Μοντελοποίηση και Πρόβλεψη Σήματος Φωνής με Μεθόδους Ανάλυσης Χαοτικών Χρονοσειρών

Ιάσονας Κόκκινος και Πέτρος Μαραγκός

Εθνικό Μετσόβιο Πολυτεχνείο

Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Τομέας Σημάτων, Ελέγχου και Ρομποτικής

Απρίλιος 2002

Τεχνική αναφορά για το ερευνητικό πρόγραμμα
ΑΡΧΙΜΗΔΗΣ
του ΕΠΙΣΕΥ–ΕΜΠ

Περιεχόμενα

1 Εισαγωγή 3

2 Ανάλυση Χαοτικών Χρονοσειρών 4
- 2.1 Βασικές έννοιες . . . 4
- 2.2 Ανακατασκευή ελκυστή χαοτικού συστήματος . . . 5
  - 2.2.1 Προσδιορισμός χρονικής καθυστέρησης T . . . 7
  - 2.2.2 Προσδιορισμός διάστασης εμβύθισης D_e . . . 9
- 2.3 Προσέγγιση της δυναμικής του συστήματος στο χώρο φάσης . . . 11
  - 2.3.1 Με local polynomials . . . 12
  - 2.3.2 Με global polynomials . . . 15
  - 2.3.3 Με τοπικά μοντέλα σε συστάδες του ελκυστή . . . 18
  - 2.3.4 Με νευρωνικά δίκτυα RBF . . . 20
  - 2.3.5 Με δίκτυα τύπου WCP,WLP . . . 23
  - 2.3.6 Με δίκτυα τύπου SVM . . . 24
  - 2.3.7 Συμπεράσματα –σύγκριση MSE πρόβλεψης . . . 27
- 2.4 Αναλλοιώτες ποσότητες ενός χαοτικού δυναμικού συστήματος . . . 28
  - 2.4.1 Κλασματική διάσταση (fractal dimension) . . . 28
  - 2.4.2 Διάσταση συσχέτισης (Correlation Dimension) . . . 29
  - 2.4.3 Εκθέτες Lyapunov . . . 31
  - 2.4.4 Προσδιορισμός του φάσματος Lyapunov ενός συστήματος . . . 33
  - 2.4.5 Τοπική διάσταση συστήματος–ψεύτικοι εκθέτες Lyapunov . . . 38
  - 2.4.6 Σύγκριση μοντέλων με βάση τους εκθέτες Lyapunov . . . 40

3 Εφαρμογή Μεθόδων Ανάλυσης Χαοτικών Χρονοσειρών στην Πρόβλεψη/Ανάλυση Σήματος Φωνής 46
- 3.1 Φωνήεντα . . . 48
- 3.2 Τυρβώδεις ήχοι . . . 62
- 3.3 Κρουστικοί ήχοι . . . 68
- 3.4 Εφαρμογή των μεθόδων ανάλυσης χαοτικών χρονοσειρών στην αναγνώριση μεμονωμένων φωνημάτων . . . 75

4 Συμπεράσματα – κατευθύνσεις μελλοντικής έρευνας 78

A Συστήματα TSK 80

B Δίκτυα Support Vector Machine 94
- B.1 Δίκτυα SVM για ταξινόμιση . . . 95
- B.2 Support Vector Machines για πρόβλεψη . . . 103

Κεφάλαιο 1

Εισαγωγή

Το καθιερωμένο μοντέλο περιγραφής του συστήματος παραγωγής φωνής βασίζεται στην παραδοχή ότι οι κανόνες που διέπουν τη συμπεριφορά του μπορούν να αποδοθούν με γραμμικές διαφορικές εξισώσεις [7] οπότε και η μοντελοποίησή του μπορεί να γίνει με γραμμικές μεθόδους (φίλτρα ARMA).

Αποτελέσματα σύγχρονης πειραματικής έρευνας έχουν αποδείξει, ωστόσο, ότι το σήμα φωνής χαρακτηρίζεται από πολύπλοκη δομή ,η οποία δεν είναι δυνατό να προέρχεται από ένα γραμμικό σύστημα, αφού αφ'ενός η μέτρηση της ροής του αέρα στην στοματική κοιλότητα κατέδειξε ότι η γραμμική σχέση που συνδέει τη ροή του αέρα με την πίεση παραβιάζεται [11], αφ'εταίρου ο ανακατασκευασμένος ελκυστής του συστήματος έχει κλασματική διάσταση[13]. Καθώς την τελευταία δεκαετία η μη-γραμμική ανάλυση χρονοσειρών έχει γνωρίσει μεγάλη πρόοδο, ενδιαφέρον έχει να εξεταστεί αν και κατά πόσο οι μη-γραμμικές μέθοδοι δίνουν καλύτερα αποτελέσματα στην πρόβλεψη και στην ανάλυση σήματος φωνής από τις γραμμικές.

Χρησιμοποιήθηκαν μεθόδοι που εφαρμόζονται στην ανάλυση μη γραμμικών χρονοσειρών και εξετάστηκε κατά πόσο μπορεί να επιτευχθεί μικρότερο σφάλμα πρόβλεψης για το σήμα φωνής με τις μεθόδους αυτές, καθώς και αν μπορούμε μέσω ορισμένων χαρακτηριστικών της δυναμικής του μη γραμμικού μοντέλου που χρησιμοποιούμε να αναγνωρίσουμε τα διάφορα φωνήματα.

Προηγούμενη δουλειά στο αντικείμενο της πρόβλεψης φωνής με μη γραμμικές μεθόδους υπάρχει στο [9] καθώς και στο [10] αλλά καθώς από τότε που δημοσιεύτηκαν τα παραπάνω άρθρα έχουν υπάρξει αρκετές εξελίξεις στο χώρο της μη γραμμικής ανάλυσης χρονοσειρών, στην εργασία αυτή, χρησιμοποιώντας πρόσφατες μεθόδους, πήραμε καλύτερα αποτελέσματα από αυτά που υπάρχουν στις παραπάνω δημοσιεύσεις, ενώ εξετάστηκαν διεξοδικότερα οι επιδόσεις διαφορετικών και σύγχρονων μοντέλων που δεν είχαν εφαρμοστεί, ως τώρα, για την πρόβλεψη σήματος φωνής.

Οι κυριότερες τυπογραφικές συμβάσεις για τους συμβολισμούς στη συνέχεια του κειμένου είναι οι παρακάτω:

- $\mathbf{X}$: πίνακας
- X :διάνυσμα
- X_i :i-οστή συνιστώσα διανύσματος X
- x :βαθμωτή ποσότητα

Κεφάλαιο 2

Ανάλυση Χαοτικών Χρονοσειρών

2.1 Βασικές έννοιες

Παρουσιάζουμε αρχικά ορισμένους όρους που χρησιμοποιούνται συχνά όταν αναφερόμαστε σε μη-γραμμικά δυναμικά συστήματα:

- Λέμε ότι ένα δυναμικό σύστημα καταλήγει σε έναν ελκυστή αν μετά από την όποια μεταβατική του κατάσταση καταλήγει σε μία περιορισμένη υποπεριοχή $\mathcal{A}$ του χώρου στον οποίο λαμβάνει χώρα η δυναμική του (χώρος φάσης). Η υποπεριοχή αυτή του χώρου φάσης αποτελεί το σταθερό σημείο της δυναμικής του συστήματος, δηλαδή έχουμε $\mathcal{A} = f(\mathcal{A})$ όπου f η συνάρτηση που περιγράφει τη δυναμική του συστήματος.

- 'Ενα χαοτικό δυναμικό σύστημα χαρακτηρίζεται (μεταξύ άλλων) από εγγενή αδυναμία πρόβλεψης της εξέλιξής του πέρα από ένα χρονικό σημείο. Η αδυναμία αυτή οφείλεται στην εξαιρετική του ευαισθησία στις αρχικές συνθήκες, η οποία εκδηλώνεται με τη μεγέθυνση με την πάροδο του χρόνου και των μικρότερων διαφορών σε τέτοιο βαθμό, ώστε δύο τροχιές που ξεκινάνε από σημεία που απέχουν ελάχιστα μεταξύ τους να καταλήγουν να μην έχουν καμία συσχέτιση.

- Δύο πολύ σημαντικές ιδιότητες των χαοτικών δυναμικών συστημάτων είναι η κλασματική διάσταση του ελκυστή τους και οι εκθέτες Lyapunov. Για λογούς συνοχής και συνέχειας του κειμένου θα παρουσιαστούν σε ξεχωριστή παράγραφο(2.4) η οποία βασίζεται σε αποτελέσματα των επόμενων παραγράφων.

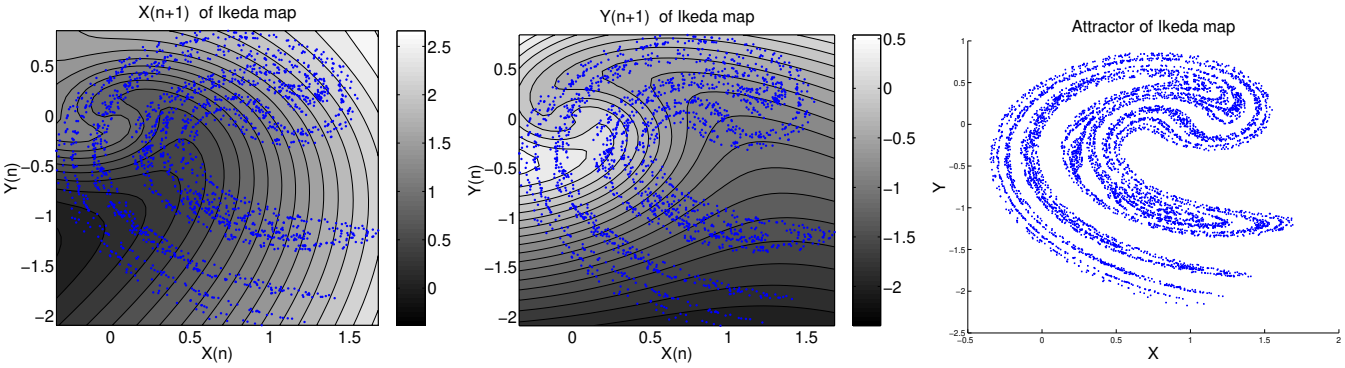
Δύο πολύ γνωστά χαοτικά δυναμικά συστήματα τα οποία εξετάστηκαν είναι:

- Η απεικόνηση Ikeda

$$Z(n+1) = 1+0.9Z(n) \cdot e^{\left\{0.4i - \frac{6 \cdot i}{(1+|Z(n)|)^2}\right\}} \tag{2.1}$$

όπου Z είναι ένα μιγαδικό διάνυσμα. Στο σχήμα 2.1 φαίνεται η δυναμική του συστήματος Ikeda καθώς και ο ελκυστής στον οποίο καταλήγει το σύστημα Ikeda. Συγκεκριμένα, για τις δύο πρώτες εικόνες έχουμε $X(n) = \mathrm{Re}(Z(n)), \quad Y(n) = \mathrm{Im}(Z(n))$ και η τιμή των συνάρτησεων

$$X(n+1) = F_1(X(n), Y(n)) \qquad Y(n+1) = F_2(X(n), Y(n))$$

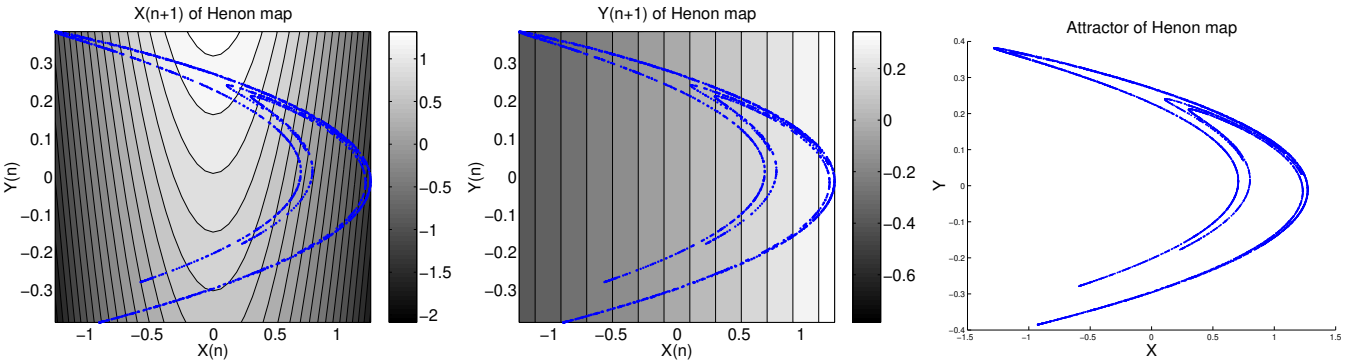


2.1: Δυναμική συστήματος Ikeda

έχει κωδικοποίηθεί στην ένταση του γκρι της εικόνας (grayscale) ενώ έχουν σχεδιαστεί και οι ισουψείς των συνάρτησεων F_1, F_2.

- Η απεικόνιση Henon

$$\begin{aligned} X(n+1) &= 1 - 1.4 \cdot X(n)^2 + Y(n) \\ Y(n+1) &= 0.3 \cdot X(n) \end{aligned} \tag{2.2}$$



2.2: Δυναμική συστήματος Henon

2.2 Ανακατασκευή ελκυστή χαοτικού συστήματος

Κεντρικής σημασίας στη μη γραμμική ανάλυση χρονοσειρών είναι η ιδέα ότι ο χώρος στον οποίο πραγματοποιείται η εξέλιξη του συστήματος από το οποίο προέρχεται η παρατηρούμενη χρονοσειρά δεν είναι αυτός ο μονοδιάστατος χώρος των παρατηρήσεων αλλά ένας χώρος τόσων διαστάσεων (έστω d) όσες είναι και οι 'εσωτερικές' μεταβλητές του συστήματος που καθορίζουν τη συμπεριφορά του (χώρος φάσης). Από το χώρο αυτό εμείς βλέπουμε μόνο μία προβολή του σε μία διάσταση που μπορεί να είναι μία οποιαδήποτε συνάρτηση $\mathcal{F} : \mathbf{R}^d \to \mathbf{R}$ των μεταβλητών αυτών (θεωρούμε πάντοτε μονοδιάστατες χρονοσειρές). Το γεγονός αυτό εξηγεί την εξαιρετικά περίπλοκη συμπεριφορά σε μία διάσταση συστημάτων τα οποία διέπονται από πολύ απλές εξισώσεις σε

δύο ή περισσότερες διαστάσεις. Στα πλαίσια της μη γραμμικής ανάλυσης χρονοσειρών επιχειρούμε να ανακατασκευάσουμε το χώρο φάσης του συστήματος.

Λέγοντας ότι με έναν μετασχηματισμό $\mathbf{R} \to \mathbf{R}^{D_e}$ ανακατασκεύάζουμε το χώρο φάσης του συστήματος εννοούμε ότι διατηρείται ο ντετερμινισμός του συστήματος στον νέο χώρο φάσης που παίρνουμε, καθώς και οι εφαπτόμενες σε κάθε σημείο, με την έννοια ότι δεν έχουμε διασταυρώσεις μεταξύ εφαπτόμενων διευθύνσεων σε σημεία του ανακατασκευασμένου χώρου φάσης που ήταν παράλληλες στον αρχικό χώρο φάσης. Ένας τέτοιος μετασχηματισμός ονομάζεται *εμβύθιση (embedding)* . Ικανή συνθήκη για την ανακατασκευή του ελκυστή είναι $D_e > 2 \cdot D_A$ όπου D_A η κλασματική (fractal) διάσταση του ελκυστή στον οποίον έχει καταλήξει το σύστημα. [1]

Το θεώρημα εμβύθισης του Takens μας εγγυάται ότι η ανακατασκευή του ελκυστή είναι εφικτή: Ας θεωρήσουμε ένα δυναμικό σύστημα του οποίου η εξέλιξη καθορίζεται από έναν αριθμό d μεταβλητών,οι οποίες συνιστούν το διάνυσμα κατάστασης

$$Z = [z_1, z_2, \ldots, z_d] \tag{2.3}$$

και έστω

$$Z(n+1) = f(Z(n)) \tag{2.4}$$

η σχέση που καθορίζει την εξέλιξή του. Αν η χρονοσειρά $x(n)$ που εμείς παρατηρούμε είναι η μονοδιάστατη προβολή h μίας διανυσματικής συνάρτησης G των d αυτών μεταβλητών, δηλαδή

$$x(n) = h(G(Z(n))) \tag{2.5}$$

ο ελκυστής του συστήματος θα είναι 'διπλωμένος' στη μία διάσταση που παρατηρούμε, αφού δεν είναι δυνατό να προβληθεί από δύο ή περισσότερες διαστάσεις σε μία διατηρώντας τις (σχετικές) αποστάσεις μεταξύ των σημείων του στον αυθεντικό χώρο φάσης. Σύμφωνα όμως με το θεώρημα του Takens, υπό την προϋπόθεση ότι $h, G \in C^1$, μπορούμε να *ανακατασκευάσουμε* το χώρο φάσης του συστήματος, όπου ξεδιπλώνεται ο ελκυστής του συστήματος, χρησιμοποιώντας διανύσματα X που προέρχονται από αυτές τις παρατηρήσεις :

$$X(n) = [h(G(Z(n)), h(G^{T_1}(Z(n)), \ldots, h(G^{T_{D_e-1}}(Z(n))] \tag{2.6}$$

όπου τα νέα διανύσματα X[2] ορίζουν τώρα μία νέα τροχιά σε ένα D_e–διάστατο χώρο. Η διάσταση D_e ονομάζεται διάσταση εμβύθισης (embedding dimension) και δεν είναι υποχρεωτικά ίση με την αρχική διάσταση d στην οποία εξελίσσεται η δυναμική του συστήματος. Η επιλογή των συναρτήσεων h, G η οποία εφαρμόζεται συνήθως για να πάρουμε τα διανύσματα X είναι να θεωρήσουμε ότι η συνάρτηση h είναι η παρατηρούμενη μεταβλητή δηλαδή ότι $h(Z(n)) = x(n)$ ενώ για την G επιλέγουμε την απεικόνιση η οποία υψωμένη σε δύναμη τ_s σε ένα διάνυσμα $Z(n)$ αντιστοιχεί το $Z(n+\tau_s)$ [3], δηλαδή:

$$G^{\tau_s}(Z(n)) = Z(n+\tau_s) \tag{2.7}$$

Αν επιπλέον πάρουμε $T_n = nT \; n = 1, \ldots, D_e$ τότε καταλήγουμε στην εξής μορφή για το διάνυσμα X:

$$X(n) = [x(n), x(n+T), \ldots, x(n+T \cdot (D_e - 1))] \tag{2.8}$$

[1] για μία σύντομη ανάφορά στην έννοια της κλασματικής διάστασης βλ.παρακάτω ,ενώ για περαιτέρω βλ. π.χ. [5] [13]

[2] Η ονομασία X δεν είναι τυχαία, όπως φαίνεται παρακάτω

[3] τ_s είναι ο ρυθμός δειγματοληψίας του σήματος

το οποίο είναι πολύ εύκολο να το αποκτήσουμε από το παρατηρούμενο σήμα (π.χ. με ένα φίλτρο απομάστευσης). Με μία κατάλληλη επιλογή δηλαδή των συναρτήσεων f, G μπορούμε αντί της 'εξωτικής' μορφής της εξίσωσης (2.6) να πάρουμε μία πολύ απλούστερη έκφραση, όπως η (2.8)

Προτού προχωρήσουμε στην περιγραφή μεθόδων για τον προσδιορισμό των παραμέτρων T και D_e αξίζει να τονίσουμε τη διαφορετική προσέγγιση της χρήσης χρονικά καθυστερημένων δειγμάτων του σήματος στην γραμμική ανάλυση χρονοσειρών και στη μη γραμμική ανάλυση[4]: Στην πρώτη περίπτωση οι καθυστερήσεις χρησιμοποιούνται για να καθορίσουν την τιμή του επόμενου δείγματος μέσω μίας γραμμικής εξίσωσης, για την οποία αρκεί να βρούμε τους βέλτιστους συντελεστές. Στη δεύτερη περίπτωση τα καθυστερημένα δείγματα χρησιμεύουν για να ανακατασκευάσουμε το χώρο φάσης, και να 'ξεδιπλώσουμε' εκεί τον ελκυστή του συστήματος, οπότε και να αναλύσουμε την δυνάμική του συστήματος σε D_e διαστάσεις. Αυτό το οποίο επιτυγχάνουμε χρησιμοποιώντας καθυστερημένα δείγματα είναι εκμεταλλευόμενοι πληροφορία σχετικά με την ιστορία του συστήματος στην πρόβλεψη που κάνουμε, κατά κάποιο τρόπο να ενσωματώνουμε τη δυναμική του συστήματος στο σημείο $\mathbf{X}(n)$ -αφού το σύστημα έχει διανύσει προηγουμένως μία συγκεκριμένη τροχία στο χώρο φάσης του, μπορούμε να πούμε πού θα πάει αργότερα; αυτό το πρόβλημα καλείται να επιλύσει η μή-γραμμική ανάλυση.

Το θεώρημα του Takens το οποίο αναφέρθηκε προηγούμένως εξασφαλίζει απλά τη δυνατότητα ανακατασκευής του ελκυστή. Ενα πρακτικό πρόβλημα με μεγάλο ενδιαφέρον είναι ποιές είναι οι κατάλληλες τιμές για τις παραμέτρους T και D_e (και ποιο είναι το κριτήριο για την καταλληλότητά τους). Για το σκοπό αυτό χρησιμοποιούμε τις μεθόδους που αναφέρονται στο [1]:

2.2.1 Προσδιορισμός χρονικής καθυστέρησης T

Η χρονική καθυστέρηση T[5] πρέπει να έχει δύο αντικρουόμενα χαρακτηριστικά:

- Να είναι αρκετά μεγάλη ώστε το δείγμα $x(n)$ να μην είναι πολύ κοντά στο δείγμα $x(n+T)$. Αλλιώς δεν παίρνουμε αρκετή πληροφορία για την ιστορία του συστήματος, πράγμα που καθιστά ανακριβή την πρόβλεψή μας, αφού δεν είμαστε σε θέση να ενσωματώσουμε ικανοποιητικά την προηγούμενη τροχία του συστήματος στην πρόβλεψη χρησιμοποιώντας μόνο σημεία που έχουν προέλθει από τα τελευταία στιγμιότυπά της.

- Να είναι αρκετά μικρή ώστε τα δείγματα $x(n), x(n+T)$ να μην είναι άσχετα μεταξύ τους, όπως είναι αναμενόμενο ότι θα είναι για μεγάλο T για ένα χαοτικό συστημά. Διαφορετικά η πρόβλεψή μας βασίζεται πρακτικά αποκλειστικά στο $x(n)$ οπότε είναι καταδικασμένη να αποτύχει (δεν παίρνει υπ'οψιν καθόλου την ιστορία του συστήματος).

Για τον προσδιορισμό της κατάλληλης τιμής της T χρησιμοποιείται η έννοια της αμοιβαίας πληροφορίας (Mutual Information-MI), η οποία ποσοτικοποιεί το ποσό πληροφορίας MI

[4] Το ότι χρησιμοποιούνται τα επόμενα δείγματα του x ως συνιστώσες του X δεν πρέπει να μπερδεύει τον αναγνώστη: στην πράξη χρησιμοποιούνται (ισοδύναμα) τα προηγούμενα δείγματα για τον ορισμό του X και μπορούμε να πούμε ότι παίρνουμε αντί των $X(n)$ τα $\dot{X}(n - T \cdot (D_e - 1)) = [x(n), x(n-T), \ldots, x(n - T \cdot (D_e - 1))]$

[5] Για λόγους απλότητας, όπου γράφουμε T, με βάση την ορολογία της προηγούμενης παραγράφου εννοείται T_1

που παίρνουμε για μία μέτρηση a από μία άλλη μέτρηση b το οποίο ισούται (σε bits) με:

$$MI(a,b) = \log_2 \left[\frac{P(a,b)}{P(a)P(b)} \right] \tag{2.9}$$

Έτσι π.χ. μία μέτρηση b η οποία είναι άσχετη με την a μας δίνει μηδενική πληροφορία, καθώς $P(a,b) = P(a)P(b)$. Παίρνοντας τη μέση τιμή της παραπάνω τιμής για όλες τις παρατηρήσεις a, b έχουμε τη συνάρτηση που μας δίνει την μέση αμοιβαία πληροφορία (Average Mutual Information –AMI) μεταξύ των μετρήσεων a, b:

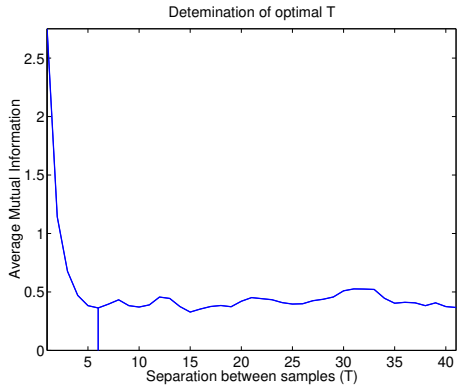
$$I_{AB} = \sum_{(a,b)} P(a,b) \log_2 \left[\frac{P(a,b)}{P(a)P(b)} \right] \tag{2.10}$$

δηλαδή για κάθε δυνατό ζεύγος παρατηρήσεων ζυγίζουμε την αμοιβαία πληροφορία που μας δίνει με την πιθανότητα να παρατηρηθεί (εφόσον έχουμε πάρει το μέσο της αμοιβαίας πληροφορίας).

Στην περίπτωση που εξετάζουμε θέλουμε να βρούμε κατά πόσο η μέτρηση $x(n)$ μας δίνει κάποια πληροφορία για την $x(n+T)$. Για το σκοπό αυτό βρίσκουμε τη μέση αμοιβαία πληροφορία για τα ζεύγη μετρήσεων $x(n), x(n+T)$ για διαφορετικές τιμές του T και σχηματίζουμε τη συνάρτηση $I(T)$:

$$I_{x(n),x(n+T)}(T) = \sum_{n=1}^{N-T} P(x(n), x(n+T)) \log_2 \left[\frac{P(x(n), x(n+T))}{P(x(n))P(x(n+T))} \right] \tag{2.11}$$

Υπολογίζοντας την $I(T)$ συναρτήσει της χρονικής καθυστέρησης T μπορούμε να βρούμε μία κατάλληλη τιμή της ώστε ναι μεν τα δείγματα $x(n), x(n+T)$ να μην είναι πολύ όμοια μεταξύ τους αλλά να μην είναι και τόσο απομακρυσμένα ώστε να μην σχετίζονται. Μία εμπειρική επιλογή για την T είναι να τεθεί ίση με το T στο οποίο παρατηρείται το πρώτο ελάχιστο της συνάρτησης $I(T)$. Στο σχήμα 2.2.1 φαίνεται η συνάρτηση μέσης αμοιβαίας πληροφορίας για ένα δείγμα του φωνήματος /aa/.



2.3: Συνάρτηση μέσης αμοιβαίας πληροφορίας για το φώνημα /aa/

Αν και δεν υπάρχει ξεκάθαρη επιστημονική απόδειξη για την αναγκαιότητα της χρήσης της παραπάνω μεθόδου για τον προσδιορισμό της παραμέτρου T, υπάρχει αρκετή πρακτική τεκμηρίωση της χρησιμότητάς της και της υπεροχής της έναντι της εναλλακτικής της, που είναι η χρήση της συνάρτησης αυτοσυσχέτισης –που είναι κατάλληλη για *γραμμικά* συστήματα, βλ. π.χ. [1] [3].

2.2.2 Προσδιορίσμός διάστασης εμβύθισης D_e

Όπως έχει ήδη αναφερθεί ο ελκυστής ενός d-διάστατου δυναμικού συστήματος βρίσκεται διπλωμένος από την προβολή του σε λιγότερες των d διαστάσεων (π.χ. αν έχουμε μόνο μία παρατήρηση), δηλαδή αλλοιώνονται οι τοπολογικές του ιδιότητες. Αυτό έχει ως συνέπεια ότι σημεία τα οποία βρίσκονται απομακρυσμένα στον αρχικό χώρο φάσης μπορεί να βρεθούν πολύ κοντά μεταξύ τους, πράγμα που μας δίνει παραπλανητικά στοιχεία για τη δυναμική του συστήματος. Για το λόγο αυτό αν θέλουμε να ανακατασκευάσουμε το χώρο φάσης του συστήματος κατά τέτοιο τρόπο ώστε να διατηρείται ο ντετερμινισμός του, οφείλουμε να χρησιμοποιήσουμε μία τέτοια διάσταση εμβύθισης D_e ώστε σημεία τα οποία βρίσκονται κοντά να είναι κοντά και στον αρχικό ελκυστή. Υπενθυμίζουμε ότι από το θεώρημα του Takens αρκεί να έχουμε $D_e > 2D_A$, όπου D_A η κλασματική διάσταση του ελκυστή, αλλά αυτό το κριτήριο δεν είναι πάντα κατάλληλο για δύο λόγους:

- Χρειάζεται να ξεδιπλώσουμε τον ελκυστή για να βρούμε την κλασματική του διάσταση.
- Ακόμα και αν γνωρίζαμε την D_A εξαρχής,η συνθήκη $D_e > 2D_A$ είναι ικανή και όχι αναγκαία. Συνήθως προσπαθούμε να βρούμε την ελάχιστη αναγκαία διάσταση εμβύθισης καθώς καθιστά το πρόβλημα της ανακατασκευής της δυναμικής του συστήματος απλούστερο.

Η μέθοδος που χρησιμοποιήσαμε είναι γνωστή ως μέθοδος των ψευδών πλησιέστερων γειτόνων (false nearest neighbors), βλ. π.χ. [1]. Σύμφωνα με τη μέθοδο αυτή το διάνυσμα κατάστασης X σχηματίζεται από τόσα επόμενα δείγματα,[6] όσα χρειάζονται για να μην υπάρχουν ψευδείς γείτονες. Για κάθε σημείο $X(n) = [x(n), x(n+T), \ldots, x(n+T\cdot M)]$, δηλαδή του M-διάστατου ελκυστή βρίσκουμε το πλησιέστερο προς αυτό στοιχείο του ελκυστή, έστω $X^{NN}(n)$, και εξετάζουμε σε ποιο βαθμό θα μεταβληθεί η απόστασή τους, αν τα θεωρήσουμε σε έναν ελκυστή με διάσταση εμβύθισης $D_e = M+1$. Ως μέτρο της απόστασης των δύο D-διάστατων σημείων $X(n), X^{NN}(n)$ χρησιμοποιούμε την ευκλείδια απόσταση,

$$d_{D+1} = \sqrt{\sum_{i=1}^{D}(X_i(n) - X_i^{NN}(n))^2} \tag{2.12}$$

όποτε η μεταβολή της απόστασής τους δηλαδή δίνεται από τη σχέση:

$$\begin{aligned} R^D(X(n), X^{NN}(n)) &= \sqrt{\frac{d_{D+1}^2(X(n), X^{NN}(n)) - d_D^2(X(n), X^{NN}(n))}{d_D^2((n), X^{NN}(n))}} = \\ &= \sqrt{\frac{\sum_{i=1}^{M+1}[X_i(n) - X_i^{NN}(n)]^2 - \sum_{i=1}^{M}[X_i(n) - X_i^{NN}(n)]^2}{\sum_{i=1}^{M}[X_i(n) - X_i^{NN}(n)]^2}} = \\ &= \frac{|X_{M+1}(n) - X_{M+1}^{NN}(n)|}{\sqrt{\sum_{i=1}^{M}[X_i(n) - X_i^{NN}(n)]^2}} \end{aligned} \tag{2.13}$$

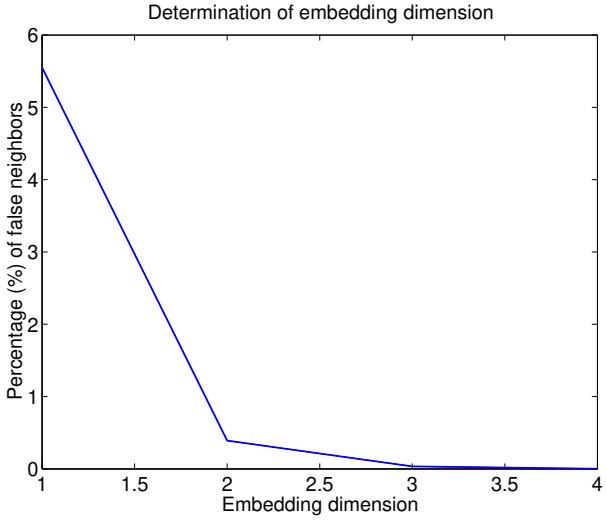
Ελέγχοντας αν $R^D > R_{thr}$, όπου $R_{thr} \sim 15 - 20$[7], συμπεραίνουμε αν η απομάκρυνσή τους ήταν τέτοια ώστε να θεωρηθούν ψευδείς γείτονες. Με βάση το παραπάνω κριτήριο,

[6] Η χρονική απόσταση μεταξύ των δειγμάτων που χρησιμοποιούνται για την ανακατασκευή του ελκυστή θεωρούμε ότι έχει υπολογιστεί σύμφωνα με τη μέθοδο της προηγούμενης παραγράφου

[7] Η επιλογή του R_{thr} είναι εμπειρική

λοιπόν, ξεκινάμε από $M = 1$ και αυξάνουμε τη διάσταση εμβύθισης του ελκυστή εως ότου το παραπάνω κριτήριο ικανοποιηθεί για όλα τα στοιχεία του ελκυστή. Πρέπει να σημειωθεί, επίσης, ότι παρουσία θορύβου είναι δύνατό ακόμα και στην ελάχιστη απαιτούμενη διάσταση εμβύθισης να υπάρχουν ορισμένα στοιχεία για τα οποία η ποσότητα R^D να ξεπερνάει το R_{thr}. Για το λόγο αυτό το κριτήριο το οποίο χρησιμοποιείται πρακτικά για την επιλογή της σωστής διάστασης εμβύθισης είναι ότι από τη διάσταση αυτή και πέρα δεν υπάρχει σημαντική διαφοροποίηση του ποσοστού των σημείων τα οποία δεν ικανοποιούν το αρχικό μας κριτήριο.

Παρακάτω φαίνεται το ποσοστό των σημείων της πρώτης συνιστώσας της απεικόνισης Ikeda που δεν ικανοποιούν τη σχέση $R^D < R_{thr}$ ως συνάρτηση της διάστασης εμβύθισης για διαφορετικές τιμές του R_{thr}.



2.4: Ποσοστό μη πραγματικών γειτόνων για το σύστημα Ikeda

2.3 Προσέγγιση της δυναμικής του συστήματος στο χώρο φάσης

Έχοντας καθορίσει τη διάσταση εμβύθισης του σήματος καθώς και την χρονική καθυστέρηση μεταξύ των δειγμάτων που χρησιμοποιούνται για την ανακατασκευή του χώρου φάσης του δυναμικού συστήματος, σκοπός μας είναι πλέον να προσεγγίσουμε τη συνάρτηση που απεικονίζει το διάνυσμα κατάστασης του συστήματος τη χρονική στιγμή n

$$X(n) = [x(n), x(n-T), x(n-2\cdot T), \ldots, x(k-D_e\cdot T)] \tag{2.14}$$

σε αυτό την στιγμή $n+1$,δηλαδή την f για την οποία έχουμε

$$X(n+1) = f(X(n)) \tag{2.15}$$

Μπορούμε, αν θέλουμε αποκλειστικά να κάνουμε πρόβλεψη για το επόμενο βήμα, αντί του διανύσματος να προβλέψουμε μόνο την επόμενη τιμη της πρώτης συνιστώσας του (X_1) και να θεωρήσουμε ότι οι υπόλοιπες συνιστώσες του προκύπτουν από ολισθήσεις προς τα δεξιά των ήδη παρατηρηθέντων διανυσμάτων. Αυτό το εκμεταλλευόμαστε στη σύγκριση των προβλεπτων μας με το LPC μοντέλο, καθώς εκτός από μικρό σφάλμα χρειαζόμαστε και μικρό προβλέπτη. Οπως θα φανεί, όμως, στην ανάλυση για τους εκθέτες Lyapunov, είναι χρήσιμο να έχουμε μία συνάρτηση που να απεικονίζει διάνυσμα σε διάνυσμα θα αναφερόμαστε στο εξής σε συναρτήσεις $R^{D_e} \to R^{D_e}$. Η πρόβλεψη για μία διάσταση είναι μία απλούστερη υποπερίπτωση, οπότε δεν θα αναφερθούμε πλέον σε αυτή.

Η μόνη πληροφορία που έχουμε για τη συνάρτηση f είναι ότι μπορεί να είναι μη γραμμική και στόχος μας είναι να προσεγγίσουμε τη συνάρτηση f με μία συνάρτηση F η οποία να βασίζεται στα ζεύγη παρατηρήσεων $< X(k), X(k+1) >$ που έχουμε. Η τακτική που ακολουθείται γενικότερα για την εκτίμηση της καταλληλότητας μίας συνάρτησης F για την προσέγγιση της f είναι αφού εκπαιδευτεί με ένα μέρος των δεδομένων να εξεταστεί η επίδοσή της με το υπόλοιπο μέρος των δεδομένων μας (cross-validation). Στα πλαίσια της επεξεργασίας φωνής, όμως, σκοπός μας δεν είναι να κατασκευάσουμε ένα μοντέλο το οποίο να μπορεί να ανταποκριθεί σε άλλα δεδομένα εκτός από αυτά τα οποία χρησιμοποιήθηκαν στην εκπαίδευσή του -πράγμα που εξετάζουμε με τη μέθοδο της cross validation, αλλά χρησιμοποιώντας όσα δεδομένα έχουμε να ελαχιστοποιήσουμε το σφάλμα πρόβλεψης. Συνεπώς το κριτήριο της καταλληλότητας ενός μοντέλου που θα χρησιμοποιούμε είναι το μέσο τετραγωνικό σφάλμα (Normalised Mean Squared Error -NMSE) της πρόβλεψης πάνω σέ όλα τα δεδομένα. Η πρόβλεψη θεωρούμε ότι γίνεται για την 'πίο πρόσφατη τιμή', δηλαδή εξετάζουμε μόνο τη διαφορά μεταξύ του $X_1(n)$ και του $F_1(X(n-1))$ οπότε το NMSE ορίζεται ως

$$\mathrm{NMSE} = \frac{\sqrt{\sum_{i=1}^{N} X_1(i) - F_1(X(i)))^2}}{\sqrt{\sum_{i=1}^{N}(X_1(i) - \overline{X}_1)^2}} \quad \text{όπου} \quad \overline{X}_1 = \sum_{i=1}^{N} X_1(i)$$

Βέβαια, όπως θα φανεί στο επόμενο μέρος αυτής της αναφοράς, υπάρχουν και άλλα κριτήρια για την καταλληλότητα μίας συνάρτησης F για να προσεγγίσει την συνάρτηση f, όπως το κατα πόσο μπορούμε να υπολογίσουμε σωστά ορισμένα χαρακτηριστικά του συστήματος, όπως οι εκθέτες Lyapunov. Πραγματικά, όπως θα δούμε, μοντέλα με σχετικά μέτριες επιδόσεις στην πρόβλεψη του σήματος με το οποίο εκπαιδεύτηκαν, είναι πολύ πιο

εύρωστα σε σχέση με άλλα που δίνουν πολύ μικρό σφάλμα. Πρόκειται για το πρόβλημα της υπερεκπαίδευσης, όπως είναι γνωστό στο χώρο της μάθησης μηχανών, που οφείλεται, στην περίπτωσή μας, στο γεγονός ότι είναι πιθανό σε μεγάλες διαστάσεις ο αριθμός των προσδιοριστέων παραμέτρων να είναι της ίδιας τάξης (ίσως και μεγαλύτερος) σπό τον αριθμό των δεδομένων που έχουμε. Το αποτέλεσμα είναι να έχουμε έναν 'προβλέπτη' ο οποίος έχει απλά μάθει τα δεδομένα (και όχι τις κανονικότητες που υπάρχουν σε αυτά).

Από τις μεθόδους που έχουν προταθεί για την προσέγγιση της f υλοποιήθηκαν και δοκιμάστηκαν οι παρακάτω

- Τοπικά πολυώνυμα (local polynomials)
- Καθολικά πολυώνυμα (global polynomials)
- Τοπικά πολυώνυμα σε συστάδες σημείων του ελκυστή
- Δίκτυα RBF (Radial Basis Functions)
- Δίκτυα WCP (Weighted Constant Predictors)
- Δίκτυα WLP (Weighted Linear Predictors)
- Δίκτυα SVM (Support Vector Machine)

Στη συνέχεια θα εξετάσουμε τις επιδόσεις των αλγορίθμων στην πρόβλεψη του συστήματος Ikeda το οποίο χαρακτηρίζεται από μεγαλύτερη πολυπλοκότητα από αυτή του συστήματος Henon. Εκτός αυτού, η δυναμική του συστήματος Henon μπορεί να εκφραστεί μόνο με πολυώνυμα, όποτε δίνεται ένα παραπλανητικό 'προβάδισμα' στα πολυώνυμα.

Εδώ χρειάζεται να τονιστεί ότι ενώ στο κύριο μέρος της βιβλιογραφίας χρησιμοποιείται μεγάλος αριθμός από σημεία (της τάξης του 10^4) από τα διάφορα test σήματα για να εξεταστεί η επίδοση των αλγορίθμων, εμείς χρησιμοποιήσαμε χρονοσειρές με περίπου 600 σημεία, καθώς τέτοια είναι και η τάξη μεγέθους των χρονοσειρών που αντιστοιχούν σε διάφορα φωνήματα δειγματοληπτημένα σε 16 KHz. Για το λόγο αυτό μοντέλα που απαιτούν μεγάλο αριθμό δεδομένων, όπως τα τοπικά πολυώνυμα, αν και θεωρούνται standard κατά κάποιο τρόπο, δεν δίνουν και τα καλύτερα αποτελέσματα. Προσπαθήσαμε, δηλαδή, να δοκιμάσουμε τους αλγορίθμους σε καταστάσεις παρόμοιες με αυτές που θα κληθούν να λειτουργήσουν. Ο ενδιαφερόμενος αναγνώστης για την επίδοση των αλγορίθμων με καθαρά, μεγάλα σήματα παραπέμπεται στη βιβλιογραφία.

2.3.1 Προσέγγιση της δυναμικής του συστήματος με local polynomials

Η κεντρική ιδέα στην οποία βασίζονται τα τοπικά μοντέλα είναι ότι ξέροντας πού απεικονίζονται τα σημεία που βρίσκονται σε μία γειτονιά στον ελκυστή μπορούμε περίπου να προβλέψουμε ποια θα είναι η επόμενη τιμή ενός σημείου $X(k)$ που βρίσκεται στη γειτονιά αυτή (παίρνοντας ως δεδομένη τη συνέχεια της συνάρτησης f).

Υπάρχουν δύο εναλλακτικές: η πρώτη και απλούστερη είναι να εντοπίσουμε το γνωστό εκ των προτέρων σημείο $X(n)$ που είναι το κοντινότερο στο σημείο $X(k)$ και να θεωρήσουμε ότι η εξέλιξη του $X(k)$ θα είναι το $X(n+1)$. Αυτή η μέθοδος είναι γνωστή ως μέθοδος των αναλογιών και έχει πολύ περιορισμένες δυνατότητες για πρόβλεψη πέρα από ένα αριθμό βημάτων. Η μέθοδος που χρησιμοποιείται ευρύτερα είναι να χρησιμοποιηθούν τα σημεία

που βρίσκονται στη γειτονιά του $X(k)$ για να βρεθεί οι πρώτοι όροι του αναπτύγματος Taylor[8] της συνάρτησης f στην περιοχή του $X(k)$. Το σκεπτικό στο οποίο βασίζεται η μέθοδος είναι ότι το σύστημα επισκέπτεται την περιοχή κάθε σημείου το οποίο ανήκει στον ελκυστή αρκετές φορές κατά την εξέλιξή του (είναι μία από τις ιδιότητες των χαοτικών συστημάτων), οπότε παρατηρώντας πώς απεικονίζει κάθε σημείο της περιοχής αυτής στο επόμενό του μπορούμε να υπολογίσουμε περίπου πού θα απεικονίσει το σημείο $X(k)$.

