τιμές με την f στα $n+1$ σημεία. Δηλαδή ψάχνουμε ένα πολυώνυμο $p_n(x)$ το οποίο να ικανοποιεί τις συνθήκες παρεμβολής $f(x_i)=p_n(x_i)$ για $i=0,1,2,\dots, n$. Τα σημεία x_0, x_1, \dots, x_n καλούνται σημεία παρεμβολής και το $p_n(x)$

βαθμού n . Το παρακάτω θεώρημα μας δίνει πληροφορίες όσον αφορά την ύπαρξη και μοναδικότητα του πολυωνύμου παρεμβολής.

Θεώρημα 4.1 (ύπαρξης και μοναδικότητα) [9], [10]

Για οποιοδήποτε σύνολο $n+1$ διακεκριμένων σημείων x_0, x_1, \dots, x_n και των αντίστοιχων τιμών της συνάρτησης f υπάρχει μοναδικό πολυώνυμο $p(x)$ τέτοιο ώστε $f(x_i)=p(x_i)$ για $i=0, 1, 2, \dots, n$

Απόδειξη

Έστω ένα πολυώνυμο $p(x)=\sum_{i=0}^n a_i x^i$ που ικανοποιεί τις συνθήκες παρεμβολής αν και μόνο αν οι συντελεστές a_i αποτελούν τη λύση του γραμμικού συστήματος:

$$a_0+a_1x_0+\dots+a_nx_0^n=f(x_0)$$

$$a_0+a_1x_1+\dots+a_nx_1^n=f(x_1)$$

.....

$$a_0+a_1x_n+\dots+a_nx_n^n=f(x_n)$$

Αυτό το σύστημα έχει μοναδική λύση αν και μόνο αν η ορίζουσα δε μηδενίζεται. Μια τέτοια ορίζουσα καλείται ορίζουσα του *Vandemorde* και συμβολίζεται με $VDM(x_0, x_1, \dots, x_n)$. Έτσι

$$VDM(x_0, x_1, \dots, x_n) = \begin{vmatrix} 1 & x_0 & x_0^2 & \cdots & x_0^{n-1} & x_0^n \\ 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} & x_1^n \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} & x_n^n \end{vmatrix} \quad (4.1.1)$$

με $VDM(x_0, x_1, \dots, x_n) = \prod_{0 \leq i < j \leq n} (x_i - x_j)$. Αφού τα σημεία παρεμβολής είναι διακεκριμένα

δηλαδή $x_i \neq x_j$ για $i \neq j$, τότε η ορίζουσα δε μηδενίζεται και το πρόβλημα έχει μία και μοναδική λύση.

4.2 Πολυώνυμο παρεμβολής Newton

Το πολυώνυμο παρεμβολής Newton ισχύει και για ανισαπέχοντα σημεία. Πριν δώσουμε το τύπο παρεμβολής του Newton θα δώσουμε τους ορισμούς των διαιρεμένων διαφορών.

Διαιρεμένη διαφορά πρώτης τάξης στα σημεία x_0, x_1 συμβολίζεται με $f[x_0, x_1]$ και είναι το πηλίκο:

$$f[x_0, x_1] = \frac{f(x_0) - f(x_1)}{x_1 - x_0} \quad (4.2)$$

Σα διαιρεμένη διαφορά δεύτερης τάξης στα σημεία x_0, x_1, x_2 ορίζεται το παρακάτω πηλίκο:

$$f[x_0, x_1, x_2] = \frac{f[x_1, x_2] - f[x_0, x_1]}{x_2 - x_0} \quad (4.2.1)$$

Και οι διαιρεμένες διαφορές τάξης n στα σημεία x_0, x_1, \dots, x_n ορίζονται από τη σχέση:

$$f[x_0, x_1, x_2, \dots, x_n] = \frac{f[x_1, x_2, \dots, x_n] - f[x_0, x_1, \dots, x_{n-1}]}{x_2 - x_0} \quad (4.2.2)$$

Ένα πολυώνυμο στη μορφή Newton με κέντρα τα σημεία παρεμβολής x_0, x_1, \dots, x_{n-1} γράφεται:

$$p_n(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) + \dots + a_n(x - x_0)(x - x_1) \dots (x - x_{n-1})$$

Έχει αποδειχθεί ότι το πολυώνυμο που προκύπτει αν συνδυάσουμε τα παραπάνω είναι της μορφής:

$$p_n(x) = f(x_0) + f[x_0, x_1](x - x_0) + \dots + f[x_0, x_1, \dots, x_n](x - x_0)(x - x_1) \dots (x - x_{n-1})$$

Το σφάλμα αποκοπής του πολυωνύμου παρεμβολής Newton δίνεται από τον τύπο:

$$E_n(x) = f(x) - p_n(x) = f[x_0, x_1, \dots, x_n, x] \prod_{i=0}^n (x - x_i) \quad (4.2.3)$$

Παράδειγμα 4.2

Να βρεθούν οι συντελεστές στο πολυώνυμο παρεμβολής Newton πέμπτου βαθμού καθώς και η τιμή της συνάρτησης στο σημείο $x=1.7$, όταν η συνάρτηση παίρνει τις παρακάτω τιμές στα αντίστοιχα σημεία:

$$\begin{bmatrix} x & 1.2 & 3.0 & 1.5 & 4.0 & 2.0 & 2.5 \\ f(x) = \ln(x) & 0.1823 & 1.0986 & 0.4055 & 1.3863 & 0.6931 & 0.9163 \end{bmatrix}$$

Λύση

Χρησιμοποιώντας τις παραπάνω τιμές και σχηματίζοντας τις διαιρεμένες διαφορές το πολυώνυμο παρεμβολής Newton είναι $p(x)=0.1823+0.5091(x-1.2)-0.1567(x-1.2)(x-3)+0.31(x-1.2)(x-3)(x-1.5)-0.0117(x-1.2)(x-3)(x-1.5)(x-4)+0.0039(x-1.2)(x-3)(x-1.5)(x-4)(x-2)$.

Για να βρούμε το $p(1.7)$ αντικαθιστούμε τη τιμή $x=1.7$ στο παραπάνω πολυώνυμο και έχουμε ότι $p(1.7)=0.5308$ ενώ η πραγματική τιμή είναι η $f(1.7)=\ln(1.7)=0.53062825$. Άρα το πραγματικό σφάλμα αποκοπής προκύπτει από τη διαφορά $f(1.7)-p(1.7)=-0.00017175$.

4.3 Νευρωνική προσέγγιση στη μέθοδο παρεμβολής

Έστω πολυώνυμο $Q(s)$ το πολυώνυμο παρεμβολής βαθμού n της μορφής $Q(s)=q_0+q_1s+q_2s^2+\dots+q_ns^n$. Ας υποθέσουμε ότι ξέρουμε τις τιμές του πολυωνύμου στα σημεία s_0, s_1, \dots, s_n οι οποίες έχουν προκύψει από κάποια είδους μέτρηση και ότι οι συντελεστές του πολυωνύμου είναι πραγματικοί αριθμοί. Από τις τιμές $Q_0=Q(s_0), Q_1=Q(s_1), \dots, Q_n=Q(s_n)$, μπορούμε να βρούμε τους συντελεστές του πολυωνύμου q_0, q_1, \dots, q_n λύνοντας το σύνολο των αλγεβρικών εξισώσεων όπως βλέπουμε παρακάτω:

$$\begin{bmatrix} 1 & s_0 & s_0^2 & \dots & s_0^n \\ 1 & s_1 & s_1^2 & \dots & s_1^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & s_n & s_n^2 & \dots & s_n^n \end{bmatrix} \begin{bmatrix} q_0 \\ q_1 \\ \vdots \\ q_n \end{bmatrix} = \begin{bmatrix} Q_0 \\ Q_1 \\ \vdots \\ Q_n \end{bmatrix} \quad (4.3)$$

Οι παραπάνω σχέσεις μπορούν να γραφούν σε μορφή πινάκων της μορφής $AX=B$, όπου $X=[q_0, q_1, \dots, q_n]^T$ και $B=[Q_0, Q_1, \dots, Q_n]^T$. Θα είναι σωστό να τονιστεί ότι το παραπάνω σύστημα έχει μοναδική λύση εφόσον ο πίνακας A είναι αντιστρέψιμος

(μορφή Vandermonde). Διάφορες λύσεις μπορούν προταθούν για την επίλυση της παραπάνω εξίσωση πινάκων, στο προτείνεται [25] μία μέθοδος με νευρωνικά δίκτυα, εφόσον η εφαρμογή του κυκλώματος μπορεί να πραγματοποιηθεί σε πραγματικό χρόνο (στηρίζεται στη μέθοδο μέγιστης κλίσης) .

Σε μία νευρωνική προσέγγιση το πρόβλημα μας τυποποιείται σε πρόβλημα ελαχιστοποίησης $\min F(X)$, όπου:

$$F(X) = (AX - B)^T (AX - B)$$

Το παραπάνω πρόβλημα επιλύεται με νευρωνική μέθοδο σύμφωνα με την οποία έχουμε τις παρακάτω σχέσεις:

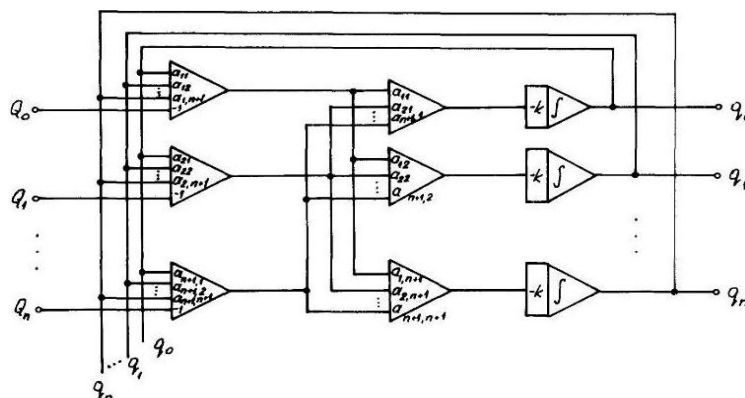
$$\frac{dX}{dt} = kp(X) \quad \text{και} \quad p(X) = -\nabla F(X)$$

όπου k αποτελεί μία σταθερά ολοκλήρωσης και $\nabla F(X) = A^T(AX - B)$.

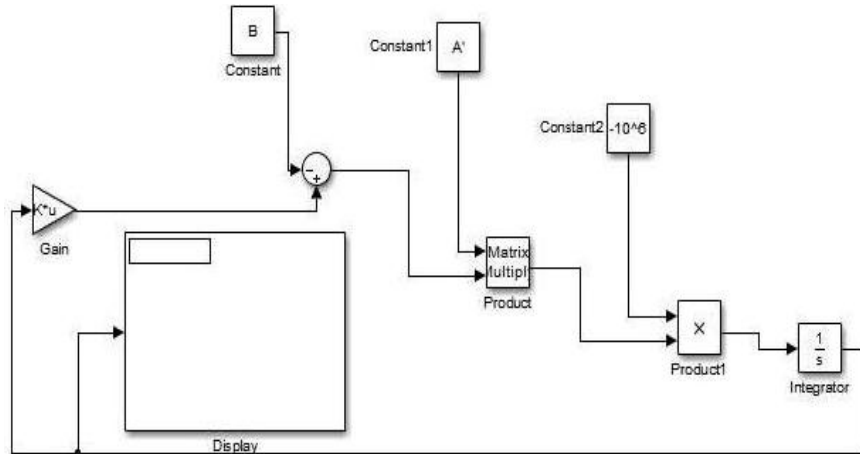
Όπως παρατηρούμε η λογική του παραπάνω μοντέλου στηρίζεται στη μέθοδο μέγιστης κλίσης που περιγράψαμε στο πρώτο κεφάλαιο.

Η εκτέλεση του νευρωνικού δικτύου παρουσιάζεται στο παρακάτω σχήμα. Με τη ρύθμιση της σταθεράς ολοκλήρωσης k περίπου $k=10^6$ παίρνουμε πρακτικά τη λύση που θέλουμε σε πραγματικό χρόνο (σε λίγα δευτερόλεπτα). Δηλαδή αντί να το λύσουμε με τη κλασσική μέθοδο $AX=B$ μπορούμε να έχουμε ένα νευρωνικό δίκτυο (τη διαφορική):

$$\frac{dX}{dt} = -kA^T(AX - B). \quad (4.3.1)$$



Νευρωνικό κύκλωμα που λύνει το πρόβλημα ελαχιστοποίησης $\min F(X)$ [25]



Νευρωνικό κύκλωμα στο Matlab

Παράδειγμα 4.3

Να βρεθεί πολυώνυμο παρεμβολής δευτέρου βαθμού στα σημεία $s_0=0, s_1=1, s_2=2$ με τιμές $Q(s_0)=2, Q(s_1)=1, Q(s_2)=3$ Δηλαδή ζητάμε στην ουσία να βρούμε τους συντελεστές q_0, q_1, q_2 του πολυωνύμου $Q(s) = q_2s^2 + q_1s + q_0$.

Λύση

Θα έχουμε

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix}, B = [2 \quad 1 \quad 3]^T \text{ και } X = [q_0 \quad q_1 \quad q_2]^T$$

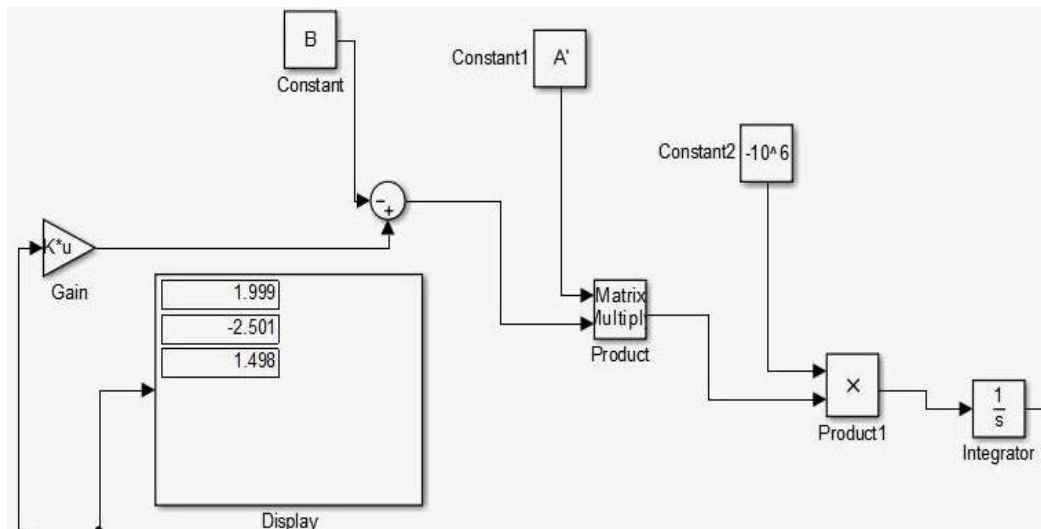
και θα παραθέσουμε παρακάτω δύο τρόπους επίλυσης του προβλήματος.