Η γειτονιά μπορεί να οριστεί είτε από όλα τα σημεία που βρίσκονται μέσα σε μία σφαίρα που έχει κέντρο το σημείο $X(k)$ (η ακτίνα της σφαίρας μπορεί να οριστεί ως ένα κλάσμα της ακτίνας του ελκυστή) ή από τα N_B πλησιέστερα σημεία του $X(k)$. Η ακτίνα και ο αριθμός N_B είναι προσδιοριστέες ποσότητες και στις δύο περιπτώσεις. Προτιμάται συνήθως να χρησιμοποιείται ένας συγκεκριμένος αριθμός από στοιχεία, καθώς η χρήση των σημείων που βρίσκονται εντός μίας σφαίρας συγκεκριμένης ακτίνας μπορεί σε αραιές περιοχές του ελκυστή να δώσει πολύ λίγους γείτονες οπότε δεν έχουμε αρκετά δεδομένα ώστε να προσεγγίσουμε ικανοποιητικά την συνάρτησή f που περιγράφει τη δυναμική του συστήματος. Σκοπός μας ,λοιπόν είναι να βρούμε τους συντελεστές $C(k)$ οι οποίοι δίνουν την καλύτερη προσέγγιση F:

$$F(x) = \sum_{k=1}^{M} C(k) \cdot p_k(x) \tag{2.16}$$

Το διάνυσμα $C(k)$ εκφράζει τον τρόπο με τον οποίο η ποσότητα $p_k(X)$ επηρεάζει την έξοδο, ενώ p_k είναι είναι ένα μονώνυμο το οποίο προκύπτει υψώνοντας την κάθε συνιστώσα του διανύσματος X σε μία δύναμη και εν συνεχεία πολλαπλασιάζοντας τις προκύπτουσες συνιστώσες. Ο αριθμός k προκύπτει από την ανάγκη για απλότητα στους τύπους μας και αποτελεί μία κωδικοποίηση του διανύσματος των εκθετών του k-οστού μονωνύμου: η βάση των πολυμεταβλητών πολυωνύμων $1, X_1^1 X_2^0 \cdots X_{D_e}^0, x_1^0 X_2^1, \ldots X_{D_e}^0, \ldots$ μπορεί να εκφραστεί απλούστερα κωδικοποιώντας με μία αντιστοιχία 1-1 το διάνυσμα των εκθετών σε έναν ακέραιο:

$$\begin{aligned} X_1^{e_1} \cdot X_2^{e_2} \cdots X_{D_e}^{e_{D_e}} &= p_k(X_1, X_2 \ldots, X_{D_e}) \leftrightarrow \\ [e_1, e_2, \ldots, e_{D_e}] &= decode(k) \end{aligned} \tag{2.17}$$

Η συνάρτηση $decode$ μπορεί να είναι π.χ. η συνάρτηση αποκωδικοποίησης του Cantor [27] αν για να πάρουμε το k φαρμόζουμε την αντίστοιχη $encode$ στο διάνυσμα των εκθετών και απλουστεύει σημαντικά τις εξισώσεις μας, όπως θα φανεί ιδιαίτερα στην επόμενη παράγραφο.Η (2.17) μας δίνει:

$$\begin{aligned} p_0(X) &= 1 \cdot 1 \cdots 1 = 1 \\ p_1(X) &= x_1 \cdot 1 \cdots 1 = x_1 \\ &\vdots \\ p_{d+2}(X) &= x_1^2 \cdot 1 \cdots 1 = x_1^2 \\ p_{d+3}(X) &= x_1 \cdot x_2 \cdot 1 \cdots 1 = x_1 \cdot x_2 \\ &\vdots \end{aligned} \tag{2.18}$$

Για τον υπολογισμό των συντελεστών $C(k)$ ζητάμε να ελαχιστοποιείται το μέσο τετραγωνικό σφάλμα μεταξύ πρόβλεψης και παρατηρούμενης τιμής για τα N_B σημεία της γειτονιάς.

[8] Το ανάπτυγμα Taylor είναι, προφανώς για συνάρτηση D_e μεταβλητών

Το πρόβλημα αυτό λύνεται με τη μέθοδο των ελαχίστων τετραγώνων, όπου πρέπει να βρούμε τους συντελέστές εκείνους για τους οποίους το μέσο τετραγωνικό σφάλμα [9] μεταξύ της πρόβλεψης F με βάση τις τιμές των $p_1(X(k)), p_2(X(k)), \ldots, p_M(X(k))$ και του $X(k+1)$ θα είναι ελάχιστο. Η λύση του προβλήματος ανάγεται στην αντιστροφή ενός πίνακα $\mathbf{M}$ διαστάσεων $M \times M$. Το m_{ij} στοιχείο αυτού του πίνακα θα είναι:

$$m_{ij} = \sum_{n=1}^{N_B} p_i(X(n)) p_j(X(n)) \text{[10]} \tag{2.19}$$

Αν θεωρήσουμε ότι

$$\mathcal{F}(n) = [p_1(n), p_2(n), \cdots, p_M(n)]^T \tag{2.20}$$

μπορούμε να γράψουμε τον πίνακα ως

$$\mathbf{M} = \sum_{n=1}^{N_B} \mathcal{F}(n) \cdot \mathcal{F}^T(n) \tag{2.21}$$

Η λύση που προκύπτει από την απαίτηση να ελαχιστοποιείται το μέσο τετραγωνικό σφάλμα πρόβλεψης είναι

$$\mathbf{C} = \mathbf{M}^{-1} \sum_{i=1}^{N_B} \mathcal{F}(X(i)) X(i+1) \tag{2.22}$$

όπου ο πίνακας $\mathbf{C}$ έχει ως στύλες τα D_e M–διάστατα διανύσματα $C(1), C(2), \ldots, C(D_e)$[11]Αν ο πίνακας $\mathbf{M}$ δεν είναι αντιστρέψιμος τη βέλτιστη λύση σύμφωνα με το MSE κριτήριο την παίρνουμε από τον ψευδοαντίστροφό του. Αυτή η διαδικασία εκετελείται για όλα τα σημεία, οπότε έχουμε για κάθε σημείο του ελκυστή έναν πίνακα από σταθερές που εκφράζουν την προσέγγιση της δυναμικής του συστήματος στην γειτονιά του. 'Ετσι, όταν θέλουμε να προσεγγίσουμε την επόμενη κατάσταση του συστήματος από την τρέχουσα, αναζητούμε το πλησιέστερο σημείο του ελκυστή και με βάση τις σταθερές που έχουν υπολογιστεί για τη δυναμική στην περιοχή του βρίσκουμε την επόμενη τιμή.

Μπορούμε, στην περίπτωση που οι συναρτήσεις p_i δεν χρησιμοποιούν δυνάμεις μεγαλύτερες της μονάδας, να πούμε ότι βρίσκουμε ένα πολυδιάστατο τοπικό AR μοντέλο για να ταιριάξει στα δεδομένα της γειτονιάς κάθε σημείου (οι εξισώσεις που μας δίνουν το διάνυσμα σταθερών $\mathbf{C}$ είναι οι ίδιες που μας δίνουν τους συντελεστές ενός AR μοντέλου). Φαίνεται λοιπόν η ευελιξία που μας παρέχεται από την μέθοδο αυτή σε αντίθεση με το AR μοντέλο το οποίο υποχρεώνει όλα τα στοιχεία του ελκυστή να συμμετάσχουν στον υπολογισμό των διανυσμάτων C.

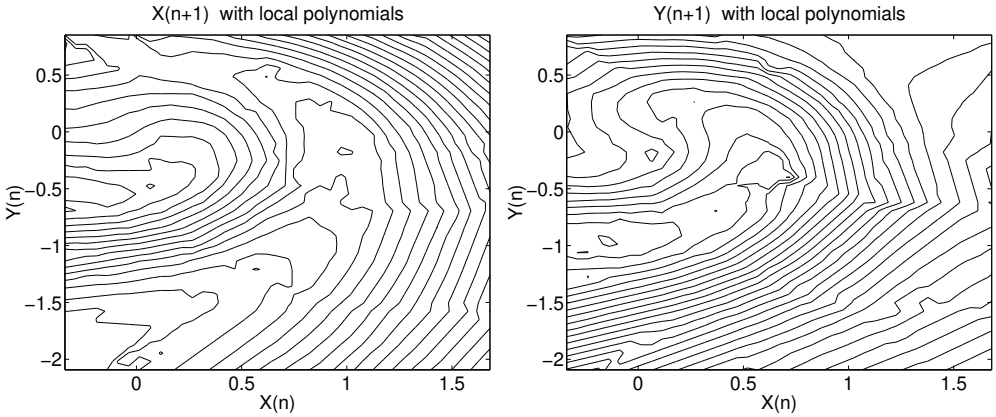
Σημειώνεται ότι στις περισσότερες περιπτώσεις τα γραμμικά τοπικά πολυώνυμα έχουν αποτελέσματα παραπλήσια με τα μοντέλα υψηλότερου βαθμού, οπότε και προτιμώνται, καθώς τα τελευταία απαιτούν έναν μεγαλύτερο αριθμό από γείτονες. Αν και η χρήση τοπικών πολυωνύμων είναι από τις πλέον καθιερωμένες μέθοδος και μπορεί να δώσει προβλέψεις πολύ μεγάλης ακρίβειας, μπορεί να θεωρηθεί μάλλον ακατάλληλη για τους σκοπούς της πρόβλεψης και ανάλυσης φωνής, καθώς απαιτεί πολύ μεγάλο χρόνο εκπαίδευσης, μεγάλο αριθμό δειγμάτων και έναν πολύ μεγάλο αριθμό από προσδιοριστέες παραμέτρους.

[9] Η μέση τιμή θεωρείται πάνω στα σημεία της γειτονιάς

[10] $X(n)$ είναι ο n–οστός πλησιέστερος γείτονας του σημείου $X(k)$

[11] Ο πίνακας $\mathbf{M}$ είναι $M \times M$ οπότε και η ποσότητα στο δεξί μέρος της εξίσωσης είναι $(M \times M) \times (M \times D_e) = M \times D_e$.

Παρακάτω βλέπουμε πώς ανακατασκευάζεται η δυναμική της απεικόνισης Ikeda με τοπικά πολυώνυμα. Παρατηρούμε ότι η ανακατασκευή είναι πολύ ακριβής αλλά σε ορισμένες περιοχές υπάρχει εμφανής ασυνέχεια: αυτό παρατηρείται στο σύνορο μεταξύ δύο περιοχών στις οποίες αντιστοιχούν διαφορετικά μοντέλα και προκαλεί προβλήματα αν χρειαζόμαστε μία συνεχώς διαφορίσιμη συνάρτηση που να εκφράζει τη δυναμική του συστήματος.



2.5: Ανακατασκευή δυναμικής της απεικόνισης Ikeda με local polynomials

2.3.2 Προσέγγιση της δυναμικής του συστήματος με global polynomials

Στην περίπτωση αυτή η συνάρτηση F εκφράζεται ως ένα ανάπτυγμα σε μία βάση Π ορθογώνιων πολυμεταβλητών πολυώνυμων ϕ. Η ορθογωνιότητα των πολυωνύμων βασίζεται στην πράξη του εσωτερικού γινομένου η οποία ορίζεται ως εξής:

$$< \phi_i \cdot \phi_j >= \int_{x_1}\int_{x_2}\cdots\int_{x_{D_e}} \rho(X)\phi_i(X)\phi_j(X)\,\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_{D_e} \tag{2.23}$$

Η ποσότητα $\rho(X)$ που υπάρχει στη σχέση (2.23) είναι η φυσική πυκνότητα του ελκυστή και ισούται με

$$\rho(X) = 1/N \sum_{j=1}^{N} \delta(X - Y(n)) \tag{2.24}$$

Στην παραπάνω σχέση η συνάρτηση δ είναι η συνάρτηση Dirac, $Y(n)$ είναι το n-οστο διάνυσμα δεδομένων ενώ N είναι το πλήθος των δεδομένων μας. Για να είναι ορθογώνια τα πολυώνυμα απαιτούμε να ισχύει:

$$< \phi_i, \phi_j >= \delta_{i,j} \tag{2.25}$$

Καθώς τα δεδομένα μας είναι πολυδιάστατα, οι δείκτες των πολυωνύμων στις παραπάνω σχέσεις είναι κωδικοποιήσεις των διανυσμάτων των οποίων κάθε συνιστώσα δείχνει σε ποια δύναμη υψώνεται η αντίστοιχη μεταβλητή (όπως στην προηγούμενη παράγραφο). Έχοντας βρει μία ορθογώνια βάση πολυωνύμων μπορούμε να προσεγγίσουμε τη συνάρτηση f στη βάση αυτή ικανοποιητικά χρησιμοποιώντας έναν σχετικά μικρό αριθμό από

παραμέτρους:

$$f(X) = \sum_{i=1}^{\infty} C(i)\phi_i(X) \rightarrow$$
$$F(X) = \sum_{i=1}^{M} C(i)\phi_i(X) \simeq f(X) \tag{2.26}$$

Τα διανύσματα $C(i)$,$i = 1 \ldots M$ είναι οι συντελεστές Fourier της συνάρτησης f ως προς την ορθογώνια βάση Π και ορίζονται ως εξής:

$$\begin{aligned} C(i) &= < f \cdot \phi_i >= \int_{x_1}\int_{x_2}\cdots\int_{x_{D_e}} f(X)\phi_i(X)\rho(X)\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_{D_e} \overset{(2.24)}{=} \\ &= \frac{1}{N}\sum_{k=1}^{N} f(X(k))\phi_i(X(k)) \overset{f(X(k))=X(k+1)}{=} \\ &= \frac{1}{N}\sum_{k=1}^{N} X(k+1)\phi_i(X(k)) \end{aligned} \tag{2.27}$$

Η σχέση (2.27) προκύπτει πολλαπλασιάζοντας και τις δύο πλευρές του πρώτου μέρους της σχέσης (2.26) με το $\phi_i(x)$ και παίρνοντας το ολοκλήρωμα τους κατα μήκος του ελκυστή. Χρησιμοποιώντας την ορθογωνιότητα των πολυωνύμων από τα M πολυώνυμα που υπάρχουν στη σχέση (2.26) μένει μόνο το i-οστό.

Ο λόγος για τον οποίο χρησιμοποιείται η οικογένεια ορθογώνιων πολυώνυμων Π φαίνεται από την απλότητα με την οποία αποκτούμε τους συντελεστές του αναπτύγματος της f σε αυτή. Αν αντί για τη βάση Π χρησιμοποιούσαμε απλά την βάση $< 1, p_1(X), p_2(X), \ldots >$ [12] το πρόβλημα θα αποκτούσε μεγάλη πολυπλοκότητα αφου έχουμε να κάνουμε με *πολυμεταβλητά* πολυώνυμα.

Δεν έχουμε ακόμη αναφερθεί στον τρόπο με τον οποίο βρίσκουμε τους συντελεστές από τα ορθογώνια πολυώνυμα: το πιο γνωστό άρθρο σχετικά είναι από τους [15], καθώς όμως οι τύποι στους οποίους καταλήγουν είναι τόσο πολύπλοκοι που δύσκολα μπορούν να κατανοηθούν (για διδιάστατα δεδομένα!) προτιμήθηκε η συστηματικότερη και απλούστερη μέθοδος που προτείνεται στο [18]. Τα αποτελέσματα στα οποία καταλήγουν και οι δύο μέθοδοι είναι τα ίδια, όπως αναφέρεται στο τελευταίο άρθρο.

Η οικογένεια πολυμεταβλητών πολυωνύμων ϕ παρουσιάζεται ως εξής:

$$\phi_k(X) = \sum_{n=1}^{k} A_n^k p_n(X) \tag{2.28}$$

Θα είναι δηλαδή

$$\begin{aligned} \phi_1 &= A_1^1 \\ \phi_2 &= A_1^2 + A_2^2 x_1, \\ &\vdots \end{aligned} \tag{2.29}$$

Κάθε μονώνυμο της βάσης των πολυωνύμων $p_k(X)$ αποδεικνύεται ότι μπορεί να εκφραστεί ως άθροισμα ορισμένων πολυωνύμων της παραπάνω οικογένειας, δηλαδή έχουμε

$$p_k(X) = \sum_{i=1}^{k} B_i^k \phi_k(X) \tag{2.30}$$

[12] Οι συναρτήσεις p_i έχουν οριστεί στην προηγούμενη παράγραφο

Η παραπάνω οικογένεια είναι δηλαδή μία πλήρης βάση του χώρου των πολυμεταβλητών πολυωνυμικών συναρτήσεων. Για την ορθογωνιοποίηση θεωρούμε ότι έχουν ήδη βρεθεί οι συντελεστές των προηγούμενων N πολυωνύμων, και ζητάμε το $N+1$–οστο πολυώνυμο να είναι κάθετο στα προηγούμενα. Το $\phi_{N+1}(X)$, όμως, είναι δυνατόν να εκφραστεί ως:

$$\phi_{N+1}(X) = \sum_{i=1}^{N+1} A_i^{N+1} p_i(X) \tag{2.31}$$

Η απόδειξη μας επιτρέπει να θέσουμε $A_{N+1}^{N+1} = A_N^N$ οπότε έχουμε

$$\phi_{N+1}(X) = \sum_{i=1}^{N} A_i^{N+1} p_i(X) + A_N^N p_{N+1}(X) \tag{2.32}$$

και απομένει να ορίσουμε N παράγοντες έχοντας τις N σχέσεις ορθογωνιότητας

$$< \phi_{N+1} \cdot \phi_i >= 0, \qquad i = 1, \ldots, N \tag{2.33}$$

Εδώ απλοποιούνται οι εξισώσεις που προκύπτουν για τους συντελεστές χάρις στη σχέση (2.30) με την οποία ο κάθε παράγοντας της μορφής $p_p(X)$ εκφραζεται συναρτήσει των πολυώνύμων της οικογένειας ϕ. Με βάση τη σχέση αυτή ,το ϕ_{N+1} μπορεί να γραφεί και ως:

$$\phi_{N+1}(X) = \sum_{i=1}^{N} A_i^{N+1} \sum_{j=1}^{i} B_j^i \phi_j(X) + A_N^N \sum_{j=1}^{N+1} \phi_j(X) \tag{2.34}$$

Από τη σχέση αυτή απαιτώντας να ισχύει ότι $< \phi_{N+1} \cdot \phi_i >= 0 \qquad i = 1, \ldots, N$ παίρνουμε τη σχέση που ορίζει τον N–οστό συντελεστή

$$A_N^{N+1} = \frac{-A_N^N < p_{N+1} \cdot \phi_N >}{B_N^N < \phi_N \cdot \phi_N >} \tag{2.35}$$

Ομοίως, από τις υπόλοιπες συνθήκες ορθογωνιότητας παίρνουμε αναδρομικά[13]:

$$A_{N-k}^N = \frac{1}{B_{N-k}^{N-k}} \frac{-A_N^N < p_{N+1} \cdot \phi_{N-k} >}{< \phi_{N-k} \cdot \phi_{N-k} >} - \sum_{i=N-k+1}^{N} (A_i^{N+1} B_{N-k}^i) \tag{2.36}$$

για τους υπόλοιπους συντελεστές. Πρέπει επίσης να προσδιοριστούν οι συντελεστές B_i^{N+1} καθώς χρησιμοποιούνται για να προσδιοριστούν τα πολυώνυμα μεγαλύτερης τάξης. Θέτοντας $B_k^k = 1/A_k^k$ παίρνουμε για τους υπόλοιπους συντελεστές τις σχέσεις:

$$B_r^k = -B_k^k \sum_{i=r}^{k-1} A_i^k B_r^i \qquad r = 1, \ldots, k-1 \tag{2.37}$$

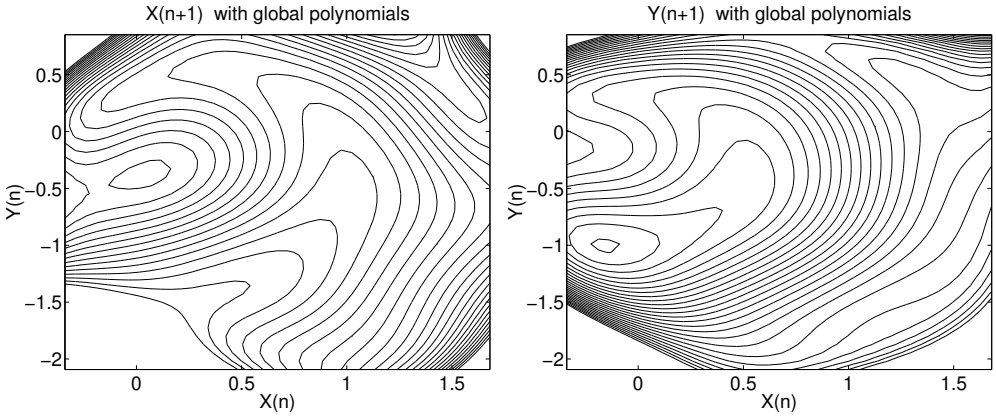
Χρησιμοποιώντας,δηλαδή την οικογένεια των πολυωνύμων ϕ_k μπορέσαμε να απλοποιήσουμε τις σχέσεις που εκφράζουν τους συντελεστές των ορθοκανονικών πολυωνύμων, εκφράζοντας το $N+1$–στο πολυώνυμο συναρτήσει των N προηγούμενων πολυωνύμων (οπότε και να απλοποιήσουμε σημαντικά την έκφραση που απαιτεί το γινόμενό τους να είναι μηδενικό).

Τα κύρια πλεονεκτήματα της μεθόδου αυτής είναι ότι είναι ανθεκτική στην παρουσία θορύβου, δεν λαμβάνει υπ'οψιν τις αποστάσεις μεταξύ των σημείων (όπου ο θόρυβος

[13]Οι αποδείξεις για τους τύπους που ακολουθούν παραλείπονται για λόγους συντομίας του κειμένου

μπορεί να αλλοιώσει σημαντικά τα αποτελέσματα), απαιτεί πολύ μικρό αριθμό δεδομένων και παραμέτρων, βρίσκει σε ελάχιστο χρόνο τις παραμέτρους, ενώ δίνει και συνεχή λύση. Τα κύριο μειονέκτημά της είναι ότι αν έχουμε ένα μεγάλο αριθμό από καθαρά δεδομένα μπορεί να χάνουμε ένα πολύτιμο μέρος από πληροφορία, περιορίζοντας την έκφραση της δυναμικής των δεδομένων σε πολυωνυμικές συναρτήσεις. Αντιθέτως, στην περίπτωση αυτή τα τοπικά πολυώνυμα θα μπορούσαν να δώσουν πολύ καλύτερα αποτελέσματα.

Παρακάτω βλέπουμε την ανακατασκευή της δυναμικής του συστημάτος Ikeda καθολικά πολυώνυμα. Παρατηρούμε ότι σε περιοχές πέρα από τον ελκυστή η πρόβλεψη δεν είναι τόσο πετυχημένη όσο και με την χρήση τοπικών μοντέλων. Βέβαια, αυτό δεν μας ενδιαφέρει άμεσα, αλλά είναι ενδεικτικό του ότι τα καθολικά πολυώνυμα δεν είναι πάντα το καταλληλότερο μοντέλο.



2.6: Ανακατασκευή δυναμικής της απεικόνισης Ikeda με global polynomials

2.3.3 Προσέγγιση της δυναμικής του συστήματος με τοπικά μοντέλα σε συστάδες του ελκυστή

Μία τεχνική η οποία έχει προταθεί στο [23] είναι να χρησιμοποιηθεί σε ένα πρώτο στάδιο ένα νευρωνικό δίκτυο SOM (Self Organizing Maps) για την ομαδοποίηση των δεδομένων πάνω στον ελκυστή του συστήματος, και στη συνέχεια, έχοντας εντοπίσει τις συστάδες του, να βρούμε ένα τοπικό πολυωνυμικό μοντέλο το οποίο να αποδίδει τη δυναμική του σε κάθε συστάδα. Η μέθοδος αυτή αποτελεί κατά κάποιο τρόπο την ενδιάμεση λύση μεταξύ των προηγούμενων μεθόδων: συνδυάζει την ακρίβεια του τοπικού μοντέλου με την ανοχή σε θόρυβο και την ευστάθεια των λύσεων σε αραιές περιοχές του ελκυστή του καθολικού μοντέλου. Η μέθοδος, για συντομία, θα καλείται SLH (SOM Local Hybrid).

Το νευρωνικό δίκτυο SOM

Τα νευρωνικά δίκτυα SOM εισήχθησαν απο τον Kohonen τη δεκαετία του 1980 και χρησιμοποιούνται για την οπτικοποίηση πολυδιάστατων δεδομένων σε μικρό αριθμό διαστάσεων -συνήθως δύο- και κατ'επέκταση για ομαδοποίηση πολυδιάστατων δεδομένων: αφού σε κάθε M-διάστατο σημείο αντιστοιχίζεται μία θέση σέ ένα διδιάστατο πλέγμα, μπορούμε να ομαδοποιήσουμε τα δεδομένα με βάση τη θέση τους στο πλέγμα. Αποτελείται από N νευρώνες ο καθένας εκ των οποίων αντιδρά διαφορετικά σε κάθε ερέθισμα (δηλαδή σε μία

M διάστατη είσοδο). Οι νευρώνες είναι τοποθετημένοι έτσι ώστε να σχηματίζουν ένα 2-διάστατο πλέγμα (αν χρησιμοποιούμε δύο διαστάσεις στην έξοδο), το οποίο μετά από τη διαδικασία εκπαίδευσης έχει τέτοια δομή ώστε να περιέχει πληροφορία για την κατανομή των δεδομένων εισόδου. Η θέση του i-οστού νευρώνα στο χώρο εισόδου δίνεται από το διάνυσμα

$$W^i = \left[w_1^i w_2^i \dots w_M^i\right] \tag{2.38}$$

Κατά τη διαδικασία μάθησης ο κάθε νευρώνας αναλαμβάνει μία πυκνή πέριοχή του χώρου εισόδου ενώ οι γειτονικοί του (στην έξοδο) νευρώνες αναλαμβάνουν μία σχετικά κοντινή περιοχή. Ακολουθεί ο αλγόριθμος εκπαίδευσης του δικτύου SOM:

- Στάδιο αρχικοποίησης:δώσε τυχαία (και διαφορετικά) βάρη σε κάθε νευρώνα
- Στάδιο ανταγωνισμού:
  - Επέλεξε ένα από τα δεδομένα εισόδου (έστω το X_j) και υπολόγισε την ευκλείδια απόσταση του X_j από κάθε νευρώνα:$|W_i - X_j|$
  - Επέλεξε τον νευρώνα k με την ελάχιστη απόσταση από το X_j
- Στάδιο συνεργασίας: Στο στάδιο αυτό η γειτονιά γύρω από τον k-οστό νευρώνα που επιλέγη στο πορηγούμενο στάδιο θα μετατεθεί προς τη μεριά του X_j.
  - Για κάθε νευρώνα i βρες την απόστασή του από τον νευρώνα k σύμφωνα με τη σχέση

    $$h_{ij} = e^{-\frac{|W_i - W_J|}{2\sigma^2}} \tag{2.39}$$

    όπου $|W_i - W_j|$ η ευκλείδια απόσταση των δύο νευρώνων και η ποσότητα σ ισούται με τη διασπορά της Gaussian. Συνήθως είναι μία φθίνουσα συνάρτηση του αριθμού των επαναλήψεων, π.χ: $\sigma(n) = \sigma_0 e^{-\frac{n}{\tau}}$ όπου σ_O η αρχική διασπορά της Gaussian και τ ο εκθετικός ρυθμός συρρίκνωσής της γειτονίας. Αυτό το οποίο επιτυγχάνουμε έτσι είναι στην αρχή, όπου πολλοί νευρώνες θα θεωρούνται ότι ανήκουν στη γειτονιά του νευρώνα-νικητή, να έχουμε τις χοντροειδείς προσαρμογές του δικτύου στα δεδομένα και σε μία δεύτερη φάση να φροντίσουμε για τη σύγκλιση του δικτύου αφήνοντας μόνο τους πολύ κοντινούς γείτονες του νικητή νευρώνα να αλλάξουν τη θέση τους.
  - Ενημέρωσε τα βάρη των νευρώνων σύμφωνα με τη σχέση

    $$W_j(n+1) = W_j(n) + a(n)h_{ij}(X - W_j(n)) \tag{2.40}$$

    Η συνάρτηση $a(n)$ καθορίζει το ρυθμό μάθησης και είναι συνήθως μία φθίνουσα συνάρτηση του χρόνου για τους ίδιους λόγους που αναφέρθηκαν προηγουμένως.
- Επίστρεψε στο δεύτερο βήμα εκτός αν η μέγιστη αλλαγή στα διανύσματα των νευρώνων δεν είναι μεγαλύτερη από ένα όριο ή αν ξεπεραστεί ένας αριθμός επαναλήψεων.

Η αναπροσαρμογή των βαρών σύμφωνα με την εξίσωση (2.40) έχει ως αποτέλεσμα τη μετακίνηση του διανύσματος των βαρών του νικητή και των γειτονικών του προς την κατεύθυνση του διανύσματος εισόδου X. Με την επαναληπτική παρουσίαση των προτύπων εκπαίδευσης, τα βάρη των νευρώνων τείνουν να ακολουθήσουν την κατανομή των

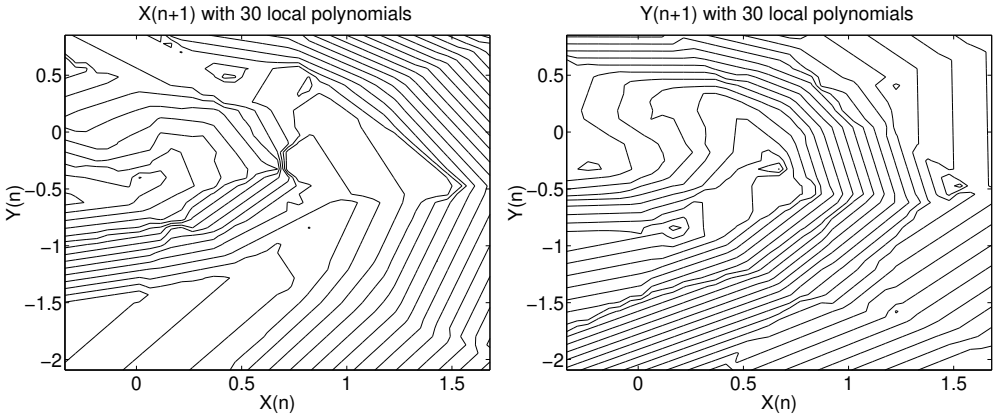
προτύπων εισόδου λόγω της διαδικασίας προσαρμογής γειτονικών νευρώνων. Συνοψίζοντας, έχουμε καταφέρει νευρώνες που βρίσκονται κοντά στο πλέγμα εξόδου να αντιστοιχούν σε περιοχές του χώρου εισόδου με παρόμοια δεδομένα και αντίστροφα.

Μεταφέροντας τη διαδικασία εκπαίδευσης του δικτύου SOM στο πρόβλημα της ανακατασκευής της δυναμικής ενός συστήματος στο χώρο φάσης του, εκπαιδεύουμε ένα δίκτυο SOM πάνω στα διανύσματα που έχουμε πάρει από την εμβύθιση των δεδομένων, οπότε οι k νευρώνες θα αναλάβουν ο καθένας και μία περιοχή του ελκυστή.

Μπορούμε να προσεγγίσουμε τη δυναμική του συστήματος στην περιοχή κάθε νευρώνα χρησιμοποιώντας όλα τα σημεία που του ανήκουν –αυτά δηλαδή από τα οποία απέχει λιγότερο από ότι οι υπόλοιποι νευρώνες– για τη δημιουργία ενός τοπικού μοντέλου βασιζόμενοι στην εξέλιξή τους (όπως στην παράγραφο 3.1). Μπορούμε να σχηματίσουμε μοντέλα δεύτερης και τρίτης τάξης στις συστάδες του ελκυστή, αλλά καθώς υπάρχει μικρός αριθμός από δεδομένα ο κίνδυνος της υπερεκπαίδευσης των δεδομένων είναι υπαρκτός. Πράγματι, χρησιμοποιώντας τοπικά μοντέλα μεγαλύτερης τάξης πήραμε σε μεγάλες διαστάσεις εμβύθισης του συστήματος Ikeda σφάλμα της τάξης μεγέθους της ακρίβειας του υπολογιστή, αλλά εκθέτες Lyapunov που διέφεραν αισθητά από τους πραγματικούς.

Η ιδέα ότι μπορούμε να σχηματίσουμε τοπικά μοντέλα σε ορισμένες μόνο περιοχές του ελκυστή είναι ελκυστική, όχι μόνο γιατί έχουμε πολύ μικρότερο αριθμό παραμέτρων να ορίσουμε, αλλά και γιατί η απαίτηση να σχηματίσουμε ένα τοπικό μοντέλο για κάθε σημείο του ελκυστή είναι υπερβολική και οδηγεί σε σφάλματα σε θορυβώδες περιβάλλον.

Στα σχήματα που ακολουθούν βλέπουμε την επίδοση του μοντέλου SLH τάξης 1 στην ανακατασκευή της δυναμικής του συστημάτος Ikeda.



2.7: Ανακατασκευή δυναμικής συστήματος Ikeda με τοπικά γραμμικά μοντέλα στις συστάδες του ελκυστή

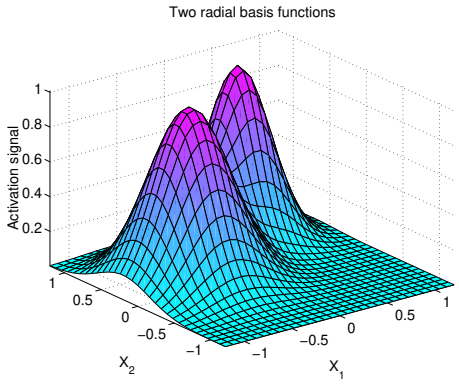
2.3.4 Προσέγγιση της δυναμικής του συστήματος με νευρωνικά δίκτυα RBF

Η χρήση νευρωνικών δικτύων με ακτινικές συναρτήσεις βάσεις (radial basis functions– RBF) αποτελεί μία από τις πλέον δημοφιλείς μεθόδους για την μοντελοποίηση της δυναμικής ενός συστήματος. Αυτό οφείλεται κυρίως στην δυνατότητα που έχουνε αυτά τα νευρωνικά δίκτυα να δίνουν μία συνεχή συνάρτηση (όπως το καθολικά πολυώνυμα), χωρίς να χάνουν πληροφορία για την τοπική συμπεριφορά του συστήματος (όπως τα τοπικά πολυώνυμα). Τα δίκτυα με RBF χρησιμοποιούν ένα εσωτερικό στρώμα από N νευρώνες το οποίο μετατρέπει το M–διάστατο σήμα X το οποίο βρίσκεται στην είσοδό του σε N–διάστατα

σήματα ενεργοποίησης. Η συνάρτηση μεταφοράς του νευρώνα i είναι μία ακτινική συνάρτηση (αμετάβλητη ως προς την περιστροφή), όπως π.χ. η πολυδιάστατη Gaussian:

$$f_i(X) = e^{\frac{|X-C_i|^2}{2\sigma^2}} \tag{2.41}$$

όπου C_i είναι το κέντρο του i-οστού νευρώνα και σ είναι η λεγόμενη διασπορά του νευρώνα που δείχνει πόσο τοπική η γενική είναι η δράση του (όπως και στην προηγούμενη παράγραφο). Ο κάθε νευρώνας 'αναλαμβάνει' μία περιοχή, δηλαδή η έξοδός του είναι μεγάλη όταν ένα σημείο στην είσοδό του βρίσκεται κοντά στην περιοχή του. Στο σχήμα 2.3.4 φαίνεται πώς αντιδρούν δύο ακτινωτές συναρτήσεις διαφορετικού κέντρου στην είσοδό τους, δίνοντας με αυτό τον τρόπο ένα διδιάστατο διάνυσμα (τα σήματα ενεργοποιήσης f_1, f_2) που δίνει κατά κάποιο τρόπο πληροφορία σχετικά με τη θέση ενός σημείου:



2.8: Δύο συναρτήσεις RBF

Το N-διάστατο διάνυσμα σημάτων ενεργοποίησης $F = [f_1, \ldots, f_N]$ πολλαπλασιαζόμενο με ένα δίανυσμα βαρών ίσων διαστάσεων δίνει την έξοδο του δικτύου, ενώ αν πρέπει να έχουμε ένα d-διάστατο σήμα εξόδου χρησιμοποιούμε τον αντίστοιχο αριθμό διανυσμάτων βαρών. Έτσι, τοποθετώντας νευρώνες σε κατάλληλες θέσεις στο N-διάστατο χώρο μπορούμε να επικεντρώσουμε στις περιοχές εκείνες που συγκεντρώνουν μεγάλο αριθμό από στοιχεία, χωρίς όμως να είμαστε αναγκασμένοι να καταφύγουμε σε μία ασυνεχή συνάρτηση για την προσέγγιση της συνάρτησης f, πράγμα που είναι διαισθητικά πιο ικανοποιητικό αλλά δίνει και καλύτερα αποτελέσματα στην περίπτωσή όπου έχουμε έναν περιορισμένο αριθμό από δεδομένα.

Ο αριθμός των νευρώνων που χρησιμοποιούνται είναι μία παράμετρος που πρέπει να καθορίσουμε. Η απλή λύση να χρησιμοποιήσουμε έναν νευρώνα για κάθε σημείο στο χώρο φάσης εκτός από υπολογιστικά απαιτητική είναι και πιο ευάλωτη στο θόρυβο, αφού έτσι 'μαθαίνουμε' και το θόρυβο. Μία απλή λύση είναι ο αριθμός των νευρώνων που θα χρησιμοποιήσουμε να είναι ένα κλάσμα του πλήθους των στοιχείων που έχουμε (π.χ. το $1/10$). Εναλλακτικά μπορεί να χρησιμοποιηθεί ένας αλγόριθμος ομαδοποίησης των δεδομένων, ο οποίος αυτόματα να καθορίζει τον κατάλληλο αριθμό συστάδων-νευρώνων. Εμείς δοκιμάσαμε διαφορετικούς αριθμούς νευρώνων (15-50) εξετάζοντας πώς και κατά πόσο επηρεάζει ο αριθμός των νευρώνων την επίδοση του δικτύου.

Αν έχουμε καταλήξει σε έναν αριθμό νευρώνων, για να ελαχιστοποιήσουμε το σφάλμα πρόβλεψης πρέπει να κατανείμουμε τους νευρώνες (δηλαδή τα κέντρα των Gaussian f_i που αναφέρθηκαν παραπάνω) με τον ίδιο τρόπο που είναι κατανεμημένα τα δεδομένα μας στο χώρο φάσης [26]. Για να προσδιορίσουμε τις θέσεις όπου είναι καταλληλότερο

να τοποθετήσουμε τους νευρώνες χρησιμοποιήθηκε ο καθιερωμένος και ιδιαίτερα απλός αλγόριθμος ομαδοποίησης δεδομένων k-means:
Αλγόριθμος k-means:

- Αρχικοποίηση των M κέντρων $C_{j,0}, (j = 1, \ldots, M)$ με τυχαίες τιμές
- Επανάλαβε:
  - Υπολόγισε τις αποστάσεις του κάθε σημείου από όλα τα κέντρα και ταξινόμησέ το στην ομάδα Ω_j του κέντρου j από το οποίο απέχει ελάχιστα.
  - Υπολόγισε τα νέα κέντρα $C_{j,t+1}$ από το μέσο όρο των συντεταγμένων των στοιχείων που έχουν καταταχθεί στην κάθε ομάδα:
    $$C_{j,t+1} = \frac{1}{N_{j,t}} \sum_{X \in \Omega_{j,t}} X \tag{2.42}$$
    όπου N_j ο αριθμός των στοιχείων που ανήκουν στην ομάδα του κέντρου C_j (από το προηγούμενο βήμα)
  - Θέσε $t = t + 1$
- Εώς ότου η διαφορά μεταξύ των διανυσμάτων $|C_{j,t+1} - C_{j,t}|$ για όλα τα διανύσματα είναι μικρότερη από ένα όριο ή ξεπεράσουμε έναν προκαθορισμένο αριθμό από επαναλήψεις.

Η τιμή της διασποράς στη συνάρτηση μεταφοράς ορίζεται ως ένα κλάσμα είτε της ακτίνας του ελκυστή είτε του μέγιστου μεταξύ των αποστάσεων δύο κέντρων (ώστε να μην υπάρχει περιοχή του χώρου φάσης που να έχει μένει ακάλυπτη). Η τιμή που χρησιμοποιήθηκε προσδιορίστηκε εμπειρικά να είναι ίση με $\sigma = 0.7 \cdot R$ όπου R η ακτίνα του ελκυστή.

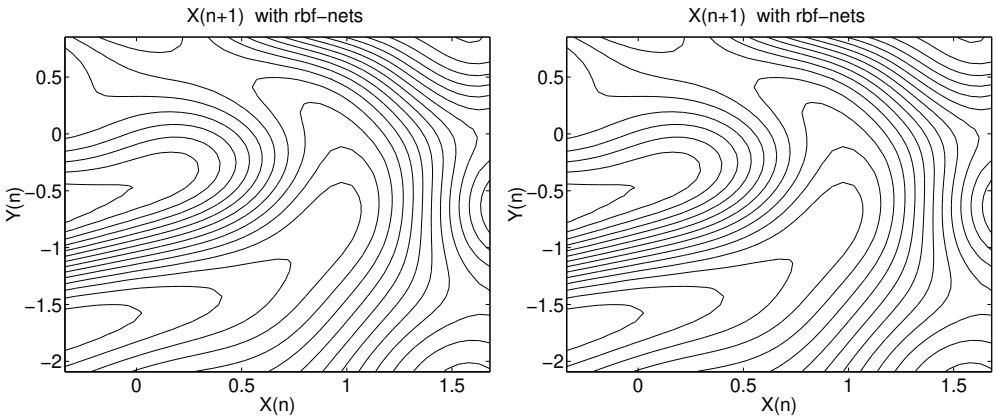
Έχοντας καθορίσει τα κέντρα και τις ακτίνες των συναρτήσεων μπορούμε να βρούμε και τα διανύσματα βαρών $W_d, d = 1 \ldots D_e$, που μας δίνουν την έξοδο του δικτύου από τα σήματα ενεργοποίησης. Στην περίπτωση των δικτύων RBF έχουμε :

$$F_j = \sum_{i=1}^{M} w_{ij} \cdot f_i \quad f_i(X) = e^{-\frac{|X - C_i|^2}{2\sigma^2}}, \qquad [w_{i1}, \ldots, w_{iM}] = W_i^T$$

Το ζητούμενο είναι να ελαχιστοποιηθεί το μέσο τετραγωνικό σφάλμα πρόβλεψης, που εκφράζεται ως η διαφορά μεταξύ της εξόδου (πρόβλεψη για την επομενη τιμή του X) και της επομενης τιμής του X. Η διαδικασία που ακολουθείται είναι όμοια με αυτή της παραγράφου (3.1) με τη μόνη διαφορά ότι αντί για τις τιμές των μονωνύμων έχουμε τα σήματα ενεργοποίησης f_i, οπότε καταλήγουμε στη σχέση

$$\mathbf{W} = \mathbf{M}^{-1} \sum_{i=1}^{N_B} \mathcal{F}(X(i))^T \cdot X(i+1) \tag{2.43}$$

όπου $\mathbf{W} = [W_1|W_2|W_3|\cdots|W_{D_e}]$ ο πίνακας με τα διανύσματα για τα οποία ελαχιστοποιείται το MSE της πρόβλεψης και $\mathbf{M}$ ο $M \times M$ πίνακας αυτοσυσχέτισης των σημάτων ενεργοποίησης. $\mathcal{F}$ είναι όπως και προηγουμένως μία συντομογραφία για τα σήματα ενεργοποίησης. Στην περίπτωση που έχουμε ένα μεγάλο αριθμό από εσωτερικούς νευρώνες μπορούμε να χρησιμοποιήσουμε και άλλες τεχνικές βελτιστοποίησης που χρησιμοποιούνται για την εκπαίδευση νευρωνικών δικτύων, όπως η gradient-descent αλλά στις περισσότερες περιπτώσεις έχουμε ένα μικρό αριθμό από νευρώνες, οπότε οι πίνακες που χρησιμοποιούνται για τον υπολογισμό του $\mathbf{W}$ δεν είναι μεγάλοι. Παρακάτω βλέπουμε την ανακτασκευή της δυναμικής του συστήματος Ikeda από το βέλτιστο, ως προς το MSE, δίκτυο RBF.



2.9: Ανακατασκευή δυναμικής της απεικόνισης Ikeda με το βέλτιστο RBF

2.3.5 Προσέγγιση της δυναμικής του συστήματος με δίκτυα τύπου WCP,WLP

Μία πιο εξελιγμένη εκδοχή των δικτύων της προηγούμενης παραγράφου κάνει χρήση αντί του σημάτος ενεργοποίησης (2.41) ενός κανονικοποιημένου σήματος ενεργοποίησης:

$$f_i(X) = \frac{e^{-\frac{|X-C_i|^2}{2\sigma^2}}}{\sum_{j=1}^{M} e^{-\frac{|X-C_i|^2}{2\sigma^2}}} = \frac{\mu_i(X)}{\sum_{j=1}^{M}\mu_j(X)} \quad i = 1 \dots R \tag{2.44}$$

Το $\mu_i(X)$ μπορεί να ερμηνευτεί ως ένα μέτρο της γειτνίασης του X με την i-οστή συστάδα δεδομένων, ενώ το $f_i(X)$ εκφράζει σε ποιο βαθμό το X ανήκει σε αυτή τη συστάδα.

Αυτό το οποίο επιτυγχάνουμε χρησιμοποιώντας κανονικοποιημένα σήματα ενεργοποίησης είναι να έχουμε μία ομαλή μετάβαση από τη μία γειτονιά του χώρου φάσης στην άλλη, χρησιμοποιώντας τις ποσότητες $f_i(X)$ ως βάρη για να υπολογιστεί ένας ζυγισμένος μέσος τοπικών μοντέλων. Τα τοπικά μοντέλα μπορεί να είναι σταθερές, οπότε έχουμε έναν Weighted Constant Predictor (WCP) ή γραμμικές συναρτήσεις, οπότε έχουμε έναν Weighted Linear Predictor (WLP). Πιο συγκεκριμένα στην περίπτωση του WCP η έκφραση στην έξοδο είναι ίση με

$$F(X) = \sum_{j=1}^{R} f_j b_j = \sum_{j=1}^{R} \left\{ \frac{\mu_j(X)}{\sum_{k=1}^{R}\mu_k(X)} \right\} \cdot b_j \tag{2.45}$$

ενώ για τον WLP έχουμε

$$F(X) = \sum_{j=1}^{R} f_j \left(A_j^T X + b_j \right) = \sum_{j=1}^{R} \left\{ \frac{\mu_j(X)}{\sum_{k=1}^{R}\mu_k(X)} \right\} \cdot (A_j X + b_j) \tag{2.46}$$

Στους παραπάνω τύπους έχουμε (αν η έξοδος είναι D_o διαστάσεων):

$$F : D_o \times 1, \quad A_j : D_o \times, b \quad b_j : D_o \times 1, \quad X : D_e \times 1 \tag{2.47}$$

Οι σταθερές A, b (θεωρούμε τη δυσκολότερη περίπτωση του WLP) υπολογίζονται απαιτώντας να ελαχιστοποιείται το MSE της πρόβλεψης, αν θεωρήσουμε ότι οι ποσότητες $f_i, f_i \cdot x_j, i = 1 \dots R, j = 1 \dots D_e$ που εμπλέκοται στις εξισώσεις (2.45,2.46) είναι σταθερές. Θεωρούμε δηλαδή ότι θέλουμε να ανακατασκευάσουμε την απεικόνιση

$$\{w_i(X(n)), w_i(X(n)) \cdot X_j(n)\} \overset{2.46}{\rightarrow} X(n+1) \tag{2.48}$$

Με βάση την 2.46 η απεικόνιση αυτή είναι γραμμική οπότε μπορούμε πάλι να καταλήξουμε σε ένα σύστημα γραμμιών εξισώσεων αν θέλουμε να ελαχιστοποιήσουμε το MSE. Θεωρώντας ότι οι παραπάνω ποσότητες συνιστούν το διάνυσμα $\mathcal{F}$,

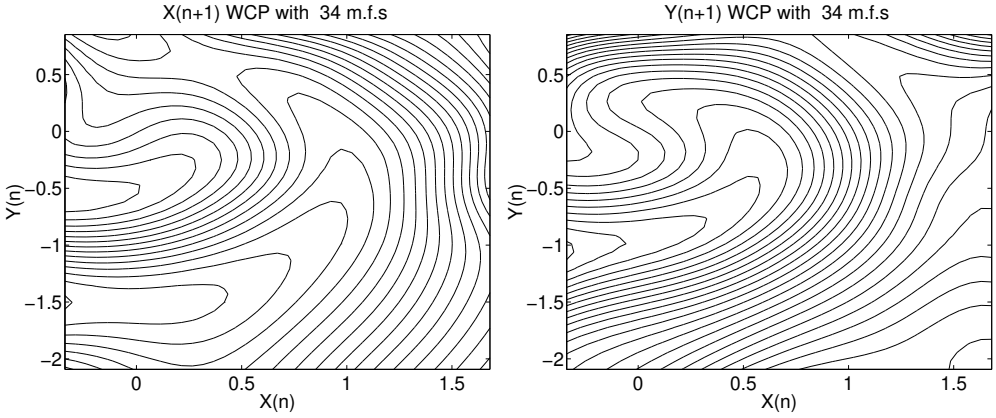
$$\mathcal{F} = [f_1, f_2, \ldots, f_R, f_1x_1, f_1x_2, \ldots, f_1x_{D_e}, f_2x_1, \ldots f_Rx_{D_e}] \tag{2.49}$$

η βέλτιστη κατά το MSE λύση βρίσκεται από την επίλυση του $(R+R\cdot D_e) \times (R+R\cdot D_e)$ συστήματος:

$$\begin{aligned} \mathbf{S}\cdot\mathbf{C} &= \sum_{i=1}^{N} \mathcal{F}^T(X(i))X(i+1) \quad \rightarrow \quad \mathbf{C} = \mathbf{S}^{-1}\sum_{i=1}^{N} \mathcal{F}^T(X(i))X(i+1) \\ \mathbf{S} &= \sum_{i=1}^{N} \mathcal{F}^T\mathcal{F}^T, \quad \mathbf{C} = [C(1)|C(2)|\ldots|C(D_o)] \end{aligned} \tag{2.50}$$

Το διάνυσμα $C(i)$ συνδέει τι διάνυσμα F με την i-οστή έξοδο του προβλέπτη, μπορούμε δηλαδή να υπολογίσουμε τα A,b από τον πίνακα $\mathbf{C}$.

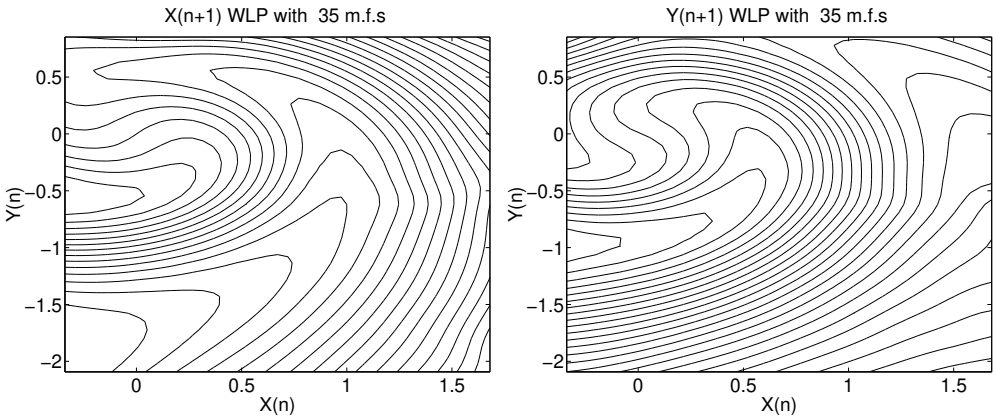
Αξίζει να σημειωθεί εδώ ότι οι προβλέπτες WCP-WLP είναι ισοδύναμοι με ένα ασαφές σύστημα για πρόβλεψη τύπου Takagi-Sugeno-Kang (TSK) [24] μηδενικής και πρώτης τάξης αντίστοιχα, με gaussian συναρτήσεις συμμετοχής. Καθώς όμως στο χώρο της πρόβλεψης χαοτικών χρονοσειρών εισήχθησαν από τον Stokbro με αυτή τη ονομασία, τη διατηρούμε αν και είναι λιγότερο διαδεδομένη και δημοφιλής. Βασιζόμενοι στην παραπάνω παρατήρηση εφαρμόσαμε διάφορες τεχνικές (SVD, εκπαίδευση με αλγόριθμο RPROP) που έχουν εφαρμοστεί με επιτυχία στην εκπαίδευση συστημάτων TSK, και πήραμε πολύ καλύτερα αποτελέσματα από αυτά που παίρνουμε λύνοντας μόνο το παραπάνω σύστημα εξισώσεων. Καθώς μία αναλυτική περιγραφή των μεθόδων αυτών θα διασπούσε τη συνέχεια του κειμένου, ή παρουσίασή τους γίνεται σε ένα ξεχωριστό παράρτημα. Εκεί φαίνονται και οι επιδόσεις τους στην ανακατσκευή της δυναμικής του συστήματος Ikeda, Παρακάτω φαίνονται οι επιδόσεις των συστημάτων WCP-WLP στην ανακατακευή της δυναμικής του συστήματος Ikeda, ενώ οι επιδόσεις τους αν εφαρμόσουμε τους αλγορίθμους SVD, RPROP φαίνονται στο παράρτημα.



2.10: Ανακατασκευή δυναμικής της απεικόνισης Ikeda με το βέλτιστο WCP

2.3.6 Προσέγγιση της δυναμικής του συστήματος με δίκτυα τύπου SVM

Τα δίκτυα SVM (Support Vector Machine) προτείνουν μία νέα μέθοδο για την εκπαίδευση ενός νευρωνικού δικτύου. Σύμφωνα με αυτήν αντί να ελαχιστοποιείται το συνολικό MSE



2.11: Ανακατασκευή δυναμικής της απεικόνισης Ikeda με τον βέλτιστο WLP

της πρόβλεψης πάνω στα δεδομένα με τα οποία γίνεται η εκπαίδευσή, σκοπός τους είναι είναι να ελαχιστοποιηθεί η πολυπλοκότητα του προβλέπτη, υπό τη συνθήκη το λάθος της πρόβλεψης να μην ξεπερνάει ένα όριο ανοχής. Έτσι, σε αντίθεση με τα δίκτυα που ανφέρθηκαν προηγουμένως, τα οποία ελαχιστοποιώντας το MSE της πρόβλεψης πάνω στα δεδομένα καταλήγουν να μαθαίνουν τα δεδομένα, ένα δίκτυο SVM μαθαίνει μόνο όσα χρειάζεται για να έχει μία ικανοποιητική απόδοση.

Αυτό το οποίο επιτυγχάνεται έτσι είναι πολύ καλή ικανότητα γενίκευσης και αποφεύγεται το φαινόμενο της υπερεκπαίδευσης. Η αντίστοιχη λύση που έχει προταθεί για τα δίκτυα που παρουσιάστηκαν προηγουμένως είναι αντί του MSE να ελαχιστοποιείται το MSE σε συνδυασμό με έναν όρο ποινής για την πολυπλοκότητα του δικτύου. Τα δίκτυα τύπου SVM, όμως χρησιμοποιούν μία πολύ πιο πειθαρχημένη προσέγγιση στο θέμα αυτό οπότε δίνουν και πολύ καλύτερα αποτελέσματα.

Η k-οστή συνιστώσα της συνάρτησης f μπορεί να προσεγγιστεί απλοϊκά με βάση τη σχέση $F_k(X) = W^T \cdot X + b$. Καθώς δεν είναι δυνατό να παρουσιαστεί σε μικρό χώρο το σκεπτικό με το οποίο καταλήγουμε στις εξισώσεις με τις οποίες προσδιορίζεται το W παραπέμπουμε τον ενδιαφερόμενο αναγνώστη στο παράρτημα και παρουσιάζουμε εδώ πολύ συνοπτικά, στην τελική τους μορφή, τις εξισώσεις που προκύπτουν για την F_k:

$$\text{Ελαχιστοποίησε την ποσότητα} \quad \frac{1}{2}W^TW + C\sum_{i=1}^{l}(\xi_i + \xi_i^*)$$

$$\text{Υπό τις συνθήκες} \quad \begin{cases} X_k(n+1) - W^TX(n) - b & \leq \epsilon + \xi_i \\ W^TX(n) + b - X_k(n+1) & \leq \epsilon + \xi_i^* \quad n = 1, 2 \dots N \\ \xi_i, \xi_i^* & \geq 0 \end{cases}$$

Ο όρος W αντιστοιχεί στην πολυπλοκότητα του δικτύου, ϵ είναι το διάστημα ανοχής της συνάρτησης κόστους, N είναι το πλήθος των ζευγών $\{X(n) \to X(n+1)\}$ που έχουμε, ενώ C είναι ένας όρος που εκφράζει πόσο σημαντικά θεωρούμε τα όποια λάθη του δικτύου: ένα μεγάλο C δεν επιτρέπει στο δίκτυο να κάνει λάθος μεγαλύτερο από ϵ ενώ ένα μικρό C αφήνει το δίκτυο να έχει έναν αριθμό από λάθη μεγαλύτερο από ϵ, δίνοντάς μας ένα απλούστερο δίκτυο. Το παραπάνω πρόβλημα είναι ένα πρόβλημα τετραγωνικού προγραμματισμού, οπότε καταλήγουμε πάντα στο καθολικό ελάχιστο της συνάρτησης κόστους, πράγμα που δεν ισχύει για τα νευρωνικά δίκτυα. [14] Το δυικό πρόβλημα του προ-

[14] Η τιμή του b προκύπτει ως ένα παραπροϊόν της επίλυσης του προβλήματος βελτιστοποίησης, όπως φαίνεται στο παράρτημα.

βλήματος αυτού (βλ.π.χ. [46]) είναι

$$\begin{aligned} \text{Μεγιστοποίησε την ποσότητα} \quad & - \quad \frac{1}{2}\sum_{i,j=1}^{l}\left[(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)X(i)^TX(j)\right]-\epsilon\sum_{i=1}^{l}(\alpha_i+\alpha_i^*)+ \\ & + \quad \sum_{i=1}^{l}X_k(i+1)(\alpha_i-\alpha_i^*) \\ \text{Υπό τις συνθήκες} \quad & \begin{cases} \sum_{i=1}^{l}(\alpha_i-\alpha_i^*) & =0 \\ \alpha_i,\alpha_i^* & \in[0,C] \end{cases} \end{aligned} \tag{2.51}$$

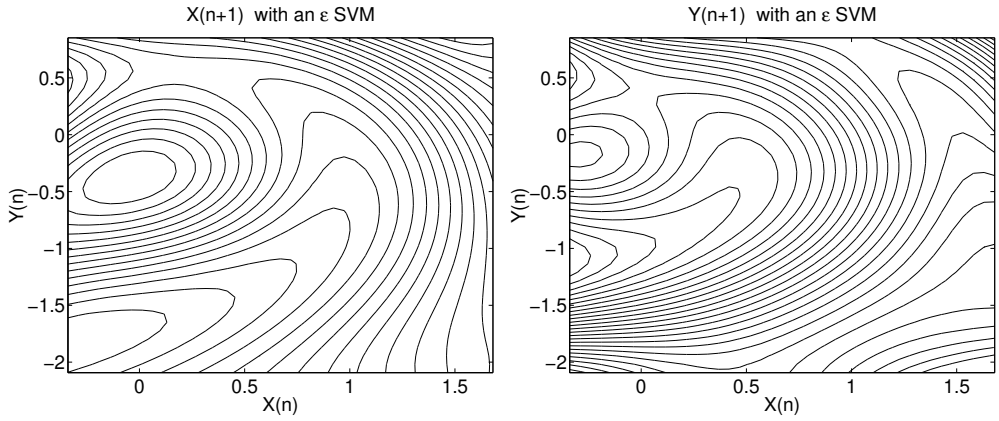
ενώ από τις συνθήκες Karush–Kuhn–Tucker προκύπτει για το βέλτιστο διάνυσμα W:

$$W=\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)X(i)\rightarrow f(X)=\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)X^TX(i)+b \tag{2.52}$$

Παρατηρούμε ότι μπορούμε αντί του εσωτερικού γινομένου δύο σημείων να χρησιμοποιήσουμε οποιαδήποτε συνάρτηση έχει ιδιότητες εσωτερικού γινομένου, συνθήκη που ικανοποιούν για παράδειγμα οι παρακάτω συναρτήσεις:

$K(x,x_i)=$	$e^{-\frac{\lvert x-x_i\rvert}{\sigma^2}}$	Gaussian
$K(x,x_i)=$	$\tanh(\beta_0x^Tx_i+\beta_1)$	Σιγμοειδής
$K(x,x_i)=$	$(x^Tx_i+1)^p$	Πολυωνυμική
$K(x,x_i)=$	$B_n^d(\lvert x-x_i\rvert)$	n–τάξης d–διάστατη spline

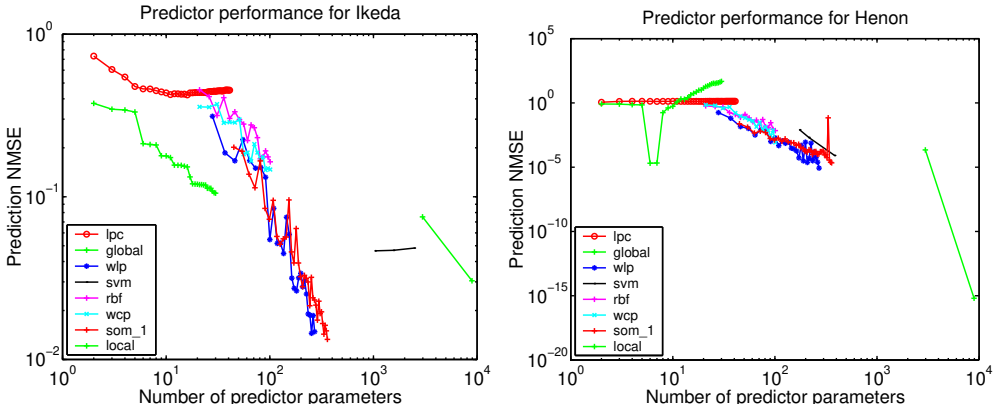
Εμείς χρησιμοποιήσαμε απλά έναν gaussian πυρήνα με διασπορά ίση με $\sigma=1.2\cdot R$, όπου R η ακτίνα του ελκυστή, ενώ θέσαμε $C=75$. Πρέπει να τονιστεί ότι αυτή είναι μόνο μία από τις επιλογές που υπάρχουν, αφού μπορεί να χρησιμοποιηθεί οποιοσδήποτε πυρήνας, καθώς και συνδυασμοί τους. Οι τιμές που θέσαμε στις ελεύθερες παραμέτρους προέκυψαν εμπειρικά μετά από διάφορες δοκιμές, αν και δεν επηρεάζουν επί της ουσίας τη δυνατότητα του SVM να μαθαίνει τη δυναμική των δεδομένων. Παρακάτω βλέπουμε πώς ανακατασκευάζει ένα δίκτυο SVM τη δυναμική του συστήματος Ikeda.



2.12: Ανακατασκευή δυναμικής της απεικόνισης Ikeda με το βέλτιστο SVM

2.3.7 Συμπεράσματα -σύγκριση MSE πρόβλεψης

Στο σχήμα (2.13) βλέπουμε τις επιδόσεις των διαφόρων μεθόδων, όπου φαίνεται το κανονικοποιημένο μέσο τετραγωνικό σφάλμα (NMSE) της πρόβλεψης για το επόμενο βήμα ως συνάρτηση του πλήθος των ελεύθερων παραμέτρων του προβλέπτη. Υπενθυμίζουμε ότι η σύγκριση γίνεται έχοντας ως απώτερο στόχο την εφαρμογή των μεθόδων αυτών στην επεξεργασία σήματος φωνής, οπότε δεν ασχολούμαστε προς το παρόν με το πρόβλημα της υπερεκπαίδευσης.



2.13: Επίδοση των προβλεπτών για τα συστήματα Ikeda/Henon

Παρατηρούμε τα εξής

- Με έναν πολύ μικρότερο αρθιμό παραμέτρων με τα συστήματα RBF,WCP,WLP,SLH, επιτυγχάνουμε εξίσου καλό NMSE πρόβλεψης όπως και με τα τοπικά πολυώνυμα. Από τους παραπάνω προβλέπτες αυτός που χρησιμοποιεί τις λιγότερες παραμέτρους σχετικά με τους άλλους είναι σχεδόν πάντα ο WLP.

- Το δίκτυο SVM για να πετύχει MSE αντίστοιχο αυτού των υπόλοιπων προβλεπτών χρειάζεται έναν πολύ μεγαλύτερο αριθμό παραμέτρων. Αυτό φυσικά δεν σημαίνει ότι έχουμε έναν πολυπλοκότερο προβλέπτη: τα περισσότερα από τα στοιχεία που συμμετέχουν στην πρόβλεψη έχουν πολύ μικρό συντελεστή βάρους (βλ. παράρτημα), οπότε δεν επηρεάζουν σε μεγάλο βαθμό την έξοδο του δικτύου. Εκτός αυτού, εξαρχής το δίκτυο SVM δεν εκπαιδεύεται με σκοπό την ελαχιστοποίηση NMSE πάνω στα δεδομένα με τα οποία έχει εκπαιδευτεί, αλλά την ελαχιστοποίηση του σφάλματός του με καινούργια δεδομένα.

- Τα καθολικά μοντέλα αν και έχουν σχετικά μέτρια αποτελέσματα είναι ελκυστικά λόγω του μικρού αριθμού προσδιοριστέων παραμέτρων του προβλέπτη. Φυσικά στο σύστημα Henon δίνουν πολύ καλά αποτελέσματα, αφού η δυναμική του μπορεί να εκφραστεί μόνο με πολυωνύμα.

- Για όλους τους προβλέπτες το NMSE είναι μεγαλύτερο για το σύστημα Ikeda, κάτι που ήταν αναμενόμενο και από την πολυπλοκότερη δυναμική του.

- Σε όλες τις περιπτώσεις το μοντέλο LPC αποτυγχάνει.

2.4 Αναλλοιώτες ποσότητες ενός χαοτικού δυναμικού συστήματος

Η μακροπρόθεσμη εξέλιξη ενός χαοτικού συστήματος είναι εξαιρετικά ευαίσθητη στις αρχικές συνθήκες, δηλαδή ακόμα και αν έχει ξεκινήσει με απειροελάχιστες διαφορές στις αρχικές συνθήκες έπειτα από ένα χρονικό σημείο οι τροχιές που διαγράφει στο χώρο φάσης είναι τελείως ασυσχέτιστες. Παρ'αυτά, υπάρχουν ορισμένες ποσότητες οι οποίες είναι ανεξάρτητες από τις αρχικές συνθήκες καθώς και από τις συντεταγμένες στις οποίες περιγράφεται το σύστημα [15]. Οι ποσότητες αυτές ονομάζονται αναλλοίωτες του δυναμικού συστήματος και μπορούν να χρησιμοποιηθούν για τους σκοπούς της αναγνώρισης και του χαρακτηρισμού ενός συστήματος. Οι βασικότερες αναλλοίωτες είναι η κλασματική διάσταση του ελκυστή στον οποίο κινείται το συστήμα και οι εκθέτες Lyapunov του.

Προτού προχωρήσουμε στην παρουσίασή τους πρέπει να τονίσουμε ότι έχουν νόημα μόνο εφόσον το σύστημα που παρατηρείται είναι εργοδικό, οπότε μέσω ενός εργοδικού θεωρήματος που παρουσιάζεται στο [12] μπορούμε να τις ορίσουμε ως μία αναλλοίωτη ποσότητα του συστήματος. Στα πλαίσια της επεξεργασίας σήματος φωνής μπορούμε να αντιστοιχίσουμε την απαίτηση αυτή με την παραδοχή, που γίνεται και για τη γραμμική ανάλυση του σήματος φωνής, ότι το σύστημα παραγωγής του παρατηρούμενου σήματος δεν αλλάζει κατά τη διάρκεια του χρονικού διαστήματος που εξετάζουμε, που είναι της τάξης των 10 msec, οπότε μπορούμε να το θεωρήσουμε ως εργοδικό.

2.4.1 Κλασματική διάσταση (fractal dimension)

Η πρώτη από τις αναλλοίωτες ποσότητες, η κλασματική διάσταση του ελκυστή του συστήματος εκφράζει το βαθμό πολυπλοκότητας του γεωμετρικού χώρου που καταλαμβάνει το σύστημα στο χώρο φασης. Την διάστασή αυτή μπορούμε να την ορίσουμε ως εξής: Θεωρούμε ένα σημείο X που ανήκει στον ελκυστή και εξετάζουμε πόσα σημεία βρίσκονται σε μία σφαίρα ακτίνας r που έχει κέντρο το X. Το πλήθος τους δίνεται από τη σχέση:

$$n(X,r) = \left[\sum_{n=1}^{K} \theta(r - |X(n) - X(k)|)\right]^{(q-1)}$$

Στην παραπάνω σχέση η συνάρτηση θ είναι η συνάρτηση Heavyside, η οποία δίνει έξοδο μονάδα αν το σημείο $X(k)$ βρίσκεται μέσα σε μία σφαίρα ακτίνας r με κέντρο το σημείο $X(k)$ και μηδέν αλλιώς. Για αρκετά μικρό r (όχι τόσο μικρό, όμως, ώστε να μην υπάρχουν σημεία μέσα στη σφαίρα) περιμένουμε για έναν d-διδιάστατο ελκυστή ότι το πλήθος αυτό θα εξαρτάται από την ακτίνα με μία σχέση της μορφής:

$$n(X,r) \sim r^{D(X)} \tag{2.53}$$

όπου D η διάσταση του ελκυστή στο σημείο X. Αν για το μέσο όρο του D της παραπάνω σχέσης βρεθεί ότι δεν παίρνει μία ακέραιη τιμή, συμπεραίνουμε ότι έχουμε να κάνουμε μέ έναν ελκυστή που έχει κλασματική διάσταση. Για παράδειγμα, για το D του ελκυστή του συστήματος Ikeda έχει υπολογιστεί ότι $D \simeq 1.56$. Από τις μεθόδους που χρησιμοποιούνται για τον υπολογισμό της κλασματικής διάστασης του ελκυστή η πιο απλή διαισθητικά είναι

[15] Αυτές εξαρτώνται, όπως είδαμε, από την επιλογή της χρονικής καθυστέρησης και τη διάσταση εμβύθισης.

αυτή της box-counting dimension, σύμφωνα με την οποία υπολογίζοντας τον αριθμό N των σφαιρών ακτίνας r που απαιτούνται για να καλύψουμε όλον τον ελκυστή βρίσκουμε το όριο της ποσότητας:

$$D_o = \lim_{r\to 0} \frac{\log[N(r)]}{\log[\frac{1}{r}]} \tag{2.54}$$

Γενικότερα πρέπει να ισχύει ότι $N(r) \sim r^{-D}$, όπου $N(r)$ ο μέσος όρος της ποσότητας $n(r)$ πάνω στον ελκυστή και D η διάσταση του ελκυστή, οπότε έχουμε ότι $D_o = D$.

2.4.2 Διάσταση συσχέτισης (Correlation Dimension)

Εκτός της Box-Counting διάστασης υπάρχουν και άλλοι τρόποι μέτρησης της κλασματικής διάστασης ενός γεωμετρικού αντικειμένου, οι οποίοι αν και δεν έχουν αυστηρή μαθηματική θεμελίωση έχουν μεγάλη πρακτική αξία. Για τον υπολογισμό τους ορίζουμε τη συνάρτηση $C(q,r)$ ως εξής:

$$\begin{aligned} C(q,r) &= \int \rho(X) n(X,r)^{(q-1)} \mathrm{d}X \\ &\simeq \frac{1}{M}\sum_{k=1}^{M}\left[\frac{1}{K}\sum_{n=1}^{K}\theta(r-|X(n)-X(k))\right]^{(q-1)} \end{aligned}$$

Η παραπάνω ποσότητα αποτελεί μία αναλλοίωτη ποσότητα υπό τη δυναμική του συστήματος, με την έννοια ότι είτε την υπολογίσουμε χρησιμοποιώντας τα σημεία $X(k)$ του ελκυστή είτε τα σημεία στα οποία απεικονίζονται υπό τη δυναμική του συστήματος $X(k) = F(X(k))$ θα πάρουμε τα ίδια αποτελέσματα. Με βάση την υπόθεση που κάναμε και στην προηγούμενη παράγραφο, ότι δηλαδή το πλήθος των σημείων του ελκυστή που βρίσκονται μέσα σε μία σφαίρα ακτίνας r θα είναι, για αρκετά μικρό r, ανάλογο του r^D όπου D η διάσταση του ελκυστή, συμπεραίνουμε ότι θα πρέπει να έχουμε και

$$C(q,r) \sim r^{(q-1)D} \tag{2.55}$$

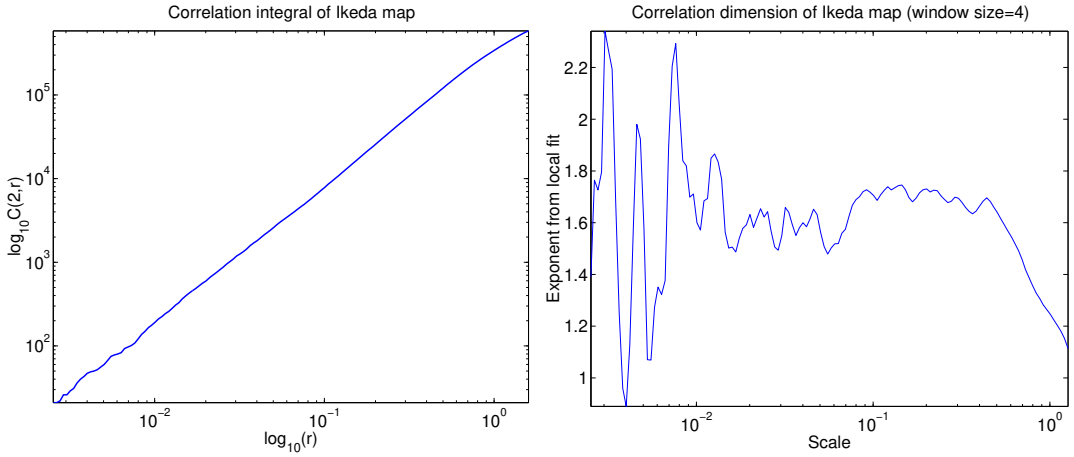
Με βάση τη σχέση (2.55) ορίζουμε την q-οστή *γενικευμένη διάσταση* ή διάσταση *Renyi [5]* του ελκυστή , D_q, ως το εξής όριο:

$$D_q = \lim_{r\to 0} \frac{\log\left[C(q,r)\right]}{(q-1)\log\left[r\right]} \tag{2.56}$$

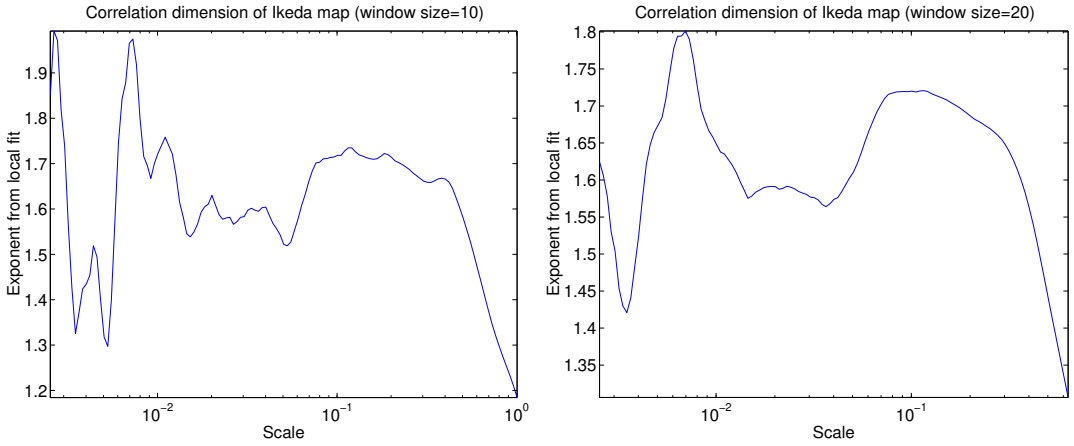
Η ποσότητα αυτή ισούται για $q = 0$ με τη Box-counting διάσταση του ελκυστή, για $q = 1$ με τη διάσταση πληροφορίας του ελκυστή, και για $q = 2$ με την αποκαλούμενη *διάσταση συσχέτισης* . Η διάσταση συσχέτισης ήταν η πρώτη διάσταση που υπολογίστηκε χρησιμοποιώντας την παραπάνω έκφραση και είναι μάλλον η πλέον συχνά χρησιμοποιούμενη κλασματική διάσταση ελκυστή κυρίως λόγω της απλότητας υπολογισμού της και του σχετικά μικρού αριθμού σημείων που χρειάζονται για να υπολογιστεί.

Η μέθοδος που ακολουθείται για τον υπολογισμό της D_q είναι να κάνουμε μετρήσεις του $C(q,r)$ για διάφορες τιμές του r και να βρίσκουμε με ελάχιστα τετράγωνα μία ευθεία η οποία να εκφράζει ικανοποιητικά την (αναμενόμενη) γραμμική σχέση μεταξύ του $\log[C(q,r)]$ και του $\log[r]$: η κλίση της ευθείας αυτής θα μας δίνει την D_q. Πρέπει να λάβουμε, όμως, υπόψιν και το γεγονός ότι έχουμε έναν πεπερασμένο αριθμό από σημεία στη διάθεσή μας, οπότε για πολύ μικρό r οι τιμές του $C(q,r)$ μπορούν να δώσουν λάθος

αποτελέσματα. Επίσης, για αρκετά μεγάλο r δεν έχει νόημα να χρησιμοποιήσουμε τις τιμές του $C(q,r)$ για τον υπολογισμό της κλίσης της ευθείας αφού μας ενδιαφέρει η κλίση της για το όριο $r \to 0$. Αυτό το οποίο κάνουμε λοιπόν είναι να χρησιμοποιούμε ένα μέρος μόνο των μετρήσεων $C(q,r)$ για τον υπολογισμό του D_q. Όσα προαναφέρθηκαν φαίνονται στα παρακάτω σχήματα, όπου παίρνουμε τα αποτελέσματα της εφαρμογής των μεθόδων αυτών για τον ελκυστή του συστήματος Ikeda.



2.14: Ολοκλήρωμα συσχέτισης του συστήματος Ikeda και εκτίμηση της διάστασης συσχέτισης του



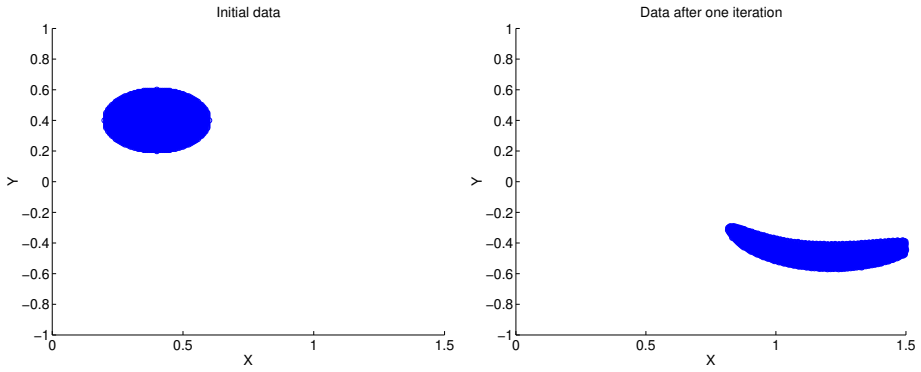
2.15: Εκτίμηση της διάστασης συσχέτισης του ελκυστή του συστήματος Ikeda με παράθυρα διαφορετικών μεγεθών

Παρατηρούμε ότι με ένα μεγάλο παράθυρο το αποτέλεσμα κυμαίνεται σε πολύ μικρότερο διάστημα αλλά το αντίτιμο της μεθόδου αυτής είναι ότι μπορούμε εύκολα να οδηγηθούμε σε ανακριβή αποτελέσματα.

2.4.3 Εκθέτες Lyapunov

Οι εκθέτες Lyapunov εκφράζουν το μέσο εκθετικό ρυθμό απόκλισης ή σύγκλισης δύο γειτονικών τροχιών στο χώρο φάσης. Με τους εκθέτες Lyapunov ποσοτικοποιείται δηλαδή η ταχύτητα με την οποία συμβαίνει η απόκλιση μεταξύ των τροχιών στο χώρο φάσης και μας δίνουν ένα άνω όριο του χρονικού διαστήματος στο οποίο μπορούμε να κάνουμε πρόβλεψη του σήματος με 'αξιώσεις' να πάρουμε σωστά αποτελέσματα (χρονικός ορίζοντας πρόβλεψης). Για το λόγο αυτό, πέραν του ότι ως αναλλοίωτες του συστήματος βρίσκουν εφαρμογή στον χαρακτηρισμό ενός συστήματος, μας ενδιαφέρει άμεσα ο υπολογισμός τους και για τους σκοπούς της πρόβλεψης και του ελέγχου συστημάτων.

Ας θεωρήσουμε μία κυκλική υποπεριοχή στο χώρο φάσης του συστήματος Ikeda. Αν εξετάσουμε την εξέλιξη της σφαίρας αυτής παρατηρούμε ότι κατά μήκος ενός άξονα η σφαίρα μπορεί να επιμηκύνεται, ενώ κατά μήκος του άλλου συρρικνώνεται και γίνεται ένα ελλιψοειδές -βλ.σχ.(2.16). Η επιμήκυνσή της d-διάστατης σφαίρας κατά μήκος



2.16: Επίδραση της δυναμικής του συστήματος Ikeda σε μία κυκλική περιοχή

του πρώτου άξονα ισοδυναμεί με απόκλιση των γειτονικών τροχιών ενώ η συρρίκνωση κατά μήκος του δεύτερου σημαίνει σύγκλιση (φυσικά, είναι δυνατόν να υπάρχουν και άλλοι συνδυασμοί ,π.χ. επιμήκυνση στην κατεύθυνση δύο αξόνων και συρρίκνωση στην κατεύθυνση των άλλων). Στους άξονες, στην κατεύθυνση των οποίων γίνεται επιμήκυνση, αντιστοιχούν θετικοί εκθέτες Lyapunov ενώ στους άλλους αντιστοιχούν αρνητικοί. Ο μεγαλύτερος θετικός εκθέτης Lyapunov λ_1 μας δίνει το ρυθμό με τον οποίο μεγεθύνεται ένα ευθύγραμμο τμήμα υπό την επίδραση της δυναμικής του συστήματος ($X(L) \sim e^{L\lambda_1}$ όπου το L είναι ο αριθμός βημάτων μπροστά), το εμβαδό μίας επιφάνειας μεταβάλλεται ανάλογα με την ποσότητα $e^{L(\lambda_1+\lambda_2)}$ κ.ο.κ.

Σημειώνεται ότι για ένα d-διάστατο σύστημα, αν θέλουμε να καταλήγει στη μόνιμη κατάσταση σε έναν ελκυστή, θα πρέπει το άθροισμα των εκθετών Lyapunov του να είναι αρνητικό, αφού η έκταση μίας περιοχής του χώρου φάσης θα μεταβάλλεται ανάλογα με τον όρο $e^{L\sum_{i=1}^{d}\lambda_i}$. Αν το άθροισμα των εκθετών δεν είναι αρνητικό, τότε αντί το σύστημα να συγκλίνει στον ελκυστή, θα έχουμε διαστολή των όγκων στο χώρο φάσης σε κάθε βήμα, οπότε το σύστημα δεν θα συγκλίνει πουθενά.

Σύμφωνα με όσα έχουν σημειωθεί στο δεύτερο κεφάλαιο για τα χαοτικά δυναμικά συστήματα, κάθε σύστημα που έχει τουλάχιστον έναν θετικό εκθέτη Lyapynov μπορεί να χαρακτηριστεί ως χαοτικό, αφού ακόμα και η μικρότερη διαφορά στις αρχικές συνθήκες μεγεθύνεται σε τέτοιο βαθμό (στην κατεύθυνση στην οποία αντιστοιχεί ο θετικός εκθέτης)

που τελικά δεν είμαστε σε θέση να προβλέψουμε μακροπρόθεσμα τη συμπεριφορά του συστήματος.

Σημειώνεται επίσης, ότι αν η δυναμική του συστήματος μπορεί να εκφραστεί με ένα σύστημα διαφορικών εξισώσεων, πρέπει να υπάρχει και ένας εκθέτης Lyapunov ίσος με μηδέν, αφού μετατοπίσεις προς μία διεύθυνση (την κατεύθυνση εξέλιξης του συστήματος με βάση τις εξισώσεις του) δεν είναι δυνατόν να μεγεθύνονται –δύο τροχιές που διαφέρουν στη διεύθυνση αυτή είναι ισοδύναμες, αφού η διαφορά της μίας από την άλλη οφείλεται μόνο σε μία χρονική καθυστέρηση.

Πρωτού προχωρήσουμε στην παρουσίαση μεθόδων για τον υπολογισμό των εκθετών Lyapunov ενός συστήματος (γνωστών και ως 'φάσμα Lyapunov') έχει ενδιαφέρον να παρουσιαστεί η σχέση μεταξύ των εκθετών Lyapunov και της κλασματικής διάστασης ενός ελκυστή.Οι Kaplan και Yorke προτείναν τον ορισμό της διάστασης Lyapunov D_L ως εξής:

$$D_L = k + \sum_{i=1}^{K} \frac{\lambda_i}{|\lambda_i|} \tag{2.57}$$

όπου το K ορίζεται ως ο αριθμός εκείνος για τον οποίο ισχύει:

$$\begin{array}{l} \sum_{i=1}^{K} \lambda_i < 0 \\ \sum_{i=1}^{K+1} \lambda_i > 0 \end{array} \tag{2.58}$$

Η παραπάνω ποσότητα ,αν και είναι αριθμητικά σχεδόν ίση με την διάσταση πληροφορίας του ελκυστή (D_I), δεν υπάρχει κάποια μαθηματική απόδειξη ότι σχετίζεται μαζί της. Μια διαισθητική ερμηνεία του τρόπου με τον οποίο συνδέεται η σχέση (4.5) με τη διάσταση του ελκυστή είναι η ακόλουθη:

Θεωρούμε την εξέλιξη στο χρόνο μίας σφαίρας ακτίνας r σε d διαστάσεις. Αυτή αλλοιώνεται σε ένα ελλειψοειδές με άξονες μήκους περίπου $re^{\lambda_i t}, i = 1 \dots d$. Επιλέγουμε έναν ακέραιο $K < d$ για τον οποίο $\lambda_{K+1} < 0$ και βρίσκουμε πόσοι K–διάστατοι κύβοι χρειάζονται για να καταλάβουν τον –διάστατο όγκο $e^{\lambda_1 t} e^{\lambda_2 t} \dots e^{\lambda_K t} \cdot r^K$, που θα είναι ο όγκος του ελλειψοειδούς μετά από χρόνο t.