1^{ος} τρόπος (Αλγεβρικά)

Όσον αφορά την αλγεβρική μέθοδο θα χρειαστεί να λύσουμε το σύστημα εξισώσεων $AX=B$ και το αποτέλεσμα που θα πάρουμε είναι $q_0=2, q_1=-2.5, q_2=1.5$.

2^{ος} τρόπος (Νευρωνική προσέγγιση)

Όσον αφορά την νευρωνική προσέγγιση θα χρειαστεί να έχουμε ένα νευρωνικό δίκτυο που να επιλύει τη διαφορική εξίσωση (4.3.1) όπως φαίνεται στο παρακάτω σχήμα:



Η λύση που παίρνουμε είναι $q_0=1.999$, $q_1=-2.501$, $q_2=1.496$ και παρατηρούμε ότι προσεγγίζει πάρα πολύ καλά τη λύση της πρώτης μεθόδου.

Παράδειγμα 4.3.1

Να βρεθεί πολώνυμο παρεμβολής πέμπτου βαθμού που να παρεμβάλει τις τιμές $s_0=0$, $s_1=1$, $s_2=2$, $s_3=3$, $s_4=4$, $s_5=5$ με $Q(s_5)=3$, $Q(s_1)=2$, $Q(s_2)=1$, $Q(s_3)=4$, $Q(s_4)=5$, $Q(s_5)=6$.

Λύση

Θα έχουμε στη προκειμένη περίπτωση

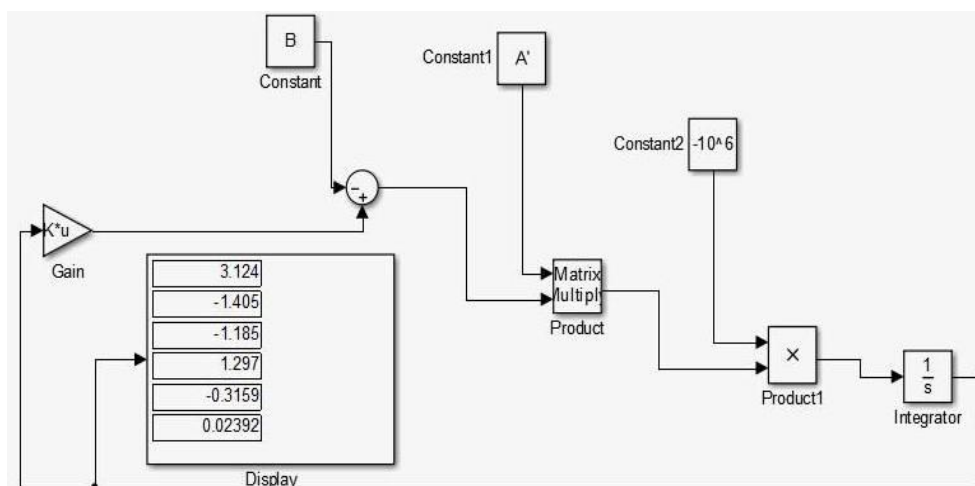
$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 & 16 & 32 \\ 1 & 3 & 9 & 27 & 81 & 243 \\ 1 & 4 & 16 & 64 & 256 & 1024 \\ 1 & 5 & 25 & 125 & 625 & 3125 \end{bmatrix} \text{ και } B = [3 \ 2 \ 1 \ 4 \ 5 \ 6]^T$$

1^{ος} τρόπος (Αλγεβρικά)

Όσον αφορά την αλγεβρική μέθοδο θα χρειαστεί να λύσουμε το σύστημα εξισώσεων $AX=B$ και το αποτέλεσμα που θα πάρουμε είναι $q_0=3$, $q_1=6.433$, $q_2=-14.08$, $q_3=8.417$, $q_4=-1.917$, $q_5=0.15$.

2^{ος} τρόπος (Νευρωνική προσέγγιση)

Όσον αφορά την νευρωνική προσέγγιση θα χρειαστεί να πάρουμε εκπαιδύσουμε ένα νευρωνικό δίκτυο που να επιλύει τη διαφορική εξίσωση (3.3.1) όπως φαίνεται στο παρακάτω σχήμα:



Η λύση που παίρνουμε είναι $q_0=3.124$, $q_1=-1.405$, $q_2=-1.185$, $q_3=1.297$, $q_4=-0.3159$, $q_5=0.02392$.

Συμπέρασμα

Παρατηρούμε ότι σε αυτό το παράδειγμα η νευρωνική προσέγγιση δε δίνει σωστά αποτελέσματα και αυτό έχει να κάνει με το γεγονός ότι ο $\text{cond}(A)$ του πίνακα A είναι πολύ μεγάλος (στη προκειμένη περίπτωση ισούται με $5.7689 \cdot 10^4$). Επομένως η συγκεκριμένη μέθοδος δε μπορεί να χρησιμοποιηθεί για μεγάλου βαθμού πολυώνυμα. Λύση στο πρόβλημα μας δίνει ο παρακάτω τρόπος σκέψης.

4.4 Βελτιωμένη Νευρωνική προσέγγιση

Στη βελτιωμένη νευρωνική προσέγγιση το πρόβλημα παρεμβολής χωρίζεται σε δύο στάδια. Στο πρώτο στάδιο ψάχνουμε για πολυώνυμο παρεμβολής της μορφής $Q(s)=C_0+(s-s_0)+C_1(s-s_0)\cdot(s-s_1)+\dots+C_n(s-s_0)\cdot(s-s_1)\cdot\dots\cdot(s-s_{n-1})$ όπου C_i ($i=1,2,\dots,n$) είναι οι συντελεστές του πολυωνύμου και s_0,s_1,s_{n-1} είναι τα σημεία στα οποία κάνω τη παρεμβολή. Σε δεύτερο στάδιο φέρνουμε το παραπάνω πολυώνυμο από δικτυωτή μορφή στη κανονική όπως στη προηγούμενη περίπτωση. Ακολουθώντας τη προηγούμενη μέθοδο τοποθετώντας $s=s_i$ για $i=0,1,2,\dots,n$ θα πάρουμε το παρακάτω σύστημα γραμμικών εξισώσεων.

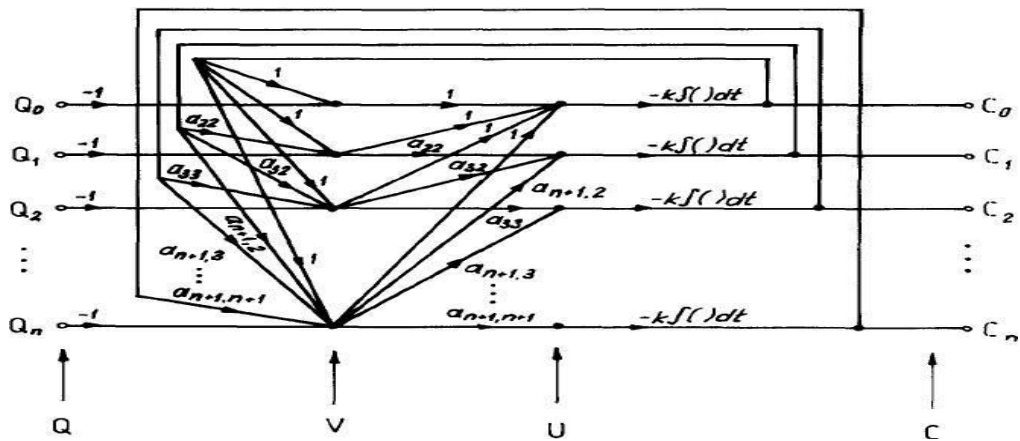
$$\begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ 1 & (s_1 - s_0) & 0 & 0 & \dots & 0 \\ 1 & (s_2 - s_0) & \prod_{i=0}^n (s_2 - s_i) & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & (s_n - s_0) & \prod_{i=0}^1 (s_n - s_i) & \dots & \dots & \prod_{i=0}^{n-1} (s_n - s_i) \end{bmatrix} \begin{bmatrix} C_0 \\ C_1 \\ C_2 \\ \vdots \\ C_n \end{bmatrix} = \begin{bmatrix} Q_0 \\ Q_1 \\ Q_2 \\ \vdots \\ Q_n \end{bmatrix} \quad (4.4)$$

Το οποίο γράφεται με μορφή πινάκων $AX=B$ όπου $X=[C_0, C_1, \dots, C_n]^T$, $B=[Q_0, Q_1, \dots, Q_n]^T$ και A ο πίνακας που δημιουργείται από τα σημεία παρεμβολής. Στη προκειμένη περίπτωση ο πίνακας A είναι κάτω τριγωνικός (ενώ στη προηγούμενη περίπτωση ήτανε Vandermonde). Όπως παρατηρούμε τα στοιχεία του πίνακα δεν είναι μεγάλα, όπως στη προηγούμενη περίπτωση, με αποτέλεσμα να έχουμε μικρότερο $\text{cond}(A)$ και να μην έχουμε πολύπλοκους υπολογισμούς όπως στη προηγούμενη περίπτωση. Κοιτώντας το προηγούμενο παράδειγμα για πολυώνυμο βαθμού 5 και $s_0=0, s_1=1, s_2=2, s_3=3, s_4=4, s_5=5$ θα πάρουμε $\text{cond}(A)=368.835$ που είναι πολύ μικρότερος από τον $5.7689 \cdot 10^4$ που είχαμε προηγουμένως.

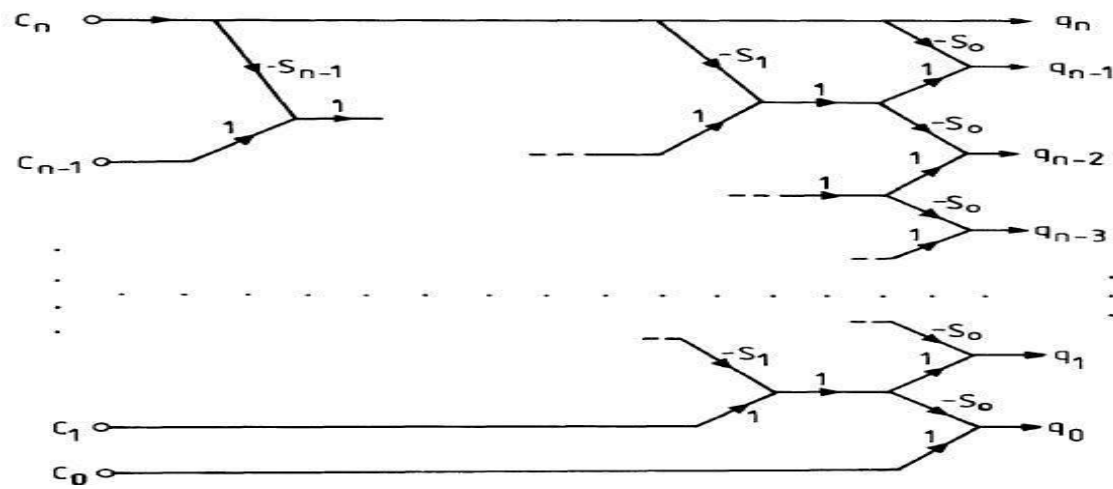
Ακολουθώντας την ίδια διαδικασία με πριν βρίσκουμε τους συντελεστές του πολυωνύμου της μορφής. Αυτή η λύση δίνεται από τη λύση της παρακάτω διαφορικής εξίσωσης:

$$\frac{dX}{dt} = -kA^T(AX - B) \quad (4.4.1)$$

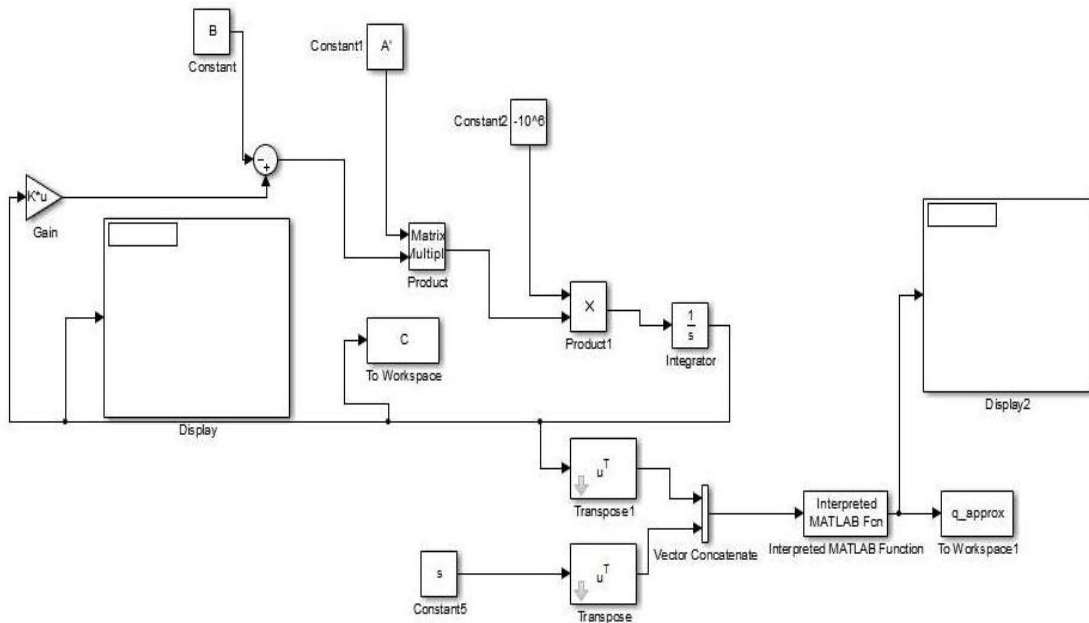
Από το παραπάνω εφόσον βρούμε τις τιμές των C_i μπορούμε εύκολα να υπολογίσουμε τους συντελεστές q_i του πολυωνύμου στη κανονική του μορφή χρησιμοποιώντας την επαναληπτική σχέση $q_i = q_i - q_{i+1}s_k$ για $k=n-1, n-2, \dots, 0$ και $i=k, k+1, \dots, n-1$ χρησιμοποιώντας σαν αρχικές τιμές $q_i = C_i$ ($j=0, 1, 2, \dots, n$) όπως φαίνεται στα παρακάτω σχήματα:



Το νευρωνικό κύκλωμα που λύνει το τροποποιημένο πρόβλημα παρεμβολής [25]



Διάγραμμα ροής σήματος για την επαναληπτική σχέση $q_i = q_i - q_{i+1}S_k$ [25]



Νευρωνικό κύκλωμα που λύνει το τροποποιημένο πρόβλημα παρεμβολής στο Matlab

Παράδειγμα 4.4

Να βρεθεί το πολυώνυμο παρεμβολής έκτου βαθμού που να παρεμβάλει τις τιμές $s_0=0, s_1=0.5, s_2=1, s_3=1.5, s_4=2, s_5=2.5, s_6=3$ με τιμές $Q(s_0)=1, Q(s_1)=1, Q(s_2)=2, Q(s_3)=4, Q(s_4)=9.25, Q(s_5)=29.75, Q(s_6)=100.37$

Λύση

Σε αυτό το παράδειγμα θα έχουμε

$$A = \begin{bmatrix} 1.0000 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1.0000 & 0.5000 & 0 & 0 & 0 & 0 & 0 \\ 1.0000 & 1.0000 & 0.5000 & 0 & 0 & 0 & 0 \\ 1.0000 & 1.5000 & 1.5000 & 0.7500 & 0 & 0 & 0 \\ 1.0000 & 2.0000 & 3.0000 & 3.0000 & 1.5000 & 0 & 0 \\ 1.0000 & 2.5000 & 5.0000 & 7.5000 & 7.5000 & 3.7500 & 0 \\ 1.0000 & 3.0000 & 7.5000 & 15.0000 & 22.5000 & 22.5000 & 11.25 \end{bmatrix}$$

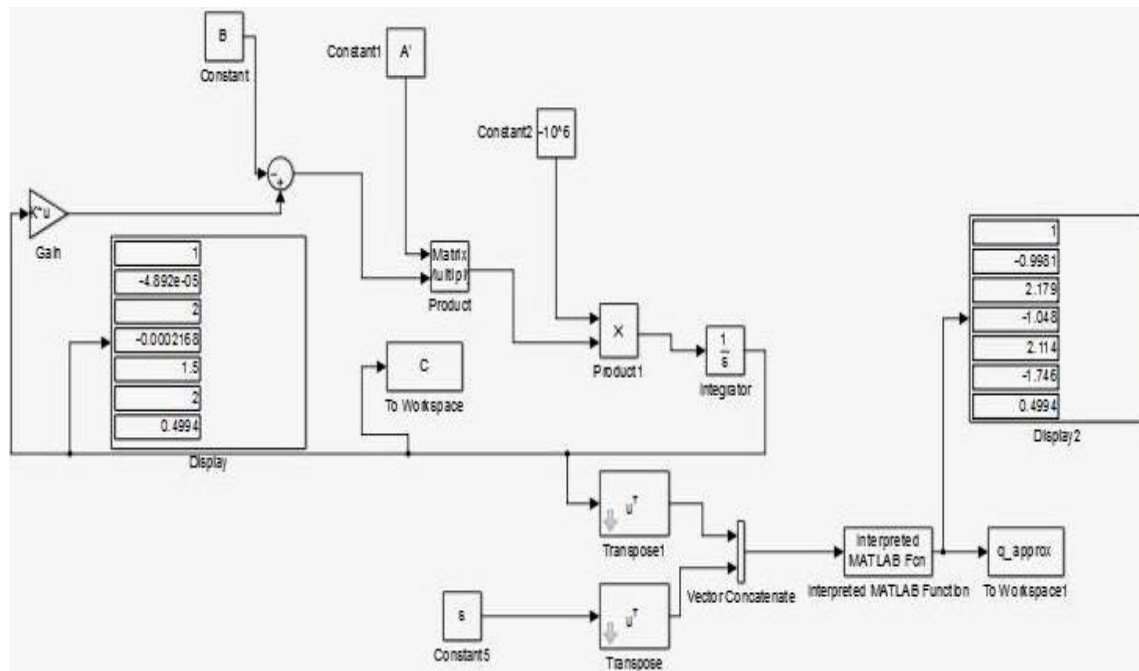
$$\text{και } B = [1.00 \quad 1.00 \quad 2.00 \quad 4.00 \quad 9.25 \quad 29.75 \quad 100.37]^T$$

1^{ος} τρόπος (Αλγεβρικά)

Όσον αφορά αυτή τη μέθοδο θα χρειαστεί να λύσουμε το σύστημα εξισώσεων $AX=B$ για να βρούμε τους συντελεστές C_i (συντελεστές πολυωνύμου στη δικτυωτή μορφή) και στη συνέχεια από την επαναληπτική σχέση $q_i = q_i - q_{i+1}s_k$ που έχουμε περιγράψει παραπάνω πως λειτουργεί θα βρούμε τους συντελεστές q_i της κανονικής μορφής του πολυωνύμου.

Στο συγκεκριμένο παράδειγμα τα αποτελέσματα που θα πάρουμε είναι $q_0=1, q_1= -0.9983, q_2=2.18, q_3= -1.05, q_4=2.116, q_5= -1.747, q_6=0.4996$.

2^{ος} τρόπος (Νευρωνική προσέγγιση)



Η λύση που παίρνουμε και φαίνεται στο παραπάνω σχήμα είναι $q_0=1$, $q_1=-0.9981$, $q_2=2.179$, $q_3=-1.048$, $q_4=2.114$, $q_5=-1.746$, $q_6=0.4994$ λύση που είναι πολύ κοντά με τη λύση της πρώτης μεθόδου.

Παράδειγμα 4.4.1

Να βρεθεί το πολυώνυμο παρεμβολής έκτου βαθμού που να παρεμβάλει τις τιμές $s_0=0$, $s_1=0.5$, $s_2=1$, $s_3=1.5$, $s_4=2$, $s_5=2.5$, $s_6=3$ με τιμές $Q(s_0)=0.5$, $Q(s_1)=1$, $Q(s_2)=2.5$, $Q(s_3)=4.25$, $Q(s_4)=10$, $Q(s_5)=28$, $Q(s_6)=73.81$.

Λύση

Σε αυτό το παράδειγμα θα έχουμε ίδιο πίνακα A και το μόνο που θα αλλάζει είναι ο πίνακας B εφόσον έχουμε διαφορετικά μετρούμενα σημεία. Άρα ο πίνακας B θα είναι ο παρακάτω:

$$B = [0.50 \quad 1.00 \quad 2.50 \quad 4.25 \quad 10.00 \quad 28.00 \quad 73.81]^T$$

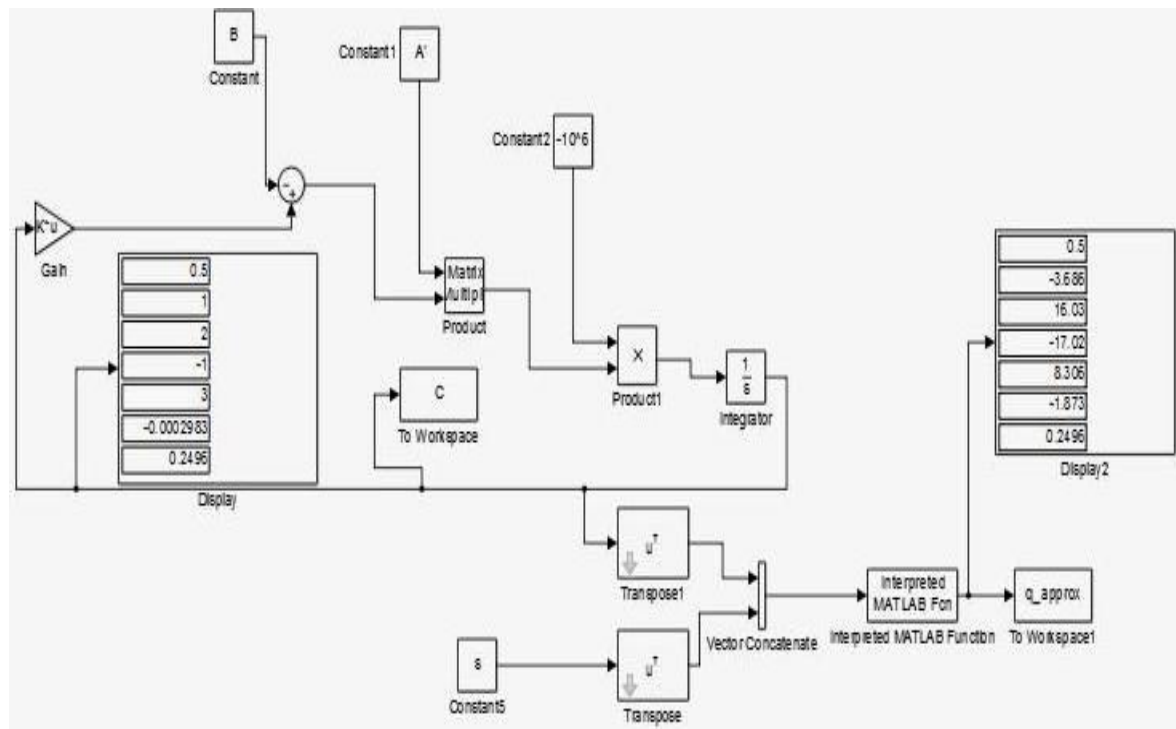
1^{ος} τρόπος (Αλγεβρικά)

Όσον αφορά αυτή τη μέθοδο θα χρειαστεί να λύσουμε το σύστημα εξισώσεων $AX=B$ για να βρούμε τους συντελεστές C_i (συντελεστές πολυωνύμου στη δικτυωτή μορφή) και στη συνέχεια από την επαναληπτική σχέση $q_i=q_{i+1}s_k$ που έχουμε περιγράψει

παραπάνω πως λειτουργεί θα βρούμε τους συντελεστές q_i της κανονικής μορφής του πολυωνύμου.

Στο συγκεκριμένο παράδειγμα τα αποτελέσματα που θα πάρουμε είναι $q_0=0.5$, $q_1=-3.687$, $q_2=16.03$, $q_3=-17.02$, $q_4=8.308$, $q_5=-1.873$, $q_6=0.2498$.

2^{ος} τρόπος (Νευρωνική προσέγγιση)



Η λύση που παίρνουμε και φαίνεται στο παραπάνω σχήμα είναι $q_0=0.5$, $q_1=-3.686$, $q_2=16.03$, $q_3=-17.02$, $q_4=8.306$, $q_5=-1.873$, $q_6=0.2496$ λύση που είναι πολύ κοντά με τη λύση της πρώτης μεθόδου.

Συμπέρασμα

Παρατηρούμε ότι αυτή η βελτιωμένη μέθοδος λειτουργεί και για πολυώνυμα μεγάλου βαθμού, εν αντιθέσει με τη προηγούμενη που για μεγαλύτερου βαθμού πολυώνυμα δεν έδινε καλά αποτελέσματα και αυτό έχει να κάνει με το γεγονός ότι ο πίνακας A στη πρώτη περίπτωση (Vandermorde) έχει μεγάλο $\text{cond}(A)$ ενώ στη βελτιωμένη νευρωνική προσέγγιση παρατηρούμε ότι ο $\text{cond}(A)$ είναι κατά πολύ μικρότερος από ότι στη πρώτη περίπτωση πράγμα που κάνει τη δεύτερη μέθοδο πιο ευσταθή σε σχέση με τη πρώτη αλλά παρ'όλα αυτά δε δίνει σε τόσο γρήγορο ρυθμό τα αποτελέσματα.

4.5 Επαναλαμβανόμενο Νευρωνικό Δίκτυο για επίλυση αλγεβρικών εξισώσεων

4.5.1 ZNN για γραμμικές χρονικά μεταβαλλόμενες εξισώσεις

Θεωρούμε τη χρονικά μεταβαλλόμενη γραμμική εξίσωση:

$$A(t)x(t) = b(t) \quad (4.5)$$

όπου $A(t) \in R^{n \times n}$ είναι ένας ομαλός χρονικά μεταβαλλόμενος πίνακας, $b(t) \in R^n$ ένα χρονικά μεταβαλλόμενο διάνυσμα και $x(t)$ ένα άγνωστο διάνυσμα το οποίο θέλουμε να βρούμε. Επίσης $\dot{A}(t)$ και $\dot{b}(t)$ τα θεωρούμε γνωστά. Η μέθοδος που θα αναπτύξουμε στηρίζεται και είναι μία παραλλαγή της μεθόδου που αναπτύχθηκε στο προηγούμενο κεφάλαιο για την αντιστροφή χρονικά μεταβαλλόμενων πινάκων. Έστω η συνάρτηση σφάλματος:

$$E(t) = A(t)x(t) - b(t)$$

όπου $E(t) \in R^n$. Το ZNN μοντέλο που επιλύει την εξίσωση (4.5) μπορεί να σχεδιαστεί όπως κάναμε στο προηγούμενο κεφάλαιο χρησιμοποιώντας τη σχέση:

$$\dot{E}(t) = -\Gamma F(E(t))$$

Συνεπώς καταλήγουμε στο μοντέλο:

$$A(t)\dot{x}(t) = -\dot{A}(t)x(t) - \Gamma F(A(t)x(t) - b(t)) + \dot{b}(t) \quad (4.5.1)$$

όπου, $\Gamma \in R^n$ θετικά ορισμένος πίνακας. Για δική μας ευκολία επιλέγουμε $\Gamma = \gamma \cdot I$, όπου γ ένα θετικός αριθμός και I κατάλληλα διαστάσεων ταυτοτικός πίνακας. Επίσης $F(\bullet): R^n \rightarrow R^n$ είναι ένα διάνυσμα συνάρτησης-ενεργοποίησης που περιέχει στη κάθε θέση μία συνάρτηση ενεργοποίησης $f(\bullet)$. Στη [16] προτείνεται να χρησιμοποιήσουμε τη γραμμική συνάρτηση ενεργοποίησης $f(u) = u$. Οπότε η σχέση (4.5.1) μπορεί να απλοποιηθεί στη:

$$A(t)\dot{x}(t) = -\dot{A}(t)x(t) - \gamma(A(t)x(t) - b(t)) + \dot{b}(t)$$

Επομένως χρησιμοποιώντας παρόμοιο αποδεικτικό συλλογισμό όπως στο προηγούμενο κεφάλαιο (αντιστροφή χρονικά μεταβαλλόμενου πίνακα με τη μέθοδο

Yunong Zhang) μπορούμε να βρούμε προσεγγιστική μέθοδο για την επίλυση της (3.3).