Καθώς έχουμε διαλέξει έναν εκθέτη $\lambda_{K+1} < 0$ δεν χρειάζεται να υπολογίζουμε και διαστάσεις που συρρικνώνονται πιο γρήγορα από το $re^{\lambda_{K+1}}$ (αυτές των υπόλοιπων εκθετών) αφού στον τελικό τύπο παίρνουμε το όριο για $t \to \infty$ [16], οπότε μηδενίζονται πιο γρήγορα από τις άλλες διαστάσεις και δεν επηρεάζουν το αποτέλεσμα.

Το πλήθος των d–διάστατων κύβων ακτίνας $re^{\lambda_{K+1} t}$ που χρειάζονται για να καλύψουν τον όγκο που καταλαμβάνει η σφάιρα μετά από χρόνο t είναι:

$$N(t) = \frac{e^{t(\lambda_1 + \lambda_2 + \lambda_3 + \dots + \lambda_K)}}{e^{K t \lambda_{K+1}}} \tag{2.59}$$

Η διάσταση Lyapunov προκύπτει από τον κανόνα box–counting ως εξής[17]:

$$\begin{aligned} D_L &= \lim_{t \to \infty} \frac{\log[N(t)]}{\log \frac{1}{e^{t \lambda_{K+1}}}} \\ &= K + \sum_{i=1}^{K} \frac{\lambda_i}{|\lambda_{K+1}|} \end{aligned} \tag{2.60}$$

[16] Για $t \to \infty$ από μία αρχική σφαίρα παίρνουμε τον ελκυστή, σύμφωνα με μία βασική ιδιότητα των χαοτικών συστημάτων που ονομάζεται πυκνότητα των τροχιών, βλ. [5]

[17] Για απλότητα πήραμε ότι η πλευρά του κύβου $re^{K t \lambda_{K+1}}$ ισούται με $e^{K t \lambda_{K+1}}$. Στο όριο $t \to \infty$ η συμπεριφορά των λογαρίθμων τους είναι ίδια

Αυτή η ποσότητα ελαχιστοποιείται όταν το K επιλεγεί σύμφωνα με τη σχέση (2.58). Το ελάχιστο μας δίνει τον ελάχιστης διάστασης υποχώρο στον οποίο το σύνολο των σημείων στην τροχία μπορεί να 'υπάρξει', που μπορεί άμεσα να θεωρηθεί ως μία κλασματική διάσταση του ελκυστή.

2.4.4 Προσδιορισμός του φάσματος Lyapunov ενός συστήματος

Αν και έχουν προταθεί γεωμετρικής φύσης μέθοδοι για τον υπολογισμό του φάσματος Lyapunov ενός συστήματος βλ. π.χ. [37] έχει αποδειχθεί ότι αποτυγχάνουν να υπολογίσουν με ένα ανεκτό λάθος τους αρνητικούς εκθέτες. Για το λόγο αυτό υλοποιήθηκε και παρουσιάζεται μόνο η μεθόδος που προτάθηκε πρώτη φορά στο [34] και στη συνέχεια επεκτάθηκε και εξελίχθηκε όπως φαίνεται στα άρθρα [30], [31], [1], [32], [33].[18] Θεωρούμε ένα δυναμικό σύστημα του οποίου η εξέλιξη στο χώρο φάσης δίνεται από τη σχέση

$$X(k+1) = F(X(k)) \tag{2.61}$$

όπου το X είναι N-διάστατο. Αν θεωρήσουμε την εξέλιξη μίας τροχιάς που τη στιγμή k βρίσκεται στο σημείο $X(k)+\Delta(k)$ έχουμε

$$X(k+1)+\Delta(k+1) = F(X(k)+\Delta(k)) \simeq \mathbf{DF}(X(k))\Delta(k) + F(X(k)) \tag{2.62}$$

οπότε έχουμε για αρκετά μικρό $\Delta(k)$:

$$\Delta(k+1) \simeq \mathbf{DF}(X(k))\Delta(k) \tag{2.63}$$

Ο πίνακας $\mathbf{DF}$ είναι ο Jacobian της συνάρτησης f στο σημείο k. Η απόσταση των τροχιών μετά από m βήματα καθορίζεται από την σύνθεση των πινάκων $\mathbf{DF}$ σύμφωνα με τη σχέση

$$\begin{aligned} \Delta(k+m) &= \mathbf{DF}(X(k+m-1))\mathbf{DF}(X(k+m-2)\cdots\mathbf{DF}(y(k))\Delta(k) \\ \Delta(k+m) &= \mathbf{DF}^m(X(k))\Delta(k) \end{aligned} \tag{2.64}$$

όπου ο πίνακας $\mathbf{DF}^m$ είναι ίσος με την σύνθεση των παραπάνω Jacobians. Η ευστάθεια της τροχιάς μπορούμε να πούμε (γενικεύοντας από τη γραμμική θεωρία ευστάθειας) ότι καθορίζεται από τις ιδιοτιμές του πίνακα $\mathbf{DF}^m$: αν μία από αυτές συμπεριφέρεται ανάλογα του $e^{[m\lambda]}$ με $\lambda > 0$ τότε η τροχιά κατά μήκος της οποίας γίνεται ο υπολογισμός του $\mathbf{DF}^m$ είναι ασταθής. Αν αντί της ποσότητας $\Delta(k+m)$ πάρουμε το τετράγωνο του μέτρου της έχουμε:

$$|\Delta(k+m)|^2 = \Delta^T(k)\left[\mathbf{DF}^m(X(k))\right]^T \mathbf{DF}^m(X(k))\Delta(k) \tag{2.65}$$

Το πολλαπλασιαστικό εργοδικό θεώρημα του Oseledec μας εγγυάται ότι το όριο του πίνακα Oseledec

$$\mathbf{OSL}(k,m) = \left([\mathbf{DF}^m(X(k))]^T\mathbf{DF}^m(X(k))\right)^{\frac{1}{2m}} \tag{2.66}$$

για $m \to \infty$ υπάρχει και δεν εξαρτάται από το m (δηλ.$\mathbf{OSL} = \lim_{m\to\infty}\mathbf{OSL}(k,m)$). Ο λογάριθμος τών ιδιοτιμών του $\mathbf{OSL}$ ($eig_1, eig_2, \ldots, eig_d$) μας δίνει τους *καθολικούς*[19] εκθέτες Lyapunov του συστήματος $\lambda_1, \lambda_2, \ldots, \lambda_d$ (γράφονται σε αύξουσα σειρά)

[18] Εκεί μπορούν να βρεθούν και οι βιβλιογραφικές πηγές αρκετών θεωρημάτων στα οποία θα ανφερθούμε στη συνέχεια

[19] Ο όρος καθολικός θα δικαιολογηθεί αργότερα

Ιδιοτιμές πίνακα OSL-τεχνική αποσύνθεσης QR

Αυτό που απομένει είναι να βρούμε τις ιδιοτιμές του πίνακα $\mathbf{OSL}$. Αυτό δεν είναι πρακτικά τόσο απλό όσο φαίνεται: ο πίνακας $\mathbf{OSL}(k,m)$ έχει δείκτη κατάστασης [20] περίπου ίσο με $e^{[m(\lambda_1-\lambda_d)]}$, ποσότητα που μπορεί να γίνει εξαιρετικά μεγάλη για μεγάλες τιμές του m, οπότε η εύρεση των ιδιοτιμών του μπορεί να γίνει πολύ ανακριβής. Ο μεγάλος δείκτης κατάστασης είναι φυσικό επόμενο της φύσης ενός χαοτικού συστήματος, ενός συστήματος δηλαδή το οποίο μεγεθύνει σε κάθε βήμα όλο και περισσότερερο τις παραμικρότερες διαφορές. Η ιδιότητα αυτή μαθηματικά ισοδυναμεί με την ύπαρξη μίας (τουλάχιστον) ιδιοτιμής μεγαλύτερης της μονάδας σε κάθε πίνακα Jacobian, οπότε όταν οι ιδιοτιμές αυτές πολλαπλασιάζονται (όπως θα δούμε παρακάτω) προκύπτει ένας αριθμός ο οποίος αυξάνεται εκθετικά με το m (δηλαδή $eig_{max} \sim e^{[m\lambda_1])}, \lambda_1 > 0$ ενώ το αντίστροφο ισχύει για τις αρνητικές ιδιοτιμές $eig_{min} \sim e^{[m\lambda_d])}, \lambda_d < 0$ (αντίστοιχες στη συστολή των όγκων στο χώρο φάσης). Η ύπαρξη των αρνητικών ιδιοτιμών επιβάλλεται, όπως είδαμε, από την απαίτηση το σύστημά μας να καταλήγει σε έναν ελκυστή.

Για το λόγο αυτό θα αναπτύξουμε την ιδέα της ανδρομικής QR αποσύνθεσης, με βάση την οποία παρακάμπτουμε το πρόβλημα της διαγωνιοποίησης του πίνακα $\mathbf{OSL}$ ανάγοντάς το σε μικρότερα, ασφαλέστερα αριθμητικά υποπροβλήματα. Συγκεκριμένα ο πίνακας Oseledec μπορεί να θεωρηθεί ως το γινόμενο των $2m$ πινάκων $\mathbf{A}(2m)\mathbf{A}(2m-1)\cdots\mathbf{A}(1)$ οι οποίοι αποσυντίθενται σύμφωνα με τον αλγόριθμο QR[21] έτσι ώστε να ικανοποιείται η σχέση:

$$\mathbf{A}(j)\mathbf{Q}(j-1) = \mathbf{Q}(j)\mathbf{R}(j) \qquad \forall j, \mathbf{Q}(0) = \mathbf{I} \tag{2.67}$$

Με βάση τη σχέση (2.67) έχουμε

$$\begin{aligned}
\mathbf{A}(1) &= \mathbf{Q}(1)\mathbf{R}(1) \\
\mathbf{A}(2)\mathbf{Q}(1) &= \mathbf{Q}(2)\mathbf{R}(2) \overset{\cdot\mathbf{R}(1)}{\Rightarrow} \\
\mathbf{A}(2)\mathbf{A}(1) &= \mathbf{Q}(2)\mathbf{R}(2)\mathbf{R}(1) \\
&\vdots \\
\mathbf{A}(2m)\mathbf{A}(2M-1),\cdots,\mathbf{A}(1) &= \mathbf{Q}(2m)\mathbf{R}(2m)\mathbf{R}(2m-1),\cdots,\mathbf{R}(1)
\end{aligned} \tag{2.68}$$

Η σχέση (2.68) μας δείχνει ότι μπορούμε να εκφράσουμε τον πίνακα Oseledec ως το γινόμενο του ορθογώνιου πίνακα $\mathbf{Q}(2m)$ και των $2m$ άνω διαγώνιων πινάκων $\mathbf{R}(2m),\cdots\mathbf{R}(1)$. Για $m \to \infty$ έχουμε $\mathbf{Q}(2m) \to \mathbf{I}$ οπότε οι ιδιοτιμές του πίνακα Oseledec θα καθορίζονται από τις ιδιοτιμές του γινομένου των υπόλοιπων πινάκων. Καθώς οι πίνακες αυτοί είναι άνω διαγώνιοι, το γινόμενό τους θα είναι άνω διαγώνιος και η διαγώνιος (από την οποία παίρνουμε άμεσα τις ιδιοτιμές του) προκύπτει από τον πολλαπλασιασμό των στοιχείων των διαγωνίων των πινάκων $\mathbf{R}(1)\ldots\mathbf{R}(2m)$:

$$diag[\mathbf{OSL}] = diag[\mathbf{R}(1)] \odot diag[\mathbf{R}(2)] \odot\cdots\odot diag[\mathbf{R}(2m)] \tag{2.69}$$

όπου με $(\odot)$ εννοούμε τον πολλαπλασιασμό ένα προς ένα των στοιχείων των διαγωνίων. Ο μεγάλος δείκτης κατάστασης του πίνακα Oseledec οφείλεται στο πολλαπλασιασμό των

[20] Λόγος μεγαλύτερης προς μικρότερη ιδιοτιμή

[21] Ο αλγόριθμος QR μας δίνει από έναν πίνακα $\mathbf{A}$ έναν ορθογώνιο πίνακα $\mathbf{Q}$ και έναν άνω διαγώνιο πίνακα $\mathbf{R}$ τέτοιους ώστε $\mathbf{A} = \mathbf{QR}$

πίνακων $\mathbf{R}(1),\ldots,\mathbf{R}(2m)$ οι οποίοι εκφράζουν το βαθμό διαστολής ή συστολής προς τις κατευθύνσεις που ορίζουν οι πίνακες $\mathbf{Q}(1),\ldots,\mathbf{Q}(2m)$ [22]. Συνεπώς αυτό το οποίο επιτυγχάνουμε είναι να ανάγουμε το αριθμητικά επισφαλές πρόβλημα της εξαγωγής των ιδιοτιμών του πίνακα $\mathbf{OSL}$ στην εξαγωγή των ιδιοτιμών των $2m$ πινάκων $A(i)$ που θα έχουν καλύτερο δείκτη κατάστασης. Οι εκθέτες Lyapunov δίνονται τώρα από τη σχέση

$$\begin{aligned}\lambda_i = log[eig_i(\mathbf{OSL})] &= \log\left[eig_i\left(\lim_{m\to\infty}([\mathbf{DF}^m(X(k))]^T\mathbf{DF}^m(X(k)))^{\frac{1}{2m}}\right)\right] \overset{(2.68)}{=} \\ &= \log\left[\lim_{m\to\infty}\left(\frac{1}{2m}diag_i[\mathbf{R}(2m)]diag_i[\mathbf{R}(2m-1)]\cdots diag_i[\mathbf{R}(1)]\right)\right] \Rightarrow \\ \lambda_i &= \lim_{m\to\infty}\frac{1}{2m}\sum_{j=1}^{m}\log\left[diag_i[\mathbf{R}(j)]\right]\end{aligned}$$

Τοπικοί εκθέτες Lyapunov

Πέρα από την εύρεση των ιδιοτιμών του πίνακα Oseledec ένα πρόβλημα το οποίο προκύπτει είναι ότι συνήθως δεν έχουμε αρκετά δεδομένα ή χρόνο για να εκτελέσουμε τόσες επαναλήψεις ώστε ο πίνακας $\mathbf{OSL}(k,m)$ να είναι ανεξάρτητος της θέσης του σημείου $X(k)$, δεν μπορούμε δηλαδή να πάρουμε στο όριο τον πίνακα $\mathbf{OSL}$. Αυτό το οποίο το οποίο έχει προταθεί αντί του υπολογισμού του πίνακα $\mathbf{OSL}$ είναι να χρησιμοποιείται η παραπάνω διαδικασία διαγωνοποίησης για την εύρεση των ιδιοτιμών του πίνακα $\mathbf{OSL}(k,m)$, οι οποίες διαφοροποιούνται από σημείο σε σημείο για σχετικά μικρές τιμές του m και μας δίνουν τους *τοπικούς* εκθέτες Lyapunov. Η διαφορά τους με τους *καθολικούς* εκθέτες Lyapunov είναι ότι μας δίνουν τους εκθετικούς ρυθμούς απόκλισης τροχιών στο σημείο $X(k)$, πληροφορία η οποία μπορεί να είναι περισσότερο χρήσιμη από τους καθολικούς εκθέτες Lyapunov, αφού οι τελευταίοι μας δίνουν το *μέσο* [23] ρυθμό με τον οποίο μία διαφορά στις αρχικές συνθήκες μεγεθύνεται. Έτσι οι τοπικοί εκθέτες μπορούν να μας δείξουν σε ποιό βαθμό μπορούμε να εμπιστευόμαστε μία πρόβλεψή μας L βήματα μπροστά στο χρόνο σε ένα συγκεκριμένο σημείο του ελκυστή (το οποίο μπορεί να χαρακτηρίζεται από πολύ μεγάλο ή μικρό ορίζοντα πρόβλεψης) ενώ οι καθολικοί μας δίνουν μία μέση εκτίμηση του ορίζοντα πρόβλεψης για το σύστημα.

Εκτός αυτού οι τοπικοί εκθέτες Lyapunov για $m\to\infty$ μας δίνουν τους καθολικούς, ενώ στο [31] αναφέρεται ότι μπορούμε, χρησιμοποιώντας ένα μοντέλο παρεμβολής, να 'παρακάμψουμε' τον υπολογισμό των εκθετών που προκύπτουν από τον πίνακα $\mathbf{OSL}$ βρίσκοντας πώς τους προσεγγίζουν οι τοπικοί εκθέτες. Το μόνο που εκκρεμεί για τον υπολογισμό των τοπικών εκθετών είναι να αντιμετωπιστεί το γεγονός ότι ο $\mathbf{Q}(2m)$ της εξίσωσης (2.68) δεν είναι πάντα μοναδιαίος για πεπερασμένο m, οπότε δεν μπορούμε να πάρουμε τις ιδιοτιμές χρησιμοποιώντας μόνο τους διαγώνιους πίνακες $\mathbf{R}(1),\ldots,\mathbf{R}(2m)$. Για το λόγο αυτό γίνεται μία μικρή διαφοροποίηση του προηγούμενου αλγορίθμου ως εξής:

1. Εκτέλεσε την διαγωνιοποίηση QR για το γινόμενο πινάκων $\mathbf{A}(2m)\mathbf{A}(2m-1),\cdots,\mathbf{A}(1)$ και πάρε τον πίνακα

$$M_1 = \mathbf{Q}_1(2m)\mathbf{R}_1(2m)\mathbf{R}_1(2m-1),\cdots,\mathbf{R}_1(1) \tag{2.70}$$

[22] κατά καποιο τρόπο μπορούμε να πούμε ότι ο πίνακας $\mathbf{Q}(i)$ αντιστοιχεί στη 'φάση' και ο πίνακας $\mathbf{R}(i)$ στο 'μέτρο' του πίνακα $A(i)$

[23] Ο μέσος εννοείται κατά μήκος του ελκυστή

2. Εάν $\mathbf{Q}(2m) \simeq \mathbf{I}$ σταμάτησε ,αλλιώς

   - Σχημάτισε τον πίνακα

   $$\mathbf{M}_2 = \mathbf{R}_1(2m)\mathbf{R}_2(2m-1), \cdots, \mathbf{R}_1(1)\mathbf{Q}_1(2m) \tag{2.71}$$

   Καθώς ισχύει ότι $\mathbf{M}_2 = \mathbf{Q}_1^T(2m)\mathbf{M}_1\mathbf{Q}_1(2m)$ και ο $\mathbf{Q}$ είναι ορθογώνιος οι $\mathbf{M}_1$ και $\mathbf{M}_2$ θα έχουν τις ίδιες ιδιοτιμές.
   - Επανάλαβε το βήμα(1),μόνο που τώρα τη θέση του γινομένου πινάκων $\mathbf{A}(2m)\mathbf{A}(2m-1)\cdots\mathbf{A}(1)$ θα πάρει το γινόμενο πινάκων που μας δίνει τον πίνακα $\mathbf{M}_2$

Από ένα θεώρημα της αριθμητικής ανάλυσης ξέρουμε ότι ο πίνακας $\mathbf{Q}_i(2m)$ συγκλίνει στον πίνακα $\mathbf{I}$, ενώ κατά κανόνα αρκούν λίγες επαναλήψεις για να πούμε ότι $\mathbf{Q}_i \simeq \mathbf{I}$.

Υπολογισμός Jacobians

Ένα σημείο στο οποίο δεν έχουμε αναφερθεί είναι ο τρόπος με τον οποίο υπολογίζονται οι Jacobians για τα διάφορα μοντέλα που παρουσιάστηκαν στο προηγούμενο μέρος

Εξαγωγή πίνακα Jacobian –εκθετών Lyapunov από τοπικά πολυώνυμα

Θεωρούμε ότι έχουμε σχηματίσει ένα τοπικό μοντέλο της δυναμικής του συστήματος γύρω από το σημείο X (βλ εξ. 2.16):

$$F(X) = \sum_{k=1}^{M} C(k) \cdot p_k(X) \simeq \mathbf{f}(X) \tag{2.72}$$

όπου f είναι η συνάρτηση που περιγράφει τη δυναμική του συστήματος, F η προσέγγισή μας, p_k είναι ένα πολυώνυμο με διάνυσμα εκθετών που κωδικοποιείται με τον αριθμό k (βλ παράγραφο 3.1), M το πλήθος των πολυωνύμων που χρησιμοποιούνται για την προσέγγιση της συνάρτησης f και $C(1), C(2), \ldots, C(k)$ είναι το διάνυσματα για τα οποία ελαχιστοποιείται το μέσο τετραγωνικό σφάλμα της πρόβλεψης. Το ζητούμενο είναι να υπολογίσουμε την ποσότητα

$$J_{ab} = \frac{\partial F_a}{\partial x_b},\ a{=}1 \ldots d,\ b{=}1 \ldots d \tag{2.73}$$

τα στοιχεία δηλαδή του Jacobian του συστήματος. Από την (2.16) έχουμε:

$$\frac{\partial F_a}{\partial x_b} = \sum_{k=1}^{M} C_a(k) \cdot \frac{\partial(p_k(X))}{\partial x_b} \tag{2.74}$$

Από τη σχέση (2.17),δηλ. $x_1^{e_1} \cdot x_2^{e_2} \cdots x_d^{e_d} = p_k(x_1, x_2 \ldots, x_d) \leftrightarrow [e_1, e_2, \ldots, e_d] = decode(k)$ η οποία χρησιμοποιήθηκε στην παράγραφο 3.1 για τον ορισμό του k μπορούμε να γράψουμε την προηγούμενη σχέση ως εξής:

$$\frac{\partial F_a}{\partial x_b} = \sum_{(k|e_b(k)\neq 1)} C_a(k) \cdot e_b(k) p_{\tilde{k}}(X) \tag{2.75}$$

$$\tilde{k} \leftrightarrow [e_1 \ldots e_2 \ldots e_d]$$

Έτσι μπορούμε να παρουμε άμεσα τον Jacobian του συστήματος σε ένα σημείο από το τοπικό μοντέλο της δυναμικής του στην περιοχή του σημείου αυτού.

Εξαγωγή πίνακα Jacobian -εκθετών Lyapunov από καθολικά πολυώνυμα

Ο τύπος ο οποίος δίνει την πρόβλεψη για την επόμενη τιμή της μετάβλητης X είναι (2.26)

$$F(X) = \sum_{i=1}^{M} C(i)\phi_i(X) \tag{2.76}$$

όπου έχουμε (2.28)

$$\phi_i(x) = \sum_{k=1}^{m} A_k^m p_k(X) \tag{2.77}$$

Ο τύπος για την εξαγωγή του πίνακα Jacobian από ένα καθολικό πολυώνυμο είναι παρόμοιος με αυτόν για τα τοπικά πολυώνυμα. Η μόνη διαφορά είναι ότι οι εκφράσεις μέσα στο άθροισμα της (2.76) δεν είναι μονώνυμα αλλά πολυώνυμα του X οπότε πρέπει να τροποποιήσουμε τον τύπο για τη μερική παράγωγο. Συγκεκριμένα έχουμε:

$$\begin{aligned}
\frac{\partial F_a}{\partial x_b} &= \sum_{i=1}^{M} C_a(i) \cdot \frac{\partial(\phi_i(X))}{\partial x_b} = \\
&= \sum_{i=1}^{M} C_a(i) \cdot \frac{\partial(\sum_{k=1}^{i} A_k^i p_k(X))}{\partial x_b} \overset{[e_1 \ldots e_d]=decode(k)}{=} \\
&= \sum_{i=1}^{M} C_a(i) \cdot \sum_{(e_b(k)\neq 1 | k<i)} A_k^i e_b(k) p_{\tilde{k}}(X) \\
\tilde{k} &\leftrightarrow [e_1, \ldots, e_b - 1, \ldots, e_{D_e}]
\end{aligned} \tag{2.78}$$

Εξαγωγή πίνακα Jacobian -εκθετών Lyapunov για το SLH

Στην περίπτωση αυτή για κάθε δεδομένο βρίσκουμε για ποιον νευρώνα ελαχιστοποιείται η ποσότητα $|X - W_i|$ όπου W_i το διάνυσμα θέσης του κάθε νευρώνα και στη συνέχεια σχηματίζεται ο πίνακας Jacobian με βάση το τοπικό μοντέλο που αντιστοιχεί στη δυναμική του συστήματος στην περιοχή του νευρώνα με βάση τον τύπο (2.75).

Εξαγωγή πίνακα Jacobian -εκθετών Lyapunov από δίκτυα με RBFs

- Για δίκτυα RBF η j-οστή συνιστώσα της F δίνεται σύμφωνα με τη σχέση (2.41) ως:

$$F_j = \sum_{i=1}^{M} w_{ij} \cdot f_i \qquad f_i(X) = e^{\frac{|X-C_i|^2}{2\sigma^2}} \rightarrow \qquad \frac{\partial f_i}{\partial x_b} = \frac{-(x_b - c_{i_b}) \cdot e^{\frac{|X-C_i|^2}{2\sigma^2}}}{2\sigma^2} \tag{2.79}$$

  οπότε έχουμε

$$\frac{\partial F_a}{\partial x_b} = \sum_{i=1}^{M} w_{ia} \cdot \frac{-(x_b - c_{i_b}) \cdot e^{\frac{|X-C_i|^2}{2\sigma^2}}}{2\sigma^2} \tag{2.80}$$

- Για δίκτυα WCP το σήμα ενεργοποίησης f_i δίνεται, όπως είναι γνωστό, από τη σχέση

$$f_i(X) = \frac{e^{\frac{|X-C_i|^2}{2\sigma^2}}}{\sum_{j=1}^{M} e^{\frac{|X-C_i|^2}{2\sigma^2}}} \tag{2.81}$$

Είναι λοιπόν:

$$\frac{\partial f_i(X)}{\partial x_b} = \frac{\frac{\partial\{e^{\frac{|X-C_i|^2}{2\sigma^2}}\}}{\partial x_b}}{\sum_{j=1}^{M} e^{\frac{|X-C_i|^2}{2\sigma^2}}} - \frac{e^{\frac{|X-C_i|^2}{2\sigma^2}} \frac{\partial \sum_{j=1}^{M} e^{\frac{|X-C_i|^2}{2\sigma^2}}}{\partial x_b}}{\left\{\sum_{j=1}^{M} e^{\frac{|X-C_i|^2}{2\sigma^2}}\right\}^2} \tag{2.82}$$

Χρησιμοποιώντας και τη σχέση (2.45) καταλήγουμε στην ακόλουθη έκφραση για την μερική παράγωγο του σήματος εξόδου:

$$\frac{\partial F_a}{\partial x_b} = \sum_{i=1}^{M} w_{ia} \cdot \left\{ \frac{\frac{-(x_b - c_{i_b}) \cdot e^{\frac{|X-C_i|^2}{2\sigma^2}}}{2\sigma^2}}{\sum_{j=1}^{M} e^{\frac{|X-C_i|^2}{2\sigma^2}}} - \frac{e^{\frac{|X-C_i|^2}{2\sigma^2}} \sum_{j=1}^{M} \frac{-(x_b - c_{j_b}) \cdot e^{\frac{|X-C_j|^2}{2\sigma^2}}}{2\sigma^2}}{\left\{\sum_{j=1}^{M} e^{\frac{|X-C_i|^2}{2\sigma^2}}\right\}^2} \right\} \tag{2.83}$$

- Για τα δίκτυα wlp η έξοδος του δικτύου σχηματίζεται μπορεί να γραφεί ως (βλ. 2.46):

$$F_i = \sum_{j=1}^{N} c_{ij} f_j + \sum_{j=1}^{M} \sum_{k=1}^{N} l_{ijk} x_j f_k \tag{2.84}$$

όπου το σήμα f_i είναι όπως και στην περίπτωση του wcp. Θα έχουμε λοιπόν

$$\frac{\partial F_a}{\partial x_b} = \sum_{j=1}^{N} c_{aj} \frac{\partial f_j}{\partial x_b} + \sum_{j=1}^{M} \sum_{k=1}^{N} l_{ajk} x_j \frac{\partial f_k}{\partial x_b} + \sum_{k=1}^{N} l_{abk} f_k \tag{2.85}$$

Για τα δίκτυα SVM με gaussian πυρήνα η έκφραση για την Ιακωβιανή του συστήματος είναι ίδια με αυτήν των RBFs (2.79) ενώ μπορεί να τροποιηθεί πολύ εύκολα ώστε να καλύπτει και άλλους πυρήνες εκτός των gaussian.

2.4.5 Τοπική διάσταση συστήματος-Ψεύτικοι εκθέτες Lyapunov

Όπως εξηγήθηκε σε προηγούμενη παράγραφο, τις περισσότερες φορές η διάσταση εμβύθισης D_e της χρονοσειράς που παίρνουμε από ένα σύστημα είναι μεγαλύτερη από τη διάσταση στην οποία λαμβάνει χώρα η δυναμική του συστήματος. Η τελευταία διάσταση ονομάζεται *τοπική διάσταση* (local dimension) D_L ενώ η διάσταση D_e ονομάζεται καθολική διάσταση -global dimension. Η χρήση της τοπικής διάστασης D_L μπορεί να γίνει αφού έχουν εμβυθιστεί τα δεδομένα μας σε ένα χώρο φάσης διάστασης D_e, οπότε έχουμε εξασφαλίσει ότι ο ελκυστής του συστήματος έχει αποδιπλωθεί πλήρως, και μπορούμε να προβάλλουμε μικρότερες γειτονιές του ελκυστή στις D_L διαστάσεις στις οποίες υπάρχουν σε μεγαλύτερο βαθμό τα δεδομένα του ελκυστή χωρίς να έχουμε αναδίπλωση του ελκυστή, αφού δουλεύουμε με μικρές γειτονιές και όχι με το σύνολο του ελκυστή.

Από μαθηματικής σκοπίας είναι απολύτως θεμιτό να έχουμε διάσταση εμβύθισης μεγαλύτερη από την D_L και δεν πρέπει να οδηγεί σε εσφαλμένα αποτελέσματα. Πρακτικά όμως, προκύπτουν διάφορα προβλήματα όταν η διάσταση εμβύθισης είναι μεγαλύτερη από τη διάσταση στην οποία λαμβάνει χώρα η δυναμική του συτήματος: Όπως είδαμε και προηγουμένως, αν $D_e > D_L$ είναι δυσκολότερο να κατασκευάσουμε τοπικά μοντέλα για απεικονίσεις $D_e \times D_e$ απότι για απεικονίσεις $D_L \times D_L$. Χρειάζονται περισσότερες παράμετροι, αντιστροφές πινάκων μεγαλύτερης διάστασης, ενώ στις υπόλοιπες των πραγματικών D_L διαστάσεων κυριαρχεί ο θόρυβος, αφού τα δεδομένα θα

συγκεντρώνονται σε έναν ελκυστή διάστασης μικρότερης του D_L. Συνεπώς, το αποτέλεσμα της δημιουργίας μοντέλων σε D_e αντί για D_L διαστάσεις είναι να επιβαρύνεται υπολογιστικά η ανακατασκευή της δυναμικής του συστήματος, ενώ η ύπαρξη του θορύβου στις $D_e - D_L$ υπόλοιπες διαστάσεις , οδηγεί σε πίνακες με μεγάλο δείκτη κατάστασης, που συνεπάγεται ανακριβή μοντέλα της δυναμικής του συστήματος και χειρότερη πρόβλεψη.

Ένα ακόμα πρόβλημα που προκύπτει είναι με τους ψευδείς εκθέτες Lyapunov (spurious Lyapunov exponents): σύμφωνα με όσα αναφέρθηκαν πορηγουμένως, οι εκθέτες Lyapunov βρίσκονται τόσοι, όση είναι και η διάσταση εμβύθισης των δεδομένων. Αν, όμως, η διάσταση εμβύθισης των δεδομένων είναι μεγαλύτερη από την D_L είναι σίγουρο ότι οι $D_e - D_L$ από αυτούς θα είναι κατασκευάσματα της διαδικασίας εμβύθισης και όχι χαρακτηριστικές ποσότητες του συστήματος.

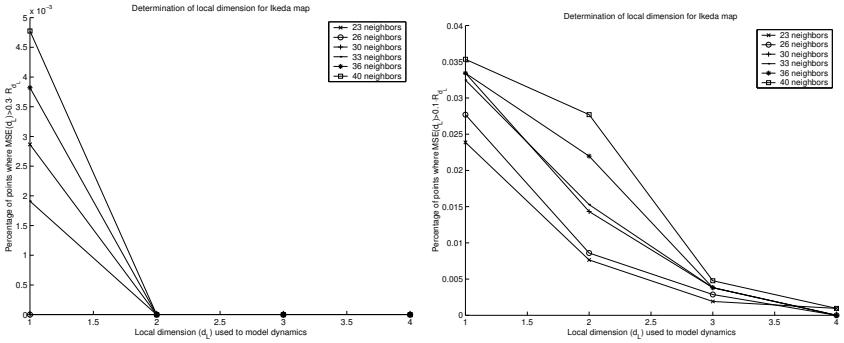
Προκύπτει λοιπόν το ερώτημα του αν και πώς είναι δυνατό να απομονώσουμε τους πραγματικούς εκθέτες Lyapunov από τους ψευδείς καθώς και με ποιον τρόπο γίνεται να προσδιοριστεί η διάσταση D_L. Για το σκοπό αυτό διάφορες μέθοδοι έχουν προταθεί ([28] [29] [1] [31]) από τις οποίες υλοποιήθηκαν οι δύο επικρατέστερες.

Η πρώτη μέθοδος εκμεταλλεύεται το γεγονός ότι μπορούμε να κάνουμε τοπικά μοντέλα από περιοχή σε περιοχή του ελκυστή σε D_L αντί για D_e διαστάσεις. Υπενθυμίζουμε ότι δεν μας απασχολεί το πρόβλημα της αναδίπλωσης του ελκυστή, αφού η επιλογή των γειτόνων των σημείων με βάση τους οποίους σχηματίζουμε τα τοπικά δεδομένα γίνεται στις D_e διαστάσεις που μας εξασφαλίζουν την αποδίπλωση του ελκυστή.

Αυτό το οποίο κάνουμε με τη μέθοδο αυτή είναι ότι δοκιμάζουμε για διαφορετικές τιμές της D_L να βρούμε $\mathbf{R}^{D_L} \to \mathbf{R}^{D_L}$ μοντέλα της δυναμικής του συστήματος, όπως αυτή παρατηρείται από γειτονιά σημείου σε γειτονιά σημείου. Με βάσή το γεγονός ότι η δυναμική λαμβάνει χώρα σε D_L διαστάσεις, αν εξετάσουμε την εξάρτηση του NMSE της πρόβλεψης από την D_L, πρέπει να παρατηρείται μία σημαντική διαφοροποίηση πριν και μετά τη D_L. Αυτό θα είναι και το κριτήριο που θα χρησιμοποιηθεί για την επιλογή της D_L

Για τη δημιουργία των τοπικών μοντέλων προβάλλονται τα δεδομένα στις D_L διαστάσεις όπου είναι πυκνότερα τα δεδομένα και εν συνεχεία ακολουθείται η γνωστή διαδικασία για τον προσδιορισμό των συντελεστών με τους οποίους ελαχιστοποιείται το σφάλμα πρόβλεψης. Η D_L επιλέγεται ως η διάσταση εκείνη για την οποία παρατηρείται για πρώτη φορά σφάλμα πρόβλεψης μικρότερο από ένα κατώφλι και για την οποία χάνεται η εξάρτηση από το πλήθος των γειτόνων που χρησιμοποιούνται για τη συγκρότηση της γειτονιάς. Το κατώφλι αυτό συνήθως ορίζεται ως ένα κλάσμα της ακτίνας του ελκυστή στις D_L διαστάσεις και θεωρούμε ότι αν το MSE της πρόβλεψης είναι μικρότερο από αυτό, η πρόβλεψη είναι σωστή. Στο σχήμα 2.17 βλέπουμε τα αποτελέσματα της μεθόδου αυτής για τα εμβυθισμένα σε τέσσερις διαστάσεις δεδομένα του συστήματος Ikeda. Είναι φανερό ότι η ποιότητα των αποτελεσμάτων της μεθόδου αυτής εξαρτάται σε μεγάλο βαθμό από το ποσοστό της ακτίνας του ελκυστή που χρησιμοποιείται για να καθορίσουμε το κατώφλι σφάλματος, ενώ βλέπουμε για παράδειγμα για ένα διαφορετικό ποσοστό πόσο διφορούμενα είναι τα αποτελέσματα. Εκτός αυτού, η μέθοδος αυτή απαιτεί την χρήση τοπικών μοντέλων, που είναι εξαιρετικά ασύμφορα, τόσο από άποψη απαιτούμενου χώρου όσο και χρόνου.

Υποτίθεται, επίσης, ότι βρίσκοντας τις D_L διευθύνσεις στις οποίες να προβάλλουμε τα δεδόμένα μας και δημιουργώντας τοπικά μοντέλα $\mathbf{R}^{D_L} \to \mathbf{R}^{D_L}$ μπορούμε να βρούμε με μεγαλύτερη ακρίβεια τους εκθέτες Lyapunov ενός συστήματος. Οι διαστάσεις αυτές είναι εκείνες στις οποίες συγκεντρώνονται τα περισσότερα δεδομένα. Η προβολή γίνεται τοπικά για κάθε δείγμα, όπου σχηματίζεται ο πίνακας αυτοσυσχέτισης των στοιχείων της



2.17: Καθορισμός τοπικής διάστασης για το συστήμα Ikeda με διαφορετικά κατώφλια

γειτονιάς του

$$Cov = \frac{1}{N_B}\sum_{i=1}^{N_B}(X(i)-\overline{X})(X(i)-\overline{X})^T \tag{2.86}$$

$$\text{όπου} \quad \overline{X} = \frac{1}{N_B}\sum_{i=1}^{N_B}X(i)$$

Οι κατάλληλες κατευθύνσεις για προβολή καθορίζονται από τα ιδιοδιανύσματα του πίνακα Cov ,$e_1, e_2, \ldots, e_{D_L}$, που έχουν τις D_L μεγαλύτερες ιδιοτιμές. Έτσι μπορούμε να κάνουμε τοπικές $\mathbf{R}^{D_L} \rightarrow \mathbf{R}^{D_L}$ απεικονίσεις χωρίς να εμπλέκονται πίνακες κακής κατάστασης στον υπολογισμό των παραμέτρων. Δοκιμάσαμε αυτή τη μέθοδο και τα αποτελέσματα που πήραμε είναι μάλλον αποθαρρυντικά, οπότε υποθέτουμε ότι η παραπάνω μέθοδος είναι κατάλληλη όταν έχουμε περισσότερα και καθαρότερα δεδομένα, σε αντίθεση με τα δεδομένα που χρησιμοποιούμε.

Η δεύτερη μέθοδος βασίζεται σε μία παρατήρηση των Eckmann και Ruelle , σύμφωνα με την οποία αν αντιστρέψουμε τη φορά με την οποία εξελίσσεται το σύστημα, παίρνουμε μία χρονοσειρά, ορισμένοι εκθέτες Lyapunov της οποίας είναι αντίθετοι από τους αντίστοιχούς τους της αρχικής χρονοσειράς.

Η ιδιότητα αυτή είναι χαρακτηριστική των πραγματικών εκθετών Lyapunov και εξηγείται αν σκεφτούμε ότι αντιστρέφοντας τη φορά του χρόνου η απόκλιση των τροχιών γίνεται σύγκλιση, οι δε προκύπτοντες ρυθμοί σύγκλισης είναι ίσοι κατά μέτρο με τους αρχικούς ρυθμούς απόκλισης και αντιστρόφως. Αντιθέτως, ένας ψευδής εκθέτης Lyapunov μπορεί να είναι μόνο αρνητικός: είτε η φορά των δεδομένων είναι η αρχική ή αντίστροφη, αυτός προέρχεται από τη διαδικασία εμβύθισης και η μόνη φυσική σημασία που ενδεχομένως να έχει είναι ότι μας δίνει τον εκθετικό ρυθμό με τον οποίο τροχιές που βρίσκονται εκτός του ελκυστή (στις υπόλοιπες $D_e - D_L$ διαστάσεις) συγκλίνουν σε αυτόν. Αυτό είναι και το κριτήριο για το διαχωρισμό των ψευδών εκθετών Lyapunov από τους πραγματικούς.

Χρησιμοποιήσαμε τη μέθοδο αυτή και πήραμε πολύ ενδιαφέροντα αποτελέσματα όπως θα φανεί παρακάτω, που μας βοήθησαν να καταλήξουμε σε ένα από όλα τα μοντέλα που έχουν προταθεί.

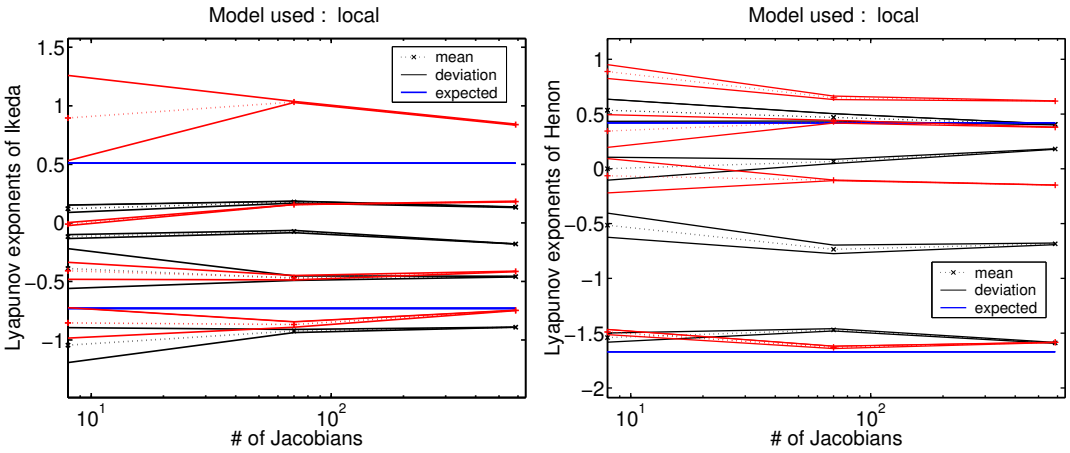
2.4.6 Σύγκριση μοντέλων με βάση τους εκθέτες Lyapunov

Οπως αναφέρθηκε και προηγουμένως, το MSE της πρόβλεψης δεν αρκεί ως μέτρο της εγκύτητας ενός μοντέλου με το αυθεντικό σύστημα το οποίο προσεγγίζει, ιδιαίτερα όταν

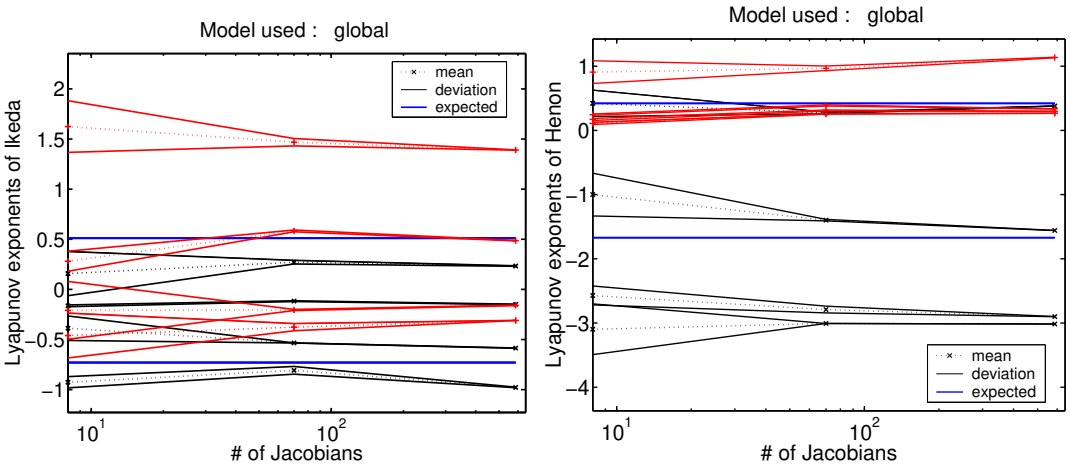
το MSE υπολογίζεται πάνω στο σύνολο εκπαίδευσης. Για την περίπτωση των χαοτικών σημάτων μία ποσότητα που μας δείχνει πόσο καλά έχει μάθει τη δυναμική του συστήματος ένα μοντέλο είναι ή ακρίβεια με την οποία υπολογίζει τους εκθέτες Lyapunov ενός συστήματος.