4.5.2 ZNN για γραμμικές σταθερές εξισώσεις

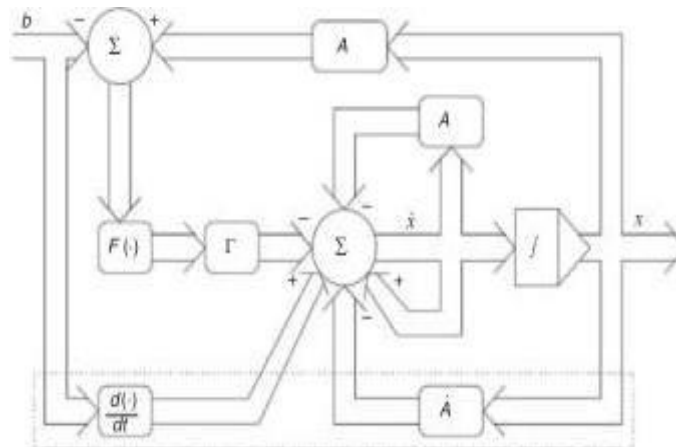
Θεωρούμε το παρακάτω σύστημα αλγεβρικών εξισώσεων:

$$Ax = b \quad (4.5.2)$$

όπου $A=[a_{ij}] \in R^{n \times n}$, $b=[b_1, b_2, \dots, b_n]^T \in R^n$ και x το σταθερό άγνωστο διάνυσμα που θέλουμε να υπολογίσουμε επιλύοντας το παραπάνω γραμμικό σύστημα.

Συνεχίζοντας εφόσον οι πίνακες A και b είναι σταθεροί τότε οι παράγωγοι τους μηδενίζονται, οπότε η σχέση (4.7.1) απλοποιείται στη:

$$A\dot{x} = -\gamma F(Ax - b) \quad (4.5.2.1)$$

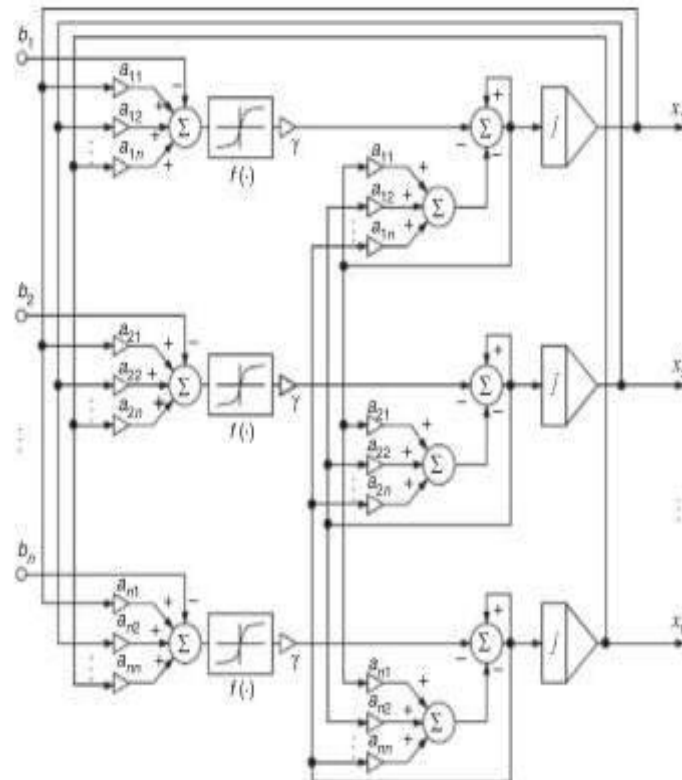


Διάγραμμα σχέσης (4.7.1), (4.7.2.1) [16]

Η παραπάνω εξίσωση πινάκων μπορεί να περιγραφεί αναλυτικά ως εξής:

$$\dot{x}_i = -\gamma f \left(\sum_{j=1}^n a_{ij} x_j - b_i \right) - \left(\sum_{j=1, j \neq i}^n a_{ij} \dot{x}_j \right) + (1 - a_{ii}) \dot{x}_i \quad (4.7.2.2)$$

για $i=1, 2, \dots, n$.



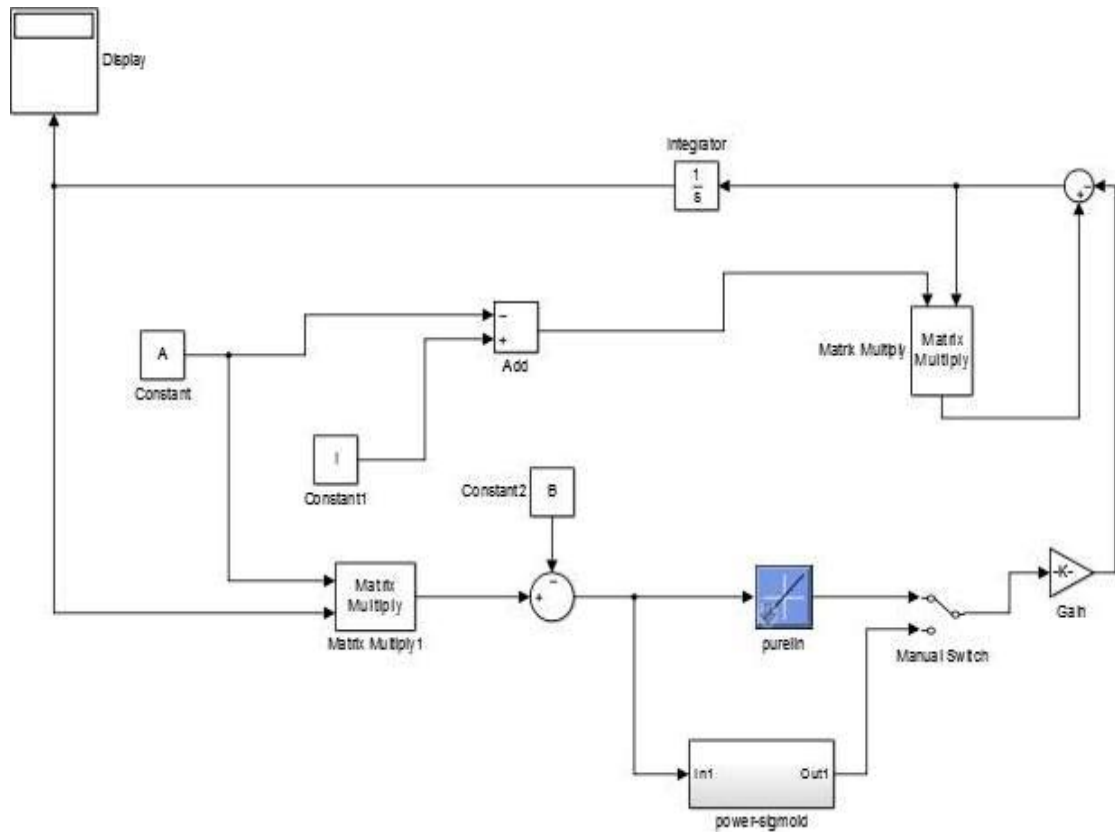
Κύκλωμα σχέσης (4.7.2.2) [16]

Όλα τα παραπάνω μας οδηγούν στη παρακάτω πρόταση της οποίας η απόδειξη δε θα μας απασχολήσει στα πλαίσια αυτού του κεφαλαίου (εφόσον είναι παρόμοια με εκείνη στο θεώρημα (3.3) του προηγούμενου κεφαλαίου).

Πρόταση 4.5.2 [16]

Έστω ένας σταθερός ομαλός πίνακας $A \in R^{n \times n}$ και ένα σταθερό διάνυσμα $b \in R^n$. Εάν χρησιμοποιήσουμε ένα διάνυσμα $n \times 1$ $F(\bullet)$, που στη κάθε θέση περιέχει μία αύξουσα περιττή συνάρτηση μίας μεταβλητής $f(\bullet)$, τότε το διάνυσμα κατάστασης $x(t)$ της σχέσης (4.5.2.1) ξεκινά από μία αρχική κατάσταση (τυχαία) $x(0) \in R^n$ και συγκλίνει στη μοναδική λύση του συστήματος γραμμικών εξισώσεων (4.5.2).

Παρακάτω βλέπουμε το νευρωνικό σύστημα που επιλύει το σύστημα (4.5.2) στο Matlab.



Προσομοίωση κυκλώματος στο Simulink

Παράδειγμα 4.5.2

Να λυθεί το σύστημα $Ax=b$ όπου:

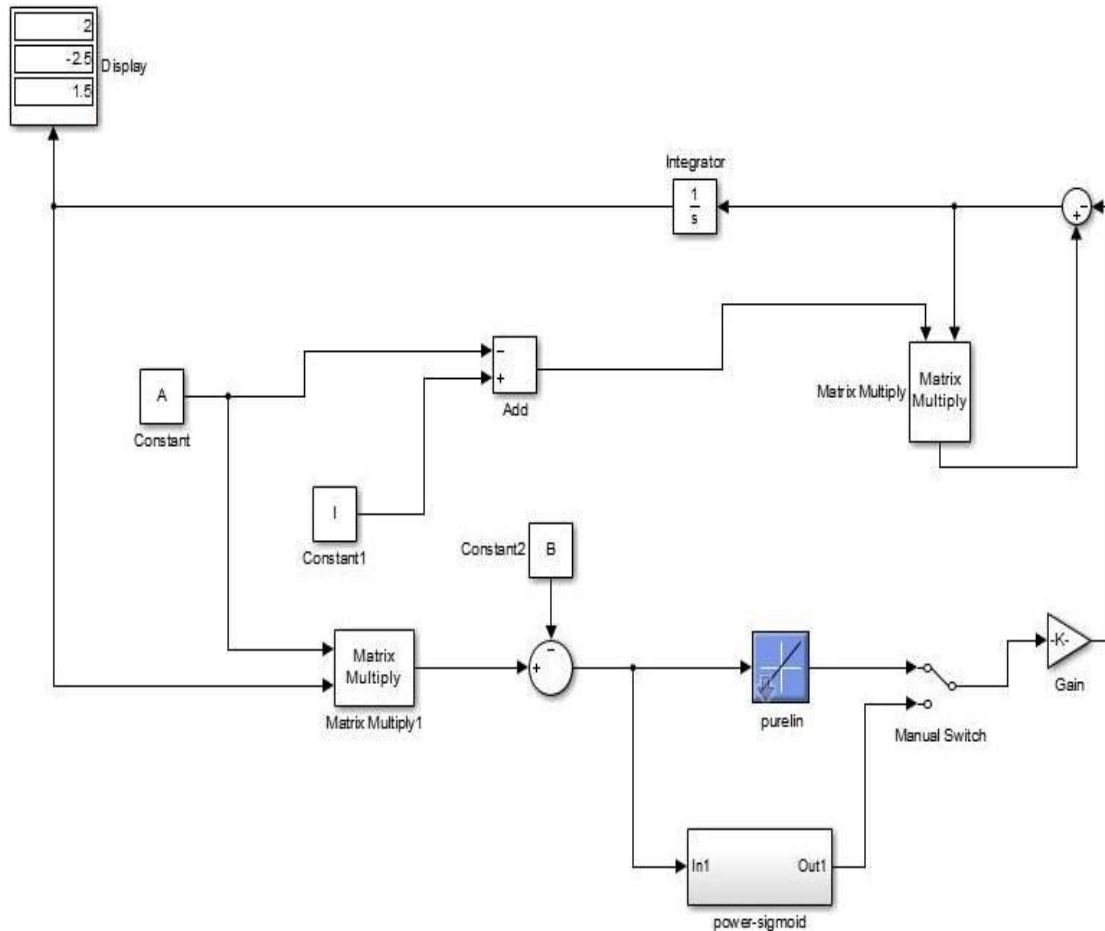
$$A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix}, b = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$$

1^{ος} τρόπος (Αλγεβρικά)

Όσον αφορά την αλγεβρική μέθοδο θα χρειαστεί να λύσουμε το σύστημα εξισώσεων $Ax=b$ και το αποτέλεσμα που θα πάρουμε είναι $x_0=2, x_1=-2.5, x_2=1.5$ (βέβαια θα πρέπει να τονιστεί ότι το Matlab υπολογίζει αριθμητικά το αποτέλεσμα.).

2^{ος} τρόπος ZNN Neural Network

Εφαρμόζοντας το κύκλωμα και τη μέθοδο του Yunong Zhang για $\gamma=100$ παίρνουμε τα παρακάτω αποτελέσματα:



Παρατηρούμε ότι το παραπάνω κύκλωμα μας δίνει ίδια αποτελέσματα με το πρώτο τρόπο. Επομένως αποτελεί κίνητρο η εφαρμογή για αρρωστημένους πίνακες.

Εφαρμογή (Osowski) [25]

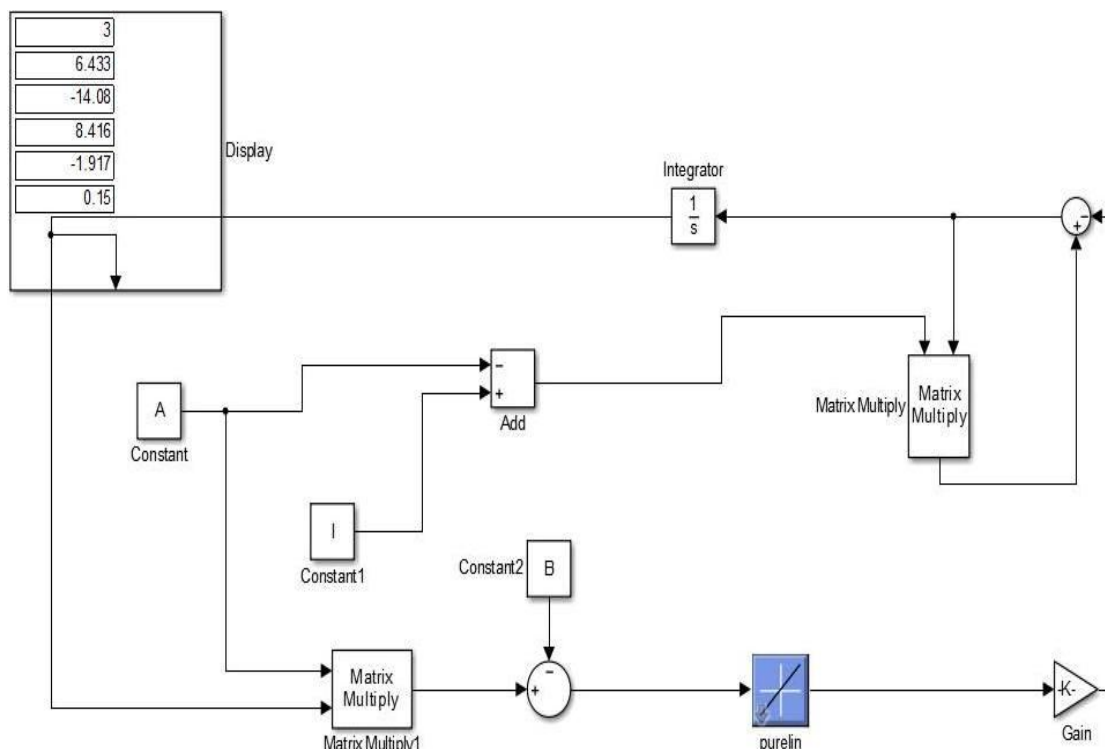
Να βρεθεί πολώνυμο παρεμβολής πέμπτου βαθμού που να παρεμβάλει τις τιμές $s_0=0, s_1=1, s_2=2, s_3=3, s_4=4, s_5=5$ με τιμές $Q(s_0)=3, Q(s_1)=2, Q(s_2)=1, Q(s_3)=4, Q(s_4)=5, Q(s_5)=6$. Θα χρειαστεί εφόσον θέλουμε να βρούμε τους συντελεστές να λύσουμε το παρακάτω γραμμικό σύστημα

$$\begin{bmatrix} 1 & s_0 & s_0^2 & s_0^3 & s_0^4 & s_0^5 \\ 1 & s_1 & s_1^2 & s_1^3 & s_1^4 & s_1^5 \\ 1 & s_2 & s_2^2 & s_2^3 & s_2^4 & s_2^5 \\ 1 & s_3 & s_3^2 & s_3^3 & s_3^4 & s_3^5 \\ 1 & s_4 & s_4^2 & s_4^3 & s_4^4 & s_4^5 \\ 1 & s_5 & s_5^2 & s_5^3 & s_5^4 & s_5^5 \end{bmatrix} \begin{bmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \\ q_4 \\ q_5 \end{bmatrix} = \begin{bmatrix} Q_0 \\ Q_1 \\ Q_2 \\ Q_3 \\ Q_4 \\ Q_5 \end{bmatrix}$$

που είναι της μορφής $Ax=b$. Είχαμε παρατηρήσει ότι η μέθοδος του Osowski δεν έδινε τα σωστά αποτελέσματα και αυτό οφειλότανε στο condition number του πίνακα A ($\text{cond}(A)=5.7689 \cdot 10^4$). Γι αυτό το λόγο για να μειώσουμε το $\text{cond}(A)$ χρησιμοποιούσαμε τη βελτιωμένη νευρωνική προσέγγιση του Osowski η οποία όμως με τη σειρά της μας έδινε, σε αργό ρυθμό, τα αποτελέσματα. Οπότε χρησιμοποιώντας τη μέθοδο του Yunong Zhang (για $\gamma=100$) για τους πίνακες:

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 & 16 & 32 \\ 1 & 3 & 9 & 27 & 81 & 243 \\ 1 & 4 & 16 & 64 & 256 & 1024 \\ 1 & 5 & 25 & 125 & 625 & 3125 \end{bmatrix} \text{ και } B = [3 \ 2 \ 1 \ 4 \ 5 \ 6]^T$$

Θα πάρουμε τα παρακάτω αποτελέσματα στο περιβάλλον του Matlab:



Ενώ λύνοντας αλγεβρικά το παραπάνω σύστημα θα είχαμε τα εξής αποτελέσματα:

$$x = \begin{bmatrix} 3.00000 \\ 6.43330 \\ -14.0830 \\ 8.41670 \\ -1.91670 \\ 0.15000 \end{bmatrix}$$

Και όπως παρατηρούμε, η παραπάνω λύση είναι πολύ κοντά στη λύση που παίρνουμε με τη μέθοδο του Yunong Zhang για γραμμικά συστήματα

Παρατηρήσεις

Από τη παραπάνω θεωρία και παραδείγματα μπορούμε να βγάλουμε τα παρακάτω συμπεράσματα για τη χρησιμότητα της μεθόδου που μόλις αναλύσαμε.

- Είναι πιο γρήγορη από τη μέθοδο Osowski (βελτιωμένη).
- Λειτουργεί ανεξαρτήτως αριθμού κατάστασης του πίνακα.
- Απλή στην εφαρμογή της.
- Ο ομαλός πίνακας A δε χρειάζεται να είναι κάποιας συγκεκριμένης μορφής όπως στη μέθοδο Osowski (Vandermonde στη πρώτη προσέγγιση και κάτω τριγωνικός στη βελτιωμένη μέθοδο).
- Μπορεί να χρησιμοποιηθεί στη θέση άλλων αριθμητικών μεθόδων όπως η LU παραγοντοποίηση.
- Γρηγορότερη σύγκλιση, αυξάνοντας τη τιμή της παραμέτρου γ .

Στην επόμενη υπό-ενότητα θα δούμε τα εισαγωγικά που αφορούν τη λύση γραμμικών συστημάτων και θα κάνουμε τη σύγκριση μεταξύ των ZNN μοντέλων και της LU παραγοντοποίησης.

4.6 Επίλυση Γραμμικών Συστημάτων

Εισαγωγή

Θεωρούμε ένα πραγματικό ομαλό γραμμικό σύστημα n εξισώσεων με n αγνώστους. Συγκεκριμένα, το

$$Ax = b \quad (4.6)$$

όπου $A \in R^{n \times n}$, $\det(A) \neq 0$, $b \in R^n$.

Από τη γραμμική άλγεβρα γνωρίζουμε ότι οι τύποι του Cramer δίνουν τη λύση τους από τις εκφράσεις:

$$x_i = \frac{\det(A_i)}{\det(A)} \quad (4.7)$$

όπου $i=1,2,\dots,n$.

Ο A_i είναι ο ίδιος πίνακας με τον A με τη μόνη διαφορά ότι η i -οστή του έχει αντικατασταθεί από τη στήλη των συνιστωσών του b . Όμως αν θελήσουμε να βρούμε τους αγνώστους x_i θα πρέπει να βρούμε τις τιμές των $n+1$ οριζουσών των τύπων (4.7). Το ανάπτυγμα όμως κάθε μίας από τις ορίζουσες αυτές αποτελείται από $n!$ όρους κάθε ένας από τους οποίους είναι γινόμενο n παραγόντων. Έτσι οι απαιτούμενοι πολλαπλασιασμοί είναι πλήθους $(n+1) \times (n-1) \times n! = (n-1) \times (n+1)!$, για την εκτέλεση μόνο των οποίων αν χρησιμοποιήσουμε έναν υπολογιστή ικανό να εκτελεί 10^6 πολλαπλασιασμούς/δευτερόλεπτο θα απαιτούνταν π.χ για $n=20$ περίπου 3×10^5 αιώνες. Αυτό δείχνει και το πρακτικά ανέφικτο της χρησιμοποίησης του κανόνα του Cramer για την εύρεση της λύσης του (4.6).

Ακόμη είναι γνωστό ότι η λύση του (4.6) δίνεται από τη κλειστή μορφή $x=A^{-1}b$. Επομένως θα μπορούσε κανείς να ισχυριστεί ότι είναι ίσως δυνατόν να βρεθεί η λύση βρίσκοντας πρώτα κατά κάποιον τρόπο τον αντίστροφο του A . Για την εύρεση του A^{-1} ο πιο προφανής τρόπος θα ήταν να εργαστεί κανείς ως εξής:

Έστω $X=A^{-1}$ ή ισοδύναμα $AX=I$. Αν x^i είναι τα διανύσματα στήλες του άγνωστου πίνακα X και e^i , $i=1,2,\dots,n$ τα αντίστοιχα διανύσματα στήλες του I , τότε θα ίσχυε $Ax^i=e^i$ και επομένως θα καταλήγαμε στην αντίφαση ότι ενώ το αρχικό μας πρόβλημα

ήταν η επίλυση ενός γραμμικού συστήματος για την εύρεση της λύσης του (4.8) τώρα έχουμε να επιλύσουμε n γραμμικά συστήματα για την εύρεση του A^{-1} και στη συνέχεια να βρούμε τη λύση του (4.6) από την έκφραση $x=A^{-1}b$.

Ως εκ τούτου σκοπός μας θα είναι η ανάπτυξη αποτελεσματικά ρεαλιστικών μεθόδων για την επίλυση του (4.6). Οι μέθοδοι διακρίνονται χοντρικά σε τρεις μεγάλες κατηγορίες. Στις άμεσες, όπου η λύση (με ακριβή αριθμητική μέθοδο) βρίσκεται μετά από ένα πεπερασμένο πλήθος πράξεων, στις έμμεσες ή επαναληπτικές, όπου κατασκευάζεται μία ακολουθία διανυσμάτων που κάτω από ορισμένες συνθήκες, και με ακριβή αριθμητική μέθοδο, συγκλίνει οριακά στην ακριβή λύση μετά από n επαναλήψεις. Οι τελευταίες μέθοδοι που έχουν γνωρίσει μεγάλη ανάπτυξη τις δύο τελευταίες δεκαετίες εφαρμόζονται από ερευνητές και χρήστες ως επαναληπτικές. Το ποια ακριβώς μέθοδο θα εφαρμοστεί σε μία συγκεκριμένη περίπτωση εξαρτιέται από τις ιδιότητες που μπορεί να έχει ο πίνακας των συντελεστών των αγνώστων A , από το υπάρχον λογισμικό, καθώς και από την εμπειρία.

Στα πλαίσια αυτού του κεφαλαίου θα ασχοληθούμε με την LU παραγοντοποίηση.

4.7 Μέθοδος LU παραγοντοποίησης

Ας υποθέσουμε ότι δίνεται για επίλυση το παρακάτω πραγματικό γραμμικό σύστημα n αλγεβρικών εξισώσεων με n αγνώστους.

$$Ax=b \quad (4.8)$$

Από τη γραμμική άλγεβρα γνωρίζουμε ότι αν:

α] Πολλαπλασιάσουμε μία εξίσωση του (4.8) επί έναν αριθμό $\lambda \in \mathbb{R} \setminus \{0\}$.

β] Εναλλάξουμε τη σειρά δύο εξισώσεων του (4.8).

γ] Αντικαταστήσουμε μία εξίσωση του (4.8) με το άθροισμα αυτής και μίας άλλης πολλαπλασιασμένης επί έναν αριθμό $\lambda \in \mathbb{R} \setminus \{0\}$.

προκύπτει το σύστημα:

$$A'x=b' \quad (4.9)$$

που είναι ισοδύναμο με το αρχικό (4.8).

Συνεχίζοντας έχουμε το παρακάτω θεώρημα:

Θεώρημα 4.8 [11]

Έστω $A \in \mathbb{R}^{n \times n}$ ένας ομαλός πίνακας. Τότε υπάρχει μοναδικός κάτω τριγωνικός πίνακας M με μοναδιαία διαγώνια στοιχεία και μοναδικός αντιστρέψιμος άνω τριγωνικός πίνακας U τέτοιος ώστε:

$$U = MA \quad (4.9)$$

Αν δε τεθεί $L = M^{-1}$ τότε:

$$LU = A \quad (4.10)$$

όπου L είναι κάτω τριγωνικός πίνακας με $l_{ii}=1$, με $i=1,2,\dots,n$ η δε συγκεκριμένη παραγοντοποίηση (4.10) είναι μοναδική.

Απόδειξη

Η απόδειξη θα γίνει με επαγωγή. Θέτουμε $A^{(i)} = A$ και παρατηρούμε ότι $\alpha_{11}^{(1)} \neq 0$. Αυτό διότι από την υπόθεση του θεωρήματος ο 1×1 κύριος υπό-πίνακας του A είναι αντιστρέψιμος. Το στοιχείο $\alpha_{11}^{(1)} \neq 0$, που θα εμφανίζεται το καλούμε οδηγός και η αντίστοιχη γραμμή οδηγός γραμμή. Βρίσκουμε στη συνέχεια τους αριθμούς:

$$m_{j1} = \frac{\alpha_{j1}^{(1)}}{\alpha_{11}^{(1)}}, \quad j = 2, 3, \dots, n$$

οι οποίοι καλούνται πολλαπλασιαστές, πολλαπλασιάζουμε τα στοιχεία της πρώτης γραμμής επί το πολλαπλασιαστή m_{j1} και αφαιρούμε τα γινόμενα που βρίσκουμε από τα αντίστοιχα στοιχεία της j -οστής γραμμής. Έτσι τα νέα στοιχεία της j -οστής γραμμής. Έτσι τα νέα στοιχεία της j -οστής γραμμής θα είναι:

$$\alpha_{jl}^{(2)} = \alpha_{jl}^{(1)} - m_{j1} \alpha_{1l}^{(1)}, \quad l = 1, 2, \dots, n$$

Η διαδικασία που μόλις εκτελέσαμε είναι εύκολο να διαπιστωθεί ότι ισοδυναμεί με το πολλαπλασιασμό του πίνακα $A^{(1)}$ από τα αριστερά επί τον πίνακα M_1 , όπου:

$$M_1 = \begin{bmatrix} 1 & & & & & \\ -m_{21} & 1 & & & & \\ -m_{31} & & 1 & & & \\ \vdots & & & \ddots & & \\ -m_{n1} & & & & & 1 \end{bmatrix}$$

Έχουμε λοιπόν ότι:

$$A^{(2)} = M_1 A^{(1)}$$

Υποθέτουμε ότι η προηγουμένως περιγραφείσα και εφαρμοσθείσα διαδικασία απαλοιφής έχει εκτελεστεί με ανάλογο τρόπο $k-1$ ($1 < k < n$) φορές και βρισκόμαστε την αρχή της εφαρμογής της για k -στη φορά, οπότε η κατάσταση παρουσιάζεται ως εξής:

$$A^{(k)} = M_{k-1} \cdots M_2 M_1 A^{(1)}$$

όπου

$$A^{(k)} = \begin{bmatrix} \alpha_{11}^{(1)} & \alpha_{12}^{(1)} & \cdots & \alpha_{1k}^{(1)} & \cdots & \alpha_{1n}^{(1)} \\ & & \cdots & \alpha_{2k}^{(2)} & \cdots & \alpha_{2n}^{(2)} \\ & & & \ddots & & \\ & & & & \alpha_{kk}^{(k)} & \cdots & \alpha_{kn}^{(k)} \\ & & & & \alpha_{k+1,k}^{(k)} & \cdots & \alpha_{k+1,n}^{(k)} \\ & & & & \alpha_{nk}^{(k)} & \cdots & \alpha_{nn}^{(k)} \end{bmatrix}$$

Καταρχάς αποδεικνύουμε ότι το στοιχείο $\alpha_{kk}^{(k)}$ που θα χρησιμεύσει ως οδηγός στο k βήμα της απαλοιφής δεν είναι 0. Επικεντρώνουμε τη προσοχή μας στο $k \times k$ κύριο υπό-πίνακα $A_{kk}^{(k)}$ του $A^{(k)}$. Αυτός έχει προκύψει από το $k \times k$ κύριο υπό-πίνακα $A_{k \times k}$ του A μετά από προσθέσεις, πολλαπλασιασμών γραμμών του σε άλλες γραμμές του. (Διότι ο πίνακας $A^{(k)}$ έχει προκύψει από τον A με τον ίδιο τρόπο.) Λόγω της αντιστρεψιμότητας του $A_{k \times k}$, από την υπόθεση του θεωρήματος, θα έχουμε $0 \neq \det(A_{k \times k}) = \det(A_{k \times k}^{(k)}) = \alpha_{11}^{(1)} \alpha_{22}^{(2)} \cdots \alpha_{k-1, k-1}^{(k-1)} \alpha_{kk}^{(k)}$ από τις οποίες βγάζουμε το συμπέρασμα ότι $\alpha_{kk}^{(k)} \neq 0$. Στη συνέχεια ορίζουμε τους πολλαπλασιαστές:

$$m_{jk} = \frac{\alpha_{jk}^{(k)}}{\alpha_{kk}^{(k)}}$$

$Ly=b$ για να βρούμε το y και έπειτα θα λύσουμε το $Ux=y$ για να βρούμε το x . Για αρχή ας βρούμε την LU παραγοντοποίηση του πίνακα A .