Στα παρακάτω σχήματα βλέπουμε τις επιδόσεις των μεθόδων στον υπολογισμό των εκθετών Lyapunov των test σημάτων μας όταν αυτά έχουνε εμβυθιστεί σε 4 διαστάσεις. Οι εκθέτες Lyapunov υπολογίστηκαν ξεκινώντας από διαφορετικά σημεία του ελκυστή κάθε σήματος και στο διάγραμμα φαίνεται η μέση τιμή και η διασπορά τους. Οι σωστές τιμές, $0.42, -1.62$ για το σύστημα Henon και $0.51, -0.73$ για το σύστημα Ikeda αντιστοιχούν στις ευθείες γραμμές ενώ στον άξονα των x είναι ο αριθμός των Jacobian που χρησιμοποιήθηκαν για τον υπολογισμό των εκθετών Lyapunov.

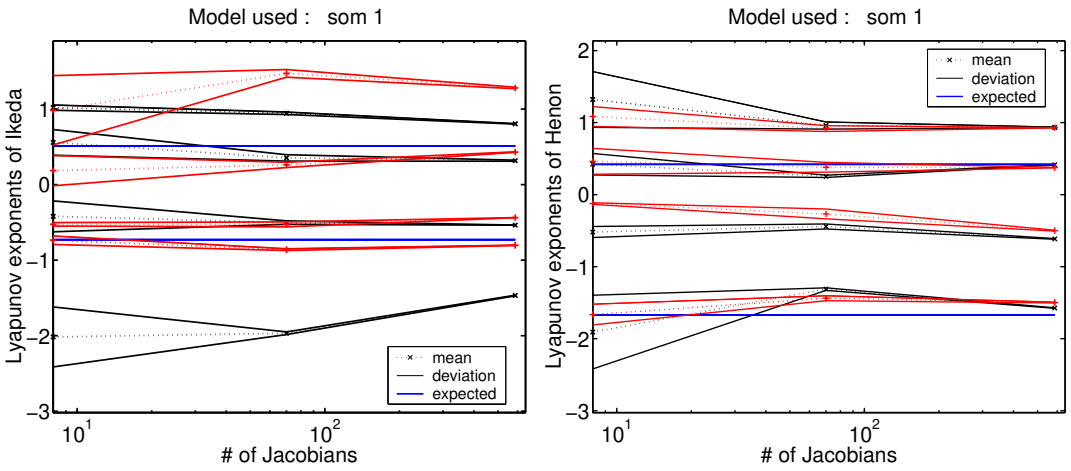
Αν και η δυναμική του συστήματος Henon είναι ιδιαίτερα απλή, (για το λόγο αυτό δεν χρησιμοποιήθηκε για να εξεταστεί η δυνατότητα πρόβλεψης) έχει ενδιαφέρον να δούμε κατά πόσο τα δίκτυα που χρησιμοποιούν ακτινικές συναρτήσεις βάσης είναι σε θέση να δώσουν σωστά αποτελέσματα, αφού είναι δύσκολο για αυτά να ανακατασκευάσουν την επίπεδη δυναμική της δεύτερης συνιστώσας του συστήματος.



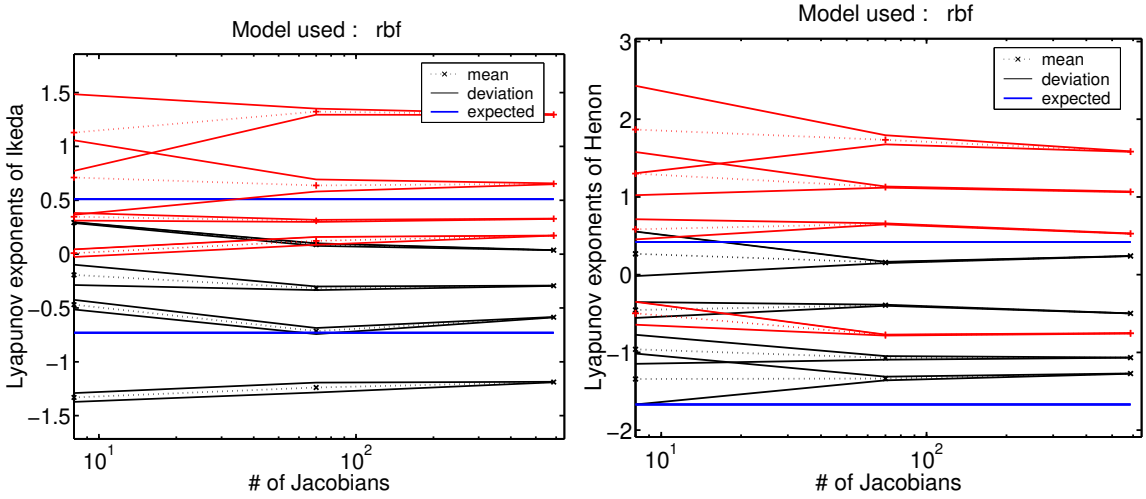
2.18: Υπολογισμός εκθετών Lyapunov με χρήση τοπικών πολυωνύμων



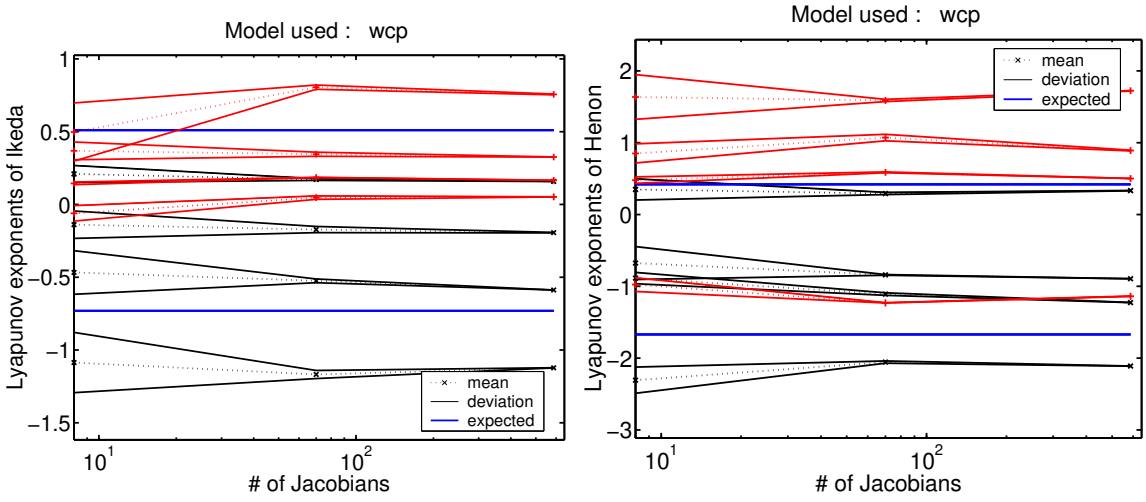
2.19: Υπολογισμός εκθετών Lyapunov με χρήση καθολικών πολυωνύμων



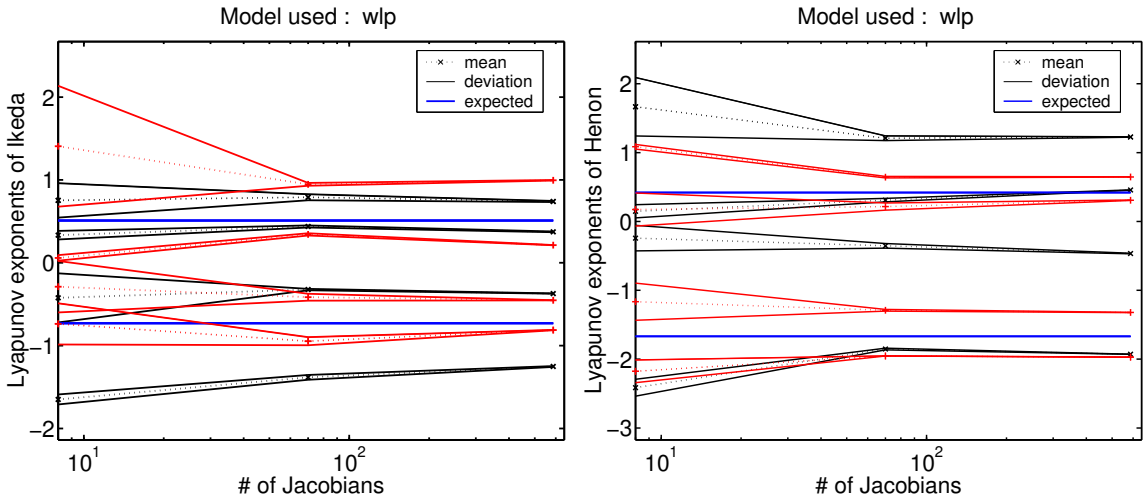
2.20: Υπολογισμός εκθετών Lyapunov με χρήση τοπικών πολυωνύμων σε συστάδες του ελκυστή



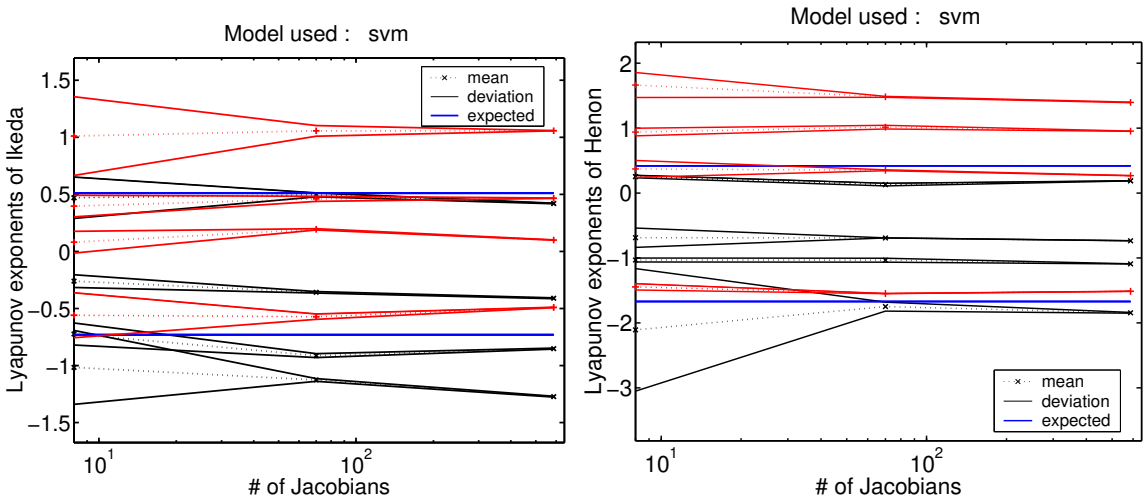
2.21: Υπολογισμός εκθετών Lyapunov με χρήση δικτύων RBF



2.22: Υπολογισμός εκθετών Lyapunov με χρήση δικτύων WCP



2.23: Υπολογισμός εκθετών Lyapunov με χρήση δικτύων WLP



2.24: Υπολογισμός εκθετών Lyapunov με χρήση δικτύων SVM

Σχολιασμός αποτελεσμάτων

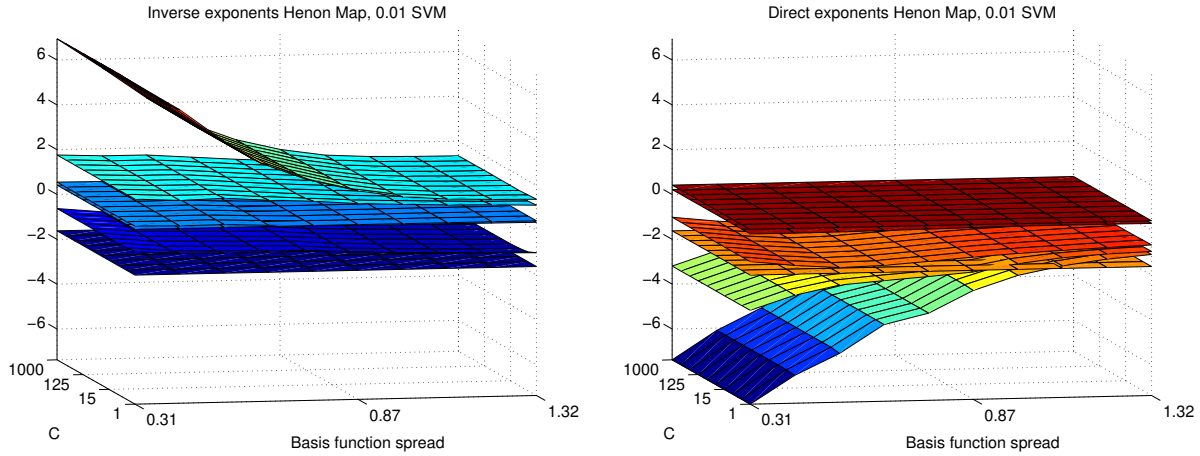
Οι προβλέπτες που απορρίπτονται απευθείας είναι

- Τα τοπικά πολυώνυμα, γιατί αποτυγχάνουν να βρουν τον θετικό εκθέτη του συστήματος Ikeda.
- Τα καθολικά πολυώνυμα γιατί δεν καταφέρνουν να βρουν τους εκθέτες του συστήματος Ikeda και τον θετικό εκθέτη του συστήματος Henon.
- Τα δίκτυα RBF και WCP γιατί δεν καταφέρνουν να βρουν τους εκθέτες του συστήματος Ikeda

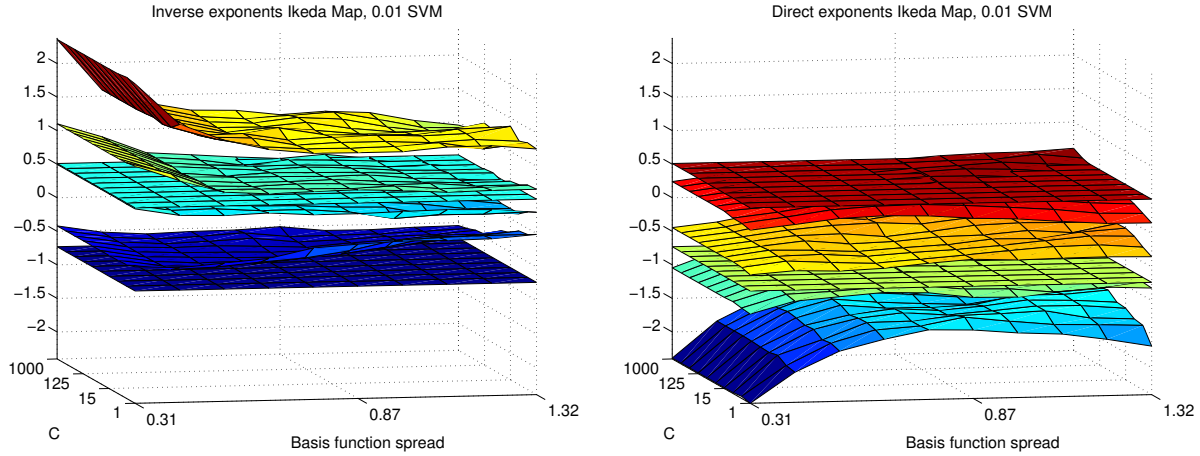
Το δίκτυο SLH έχει παραπλήσιες επιδόσεις με αυτές του WLP οπότε από τα δύο αυτά δίκτυα προτιμάται το δίκτυο WLP καθώς έχει την επιθυμητή ιδιότητα της συνέχειας που είναι ιδιαίτερα σημαντική για τον υπολογισμό των εκθετών Lyapunov. Απομένει λοιπόν να συγκρίνουμε το δίκτυο WLP με το δίκτυο SVM αφού και τα δύο έχουν καταφέρει να

υπλογίσουν επιτυχώς τους δύο σωστούς εκθέτες Lyapunov και έχουν απορρίψει του ψεύτικους. Εδώ χρειάζεται να κοιτάξουμε κάπως πιο προσεκτικά τα σχήματα για τα δύο αυτά μοντέλα έχοντας κατά νου ότι οι ψεύτικοι εκθέτες πρέπει να παίρνουν αρνητικές τιμές μόνο, οπότε και οι ανεστραμμένοι ψεύτικοι θα παίρνουν θετικές τιμές. Βλέπουμε όμως ότι αυτό ισχύει μόνο για το δίκτυο SVM αφού το δίκτυο WLP δίνει και στις δύο περιπτώσεις θετικούς ψεύτικους εκθέτες. Έτσι, αν δεν χρησιμοποιήσουμε ένα αρκετά μικρό κατώφλι για να αποφασίσουμε αν έχουμε ένα αντεστραμμένο εκθέτη κοντά στον ευθύ μπορούμε να καταλήξουμε σε εσφαλμένα αποτελέσματα (π.χ. για το σύστημα Ikeda αν δεν γνωρίζαμε εκ των προτέρων τον αριθμό των εκθετών του θα μπορούσαμε να πούμε ότι έχει 2 θετικούς εκθέτες, ενώ γνωρίζουμε ότι έχει μόνο έναν). Αξίζει να σημειωθεί ότι το δίκτυο WLP έχει πολύ μικρότερο σφάλμα πρόβλεψης για το σύστημα Ikeda από το αντίστοιχο SVM, που αποδεικνύει αυτό που είχαμε αναφέρει και εξαρχής ότι ένα μικρό MSE δεν σημαίνει ότι έχουμε μάθει τη δυναμική του συστήματος.

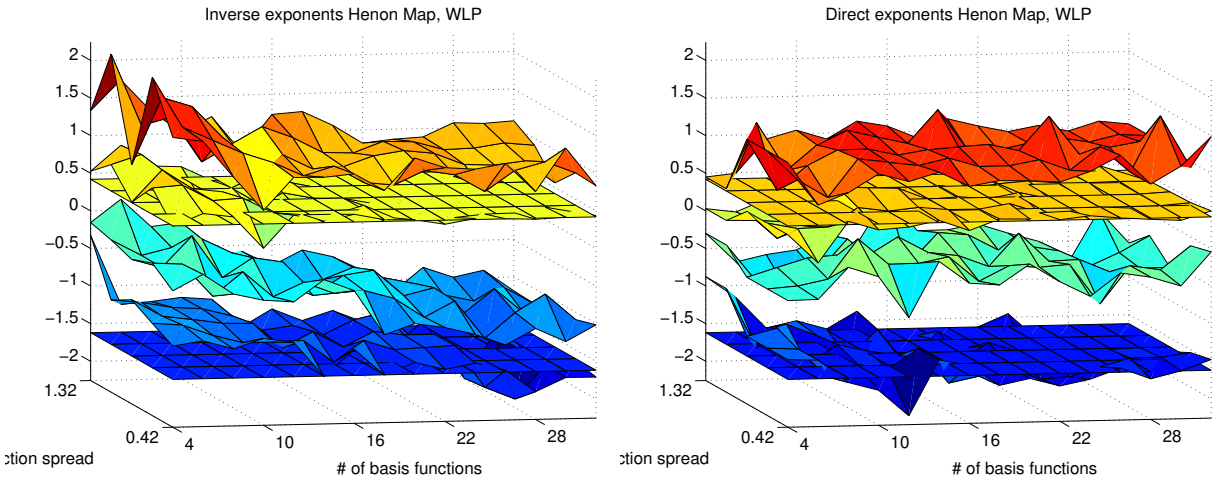
Για να σιγουρευτούμε ότι η παραπάνω παρατήρησή δεν οφείλεται στο συγκεκριμένο δίκτυο WLP/SVM που διαλέξαμε, δοκιμάζοντας δίκτυα διαφορετικής τάξης και διασποράς, πήραμε τις παρακάτω εικόνες:



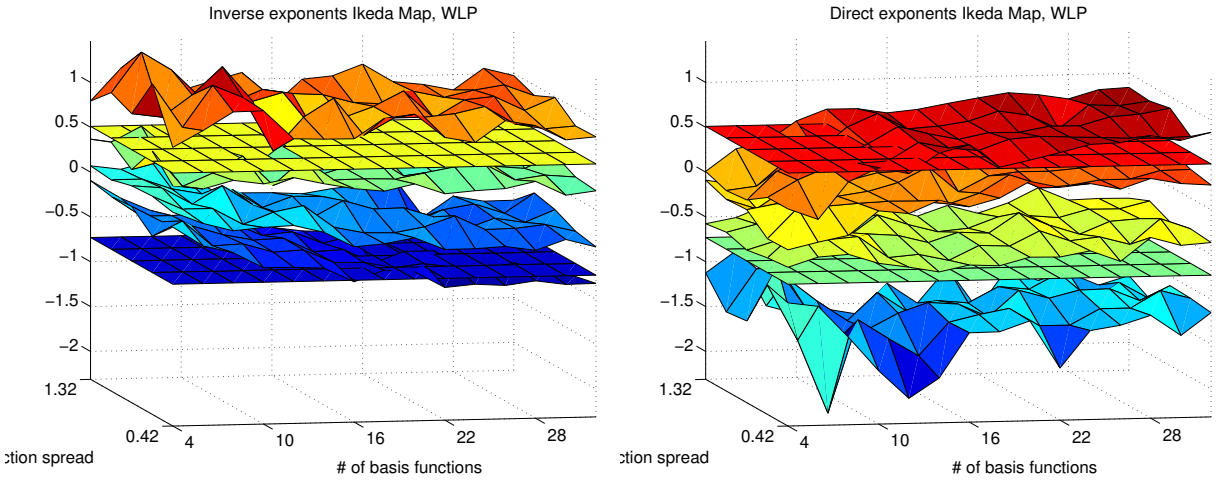
2.25: Υπολογισμός εκθετών Lyapunov συστήματος Henon από ένα δίκτυο SVM



2.26: Υπολογισμός εκθετών Lyapunov συστήματος Ikeda από ένα δίκτυο SVM



2.27: Υπολογισμός εκθετών Lyapunov συστήματος Henon από ένα δίκτυο WLP



2.28: Υπολογισμός εκθετών Lyapunov συστήματος Ikeda από ένα δίκτυο WLP

Από τα σχήματα αυτά πειστήκαμε απόλυτα για τη χρήση ενός SVM για το σκοπό του υπολογισμού των εκθετών Lyapunov, καθώς τα αποτελέσματά του συνδυάζουν ακρίβεια, σχετική ανεξαρτησία από κάποια συγκεκριμένη επιλογή παραμέτρων και επιπλέον σε καμία περίπτωση δεν έχουμε έναν θετικό ψεύτικό εκθέτη, σε αντίθεση με το WLP. Επίσης εκπαιδεύοντας το δίκτυο SVM και το WLP με δείγματα μολυσμένα με θόρυβο πήραμε σχεδόν τα ίδια αποτελέσματα από το SVM ενώ από το WLP πήραμε πολύ χειρότερα. Τέλος, αναφέρεται ότι τα δίκτυα RBF,WCP, SLH δώσανε χειρότερα αποτελέσματα από το WLP.

Κεφάλαιο 3

Εφαρμογή Μεθόδων Ανάλυσης Χαοτικών Χρονοσειρών στην Πρόβλεψη/Ανάλυση Σήματος Φωνής.

Όπως αναφέρθηκε και στο 1^o κεφάλαιο, από απότελέσματα σύγχρονης έρευνας έχει προκύψει ότι το σύστημα παραγωγής φωνής παρουσιάζει έντονη μη-γραμμική συμπεριφορά για ορισμένους ήχους. Προκύπτει λοιπόν το ερώτημα κατά πόσον οι μέθοδοι που αναφέρθηκαν στα προηγούμενα κεφάλαια, οι οποίες έχουν εφαρμοστεί με επιτυχία σε άλλα φυσικά μεγέθη όπως οικονομικές,φυσικές χρονοσειρές κ.α. θα μπορούσαν να χρησιμοποιηθούν για την πρόβλεψη και τη ανάλυση του σήματος φωνής.

Οι 3 βασικές κατηγορίες ήχων με τις οποίες ασχοληθήκαμε είναι :

Φωνήεντα	
Φωνητικός κωδικός	**Χαρακτηριστική Λέξη**
iy	b*ee*t
ih	b*i*t
eh	b*e*t
ey	b*ai*t
ae	b*a*t
aa	b*o*tt
aw	b*ou*t
ay	b*i*te
ah	b*u*t
ao	b*ou*ght
oy	b*oy*
ow	b*oa*t
uh	b*oo*k
ux	t*oo*t
er	b*ir*d
ax	ab*ou*t
ix	deb*i*t

Πίνακας 3.1: Φωνητικοί κωδικοί φωνηέντων

Τυρβώδεις Ήχοι	
Φωνητικός κωδικός	**Χαρακτηριστική Λέξη**
s	***s***ea
sh	***sh***e
z	***z***one
f	***f***in
v	***v***an
dh	***th***en

Πίνακας 3.2: Φωνητικοί κωδικοί τυρβωδών ήχων

Κρουστικοί Ήχοι	
Φωνητικός κωδικός	**Χαρακτηριστική Λέξη**
b	***b***ee
d	***d***ay
p	***p***ea
t	***t***ea
k	***k***ey
dx	mu***dd***y,dir***t***y
q	ba***t***

Πίνακας 3.3: Φωνητικοί κωδικοί κρουστικών ήχων

Από τις τρεις κατηγορίες αυτές τα ευκολότερα στην πρόβλεψη και αναγνώριση φωνήματα είναι τα φωνήεντα, στα οποία λειτρουργεί σχετικά ικανοποιητικά και το γραμμικό μοντέλο, και τα δύσκολότερα είναι τα φωνήματα των κρουστικών ήχων, για τα οποία μπορούμε να πούμε ότι θέτουν κατά κάποιο τρόπο το 'άνω όριο' της δυσκολίας. Οι τυρβώδεις ήχοι βρίσκονται κάπου ενδιάμεσα στο βαθμό δυσκολίας τους.

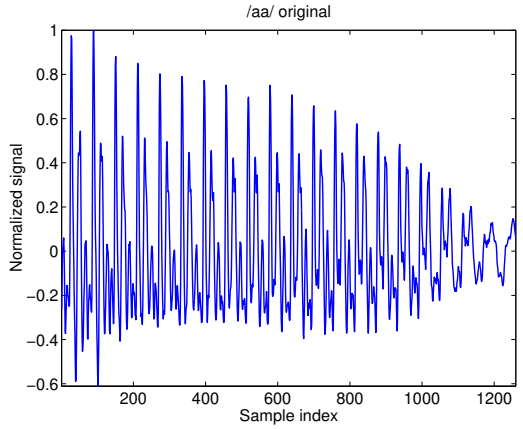
Στη συνέχεια δείχνουμε τα αποτελέσματα της πρόβλεψης και του υπολογισμου των εκθετών Lyapunov για μία χρονοσειρά από κάθε φώνημα, καθώς τα αποτελέσματα είναι ποιοτικά όμοια για διαφορετικούς ομιλητές. Στο επόμενο μέρος προσπαθούμε να αναλύσουμε τα αποτελέσματα που πηραμε και να καταλήξουμε σε κάποια συμπεράσματα.

Στα διαγράμματα των ανακατασκευασμένων ελκυστών που ακολουθούν η εξέλιξη του συστήματος φαίνεται με γεωμετρικό τρόπο μόνο στις τρείς πρώτες διαστάσεις εμβύθισης ενώ πληροφορία για τις υπόλοιπες διαστάσεις του ελκυστή έχει κωδικοποιηθεί στο χρώμα του ελκυστή (όπου αυτό είναι δυνατό).

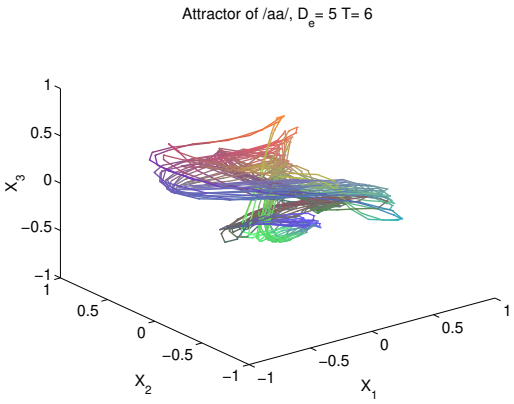
Οι εκθέτες Lyapunov που φαίνονται είναι αυτοί που παίρνουμε χρησιμοποιώντας ένα δίκτυο SVM και ένα σύστημα TSK-1 εκπαιδεύοντάς το με τους αλγορίθμους που παρουσιάζονται στο παράρτημα.

3.1 Φωνήεντα

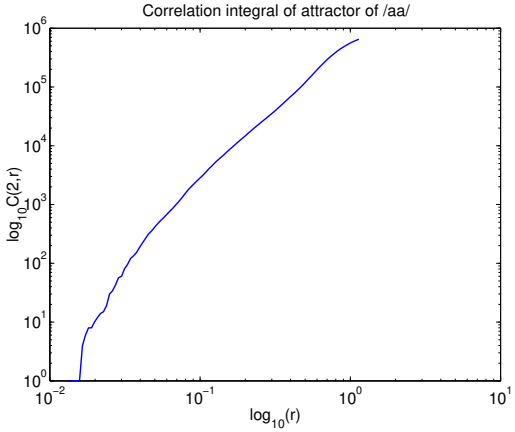
Φώνημα: /aa/



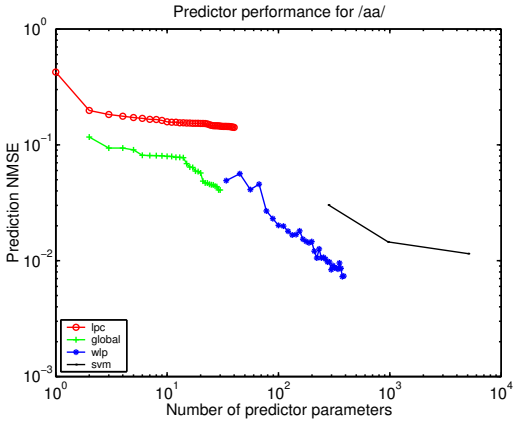
Αρχικό σήμα



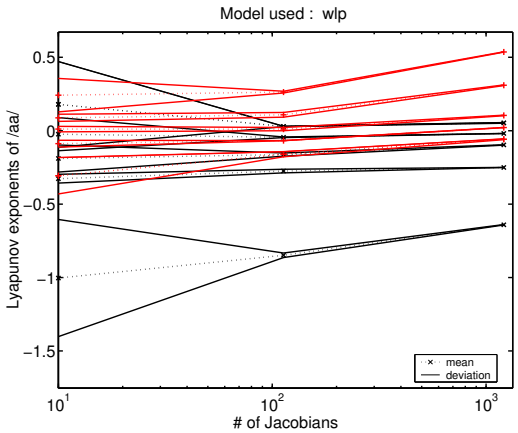
Ανακατασκευασμένος ελκυστής

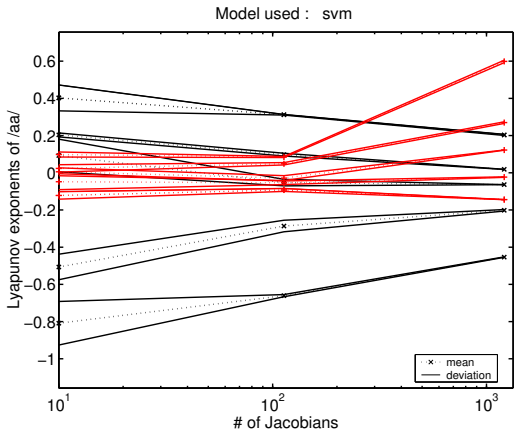


Ολοκλήρωμα συσχέτισης



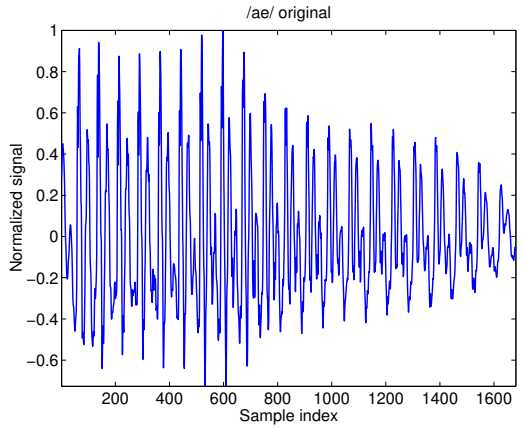
NMSE πρόβλεψης για την επόμενη χρονική στιγμή



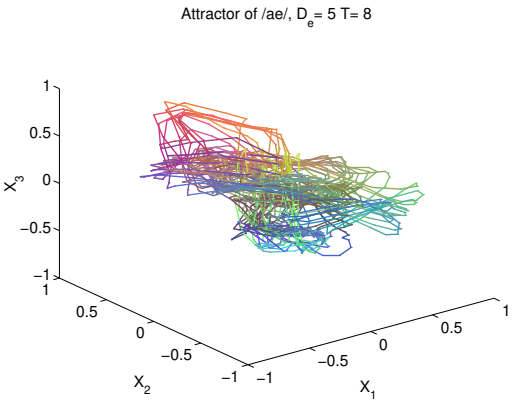


Εκθέτες Lyapunov: (a) από ένα WLP (b) από ένα SVM

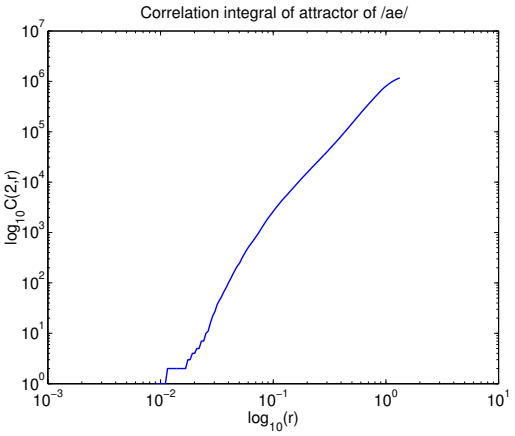
Φώνημα: /ae/



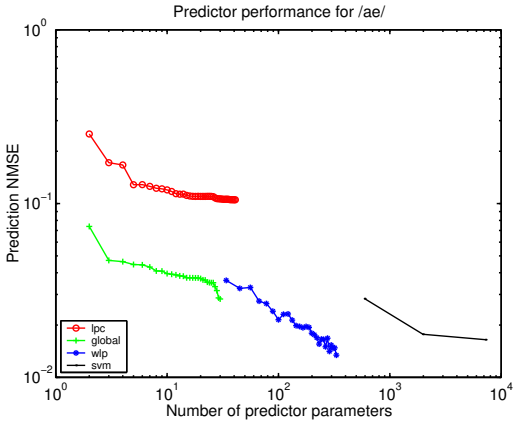
Αρχικό σήμα



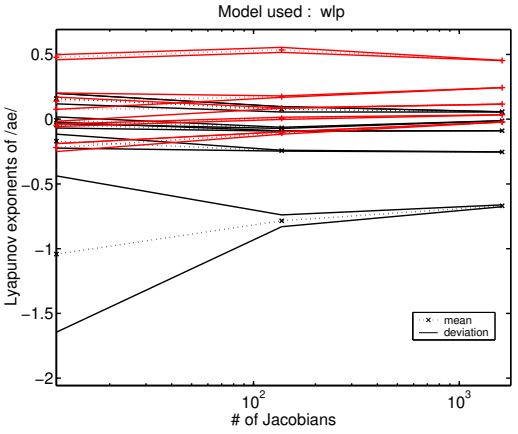
Ανακατασκευασμένος ελκυστής

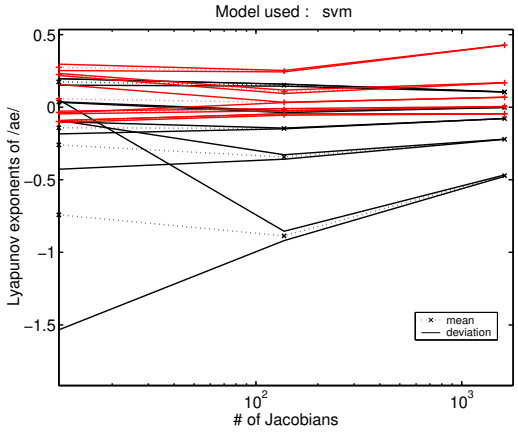


Ολοκλήρωμα συσχέτισης



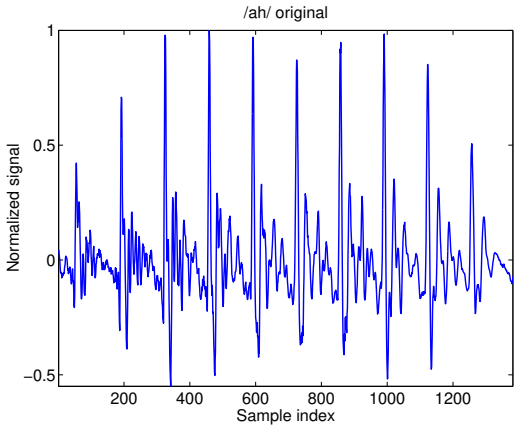
NMSE πρόβλεψης για την επόμενη χρονική στιγμή



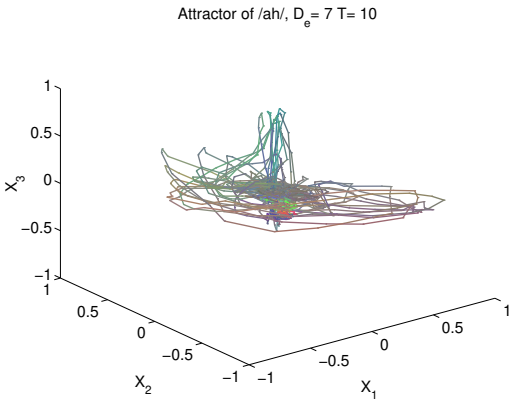


Εκθέτες Lyapunov: (a) από ένα WLP (b) από ένα SVM

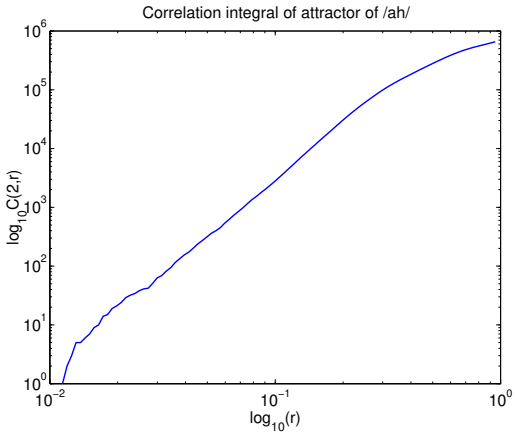
Φώνημα: /ah/



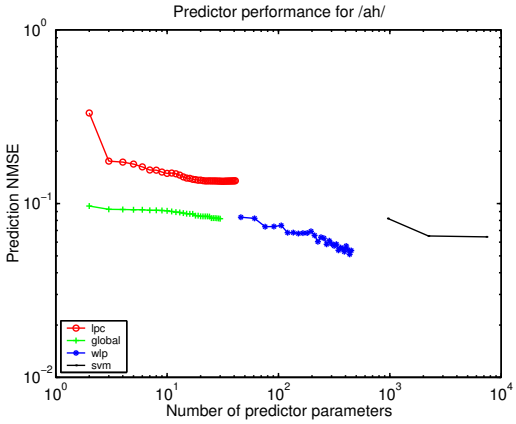
Αρχικό σήμα



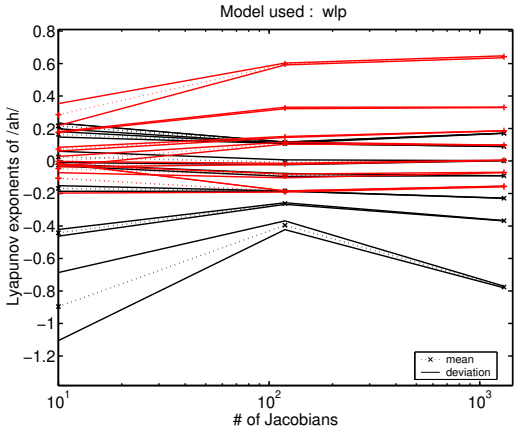
Ανακατασκευασμένος ελκυστής

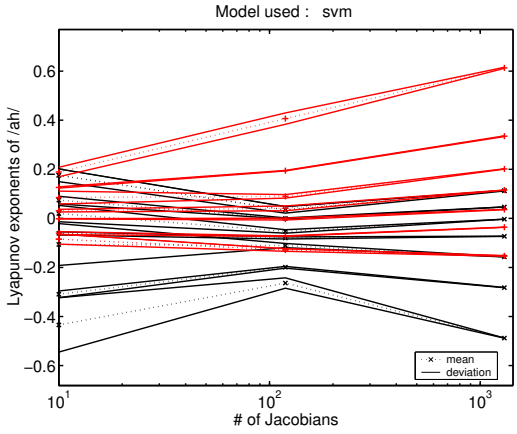


Ολοκλήρωμα συσχέτισης



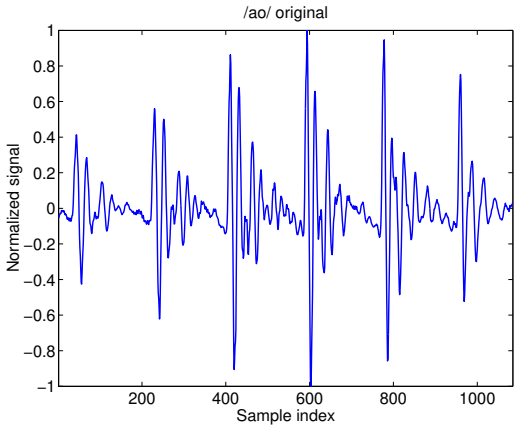
NMSE πρόβλεψης για την επόμενη χρονική στιγμή