Συνεχίζοντας έχουμε:

$$\begin{bmatrix} 1 & 1 & -1 \\ 1 & -2 & 3 \\ 2 & 3 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 4 \\ -6 \\ 7 \end{bmatrix} \rightarrow A = \begin{bmatrix} 1 & 1 & -1 \\ 1 & -2 & 3 \\ 2 & 3 & 1 \end{bmatrix}$$

οπότε πρέπει να βρούμε τους πίνακες L, U τέτοιους ώστε $A=LU$. Κάνοντας τις στοιχειώδεις πράξεις έχουμε

$$A = \begin{bmatrix} 1 & 1 & -1 \\ 1 & -2 & 3 \\ 2 & 3 & 1 \end{bmatrix} \rightarrow -1\gamma_1 + \gamma_2$$

$$\begin{bmatrix} 1 & 1 & -1 \\ 0 & -3 & 4 \\ 2 & 3 & 1 \end{bmatrix} \rightarrow -2\gamma_1 + \gamma_3$$

$$\begin{bmatrix} 1 & 1 & -1 \\ 0 & -3 & 4 \\ 0 & 1 & 3 \end{bmatrix} \rightarrow \frac{1}{3}\gamma_2 + \gamma_3$$

$$\begin{bmatrix} 1 & 1 & -1 \\ 0 & -3 & 4 \\ 0 & 0 & 13/3 \end{bmatrix} = U$$

και

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

συνεχίζοντας πρώτα θα πρέπει να επιλύσουμε τα σύστημα $Ly=b$. Οπότε:

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 2 & -1/3 & 1 \end{bmatrix}}_L \cdot \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}}_y = \underbrace{\begin{bmatrix} 4 \\ -6 \\ 7 \end{bmatrix}}_b \rightarrow \begin{cases} y_1 = 4 \\ y_1 + y_2 = -6 \\ 2y_1 - \frac{1}{3}y_2 + y_3 = 7 \end{cases} \rightarrow \begin{cases} y_1 = 4 \\ y_2 = -10 \\ y_3 = \frac{-13}{3} \end{cases}$$

τελειώνοντας θα χρειαστεί να λύσουμε το σύστημα $Ux=y$. Άρα

$$\underbrace{\begin{bmatrix} 1 & 1 & -1 \\ 0 & -3 & 4 \\ 0 & 0 & 13/3 \end{bmatrix}}_U \cdot \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}}_x = \underbrace{\begin{bmatrix} 4 \\ -10 \\ -13/3 \end{bmatrix}}_y \rightarrow \begin{cases} x_1 + x_2 - x_3 = 4 \\ -3x_2 + 4x_3 = -10 \\ (13/3)x_3 = -(13/3) \end{cases} \rightarrow \begin{cases} x_1 = 1 \\ x_2 = 2 \\ x_3 = -1 \end{cases}$$

Οπότε το σύστημα λύθηκε με τη μέθοδο LU παραγοντοποίησης.

Επίσης η LU παραγοντοποίηση μπορεί να χρησιμοποιηθεί για την εύρεση αντιστρόφων πινάκων (σταθερών και μη σταθερών πινάκων) .

Παρατήρηση

Θα πρέπει να τονιστεί ότι η συγκεκριμένη μέθοδος χρησιμοποιείται πολλές φορές (λόγω χαμηλής πολυπλοκότητας πράξεων) για την εύρεση της ορίζουσας και του αριθμού κατάστασης ενός πίνακα A. Παρακάτω έχουμε τη σύγκριση της LU παραγοντοποίησης με το ZNN μοντέλο.

Συγκριτική Μελέτη LU παραγοντοποίησης-ZNN μοντέλου

Πατώντας τις παρακάτω εντολές:

A=rand(100,100);

B=rand(100,1);

x=inv(A)*B

format long;x

Εξάγουμε τα παρακάτω αποτελέσματα:

<u><i>Ακριβής Απόση</i></u>	<u><i>LU Απόση</i></u>	<u><i>ZNN Απόση</i></u>
-0.768885311367680	-0.768885311367655	-0.768885301424709
-0.638081967132791	-0.638081967132783	-0.638081958881403
-2.936304604638162	-2.936304604638046	-2.936304566667275
0.133489031174589	0.133489031174563	0.133489029448315
-0.202540519776020	-0.202540519775999	-0.202540502432915
0.739389078298522	0.739389078298501	0.739389014988051
0.168141110859261	0.168141110859209	0.168141108684857
-0.911915944891469	-0.911915944891434	-0.911915933098944
-0.488220325458479	-0.488220325458451	-0.488220319145194
0.596064475758032	0.596064475758022	0.596064468050093
-1.414304392721639	-1.414304392721607	-1.414304374432465
1.832077505632643	1.832077505632630	1.832077348763224
0.622637293668967	0.622637293668928	0.622637285617278
1.508119887445332	1.508119887445261	1.508119867943209
-0.687373044282213	-0.687373044282171	-0.687373035393361
-0.374326128908323	-0.374326128908346	0.374326124067864
-0.417360336788270	-0.417360336788290	-0.417360331391400
1.236322252821821	1.236322252821760	1.236322236834415
1.583247716035479	1.583247716035452	1.583247695561752
0.948244381910597	0.948244381910538	0.948244369648386
1.133383893982096	1.133383893982091	1.133383879325760
-0.371170391997666	-0.371170391997662	-0.371170387197993
0.391047121538251	0.391047121538232	0.391047116481473
0.254143519705536	0.254143519705535	0.254143516418980
-0.830495785045032	-0.830495785045015	0.830495774305615
1.074149238990523	1.074149238990515	1.074149225100163
-0.750291142988433	-0.750291142988394	-0.750291133286103
1.042414634200892	1.042414634200843	1.042414620721007
0.806416663617952	0.806416663617948	0.806416653189739

0.062427500058038	0.062427500058050	0.062427499250815
-0.613643150479496	-0.613643150479463	-0.613643142544116
0.692004175233775	0.692004175233748	0.692004166285136
-3.327043896512127	-3.327043896512070	-3.327043853488909
1.995087201263468	1.995087201263413	1.995087175464201
1.354494732061610	1.354494732061630	1.354494714546199
-1.248630083806305	-1.248630083806250	-1.248630067659390
-0.291771655000281	-0.291771655000232	-0.291771651227212
-0.247447183535930	-0.247447183535955	-0.247447180335922
-0.499335298897667	-0.499335298897682	-0.499335292440546
-0.669614717506032	-0.669614717506028	-0.669614708846972
-0.503017102942686	-0.503017102942641	-0.503017096437811
-1.206081556795785	-1.206081556795748	-1.206081541199540
-2.302360062689070	-2.302360062689025	-2.302360032916332
0.349191995204744	0.349191995204732	0.349191990689266
-0.910827527891126	-0.910827527891088	-0.910827516112888
1.692427080073649	1.692427080073551	1.692427058187931
0.616416219991552	0.616416219991535	0.616416212020349
0.872338013912256	0.872338013912245	0.872338002631670
-1.099597090478270	-1.099597090478191	-1.099597076258778
0.737911672651606	0.737911672651603	0.737911663109364
1.416209947494207	1.416209947494167	1.416209929180514
-0.081620786199130	-0.081620786199151	-0.081620785143715
-0.683011583860260	-0.683011583860276	-0.683011575028062
0.183242428707527	0.183242428707515	0.183242426338020
-1.669653703499711	-1.669653703499662	-1.669653681908793
0.672220715037398	0.672220715037370	0.672220706344583
-0.745412358446236	-0.745412358446225	-0.745412348806926
-1.588127988573550	-1.588127988573483	-1.588127968036631
-0.895753922748518	-0.895753922748488	-0.895753911165096
-1.886456503886784	-1.886456503886762	-1.886456479492299
0.279329128749032	0.279329128749028	0.279329125136894
-1.819100177508770	-1.819100177508711	-1.819100153985136

-0.842218245981276	-0.842218245981260	-0.842218235090106
0.868495629805855	0.868495629805819	0.868495618574967
-1.658716461159888	-1.658716461159847	-1.658716439710365
-0.051367791481841	-0.051367791481784	-0.051367790817426
0.215277693515284	0.215277693515276	0.215277690731533
1.079293412970391	1.079293412970373	1.079293399013644
2.353230481138164	2.353230481138087	2.353230450707353
2.312536534049684	2.312536534049601	2.312536504145309
0.707006052274594	0.707006052274620	0.707006043131860
0.732233206434523	0.732233206434513	0.732233196965686
-1.802005300499628	-1.802005300499586	-1.802005277197024
1.713543877468143	1.713543877468090	1.713543855309385
1.568077172293000	1.568077172292992	1.568077152015518
-2.489975178613585	-2.489975178613520	-2.489975146414492
1.280125087418534	1.280125087418535	1.280125070864721
-1.465133565765740	-1.465133565765733	-1.465133546819502
-1.705014640871191	-1.705014640871176	-1.705014618823054
-1.008637293802753	-1.008637293802732	-1.008637280759560
-0.461843905333916	-0.461843905333870	-0.461843899361657
0.157147716034859	0.157147716034831	0.157147714002658
0.216536864796896	0.216536864796937	0.216536861996912
-0.823715686346537	-0.823715686346558	-0.823715675694838
0.765307592367977	0.765307592367998	0.765307582471750
2.074154246212273	2.074154246212240	2.074154219390415
0.460879793010106	0.460879793010114	0.460879787050244
0.818291938648547	0.818291938648497	0.818291928066942
0.965588831888525	0.965588831888498	0.965588819402069
0.520897287050935	0.520897287050962	0.520897280315121
-0.448011209492198	-0.448011209492194	-0.448011203698748
0.682937894846822	0.682937894846806	0.682937886015336
0.240126868203242	0.240126868203256	0.240126865098009
-0.064184414972192	-0.064184414972206	-0.064184414142168
-0.042338977642572	-0.042338977642543	-0.042338977095187

0.059520308691097	0.059520308691068	0.059520307921411
0.848699116890850	0.848699116890804	0.848699105915817
0.480827307914308	0.480827307914307	0.480827301696613
0.918954239703518	0.918954239703438	0.918954227819934
-0.753093850351228	-0.753093850351223	-0.753093840612601

<i>Σχετικό Σφάλμα επί της % (για LU παραγοντοποίηση)</i>	<i>Σχετικό Σφάλμα επί της % (για ZNN μοντέλο)</i>
0.003162225091287e ⁻⁹	0.129316691319011e ⁻⁵
0.001252755318133e ⁻⁹	0.129315486760691e ⁻⁵
0.003932262149459e ⁻⁹	0.129315216054601e ⁻⁵
0.019170594386995e ⁻⁹	0.129319558860833e ⁻⁵
0.010140749109501e ⁻⁹	0.129320999316113e ⁻⁵
0.002822897101886e ⁻⁹	0.129317157985515e ⁻⁵
0.031017236840680e ⁻⁹	0.129320200225198e ⁻⁵
0.003798481489698e ⁻⁹	0.129315915337432e ⁻⁵
0.005741901741895e ⁻⁹	0.129312207104743e ⁻⁵
0.001732207596482e ⁻⁹	0.129313843656709e ⁻⁵
0.002292187769895e ⁻⁹	0.129315683718438e ⁻⁵
0.000739309382861e ⁻⁹	0.129314336079336e ⁻⁵
0.006276503329334e ⁻⁹	0.129315871518978e ⁻⁵
0.004682000745032e ⁻⁹	0.129314139716962e ⁻⁵
0.006073032114534e ⁻⁹	0.129316269232464e ⁻⁵
0.006347078360858e ⁻⁹	0.129311262269851e ⁻⁵
0.004774891511105e ⁻⁹	0.129309609773223e ⁻⁵
0.004921065006360e ⁻⁹	0.129314228609421e ⁻⁵
0.001725029262898e ⁻⁹	0.129314740243269e ⁻⁵
0.006181927024139e ⁻⁹	0.129314883054358e ⁻⁵
0.000470191128222e ⁻⁹	0.129314845963402e ⁻⁵
0.001046899394446e ⁻⁹	0.129311849648126e ⁻⁵
0.004727106361835e ⁻⁹	0.129313774793545e ⁻⁵
0.000720800590451e ⁻⁹	0.129318908197191e ⁻⁵
0.002085438541174e ⁻⁹	0.129313320954615e ⁻⁵

0.000702836839930e ⁻⁹	0.129314989454862e ⁻⁵
0.005223422018974e ⁻⁹	0.129314203588555e ⁻⁵
0.004750120084053e ⁻⁹	0.129314050589129e ⁻⁵
0.000509392399294e ⁻⁹	0.129315442106478e ⁻⁵
0.019395890995252e ⁻⁹	0.129305676169412e ⁻⁵
0.005355327685680e ⁻⁹	0.129315870168307e ⁻⁵
0.003946722777583e ⁻⁹	0.129314809797484e ⁻⁵
0.001708526266227e ⁻⁹	0.129313646634764e ⁻⁵
0.002771262659162e ⁻⁹	0.129313981507556e ⁻⁵
0.001524564678564e ⁻⁹	0.129313241615200e ⁻⁵
0.004392415185873e ⁻⁹	0.129317047473022e ⁻⁵
0.016704429626996e ⁻⁹	0.129315814206026e ⁻⁵
0.010139957949000e ⁻⁹	0.129320874599197e ⁻⁵
0.003090528563944e ⁻⁹	0.129314336406128e ⁻⁵
0.000547140898412e ⁻⁹	0.129314065375320e ⁻⁵
0.008850582420960e ⁻⁹	0.129317177407938e ⁻⁵
0.003019307857128e ⁻⁹	0.129313352602813e ⁻⁵
0.001948131872235e ⁻⁹	0.129313998185103e ⁻⁵
0.003592728456134e ⁻⁹	0.129312193036559e ⁻⁵
0.004217452314048e ⁻⁹	0.129313598183627e ⁻⁵
0.005785872367847e ⁻⁹	0.129315572530468e ⁻⁵
0.002809705296916e ⁻⁹	0.129315272432448e ⁻⁵
0.001272698205190e ⁻⁹	0.129314396185960e ⁻⁵
0.007208997244960e ⁻⁹	0.129315477218558e ⁻⁵
0.000376136828354e ⁻⁹	0.129314148972247e ⁻⁵
0.002759467304685e ⁻⁹	0.129314813773533e ⁻⁵
0.024705012808330e ⁻⁹	0.129307189826043e ⁻⁵
0.002389458516904e ⁻⁹	0.129312556720576e ⁻⁵
0.006740377343868e ⁻⁹	0.129309978484445e ⁻⁵
0.002939044041622e ⁻⁹	0.129313747366318e ⁻⁵
0.004178485115778e ⁻⁹	0.129314900195608e ⁻⁵
0.001504301937264e ⁻⁹	0.129315131075595e ⁻⁵
0.004222422321742e ⁻⁹	0.129315262875246e ⁻⁵

0.003371245774414e ⁻⁹	0.129314780848113e ⁻⁵
0.001165275522774e ⁻⁹	0.129313791015967e ⁻⁵
0.001172508551941e ⁻⁹	0.129314747989739e ⁻⁵
0.003283491446119e ⁻⁹	0.129314673741637e ⁻⁵
0.001885044562724e ⁻⁹	0.129315294804858e ⁻⁵
0.004154568781006e ⁻⁹	0.129314273128858e ⁻⁵
0.002503281428337e ⁻⁹	0.129313981918367e ⁻⁵
0.110956832837250e ⁻⁹	0.129344702107967e ⁻⁵
0.004009697370642e ⁻⁹	0.129309783857193e ⁻⁵
0.001666424790774e ⁻⁹	0.129313738780912e ⁻⁵
0.003302507440163e ⁻⁹	0.129315049640236e ⁻⁵
0.003591064660783e ⁻⁹	0.129314173142723e ⁻⁵
0.003690242961111e ⁻⁹	0.129316205779245e ⁻⁵
0.001394917867258e ⁻⁹	0.129314500839234e ⁻⁵
0.002378162191790e ⁻⁹	0.129314848764093e ⁻⁵
0.003084053852781e ⁻⁹	0.129315384900003e ⁻⁵
0.000523931508435e ⁻⁹	0.129314315266002e ⁻⁵
0.002603922528827e ⁻⁹	0.129314916253368e ⁻⁵
0.000121418777723e ⁻⁹	0.129314019868518e ⁻⁵
0.000424346904826e ⁻⁹	0.129314061792316e ⁻⁵
0.000885566186528e ⁻⁹	0.129313479247642e ⁻⁵
0.002047331422793e ⁻⁹	0.129314998069871e ⁻⁵
0.009879997502425e ⁻⁹	0.129313368926311e ⁻⁵
0.017803389655595e ⁻⁹	0.129317891527443e ⁻⁵
0.018701381338086e ⁻⁹	0.129307501903193e ⁻⁵
0.002520429185047e ⁻⁹	0.129312805507907e ⁻⁵
0.002712787742802e ⁻⁹	0.129310448984070e ⁻⁵
0.001584385615916e ⁻⁹	0.129314674243179e ⁻⁵
0.001854869191331e ⁻⁹	0.129314878371508e ⁻⁵
0.006159675164493e ⁻⁹	0.129313322371172e ⁻⁵
0.002874471482996e ⁻⁹	0.129314424649622e ⁻⁵
0.005285788906021e ⁻⁹	0.129311740145354e ⁻⁵
0.000755824888610e ⁻⁹	0.129314838495365e ⁻⁵

0.002422229487006e ⁻⁹	0.129316097212728e ⁻⁵
0.005929619786973e ⁻⁹	0.129316332739745e ⁻⁵
0.020864979824686e ⁻⁹	0.129318623858435e ⁻⁵
0.067522279648417e ⁻⁹	0.129286187177244e ⁻⁵
0.049301797932397e ⁻⁹	0.129314779319856e ⁻⁵
0.005428809174532e ⁻⁹	0.129315953442237e ⁻⁵
0.000150084022710e ⁻⁹	0.129312423983805e ⁻⁵
0.008686508209190e ⁻⁹	0.129316387196838e ⁻⁵
0.000678139372788e ⁻⁹	0.129314926240470e ⁻⁵

Συμπέρασμα

1] Τα ZNN μοντέλα όπως παρατηρούμε δίνουν πολύ καλή προσέγγιση (μέχρι και 7 δεκαδικά ψηφία) για τη λύση συστημάτων.

2] Η LU παραγοντοποίηση δίνει καλύτερα αποτελέσματα σε μικρότερο χρόνο, από τα ZNN μοντέλα. Βέβαια αυτό έχει να κάνει και με το γεγονός ότι η συγκεκριμένη μέθοδος είναι πιο παλιά και έχει υποστεί βελτιώσεις όσον αφορά τη λειτουργία της και τη προγραμματιστική της υλοποίηση.

Γενικά Συμπεράσματα

Στη παρούσα διπλωματική εργασία παρουσιάσαμε και αποδείξαμε μεθόδους σύμφωνα με τις οποίες μπορούμε να κατασκευάσουμε νευρωνικά δίκτυα για να επιλύουν συγκεκριμένες εξισώσεις πινάκων και εξετάσαμε κατά πόσο μπορεί να γίνει εφαρμογή σε πολυωνυμικούς πίνακες. Συγκρίναμε τα δύο μοντέλα (ZNN και GNN) και καταλήξαμε στο συμπέρασμα ότι τα ZNN δίνουν καλύτερα αποτελέσματα σε λιγότερο χρόνο (εξηγήσαμε τους λόγους) και επιλύουν το πρόβλημα που σχετίζεται με τον αριθμό κατάστασης ενός πίνακα. Στη συνέχεια είδαμε μία εφαρμογή για τον υπολογισμό συναρτήσεων μεταφοράς ηλεκτρικών κυκλωμάτων (μέθοδος Osowski) και συγκρίναμε τα ZNN μοντέλα, που αφορούν την επίλυση αλγεβρικών συστημάτων, με τη μέθοδο της LU παραγοντοποίησης. Καταλήξαμε στο συμπέρασμα ότι το συγκεκριμένο μοντέλο δίνει αποτελέσματα με την ακρίβεια 7 δεκαδικών ψηφίων. Οπότε από τη παρούσα διπλωματική, όποιος ερευνητής θέλει να κατασκευάσει ένα επαναλαμβανόμενο νευρωνικό δίκτυο (σύστημα διαφορικών εξισώσεων) για να επιλύσει συγκεκριμένη εξίσωση πινάκων θα πρέπει να ακολουθήσει τα παρακάτω βήματα:

- 1] Να αποδείξει ότι το σύστημα διαφορικών είναι ασυμπτωτικά ευσταθές.
- 2] Να αποδείξει ότι όντως το συγκεκριμένο σύστημα, επιλύει την εξίσωση πινάκων που θέλει ο ερευνητής.

Μελέτη για περαιτέρω έρευνα.

Στη παρούσα εργασία μελετήθηκε η κατασκευή επαναλαμβανόμενων νευρωνικών δικτύων (δυναμικών συστημάτων) για την αντιμετώπιση χρονικά μεταβαλλόμενων προβλημάτων, επίλυση συστημάτων και παρεμβολής. Όμως κατά τη διάρκεια της εκπόνησης της δημιουργήθηκαν οι παρακάτω προβληματισμοί και ερωτήματα για περαιτέρω έρευνα:

- Η εφαρμογή των ZNN μοντέλων για επίλυση και άλλων εξισώσεων πινάκων όπως Riccati ή Διοφαντικές Εξισώσεις.
- Η εφαρμογή του επαναλαμβανόμενου νευρωνικού δικτύου του Stanimirovic για τον υπολογισμό Drazin inverse χρονικά μεταβαλλόμενων πινάκων.
- Προγραμματισμός των παραπάνω μεθόδων σε άλλα λογισμικά όπως το Gnu Octave ή Scilab.
- Γενίκευση των παραπάνω μοντέλων και στη περίπτωση των δύο μεταβλητών (t =χρόνος, x =θέση).
- Επίλυση του ZPS (Zhang Open Problem) για την εφαρμογή του ZNN μοντέλου και για πολυωνυμικούς πίνακες.
- Η εφαρμογή των παραπάνω μοντέλων για την επίλυση μη γραμμικών εξισώσεων πινάκων.
- Την εφαρμογή και άλλων προσεγγιστικών μεθόδων επίλυσης διαφορικών εξισώσεων εκτός της ode45, όπως για παράδειγμα της ode15s [27] (για τη συγκριτική μελέτη σφάλματος μεταξύ των μεθόδων και, γιατί όχι, την επίλυση του προβλήματος του conditioning).
- Την εφαρμογή του νευρωνικού δικτύου του Stanimirovic για τον υπολογισμό $\{1\}$ -αντίστροφου ενός πίνακα.

Παράρτημα

• Δημιουργία Vandemorde πίνακα στη μέθοδο Osowski (στο Matlab)

```
n=input('give degree');
k=n+1;
for i=1:k
    str='please put the interpolation points';
    z=input(str);
    s(i)=z;
end
for i=1:k
    for j=1:k
        A(i,j)=s(i)^(j-1);
    end
end
for i=1:k
    str1='please put the polynomial value';
    Q=input(str1);
    B(i)=Q;
end
A
B
cond(A)
```

• Δημιουργία κάτω τριγωνικού πίνακα στη βελτιωμένη μέθοδο Osowski (στο Matlab)

```
n=input('give degree')
k=n+1;
for i=1:k
    str='please put the interpolation points';
    z=input(str);
    s(i)=z;
end
for i=1:k
```



```

        A(i,1)=1;
end
for j=2:k
    A(1,j)=0;
end
for i=2:k
    for j=2:k
        if (i>=j)
            v=1;
            for p=1:(j-1)
                v=v*(s(i)-s(p));
            end
            A(i,j)=v;
        else
            A(i,j)=0;
        end
    end
end
end
for i=1:k
    str1='please put the polynomial value';
    Q=input(str1);
    B(i)=Q;
end

```

A

cond(A)

B

• **Συνάρτηση για τη σχέση $q_i = q_i - q_{i+1} s_k$**

```
function[coeffs]=OriginalCoefficients( CoeffsAndPoints )
```

```
%UNTITLED2 Summary of this function goes here
```

```
% This function calculates the original coefficients
for an interpolation
```

```
% that has been carried out using the coefficients of
the Newton version
```

```

% of a polynomial.

coeffs = CoeffsAndPoints(1,:);
n = length(coeffs);
for k = n-1:-1:1
    for i = k:n-1
        coeffs(i) = coeffs(i) -
CoeffsAndPoints(2,k)*coeffs(i+1);
    end
end

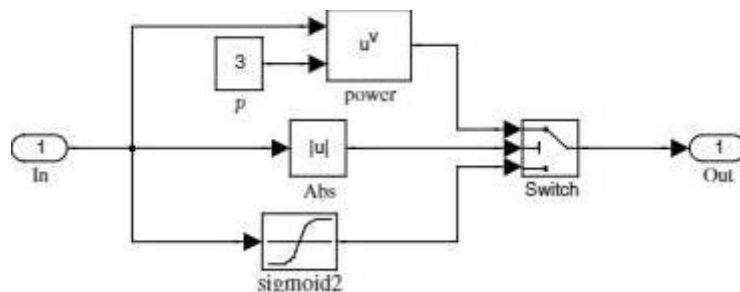
end

```

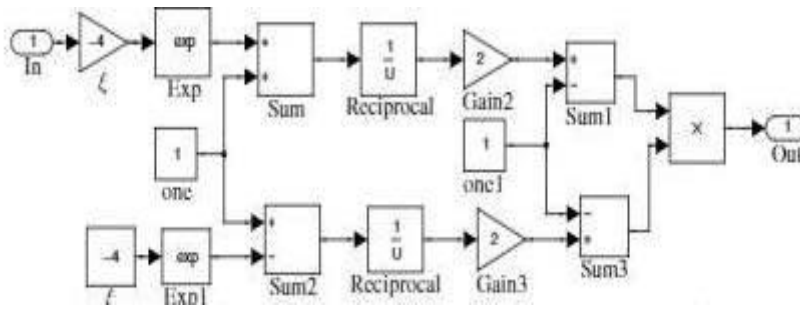
• **Butcher Tableau για τη μέθοδο Dormand-Price**

c_i	a_{ij}					\hat{b}_i	b_i
0	$\frac{1}{5}$					$\frac{35}{384}$	$\frac{5179}{57600}$
$\frac{1}{5}$	$\frac{3}{40}$	$\frac{9}{40}$				0	0
$\frac{3}{10}$	$\frac{44}{45}$	$-\frac{56}{15}$	$\frac{32}{9}$			$\frac{50}{1113}$	$\frac{7571}{16695}$
$\frac{4}{5}$	$\frac{19372}{6561}$	$-\frac{25360}{2187}$	$\frac{64448}{6561}$	$-\frac{212}{729}$		$\frac{125}{192}$	$\frac{393}{640}$
$\frac{8}{9}$	$\frac{9017}{3168}$	$-\frac{355}{33}$	$\frac{46732}{5247}$	$\frac{49}{176}$	$-\frac{5103}{18656}$	$\frac{2187}{6784}$	$-\frac{92097}{339200}$
1	$\frac{35}{384}$	0	$\frac{500}{1113}$	$\frac{125}{192}$	$-\frac{2187}{6784}$	$\frac{11}{84}$	$\frac{1}{40}$