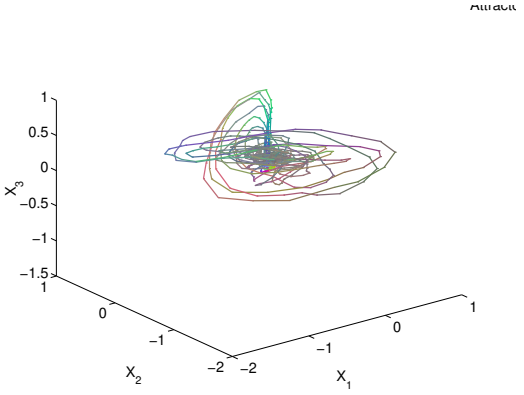


Εκθέτες Lyapunov: (a) από ένα WLP (b) από ένα SVM

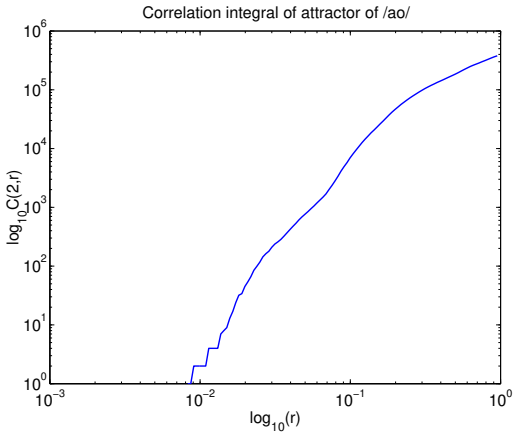
Φώνημα: /ao/



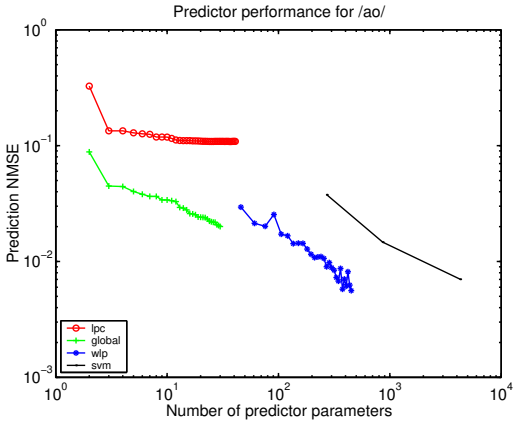
Αρχικό σήμα



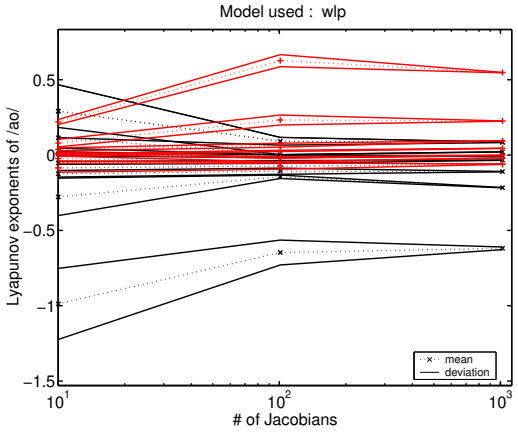
Ανακατασκευασμένος ελκυστής

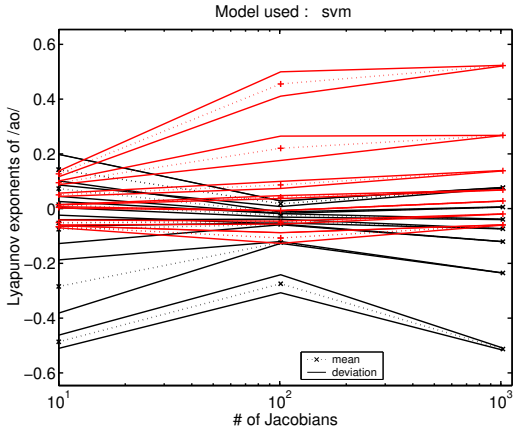


Ολοκλήρωμα συσχέτισης



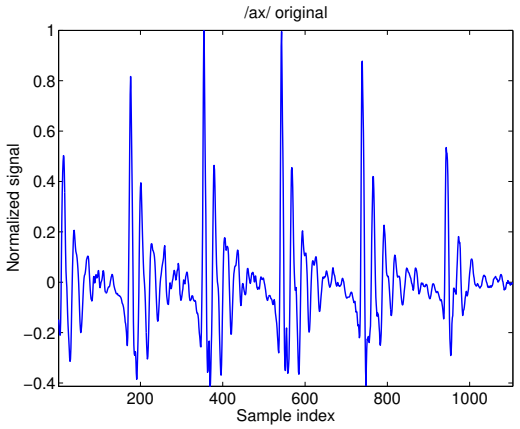
NMSE πρόβλεψης για την επόμενη χρονική στιγμή



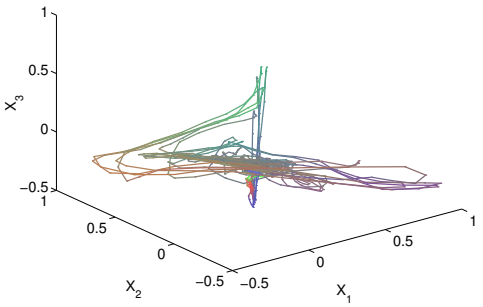


Εκθέτες Lyapunov: (a) από ένα WLP (b) από ένα SVM

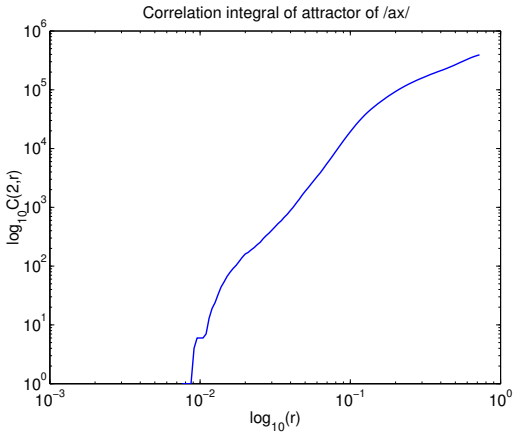
Φώνημα: /ax/



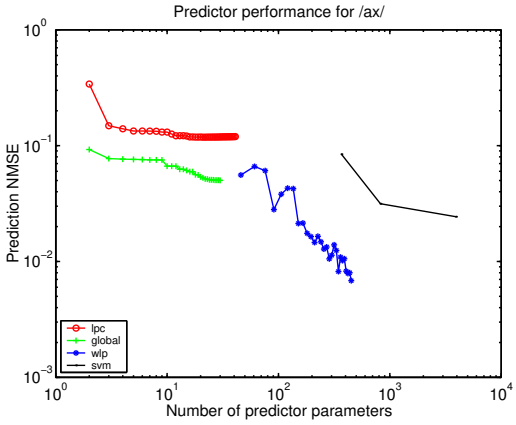
Αρχικό σήμα



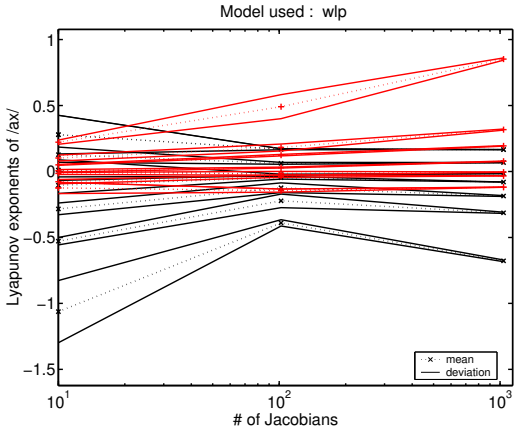
Ανακατασκευασμένος ελκυστής

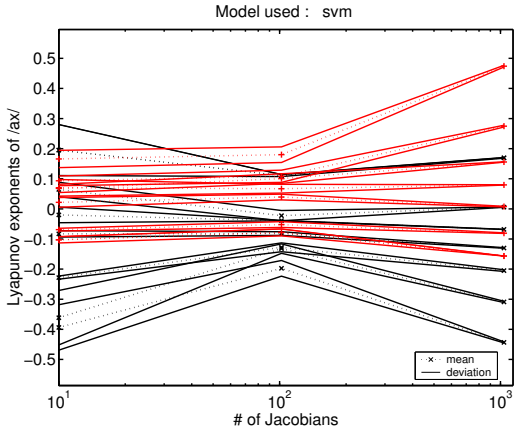


Ολοκλήρωμα συσχέτισης



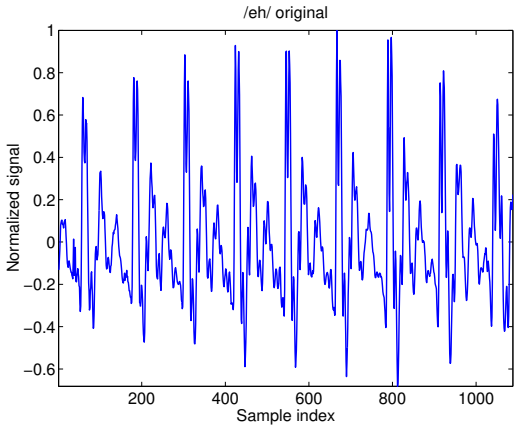
NMSE πρόβλεψης για την επόμενη χρονική στιγμή



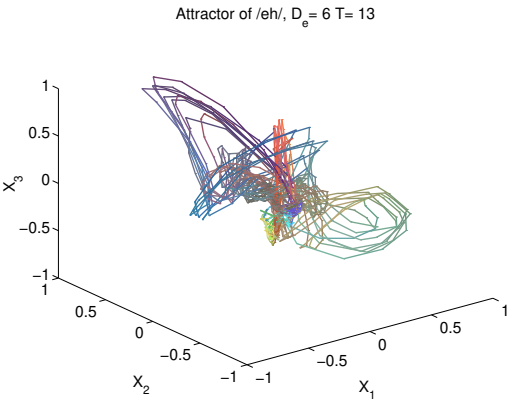


Εκθέτες Lyapunov: (a) από ένα WLP (b) από ένα SVM

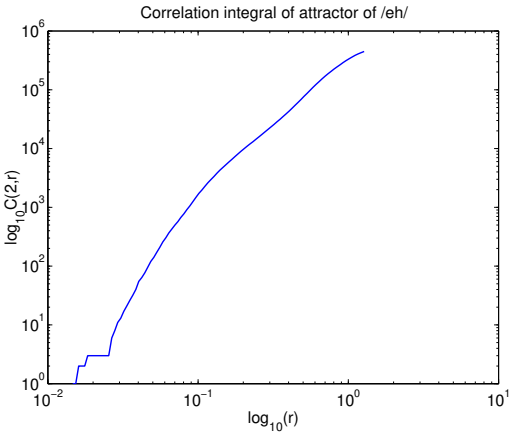
Φώνημα: /eh/



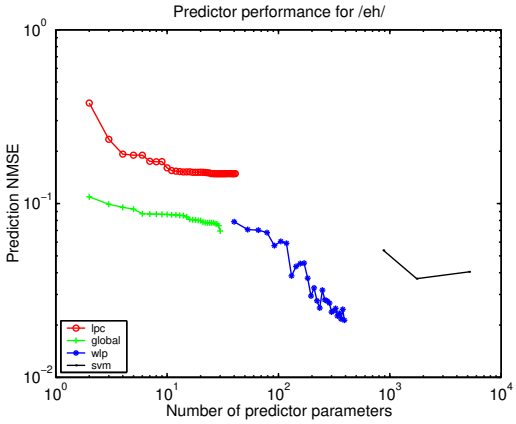
Αρχικό σήμα



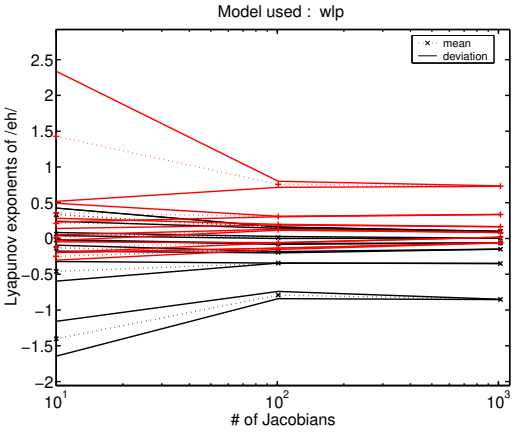
Ανακατασκευασμένος ελκυστής

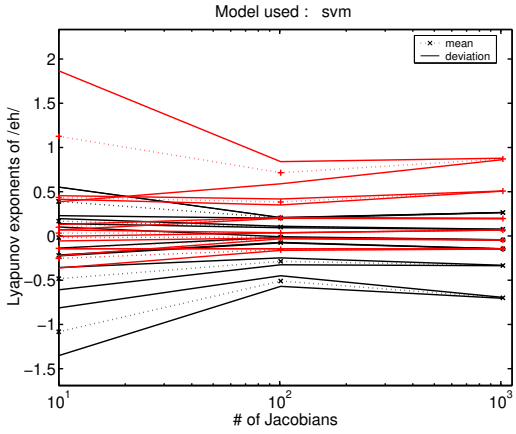


Ολοκλήρωμα συσχέτισης



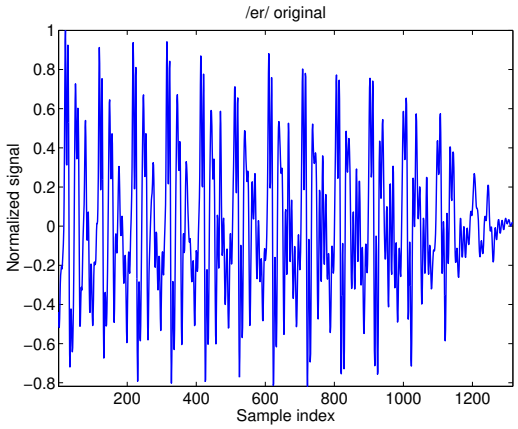
NMSE πρόβλεψης για την επόμενη χρονική στιγμή



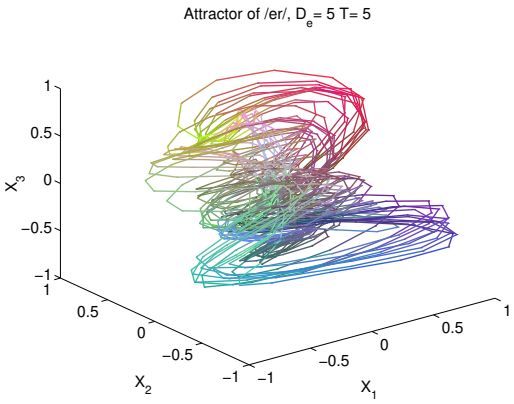


Εκθέτες Lyapunov: (a) από ένα WLP (b) από ένα SVM

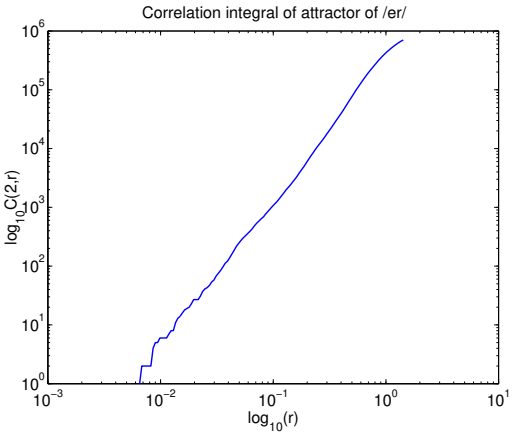
Φώνημα: /er/



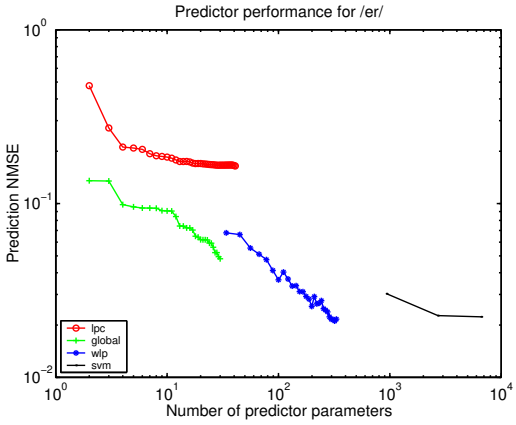
Αρχικό σήμα



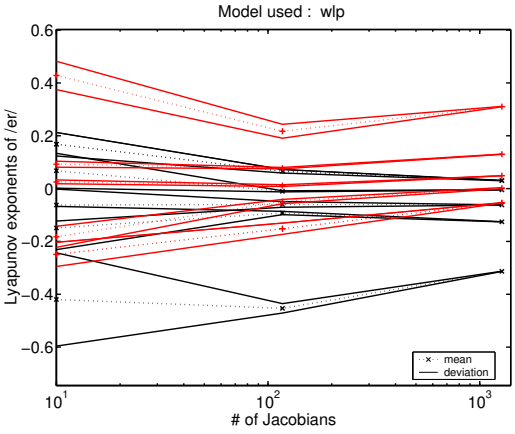
Ανακατασκευασμένος ελκυστής

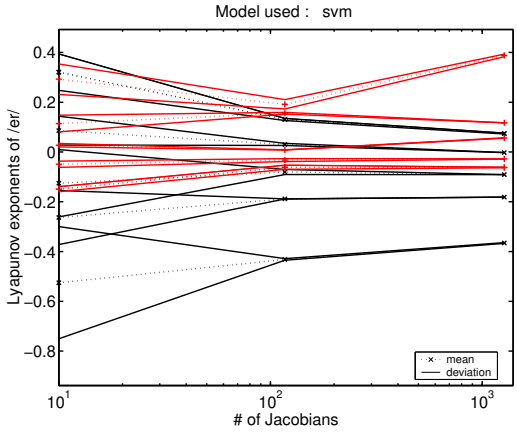


Ολοκλήρωμα συσχέτισης



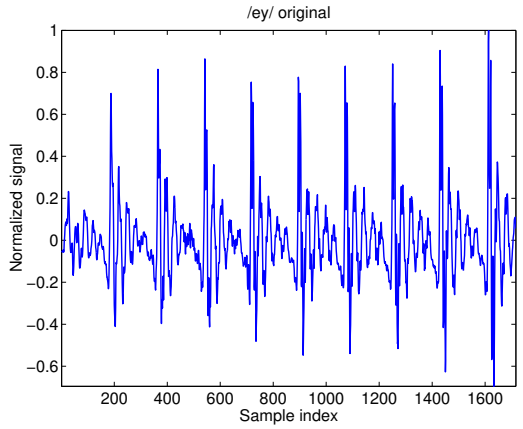
NMSE πρόβλεψης για την επόμενη χρονική στιγμή



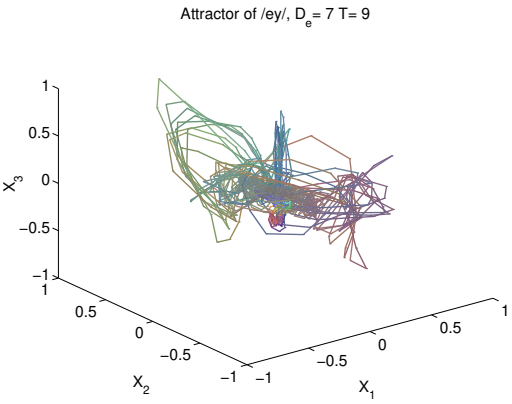


Εκθέτες Lyapunov: (a) από ένα WLP (b) από ένα SVM

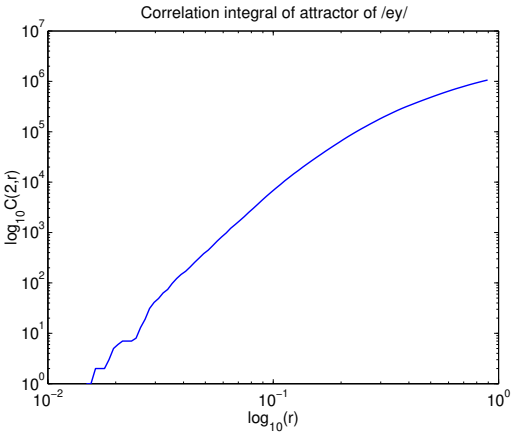
Φώνημα: /ey/



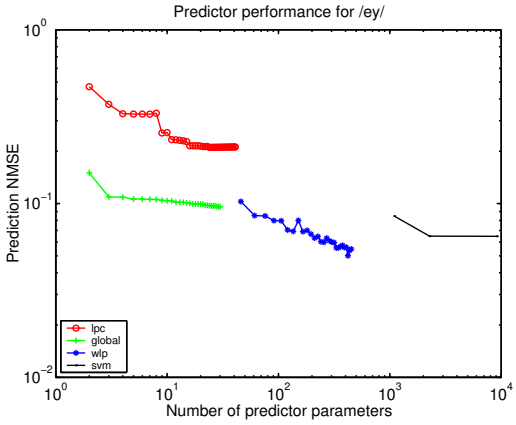
Αρχικό σήμα



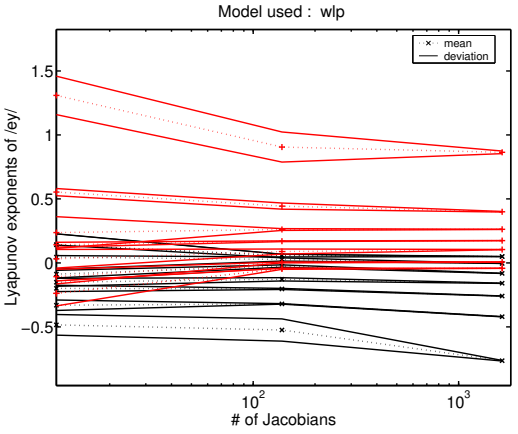
Ανακατασκευασμένος ελκυστής

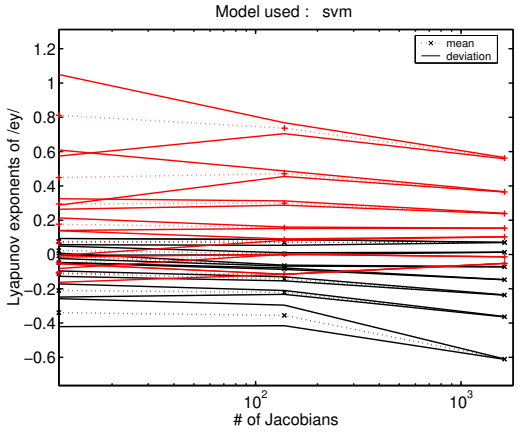


Ολοκλήρωμα συσχέτισης



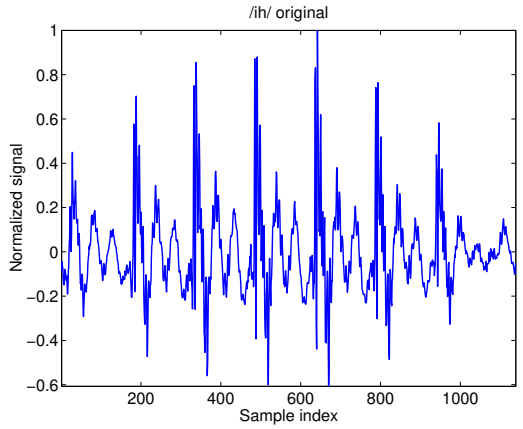
NMSE πρόβλεψης για την επόμενη χρονική στιγμή



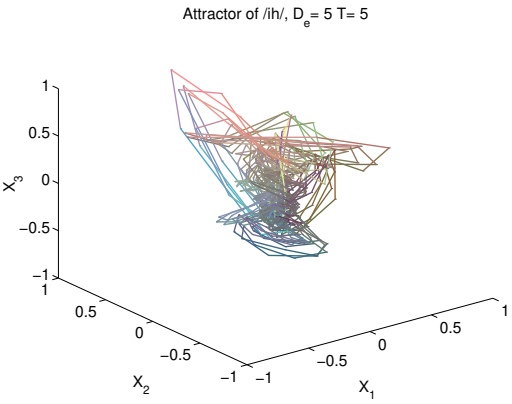


Εκθέτες Lyapunov: (a) από ένα WLP (b) από ένα SVM

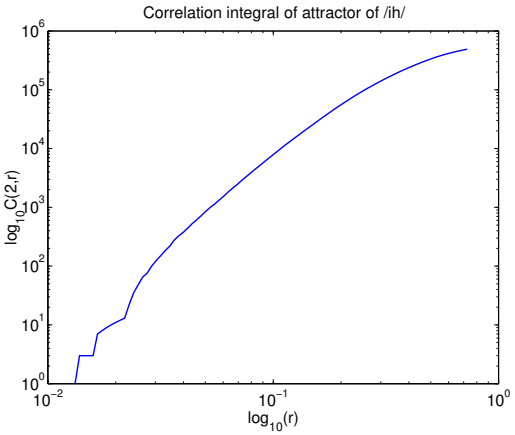
Φώνημα: /ih/



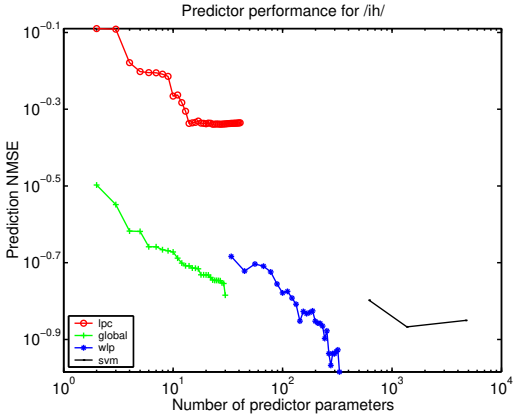
Αρχικό σήμα



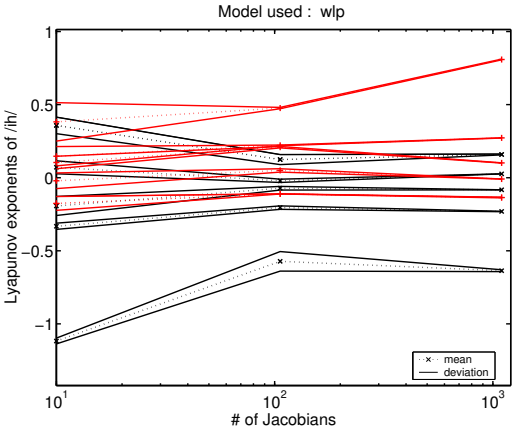
Ανακατασκευασμένος ελκυστής

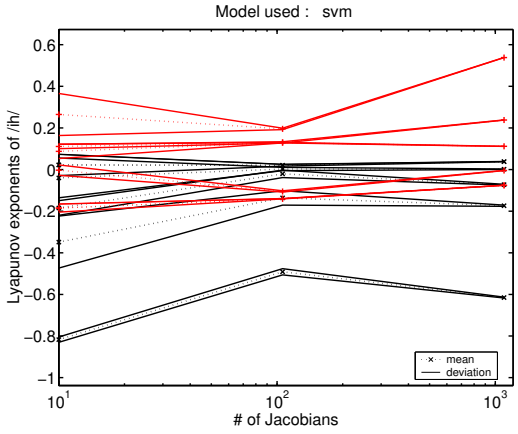


Ολοκλήρωμα συσχέτισης



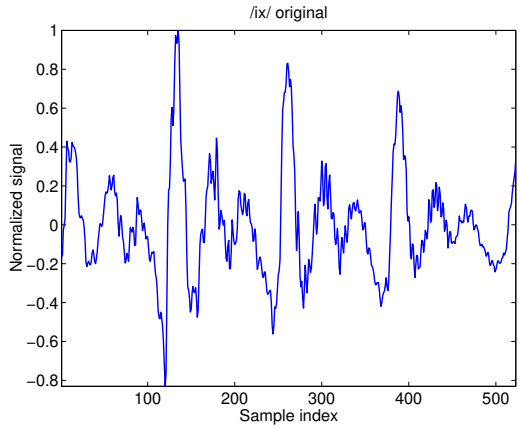
NMSE πρόβλεψης για την επόμενη χρονική στιγμή



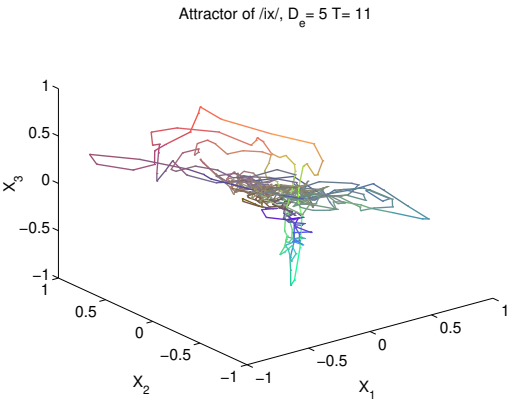


Εκθέτες Lyapunov: (a) από ένα WLP (b) από ένα SVM

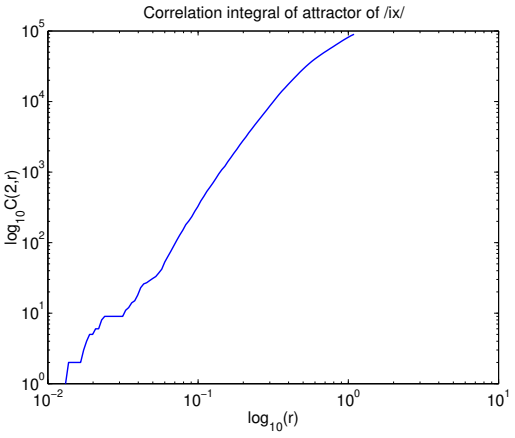
Φώνημα: /ix/



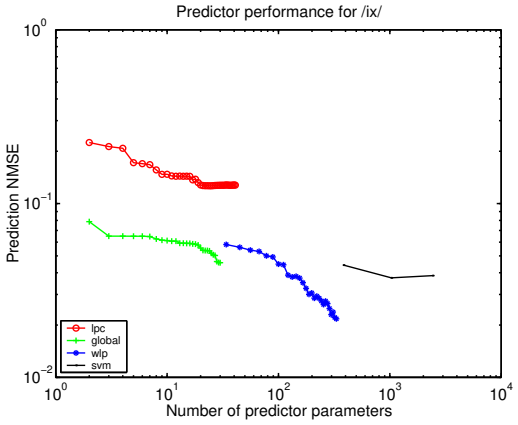
Αρχικό σήμα



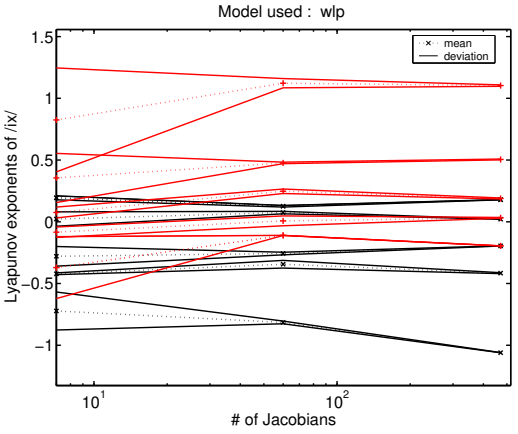
Ανακατασκευασμένος ελκυστής

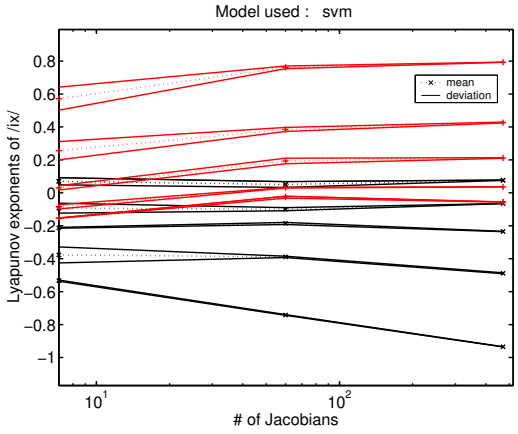


Ολοκλήρωμα συσχέτισης



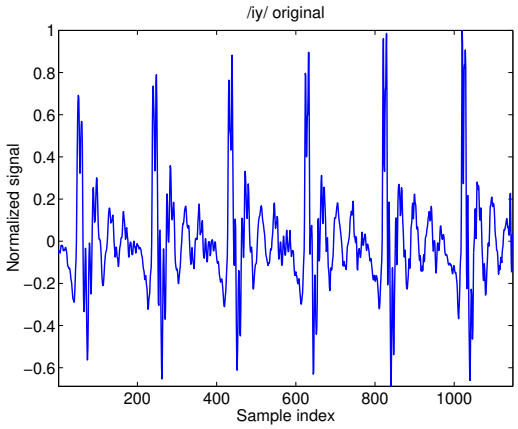
NMSE πρόβλεψης για την επόμενη χρονική στιγμή



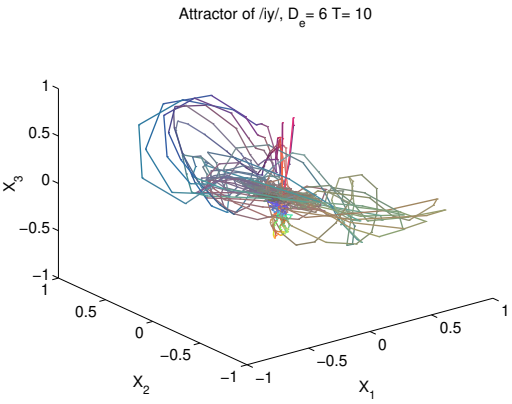


Εκθέτες Lyapunov: (a) από ένα WLP (b) από ένα SVM

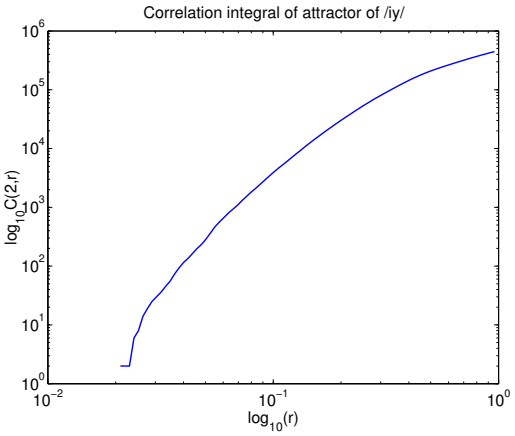
Φώνημα: /iy/



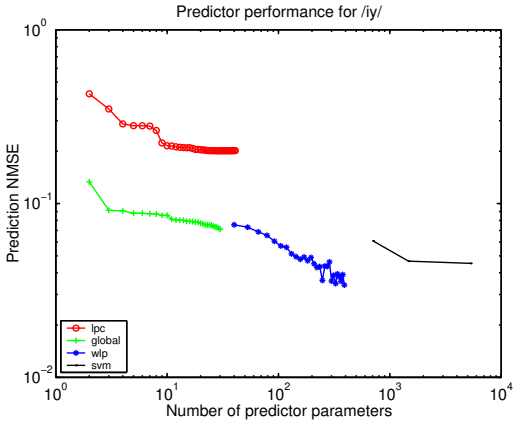
Αρχικό σήμα



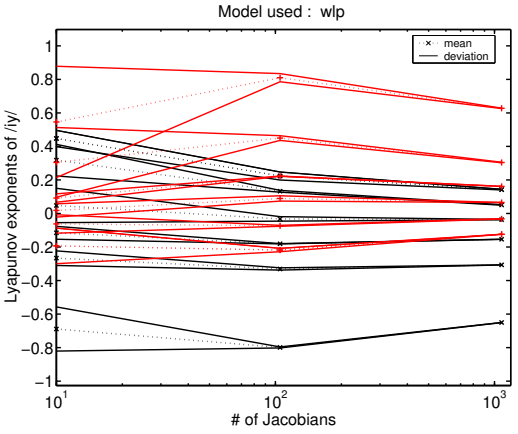
Ανακατασκευασμένος ελκυστής

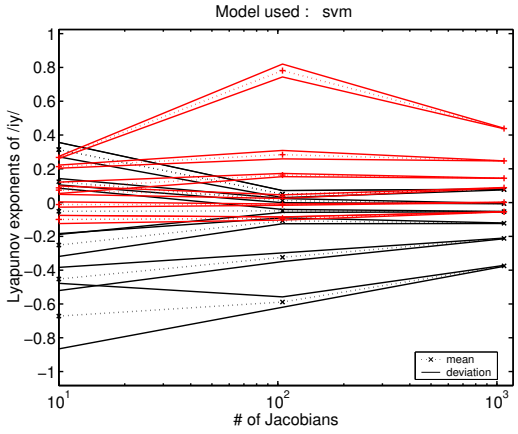


Ολοκλήρωμα συσχέτισης



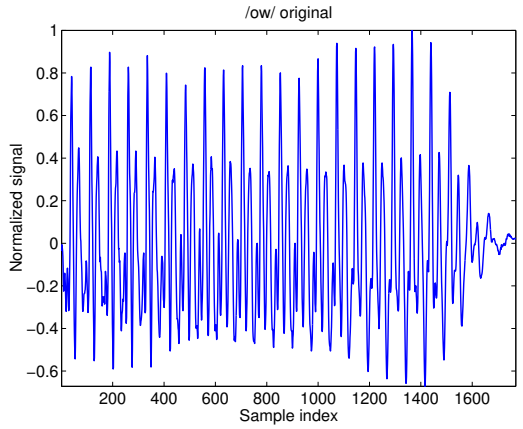
NMSE πρόβλεψης για την επόμενη χρονική στιγμή



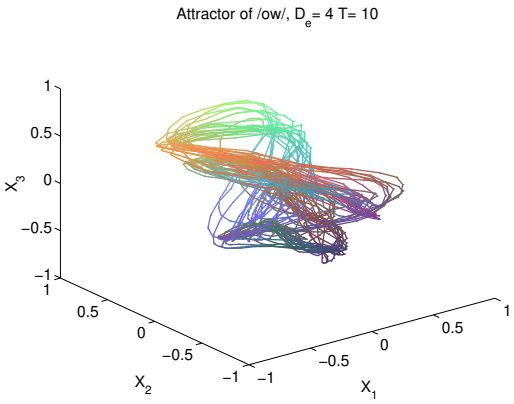


Εκθέτες Lyapunov: (a) από ένα WLP (b) από ένα SVM

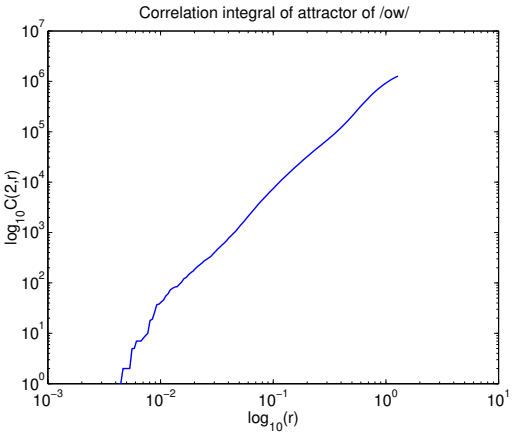
Φώνημα: /ow/



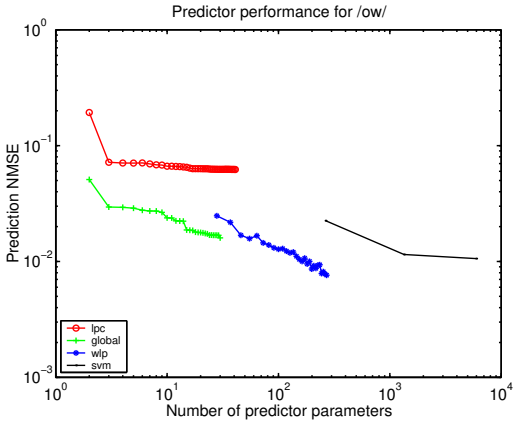
Αρχικό σήμα



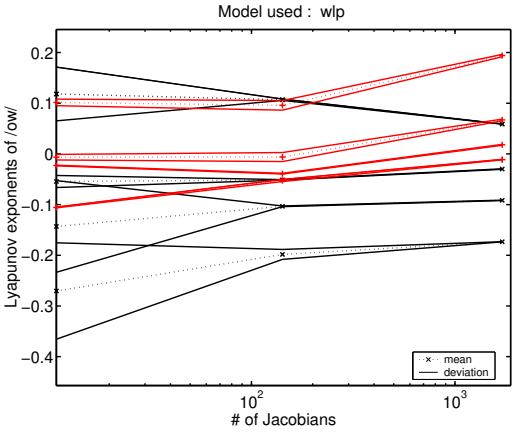
Ανακατασκευασμένος ελκυστής

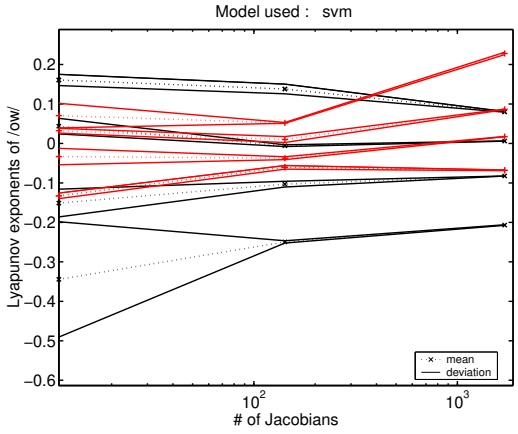


Ολοκλήρωμα συσχέτισης



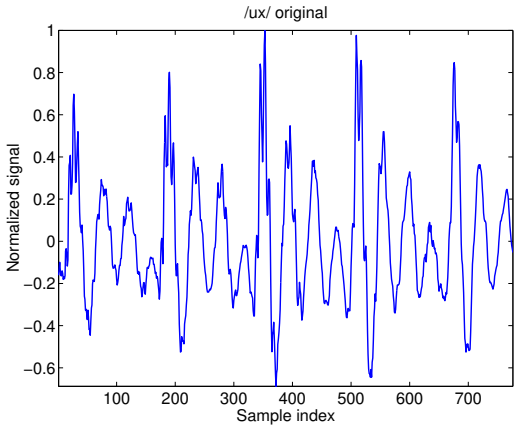
NMSE πρόβλεψης για την επόμενη χρονική στιγμή



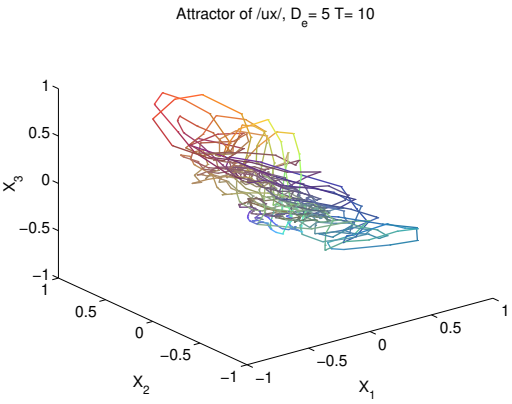


Εκθέτες Lyapunov: (a) από ένα WLP (b) από ένα SVM

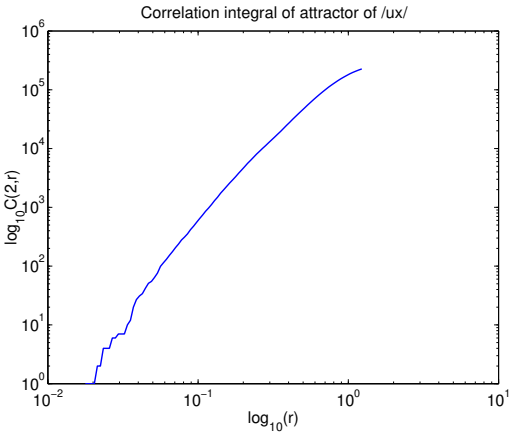
Φώνημα: /ux/



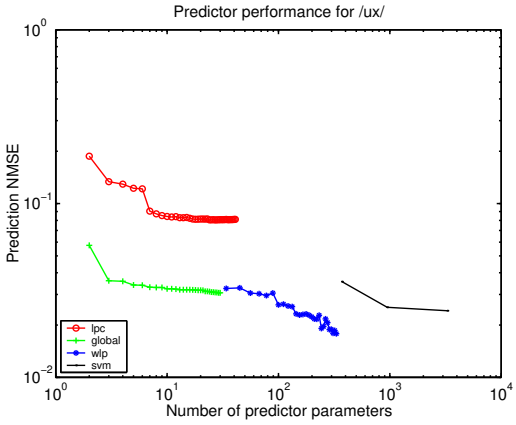
Αρχικό σήμα



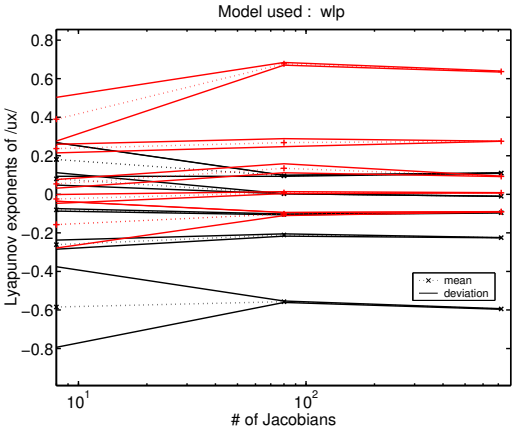
Ανακατασκευασμένος ελκυστής

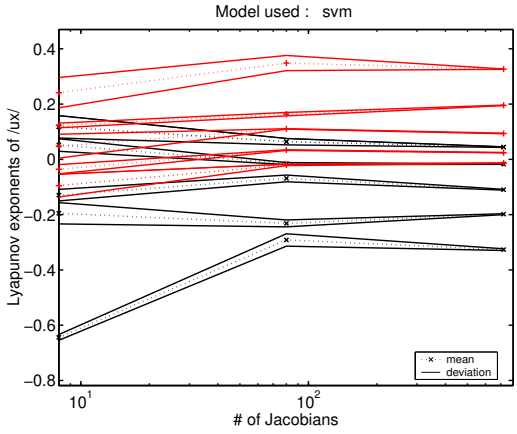


Ολοκλήρωμα συσχέτισης



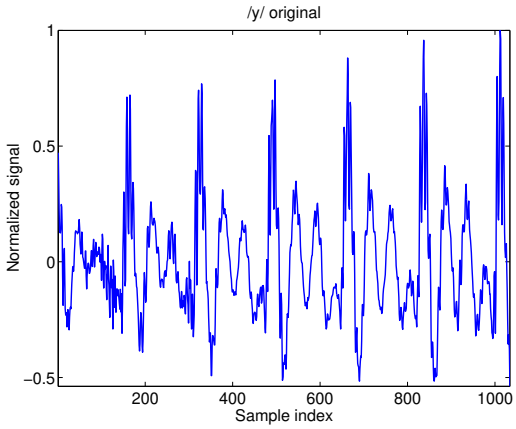
NMSE πρόβλεψης για την επόμενη χρονική στιγμή