• **Υποσυστήματα της Power-Sigmoid**

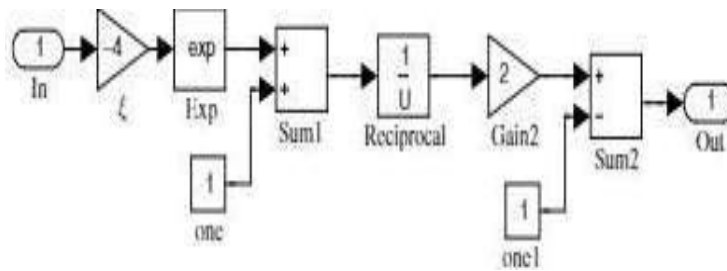


Power-Sigmoid στο Matlab



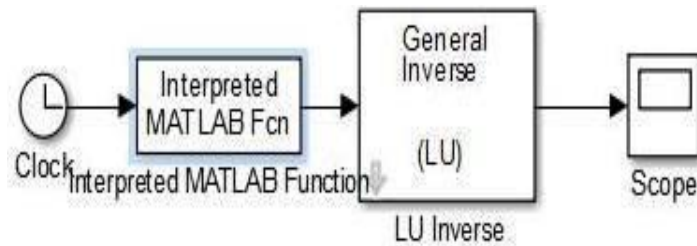
Υποσύστημα της Power-Sigmoid

• **Υποσύστημα της Bipolar-Sigmoid**

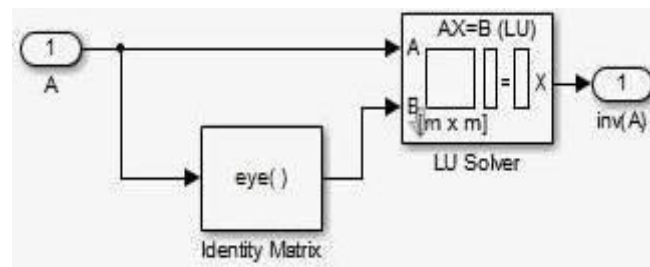


Διπολική σιγμοειδή στο Matlab

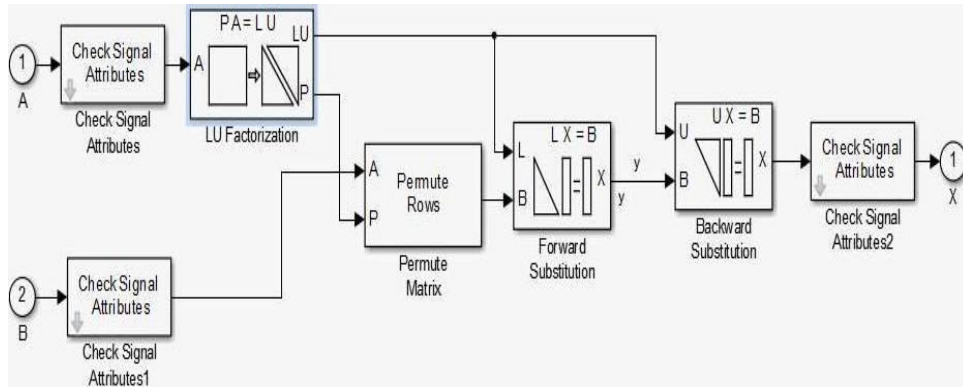
• **Simulink-Block της LU παραγοντοποίησης (για αντιστροφή πινάκων)**



LU Factorization

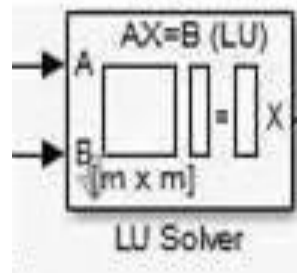


Υποσύστημα LU παραγοντοποίησης

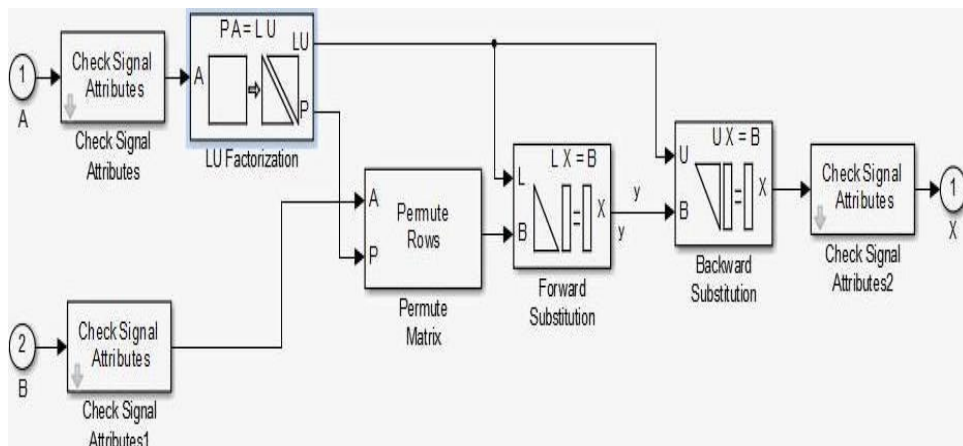


Υποσύστημα LU Solver

• **Simulink-Block της LU παραγοντοποίησης (για επίλυση συστημάτων $Ax=b$)**

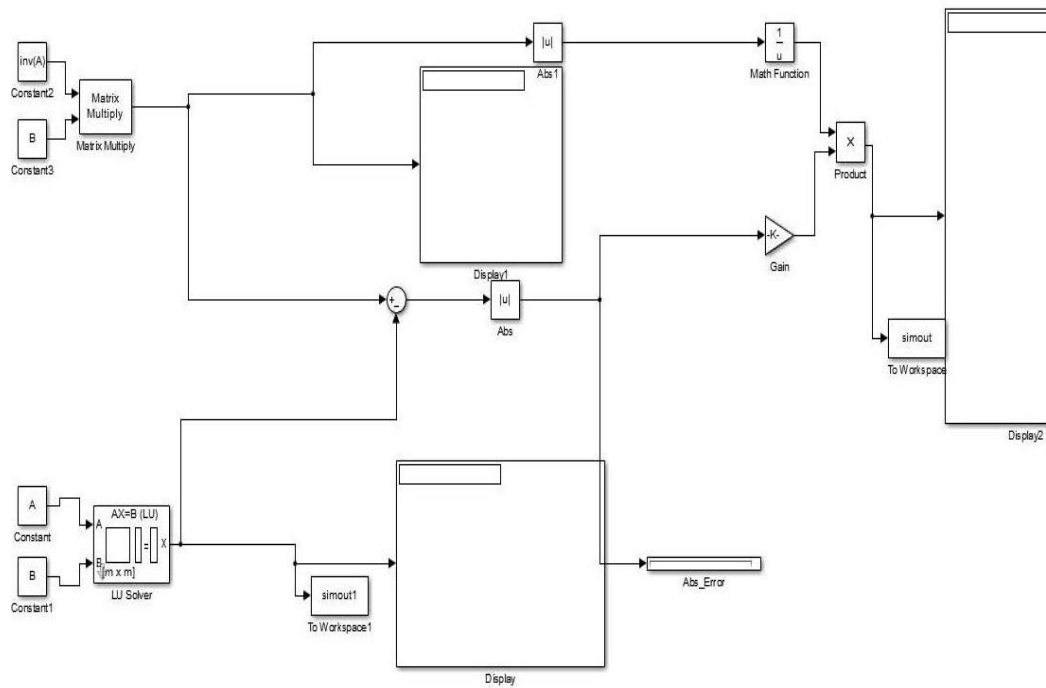


Lu Solver

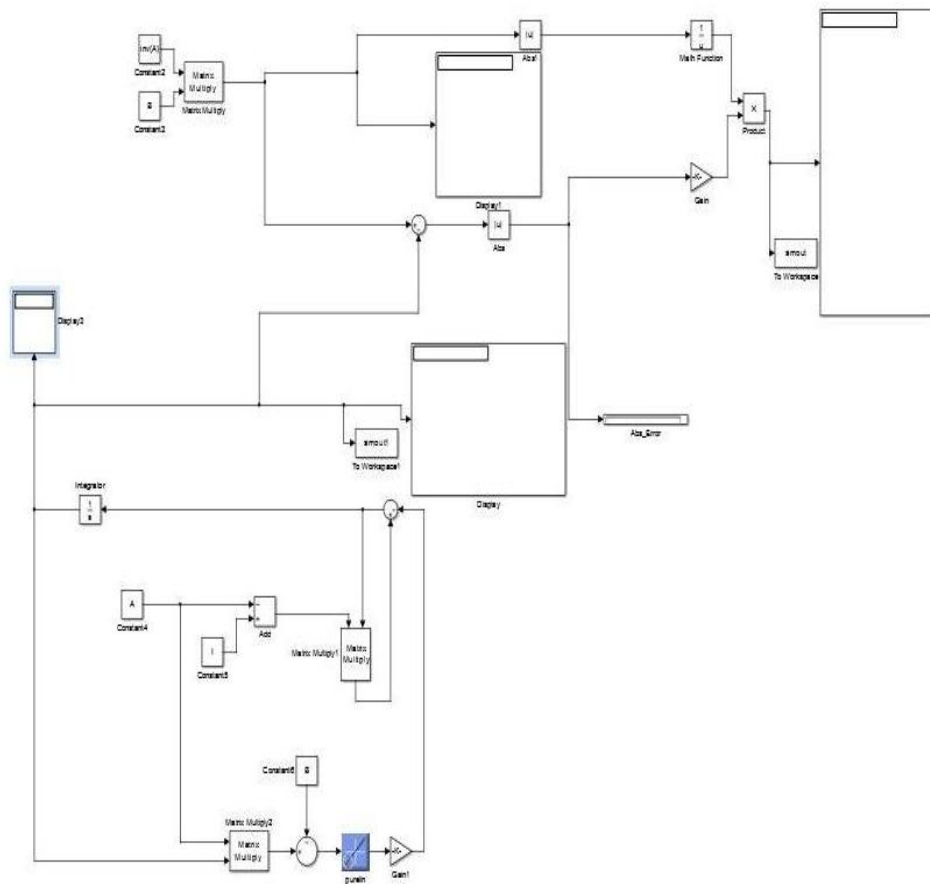


Υποσύστημα LU Solver

• Simulink-Block για συγκριτική μελέτη για LU παραγοντοποίησης



• Simulink-Block για συγκριτική μελέτη για ZNN μοντέλο



Βασικές εντολές στο Matlab

$\text{inv}(A)$ —επιστρέφει τον αντίστροφο του πίνακα A .

$\text{cond}(A)$ —επιστρέφει τον αριθμό κατάστασης του πίνακα A .

$\text{norm}(A)$ —επιστρέφει τη νόρμα του πίνακα A .

$\text{norm}(A, 'fro')$ —επιστρέφει τη Frobenius-νόρμα του πίνακα A

$\text{transpose}(A)$ —επιστρέφει τον ανάστροφο του πίνακα A .

$\text{eig}(A)$ —επιστρέφει τις ιδιοτιμές του πίνακα A .

$\text{min}(A)/\text{max}(A)$ —επιστρέφει το ελάχιστο/μέγιστο του συνόλου A .

$\text{abs}(a)$ —επιστρέφει την απόλυτη τιμή του αριθμού a .

$I=\text{eye}(n,n)$ —επιστρέφει το μοναδιαίο πίνακα $n \times n$ διαστάσεων.

$\text{rank}(A)$ —επιστρέφει τη βαθμίδα του πίνακα A .

Βασικές εντολές στο Mathematica

$\text{NSolve}[\dots]$ —Επιλύει προσεγγιστικά εξισώσεις

$\text{Arg}[z]$ —επιστρέφει το όρισμα ενός του μιγαδικού αριθμού z .

Βιβλιογραφία

- [1] S.S Ge, T.H.Lee and C.J.Harris, *Adaptive Neural Network Control of Robotic Manipulators*. London U.K World Scientific, 1998.
- [2] K.M Abadir and T.R.Magnus, *Matrix Algebra*, vol Cambridge ,2005.
- [3] D.L Kleinman and M, Athans, "The design of suboptimal linear time varying systems," IEEE Trans, Autim.Control vol 13 no.2 pp 150-159 Apr 1968.
- [4] Ροδάνθη Βιδάλη, *Αυτόματη ταξινόμηση μελωδίας σε μουσικά είδη με τη χρήση τεχνητών νευρωνικών δικτύων*, Μάρτιος 2011.
- [5] Ελένη Ανδριέλου, *Νευρωνικά δίκτυα και εφαρμογές στη θεωρία ελέγχου*, Σεπτέμβριος 2004.
- [6] Ευαγγελία Αραμπατζή, *Διαίρεση πολυωνυμικών πινάκων*, Ιούλιος 2015.
- [7] Σταύρος Βολογιαννίδης, *Ευφύης Έλεγχος, Θεωρία και εφαρμογές*.
- [8] Π.-Χ.Γ. Βασιλείου, Γεώργιος Τσακλίδης, *Εφαρμοσμένη Θεωρία Πινάκων*, Εκδόσεις Ζήτη, Θεσσαλονίκη 2003.
- [9] Μαρία Γουσιδου-Κουτίτα, *Αριθμητική Ανάλυση*, Θεσσαλονίκη 2004, Εκδόσεις Χριστοδουλίδη.
- [10] Μαρία Γουσιδου-Κουτίτα, *Υπολογιστικά Μαθηματικά* Εκδόσεις Χριστοδουλίδη Θεσσαλονίκη 2014 .
- [11] Β Δουγάλης, Δ Νούτσος, Α. Χατζηδήμος, *Αριθμητική Γραμμική Άλγεβρα*, Ιωάννινα 2012.
- [12] Αχιλλέας-Κωσταντίνος Θεοδωρακόπουλος, *Μεταφορά κωδίκων της Matlab σε προγράμματα C++ για την επίλυση Κανονικών Διαφορικών Εξισώσεων*, Ιούνιος 2015.
- [13] Yunong Zhang and Shizhi Ge, 'Design and Analysis OF A General Recurrent Neural Network Model for Time-Varying Matrix Inversion', IEEE Transactions on Neural Networks, vol 16, No 6 November 2005 1477-1490.
- [14] Yunong Zhang, Binghuang Cai, 'Matlab Simulink Modeling and Simulation of Zhang Neural Network for Online Time-Matrix Inversion', conference paper, April 2008.
- [15] Yunong Zhang, Ke Chen, Weimu Ma and Xiao-Dong-Li, 'Matlab Simulation of Gradient-Based Neural Network for Online Matrix Inversion, 2008.
- [16] C. Yi and Y Zhang, ' Analogue Recurrent Neural Network for Linear Algebraic Equation Solving', conference paper, February 2008.

- [17] Predrag S. Stanimirovic, Ivan S. Zivkovic and Yimin Wei, 'Recurrent *Neural Network for Computing the Drazin Inverse*, IEEE Transactions on Neural Networks and Learning Systems, vol 26, No 11 November 2015 2830-2843.
- [18] J Wang, *A recurrent neural network for real-time matrix inversion*, Appl.Math.Comput, vol55.
- [19] *Matlab Simulation of Gradient-Based Neural Network for online matrix inversion*, September 2009 .
- [20] *Performance Analysis of Gradient Neural Network exploited for on-line time-varying matrix inversion*, September 2009.
- [21] Luenberger, D. G., & Ye, Y. (1984). *Linear and nonlinear programming* (Vol. 2). Reading, MA: Addison-wesley.
- [22] C. Mead, *Analog VLSI and Neural Systems*, Reading, Mass.: Addison-Wesley, 1989.
- [23] D. Tank and J. Hopfield, *Simple neural optimization networks*, *IEEE Trans. CAS* 33 (1986) 533-541.
- [24] N.J.Higham *Functions of Matrices: Theory and Computation*, Philadelphia, PA, USA:SIAM, 2009 .
- [25] Stanislaw Osowski, *Neural networks in interpolation problems*, *Neurocomputing* 5 (1993) 105-118.
- [26] Minsky M.L. and Papert S.A. 1969. *Perceptrons*. Cambridge, MA:MIT Press.
- [27] <http://www.mathworks.com/help/matlab/ref/ode15s.html>.
- [28] <http://www4.ncsu.edu/~zhilin/TEACHING/MA302/ode45.pdf>.