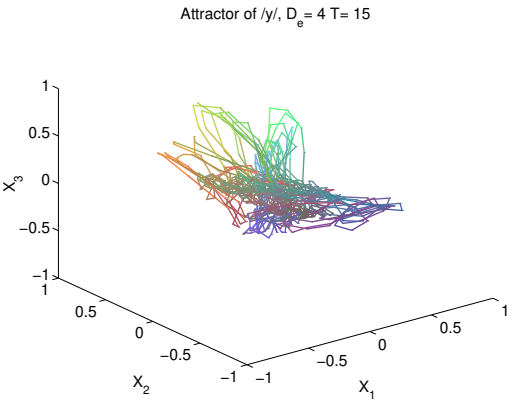


Εκθέτες Lyapunov: (a) από ένα WLP (b) από ένα SVM

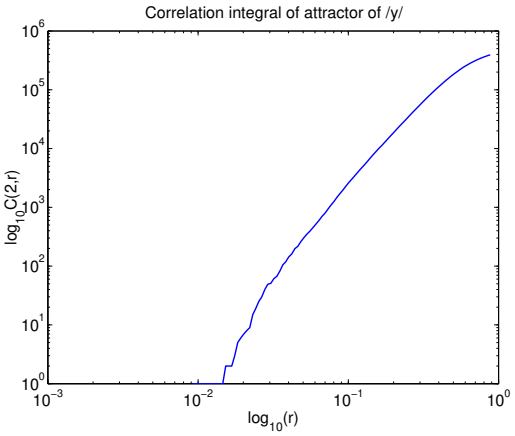
Φώνημα: /y/



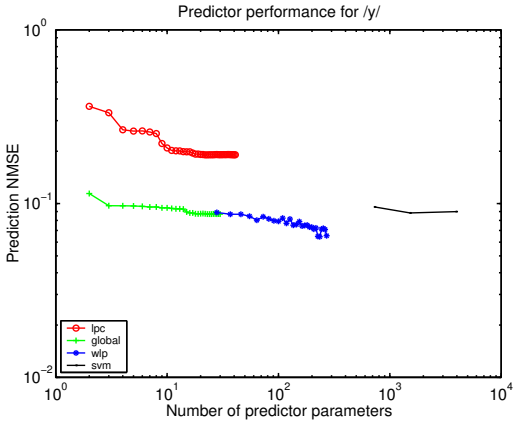
Αρχικό σήμα



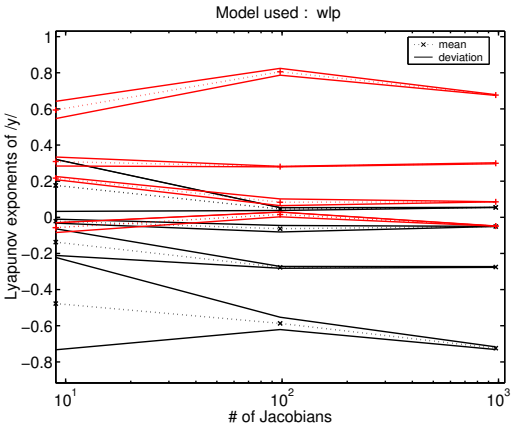
Ανακατασκευασμένος ελκυστής

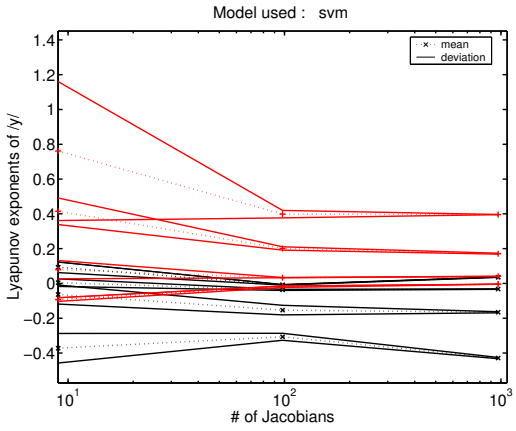


Ολοκλήρωμα συσχέτισης



NMSE πρόβλεψης για την επόμενη χρονική στιγμή

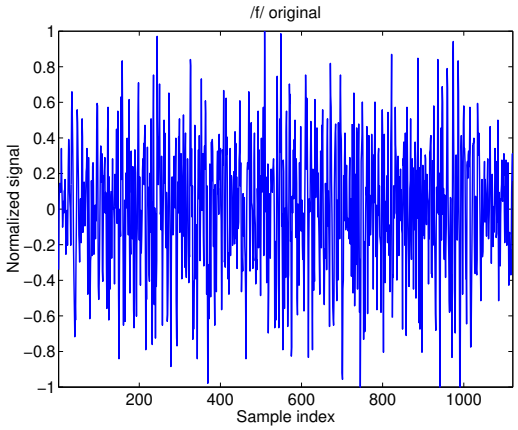




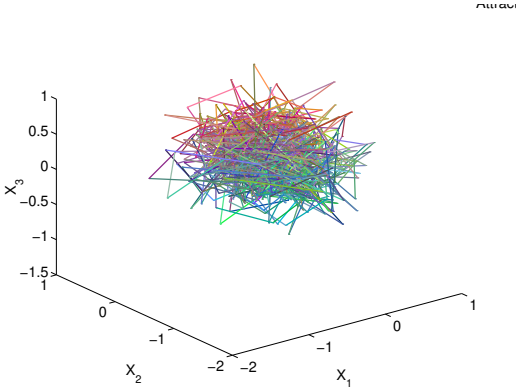
Εκθέτες Lyapunov: (a) από ένα WLP (b) από ένα SVM

3.2 Τυρβώδεις ήχοι

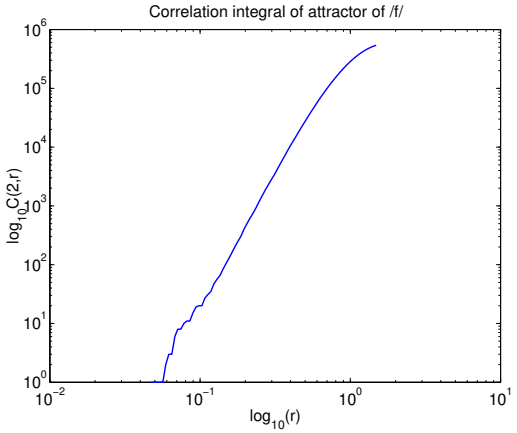
Φώνημα: /f/



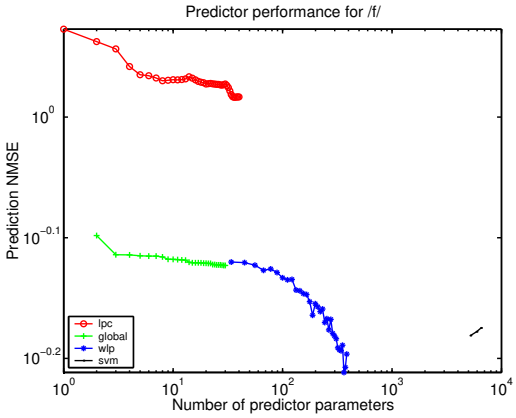
Αρχικό σήμα



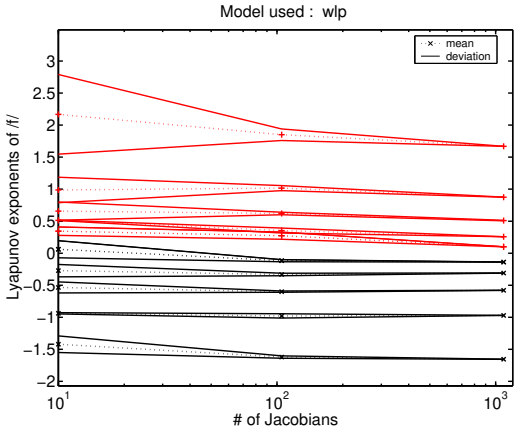
Ανακατασκευασμένος ελκυστής

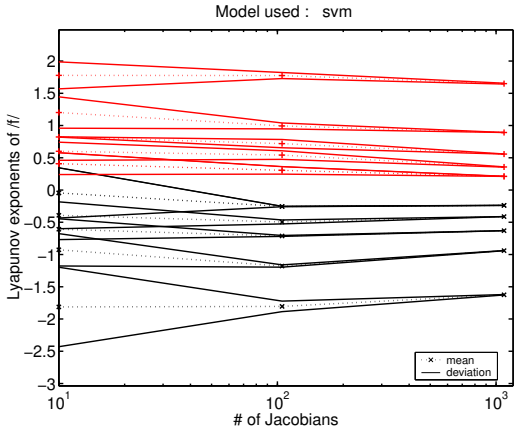


Ολοκλήρωμα συσχέτισης



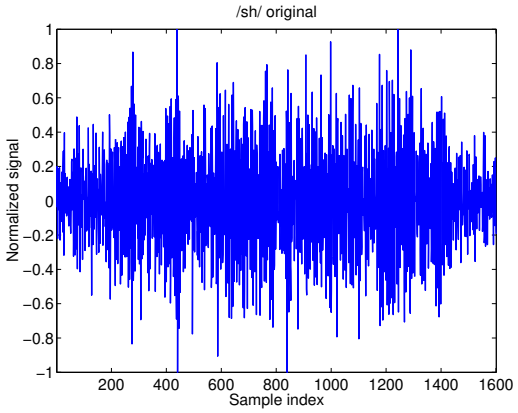
NMSE πρόβλεψης για την επόμενη χρονική στιγμή



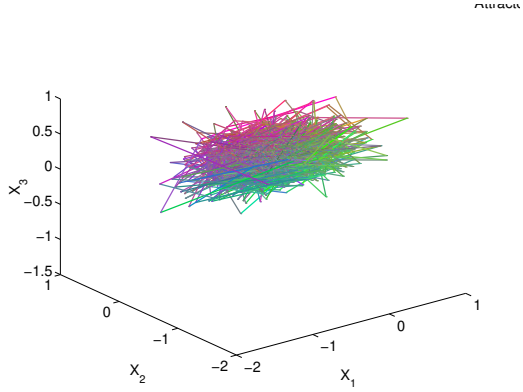


Εκθέτες Lyapunov: (a) από ένα WLP (b) από ένα SVM

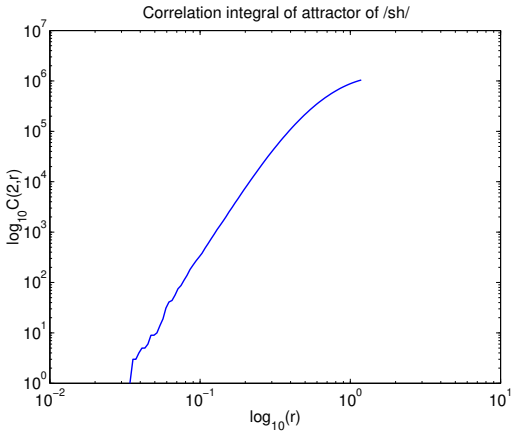
Φώνημα: /sh/



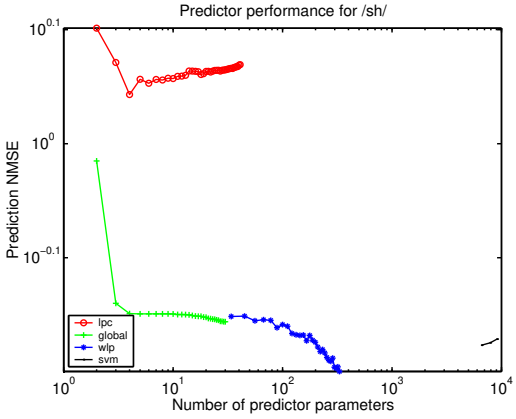
Αρχικό σήμα



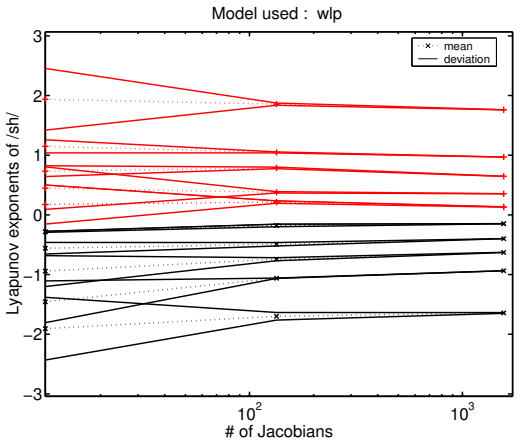
Ανακατασκευασμένος ελκυστής

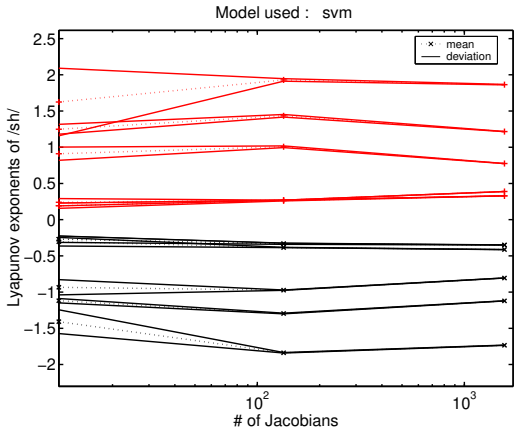


Ολοκλήρωμα συσχέτισης



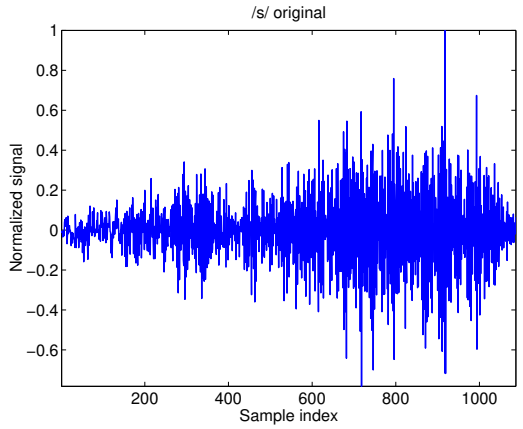
NMSE πρόβλεψης για την επόμενη χρονική στιγμή



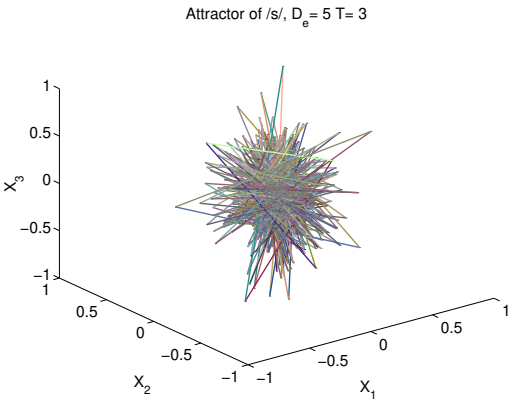


Εκθέτες Lyapunov: (a) από ένα WLP (b) από ένα SVM

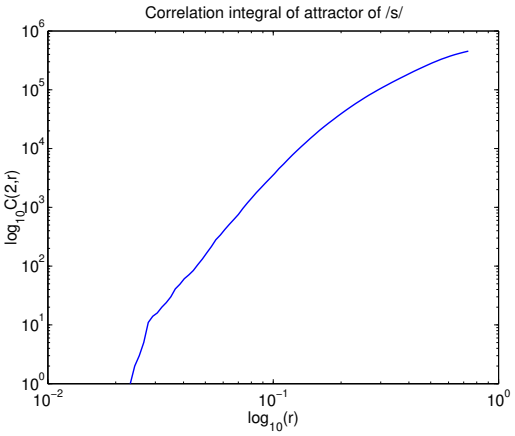
Φώνημα: /s/



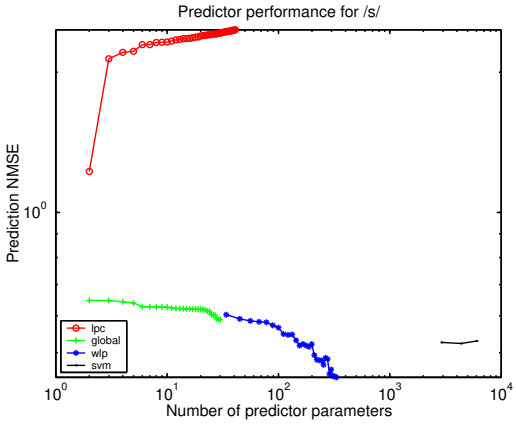
Αρχικό σήμα



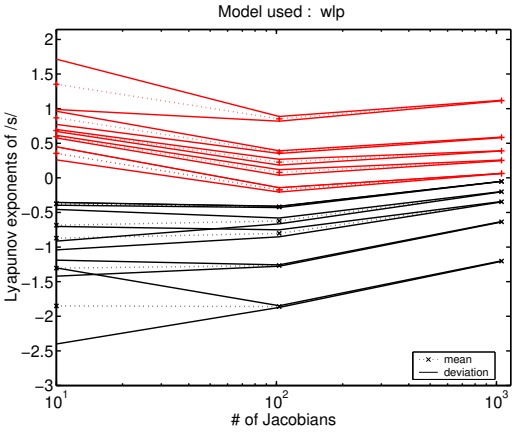
Ανακατασκευασμένος ελκυστής

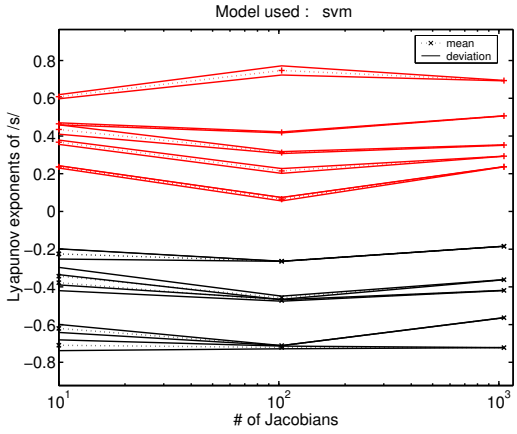


Ολοκλήρωμα συσχέτισης



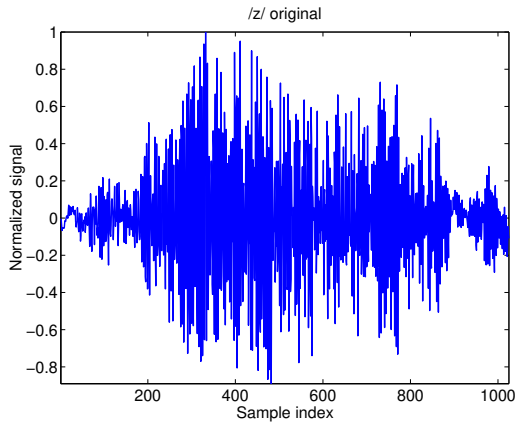
NMSE πρόβλεψης για την επόμενη χρονική στιγμή



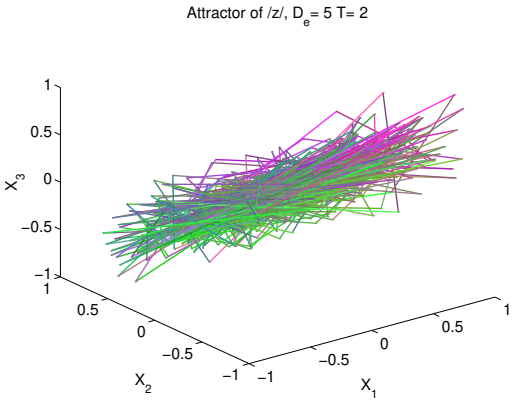


Εκθέτες Lyapunov: (a) από ένα WLP (b) από ένα SVM

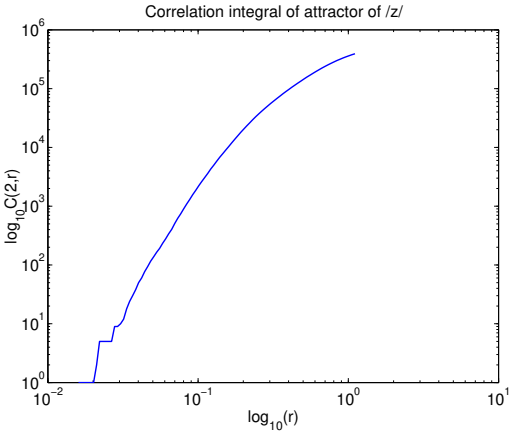
Φώνημα: /z/



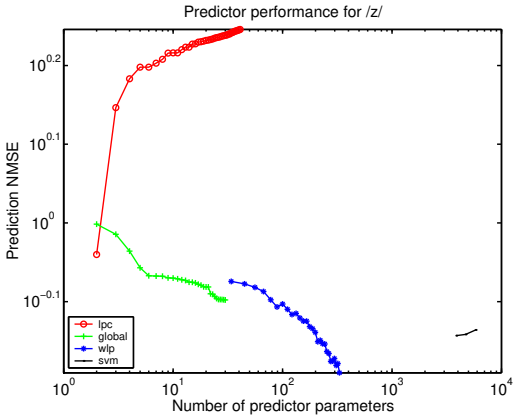
Αρχικό σήμα



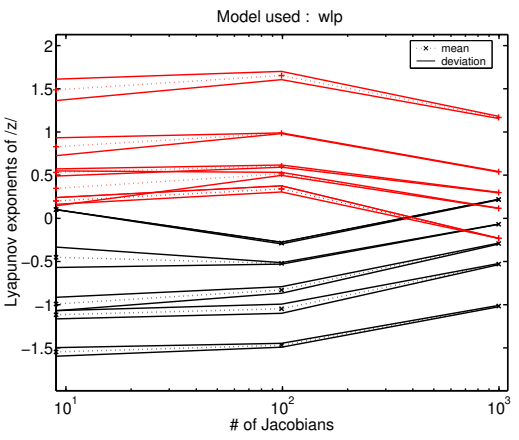
Ανακατασκευασμένος ελκυστής

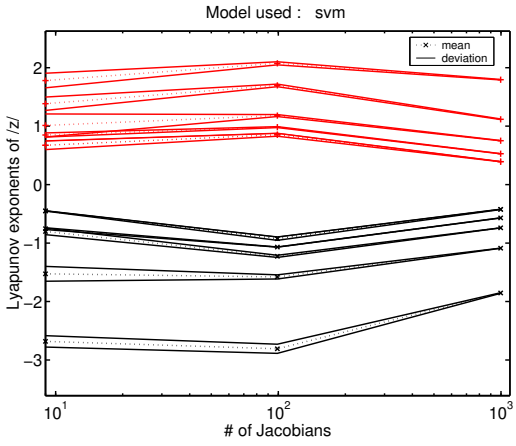


Ολοκλήρωμα συσχέτισης



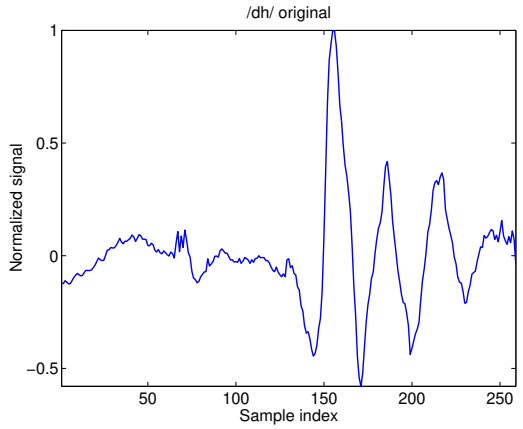
NMSE πρόβλεψης για την επόμενη χρονική στιγμή



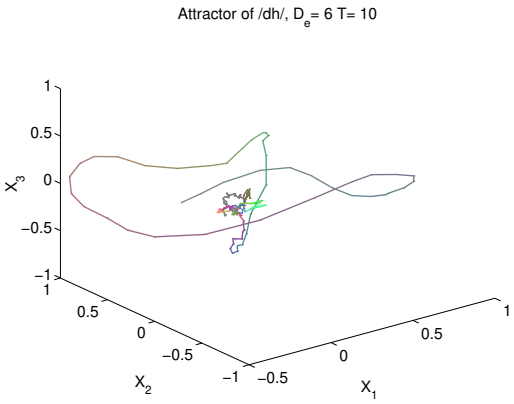


Εκθέτες Lyapunov: (a) από ένα WLP (b) από ένα SVM

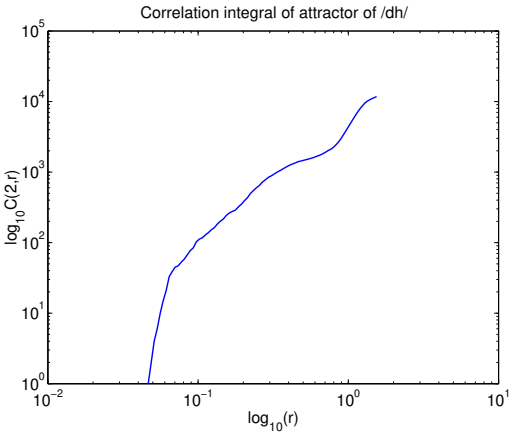
Φώνημα: /dh/



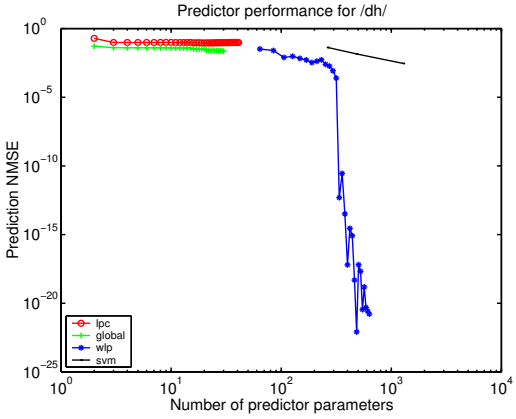
Αρχικό σήμα



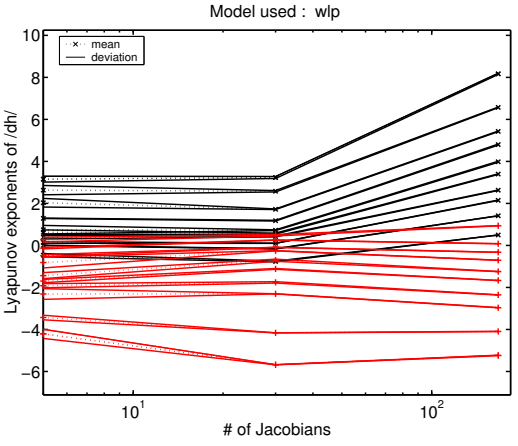
Ανακατασκευασμένος ελκυστής

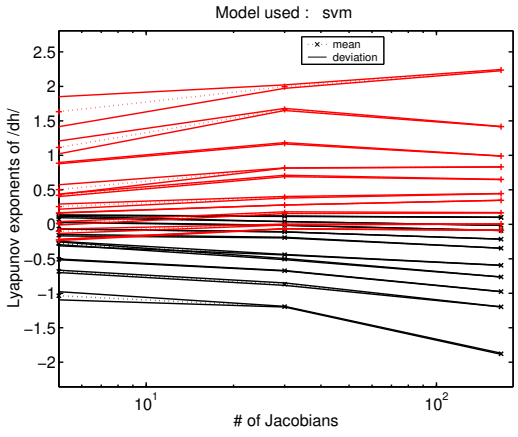


Ολοκλήρωμα συσχέτισης



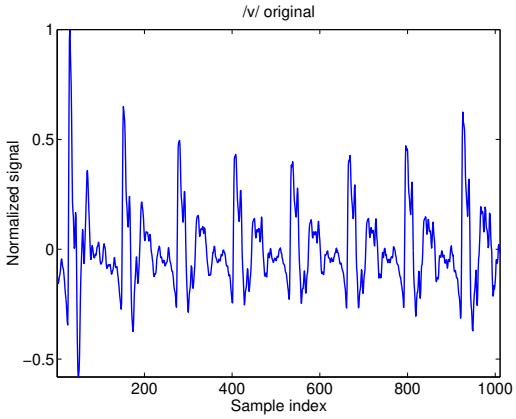
NMSE πρόβλεψης για την επόμενη χρονική στιγμή



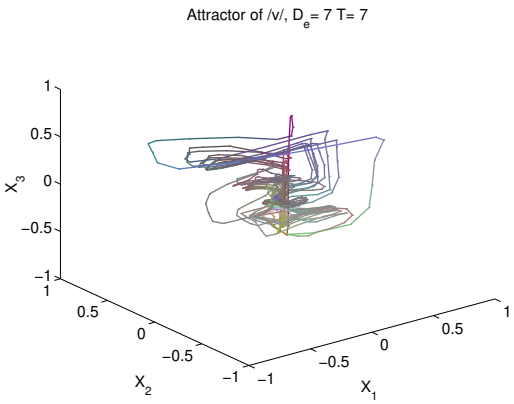


Εκθέτες Lyapunov: (a) από ένα WLP (b) από ένα SVM

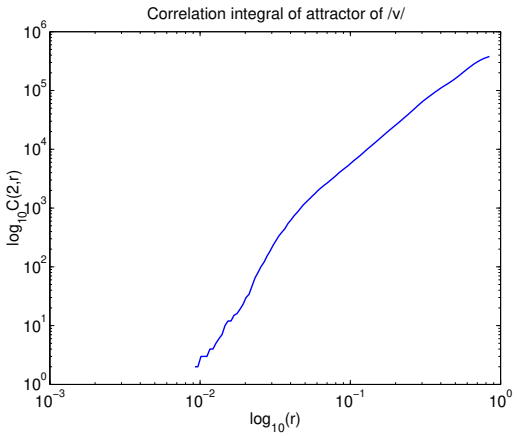
Φώνημα: /v/



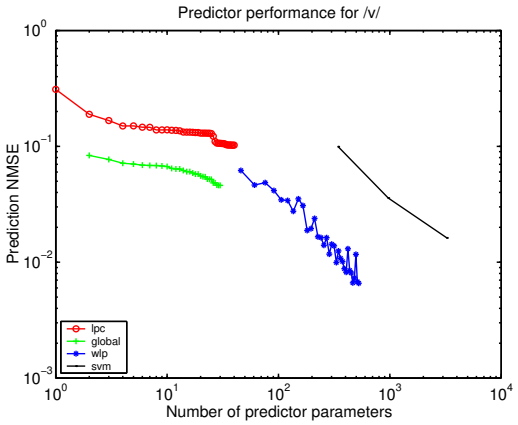
Αρχικό σήμα



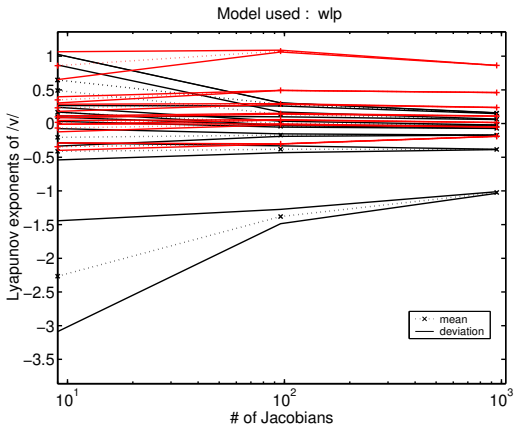
Ανακατασκευασμένος ελκυστής

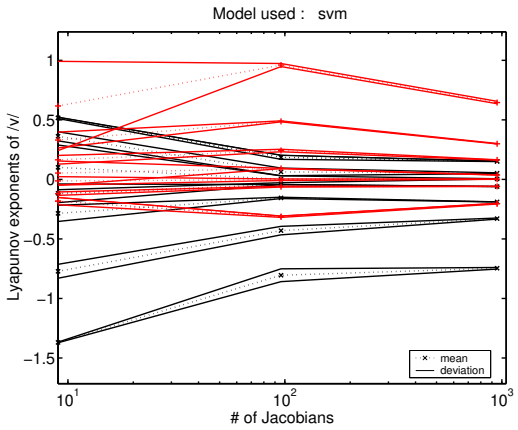


Ολοκλήρωμα συσχέτισης



NMSE πρόβλεψης για την επόμενη χρονική στιγμή

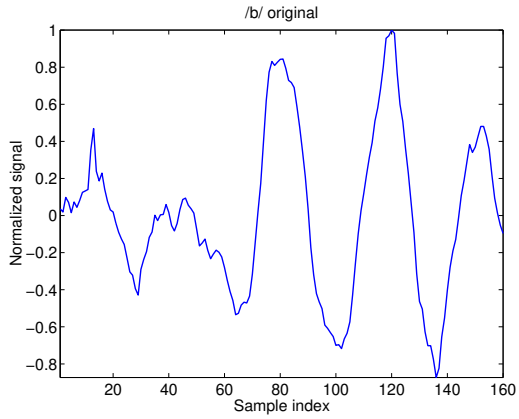




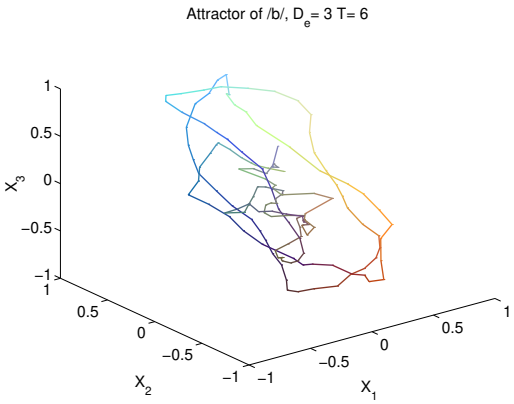
Εκθέτες Lyapunov: (a) από ένα WLP (b) από ένα SVM

3.3 Κρουστικοί ήχοι

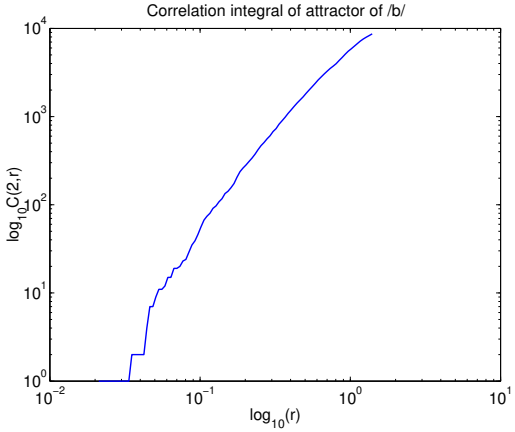
Φώνημα: /b/



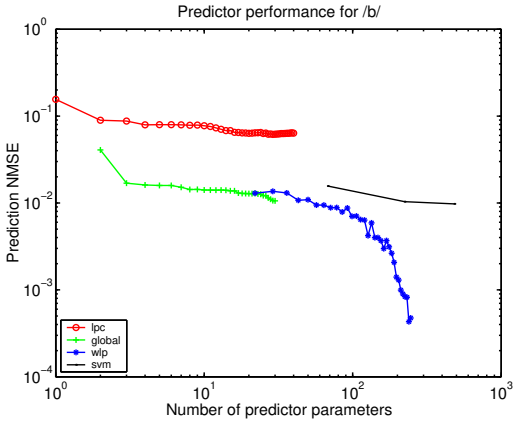
Αρχικό σήμα



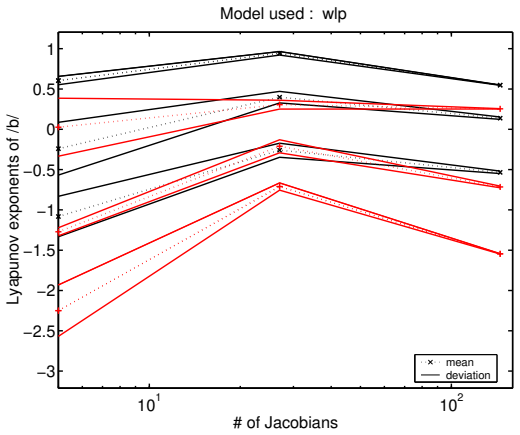
Ανακατασκευασμένος ελκυστής

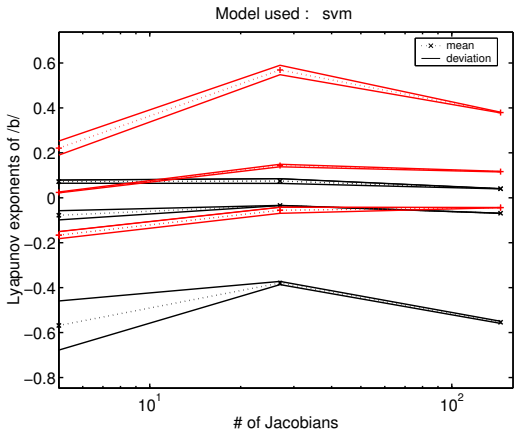


Ολοκλήρωμα συσχέτισης



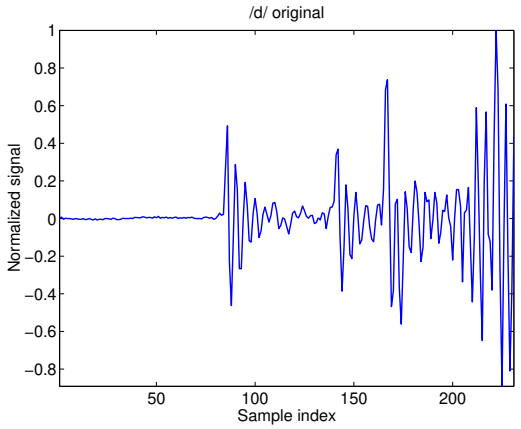
NMSE πρόβλεψης για την επόμενη χρονική στιγμή



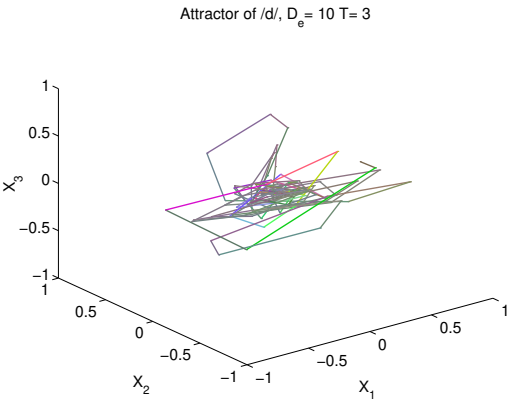


Εκθέτες Lyapunov: (a) από ένα WLP (b) από ένα SVM

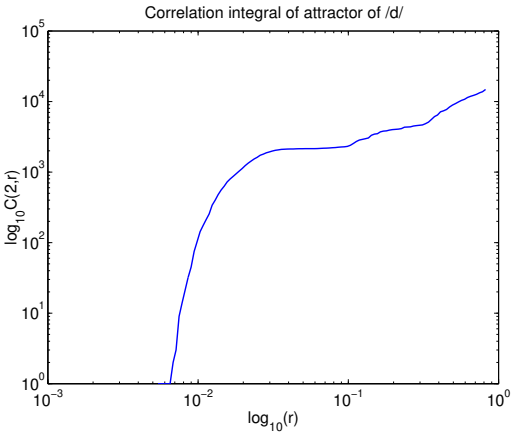
Φώνημα: /d/



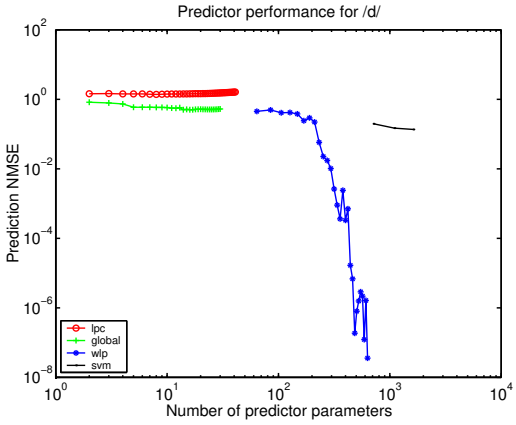
Αρχικό σήμα



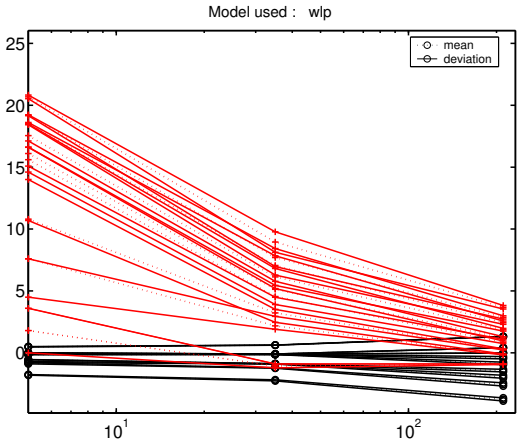
Ανακατασκευασμένος ελκυστής

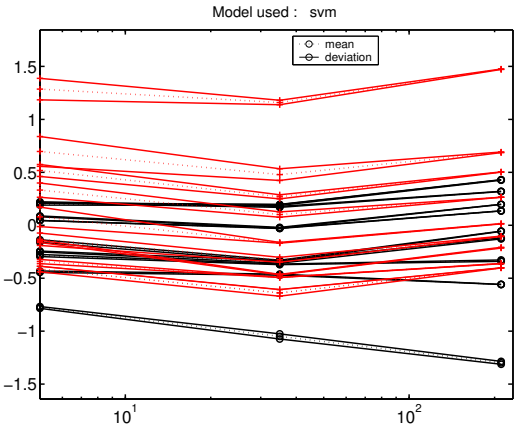


Ολοκλήρωμα συσχέτισης



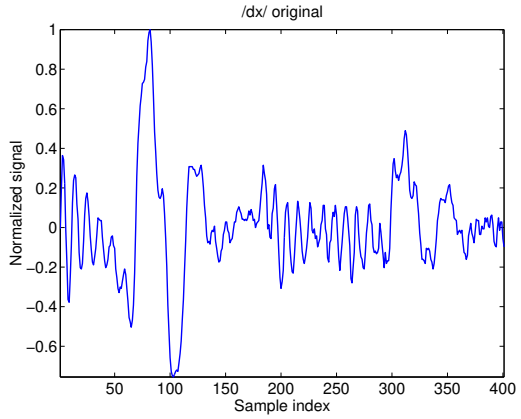
NMSE πρόβλεψης για την επόμενη χρονική στιγμή



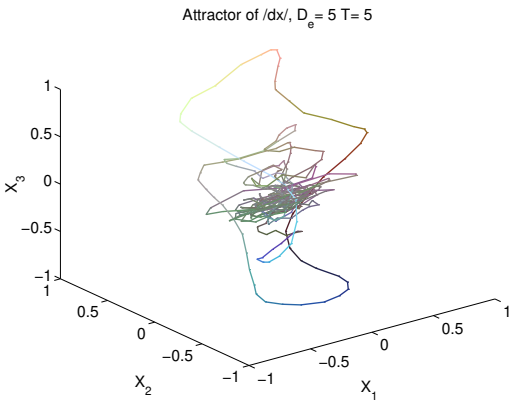


Εκθέτες Lyapunov: (a) από ένα WLP (b) από ένα SVM

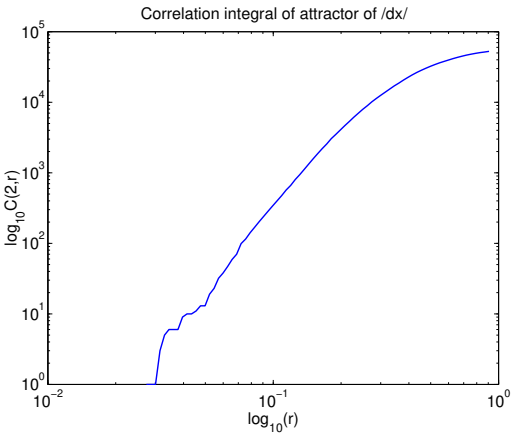
Φώνημα: /dx/



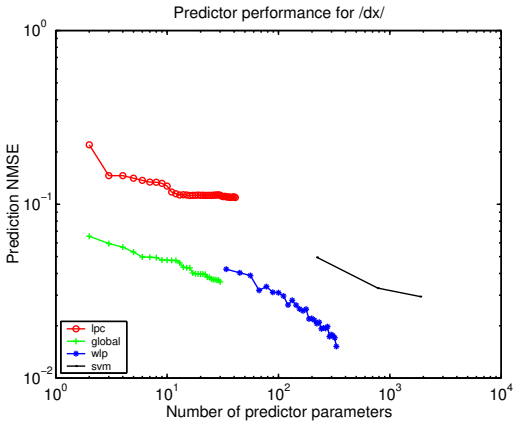
Αρχικό σήμα



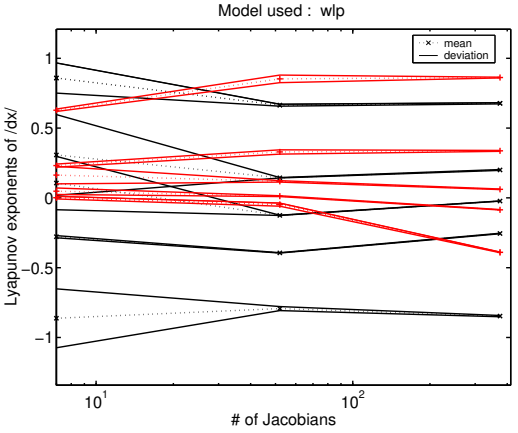
Ανακατασκευασμένος ελκυστής

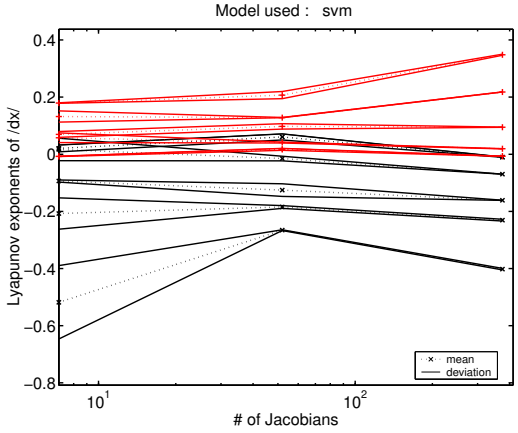


Ολοκλήρωμα συσχέτισης



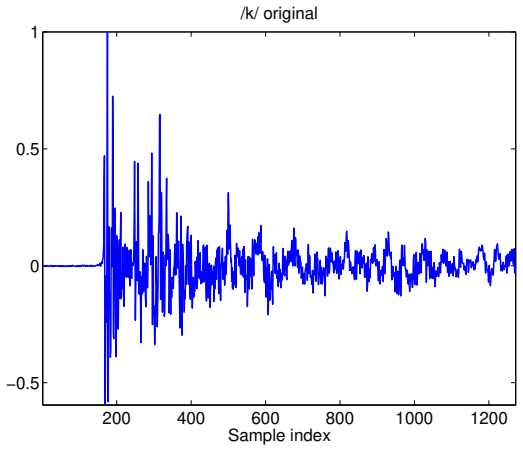
NMSE πρόβλεψης για την επόμενη χρονική στιγμή



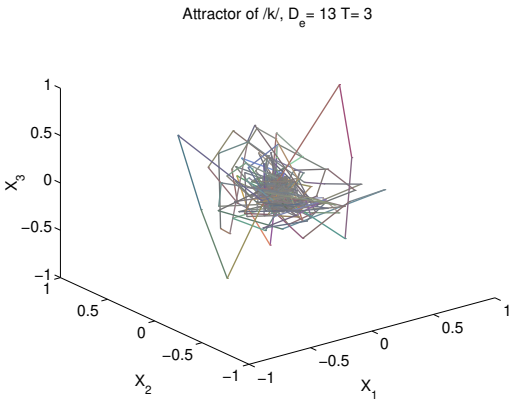


Εκθέτες Lyapunov: (a) από ένα WLP (b) από ένα SVM

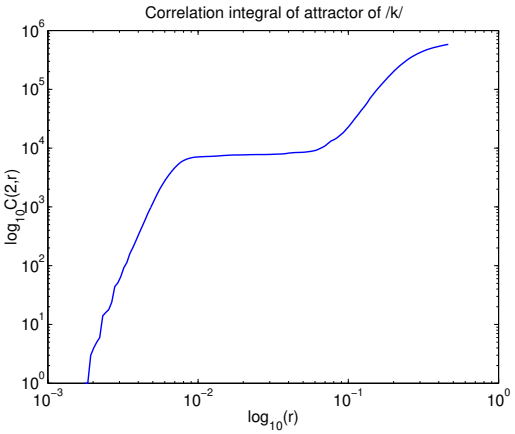
Φώνημα: /k/



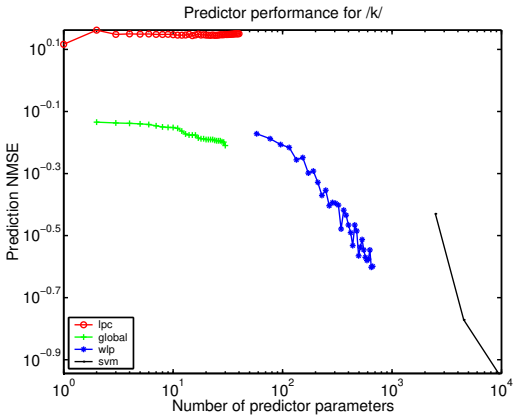
Αρχικό σήμα



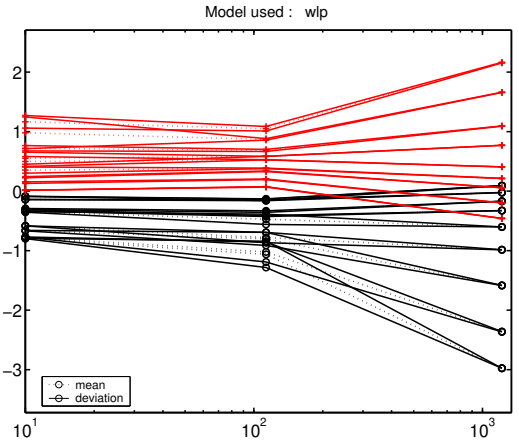
Ανακατασκευασμένος ελκυστής

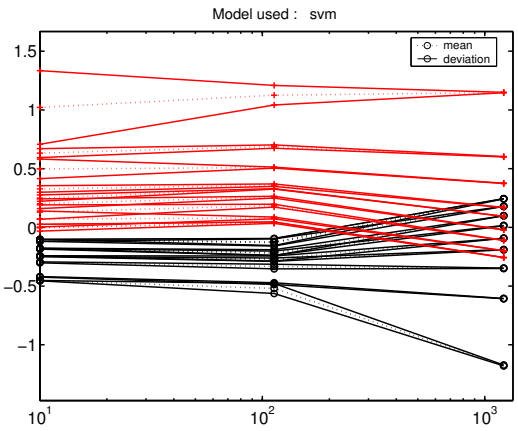


Ολοκλήρωμα συσχέτισης



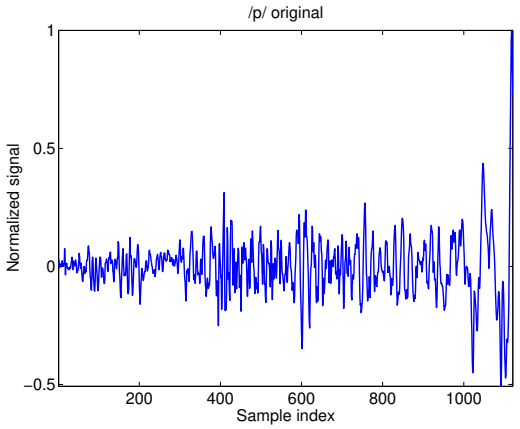
NMSE πρόβλεψης για την επόμενη χρονική στιγμή



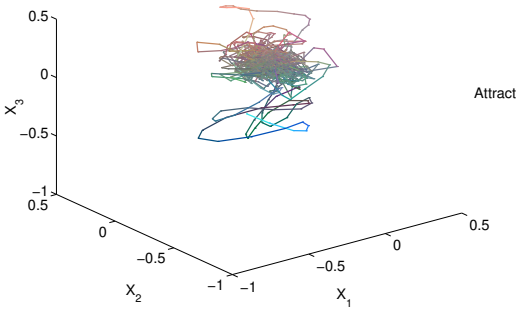


Εκθέτες Lyapunov: (a) από ένα WLP (b) από ένα SVM

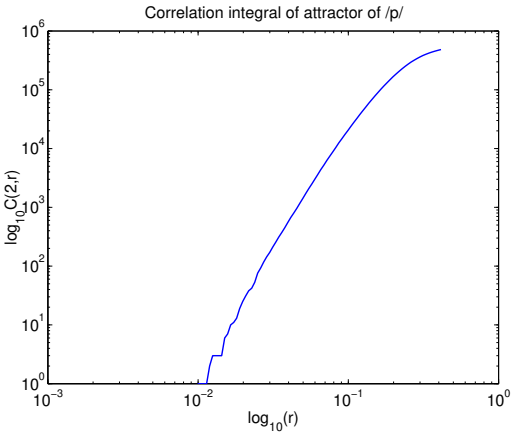
Φώνημα: /p/



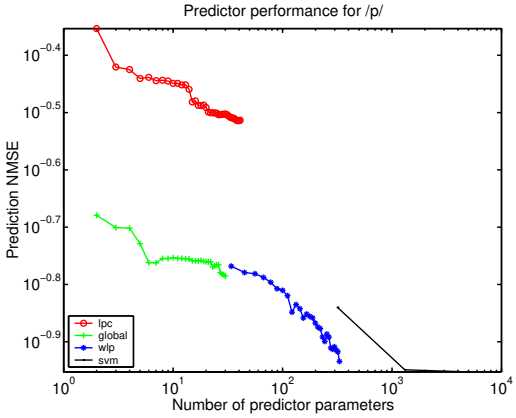
Αρχικό σήμα



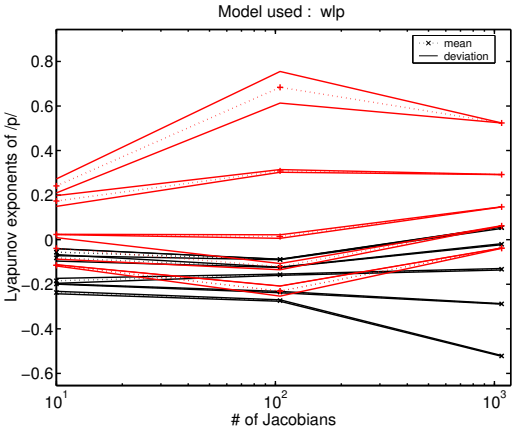
Ανακατασκευασμένος ελκυστής

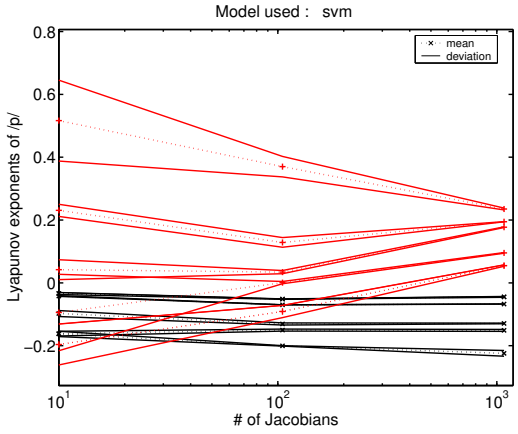


Ολοκλήρωμα συσχέτισης



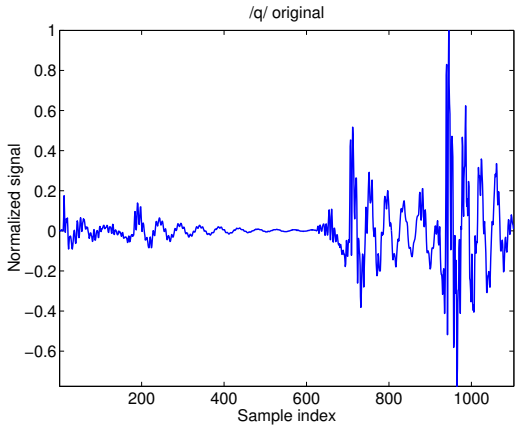
NMSE πρόβλεψης για την επόμενη χρονική στιγμή



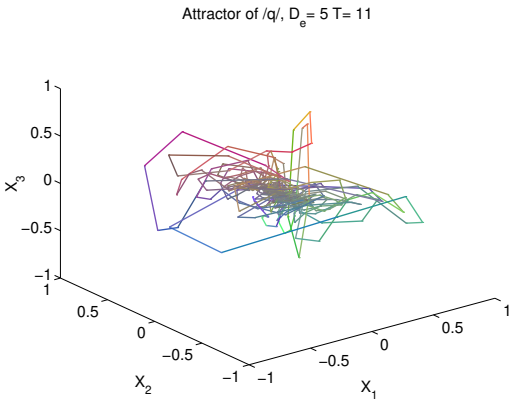


Εκθέτες Lyapunov: (a) από ένα WLP (b) από ένα SVM

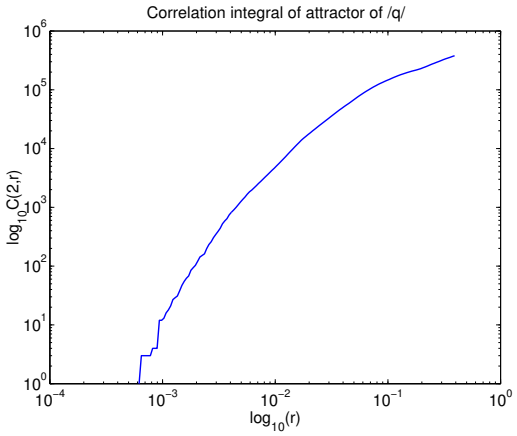
Φώνημα: /q/



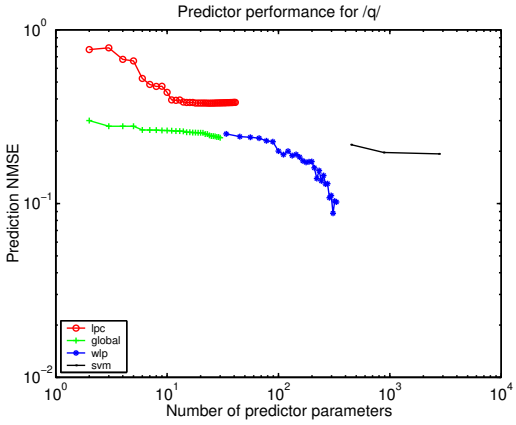
Αρχικό σήμα



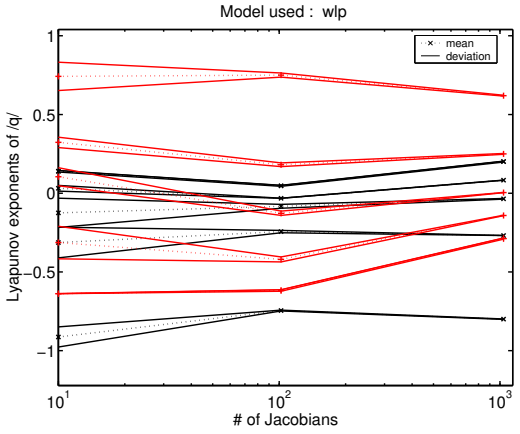
Ανακατασκευασμένος ελκυστής

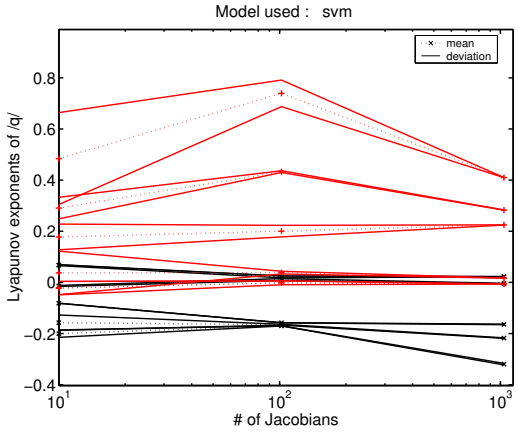


Ολοκλήρωμα συσχέτισης



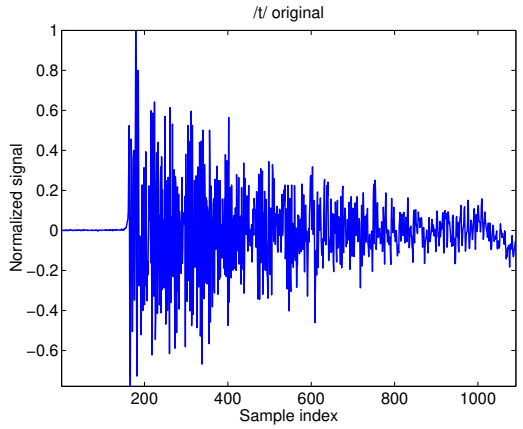
NMSE πρόβλεψης για την επόμενη χρονική στιγμή



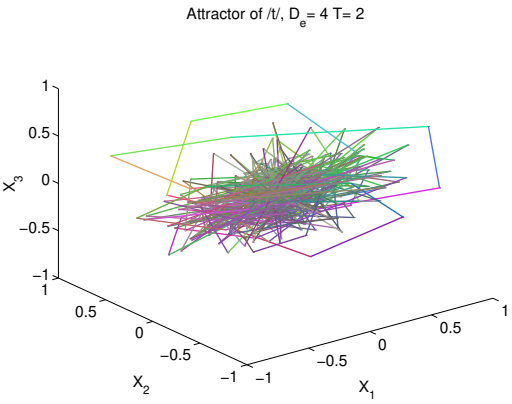


Εκθέτες Lyapunov: (a) από ένα WLP (b) από ένα SVM

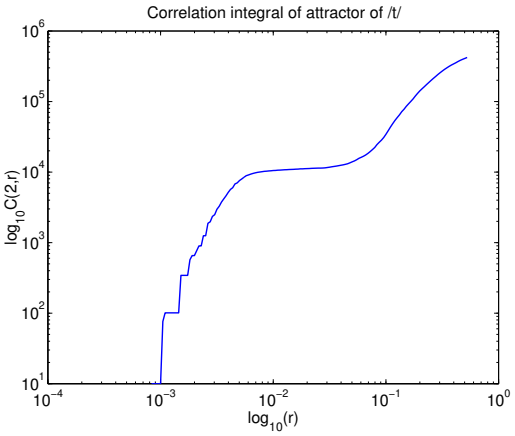
Φώνημα: /t/



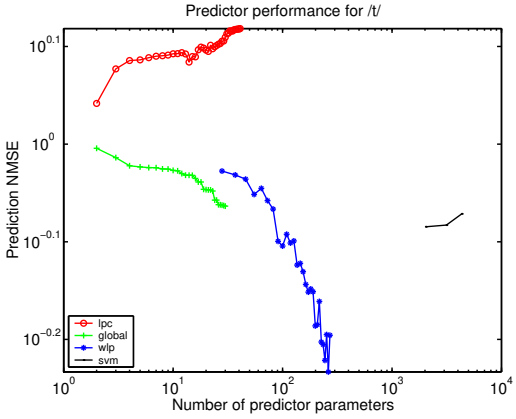
Αρχικό σήμα



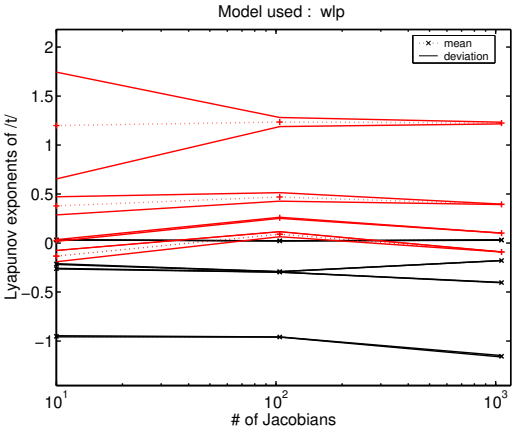
Ανακατασκευασμένος ελκυστής

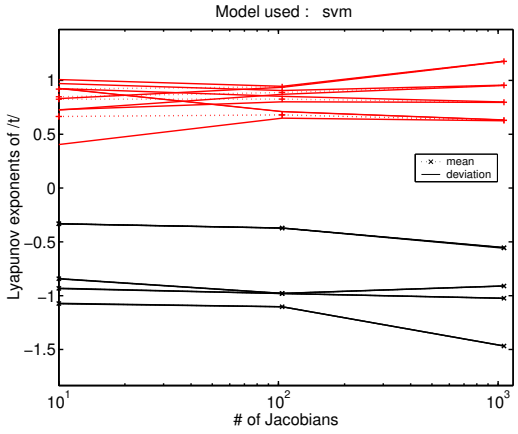


Ολοκλήρωμα συσχέτισης



NMSE πρόβλεψης για την επόμενη χρονική στιγμή





Εκθέτες Lyapunov: (a) από ένα WLP (b) από ένα SVM

Συμπεράσματα- σχολιασμός

- Για όλα τα φωνήματα τα καθολικά πολυώνυμα κάνουν καλύτερη πρόβλεψη από το μοντέλο LPC, έχοντας τον ίδιο αριθμό παραμέτρων στον προβλέπτη τους. Αυτό αποδεικνύει άμεσα ότι κάνοντας πρόβλεψη πάνω στον ανακατασκευασμένο ελκυστή του συστήματος έχουμε μεγαλύτερη δυνατότητα να συλλάβουμε τη δυναμική του συστήματος, ακόμα και με πολύ απλά μοντέλα, όπως τα καθολικά πολυώνυμα. Επίσης το γεγονός ότι μπορούμε πολύ γρήγορα να βρούμε τους συντελεστές πρόβλεψης, τα καθιστά κατάλληλα και για εφαρμογές πραγματικού χρόνου.

- Όλα τα φωνήεντα έχουν ένα μικρό θετικό εκθέτη Lyapunov και τους υπόλοιπους αρνητικούς. Χαρακτηρίζονται δηλαδή από βραχυπρόθεσμη προβλεψιμότητα.

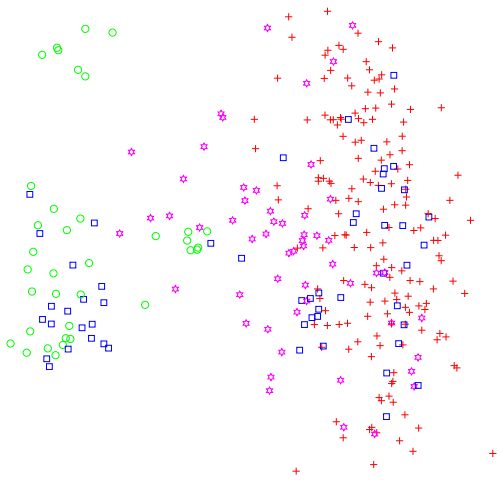
- Οι άφωνοι τυρβώδεις ήχοι δεν δίνουν κανέναν εκθέτη, ο οποίος να επαληθεύεται από τους αντίστροφούς του, ενώ οι έμφωνοι τυρβώδεις ήχοι δίνουν εκθέτες που επαληθεύονται. Η ερμηνεία της παρατήρησης αυτής είναι ότι καθώς στους έμφωνους τυρβώδεις ήχους υπάρχει κάποια ενέργεια από το σήμα, πάνω στην οποία προστίθεται ο 'θόρυβος' του συστήματος παραγωγής φωνής, δεν χάνεται τελείως ο ντετερμινισμός του σήματος, ενώ για τους άφωνους τυρβώδεις ήχους ο θορύβος κυριαρχεί στο σήμα, με αποτέλεσμα να παίρνουμε εκθέτες Lyapunov που θα παίρναμε από ένα σήμα θορύβου.

- Οι κρουστικοί ήχοι δίνουν διφορούμενα αποτελέσματα στα οποία δεν μπορέσαμε να εντοπίσουμε καμία κανονικότητα. Αυτό μπορεί να ερμηνευθεί από τη φύση των κρουστικών ήχων που ξεκινάνε με ένα διάστημα σιωπής, το οποίο ακολουθείται από μία εκτόνωση και εισπνοή. Προκύπτουν από ένα σύστημα, δηλαδή, του οποίου η δυναμική αλλάζει με το χρόνο, οπότε η παραδοχή που κάναμε αρχικά, ότι το σύστημά μας είναι εργοδικό, αναιρείται. Εκτός από την εγγενή τους μη-στασιμότητα, που αναιρεί τις προϋποθέσεις της ανάλυσης τους με τις μεθόδους της ανάλυσης χαοτικών χρονοσειρών αυτό μπορεί να οφείλεται και στις πολύ μεγάλες διαστάσεις εμβύθισης που συνήθως βρίσκονται για αυτά, που οδηγούν σε αριθμητικά σφάλματα.

3.4 Εφαρμογή των μεθόδων ανάλυσης χαοτικών χρονοσειρών στην αναγνώριση μεμονωμένων φωνημάτων

Έχοντας καταλήξει στην χρήση των δικτύων SVM για την εξαγωγή των εκθετών Lyapunov τα χρησιμοποιήσαμε για ένα δείγμα 310 φωνημάτων. Σε περίπτωση που δεν επαληθεύτηκε κανένας (θετικός) εκθέτης Lyapunov χρησιμοποιήσαμε τους αρνητικούς γιατι και αυτό αποτελεί μία πληροφορία, αφού διαφοροποιεί τα φωνήεντα και τους έμφωνους τυρβώδεις ήχους από τους άφωνους τυρβώδεις ήχους και τους κρουστικούς ήχους.

Αυτό το οποίο θέλουμε να εξετάσουμε είναι κατά πόσο χρησιμοποιώντας τους εκθέτες Lyapunov ενός φωνήματος μπορούμε να εμπλουτίσουμε το διάνυσμα χαρακτηριστικών που χρησιμοποιείται για την ανάλυση φωνής. Ένας πολύ γενικός διαχωρισμός των τάξεων μπορεί να επιτευχθεί χρησιμοποιώντας μόνο τους 3 πρώτους εκθέτες Lyapunov, όπως φαίνεται παρακάτω, όπου έχει γίνει μία PCA των τριών εκθετών σε δύο σύνιστώσες (οι εκθέτες αυτοί έχουν υπολογιστεί με ένα δίκτυο SVM). Είναι λοιπόν αναμενόμενο να



3.1: Πρώτες 2 principal components των εκθετών Lyapunov φωνημάτων τεσσάρων τάξεων

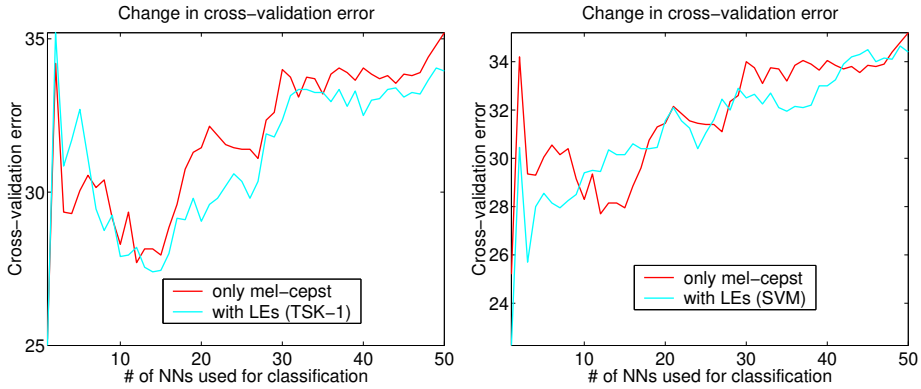
γίνεται καλύτερος διαχωρισμός των τάξεων αν επισυνάψουμε στο διάνυσμα των 12 mel-cepstrum συνιστωσών (που χρησιμοποιείται γενικότερα για την αναγνώριση φωνής) τους εκθέτες Lyapunov. Για να το ελέγξουμε αυτό, τα δεδομένα μας χωρίστηκαν *κυκλικά* σε ενα μεγάλο training set -19/20 των δεδομένων και ένα μικρότερο test set -1/20 των δεδομένων, με βάση το οποίο υπολογίστηκε η μέση επίδοση του εκάστοτε ταξινομητή στο σύνολο των test-sets. Οι ταξινομητές που χρησιμοποιήθηκαν είναι:

- Ένας k-NN ταξινομητής. Το k κυμαίνεται μεταξύ 1 και μίας σχετικά μεγάλης τιμής (50), πέραν της οποίας δεν παρατηρείται μείωση του validation error.
- Ενα δίκτυο SVM για ταξινόμηση. Για n τάξεις σχηματίζονται $\frac{n\cdot(n-1)}{2}$ δίκτυα- ένα για κάθε ζεύγος τάξεων- όπου το κάθε δίκτυο αποφασίζει για το αν ένα σημείο ανήκει στην μία τάξη ή στην άλλη. Η αρχιτεκτονική αυτή ονομάζεται DAG-SVM, όπου DAG (Directed Acyclic Graph), είναι ο γράφος που στις κορυφές του βρίσκονται οι τάξεις και στις ακμές του οποίου είναι τα δίκτυα που αποφασίζουν για την κατάταξη του σημείου στην μία κατηγορία ή στην άλλη. Ο κανόνας με τον οποίο επιλέγεται μία κατηγορία είναι ο max-wins: για κάθε σύγκριση μεταξύ δύο τάξεων αποδίδεται ένας βαθμός στην τάξη που υπερίσχυσε και τελικά επιλέγεται η τάξη με το μεγαλύτερο αριθμό βαθμών

Στο σχήμα (3.2) φαίνεται η μεταβολή του cross validation λάθους για τον k-NN ταξινομητή για διάφορα k για τους δύο τύπους εκθετών. Παρατηρούμε ότι:

- Χρησιμοποιώντας τους εκθέτες Lyapunov του TSK-1 μετά από back-propagation επιτυγχάνεται μία μείωση του ελάχιστου λάθους κατά 0.79% (από 25.2% σε 25%)
- Καλύτερα αποτελέσματα παίρνουμε χρησιμοποιώντας τους εκθέτες Lyapunov που υπολογίστηκαν από ένα δίκτυο SVM, όπου πετύχαμε μείωση του σφάλματος κατά 11.71% (από 25.2% σε 22.25%).

Συνεπώς οι εκθέτες Lyapynov που παίρνουμε από το δίκτυο SVM είναι περισσότερο χρήσιμοι. Παρακάτω φαίνεται για διαφορετικές τιμές των παραμέτρων του δικτύου SVM σ, C που



3.2: Cross validation error για τους L.E.s των TSK-1/SVM

χρησιμοποιείται για την ταξινόμηση, το cross-validation λάθος αν χρησιμοποιηθούν μόνο οι 12 πρώτοι συντελεστές του mel-cepstrum και αν προστεθούν οι εκθέτες Lyapunov του μοντέλου SVM

cross validation λάθος (%) χωρίς L.E.s

C,σ	0.1	0.12	0.13	0.15	0.2
1	16.72	15.76	15.76	15.11	16.08
2	14.11	13.71	14.15	13.72	14.11
5	15.43	14.15	15.11	15.43	16.08
10	16.40	16.08	16.08	16.78	17.12

(3.1)

cross validation λάθος (%) με L.E.s

C,σ	0.1	0.12	0.13	0.15	0.2
1	15.43	15.43	15.11	15.11	16.40
2	13.83	13.50	14.47	14.47	16.08
5	15.11	14.47	15.11	14.79	17.04
10	15.11	15.11	15.76	15.11	16.40

Παρατηρούμε ότι τα αποτελέσματα είναι παρόμοια και στις δύο περιπτώσεις, με τα αποτελέσματα του DAGSVM που χρησιμοποιεί τους εκθέτες Lyapunov να είναι ελαφρά καλύτερα αφού παρατηρείται βελτίωση του ελάχιστου λάθους κατά 1.4%(από 13.71% 13.5%). Είναι φανερό ότι οι επιδόσεις του DAGSVM είναι καλύτερες από αυτές του k-NN , πράγμα αναμενόμενο λόγω της αντοχής του SVM σε θόρυβο και της γενικότερα αναγνωρισμένης δυνατότητας γενίκευσης που έχει. Σε κάθε περίπτωση οι εκθέτες Lyapunov αποδεικνύονται χρήσιμοι στην αναγνώριση της κατηγορίας στην οποία ανήκει ένα φώνημα. Αξίζει να αναφερθεί εδώ ότι χρησιμοποιώντας τους εκθέτες που πήραμε χρησιμοποιώντας ένα δίκτυο TSK-1 χωρίς εφαρμογή του αλγορίθμου rprop προκλήθηκε αύξηση του λάθους κατά $5-10\%$ που δείχνει ότι είναι καλύτερο να μην χρησιμοποιούνται οι εκθέτες Lyapunov, αν δεν είμαστε σίγουροι για αυτούς.

Κεφάλαιο 4

Συμπεράσματα - κατευθύνσεις μελλοντικής έρευνας

Ακολουθούν τα βασικότερα συμπεράσματα της έρευνάς μας:

- Η ανάλυση τόυ σήματος φωνής με μη γραμμικές μεθόδους μας δίνει τη δυνατότητα να κάνουμε πρόβλεψη που να διαφέρει από το σήμα ακόμη και στα όρια της ακρίβειας του υπολογιστή. Ιδιαίτερα για τα φωνήεντα η πρόβλεψη ήταν εξαιρετικά επιτυχής, ενώ ακόμα κα για τους κρουστικούς ήχους έχουμε πολύ καλύτερα αποτελέσματα από ότι με το γραμμικό μοντέλο πρόβλεψης. Αυτό σημαίνει ότι με μη-γραμμικές μεθόδους είμαστε σε θέση να κατασκεύασουμε ένα μοντέλο το οποίο να ακολουθεί πολύ πιστά το σήμα φωνής. Αυτό οφείλεται αφ'ενός στο γεγονός ότι στον ανακατασκευασμένο χώρο φάσης του αναιρείται σε μεγάλο βαθμό η στοχαστικότητα που του απέδιδε το γραμμικό μοντέλο και αφ'εταίρου στις εξελιγμένες μεθόδους που εφαρμόστηκαν, οι οποίες μας δίνουν τη δυνατότητα να ανακατασκευάσουμε με μεγάλη ακρίβεια τη δυναμική ενός μη-γραμμικού συστήματος. Κατά κάποιο τρόπο μπορούμε να πούμε ότι αναλύοντας το σήμα φωνής πάνω στον ανακατασκευασμένο ελκυστή τού θέτουμε το πρόβλημα της μοντελοποίησής του σε μία υγιή βάση: καθώς έχουμε έναν αποδιπλωμένο ελκυστή, αυξάνοντας τον αριθμό των παραμέτρων του προβλέπτη μπορούμε να 'μάθουμε' πλήρως το σήμα μας. Αντιθέτως, χρησιμοποιώντας τη μέθοδο LPC από ένα σημείο και πέρα δεν παρατηρείται μείωση του MSE της πρόβλεψης όσο και αν αυξάνεται η τάξη του μοντέλου.

- Οι κρουστικοί και οι τυρβώδεις ήχοι 'αντιστέκονται' ακόμα, χωρίς αυτό να σημαίνει ότι τα μοντέλα μας απέτυχαν -έχουν όμως χαμηλότερες επιδόσεις και δίνουν εκθέτες Lyapunov που έχουν νόημα μόνο σε ορισμένες περιπτώσεις. Παρ' αυτά, το MSE της πρόβλεψης είναι σαφώς μικρότερο από αυτό της μεθόδου LPC, καθώς οι μη-γραμμικές μέθοδοι δεν υποχρεώνουν το σήμα να μοντελοποιείται με ένα καθολικό AR-μοντέλο (κεφάλαιο 2), οπότε είναι πιο ευέλικτες και μπορούν να μάθουν όποιο κομμάτι της δυναμικής του συστήματος προσφέρεται για μάθηση.

- Θεωρώντας ότι το σήμα φωνής έχει προέλθει από ένα χαοτικό σύστημα και υπολογίζοντας τους εκθέτες Lyapunov μπορούμε να εξάγουμε χρήσιμα χαρακτηριστικά για την ανάλυση-αναγνώριση του σήματος φωνής. Η παραπάνω εικασία μας, ότι δηλαδή το σήμα φωνής έχει χαοτική φύση επιβεβαιώνεται και από την παρατήρηση θετικών εκθετών Lyapunov που σημαίνουν την ύπαρξη χάους.

Οι κατευθύνσεις έρευνας που φαίνονται πιο ενδιαφέρουσες αυτή τη στιγμή είναι κατα τη γνώμη μας οι εξής:

- Αναγνώριση συνεχούς σήματος φωνής χρησιμοποιώντας τους εκθέτες Lyapunov. Εως τώρα έχουν εξεταστεί μόνο μεμονωμένα φωνήματα, τα οποία σπανίως έχουμε στη διάθεσή μας σε πραγματικές εφαρμογές. Συνήθως η ανάλυση σήματος φωνής γίνεται χρησιμοποιώντας frames συγκεκριμένου μεγέθους με σκοπό την ανανώριση συνεχόμενων λέξεων. Θα είχε ενδιαφέρον να εξεταστεί κατά πόσο μπορεί να γίνει ανάλυση κάτω από αυτές της συνθήκες του σήματος φωνής και αν παίρνουμε καλύτερα αποτελέσματα.

- *Κατασκευή αποδοτικών αλγορίθμων για την προσαρμοστική μάθηση του προβλέπτη.* Ιδιαίτερο ενδιαφέρον έχει να εξεταστεί αν γίνεται με κάποιο τρόπο ο προβλέπτης του σήματος φωνής αντί να κατσκευάζεται για κάθε frame από την αρχή, να χρησιμοποιεί όποια προηγούμενη γνώση έχει για να επιτύχει σε ικανοποιητικό χρόνο την εκμάθηση των νέων δεδομένων. Αυτό θα μπορούσε να γίνει αν διατηρούσαμε σταθερή τη διάσταση εμβύθισης και τη χρονική καθυστέρηση και αφαιρούσαμε τα παλιά σημεία και προσθέταμε τα καινούργια σημεία κάθε frame. Ο προβλέπτης μας γνωρίζει τη δυναμική των κοινών στο προηγούμενο και στο παρόν frame σημείων και οφείλει να μάθει μόνο αυτή των καινούργιων.

- *Αποθορυβοποίηση σήματος φωνής με τη δυναμική του συστήματος*
Έχουν ήδη αναπτυχθεί μέθοδοι για την αποθορυβοποίηση δεδομένων που προέρχονται από χαοτικές χρονοσειρές (βλ.π.χ. [39], [1], [38]) και θα είχε ενδιαφέρον να δούμε κατά πόσο είναι εφικτό αυτό και για το σήμα φωνής. Ενδιαφέρον θα είχε να εξετάσουμε τις επιδόσεις του συστήματος αποθορυβοποίησης για φωνήματα τα οποία χαρακτηρίζονται από συχνοτικό περιεχόμενο και σε μεγάλες συχνότητες, όπου τα γραμμικά φίλτρα αλλοιώνουν το σήμα (π.χ. άφωνα τυρβώδη φωνήματα).

- *Εφαρμογή των προβλεπτών σε συστήματα παραγωγής φωνής.*
Όπως είδαμε τα μοντέλα που δοκιμάσαμε είναι σε θέση να προβλέψουν πολύ καλύτερα το σήμα φωνής, οπότε ενδεχομένως αν μπορέσουμε να εξάγουμε τα σημαντικότερα χαρακτηριστικά των προβλεπτών αυτών να κατασκευάσουμε συστήματα τα οποία να παράγουν σήμα φωνής που να ακούγεται περισσότερο αληθοφανές από αυτό που παράγεται με τις σύγχρονες μεθόδους.

- *Αναγνώριση διαστημάτων σιωπής.* Θα ήταν άξιο διερεύνησης κατά πόσο η εφαρμογή των αντίστοιχων μεθόδων από τη μη-γραμμική ανάλυση σημάτων φωνής μπορεί να δώσει καλύτερα αποτελέσματα από τις ήδη υπάρχουσες. Για το σκοπό αυτό στα πλαίσια της μη-γραμμικής ανάλυσης χρονοσειρών, δεδομένης μιας χρονοσειράς, για την οποία θέλουμε να αποφασίσουμε αν προέρχεται από ένα στοχαστικό ή από ένα ντετερμινιστικό σύστημα, παράγουμε τυχαία δεδομένα με όμοια χαρακτηριστικά με την αρχική μας χρονοσειρά (surrogate data). Μια απλή μέθοδος για να πάρουμε surrogate data είναι π.χ. να διατηρήσουμε το μέτρο του φάσματος Fourier της αρχικής χρονοσειράς, να δώσουμε τυχαίες τιμές στη φάση του και να πάρουμε τον αντίστροφο μετασχηματισμό του. Έτσι παίρνουμε τα ίδια δεδομένα, σε ότι αφορά την ανάλυση Fourier, αλλά δεν υπάρχει κάποιος ντετερμινισμός σε αυτά. Αν τα αποτελέσματα της πρόβλεψης είναι τα ίδια (ή καλύτερα) για το 'τυχαιοποιημένο' σήμα συμπεραίνουμε ότι δεν υπήρχε κάποιος ντετερμινισμός εξαρχής στο σήμα μας ενώ αν είναι χειρότερα μπορούμε να πούμε ότι το αρχικό μας σήμα ήταν ντετερμινιστικό -οπότε δεν ήταν θόρυβος.

Παράρτημα Α

Συστήματα TSK

Το σύστημα Takagi Sugeno Kang

Το σύστημα Takagi Sugeno Kang (TSK)[24] χρησιμοποιείται κυρίως για αναγνώριση συστημάτων, προσαρμοστικό έλγχο, και για πρόβλεψη χρονοσειρών (που μας ενδιαφέρει στήν περίπτωσή μας). Από εδώ και στο εξής θεωρούμε ότι έχουμε R κανόνες, τα δεδομένα μας είναι D_e-διάστατα και έχουν πλήθος N.

Ο χώρος εισόδου του συστήματος TSK υφίσταται μία ασαφή κατάτμηση, έτσι ώστε κάθε διάνυσμα που βρίσκεται στην είσοδό του να μπορεί να ανήκει μέχρι ενός βαθμού σε διαφορετικές περιοχές του χώρου εισόδου. Για κάθε περιοχή που ορίζεται από την κατάτμηση δημιουργείται ένα μοντέλο εισόδου – εξόδου, το οποίο προσεγγίζει τοπικά τη δυναμική του συστήματος.

Η γενική μορφή ενός κανόνα του συστήματος TSK είναι:

$$\begin{aligned} \text{Κανόνας}i\text{: Αν}(\mathsf{x}_1 \text{ είναι } R_{1i}) \text{ και } (x_2 \text{ είναι } R_{1i}) \ldots \text{ και } (x_{D_e} \text{ είναι } R_{D_e i}) \\ \Rightarrow y = (A_i^T X) + b_i \end{aligned} \tag{A.1}$$

Οταν η έξοδος έιναι ένας γραμμικός συνδυασμός της εισόδου το σύστημα ονομάζεται TSK πρώτης τάξης (TSK–1), ενώ αν το διάνυσμα A_i παραληθφεί έχουμε ένα σταθερό όρο στην έξοδο, οπότε έχουμε ένα σύστημα TSK μηδενικής τάξης (TSK–0).

Για να υπολογιστεί η έξοδος του συστήματος, δοθέντος ενός διανύσματος εισόδου X υπολογίζεται αρχικά ο βαθμός συμμετοχής $\mu_i(X), i = 1 \cdots R$ του διανύσματος σε κάθε μία από τις R υποπεριοχές του χώρου εισόδου καθώς και οι έξοδοι των τοπικών μοντέλων $y_i = (A_i^T X) + b$. Για να υπολογιστεί η έξοδος του συστήματος παίρνουμε το ζυγισμένο μέσο των επιμέρους εξόδων, όπου τα βάρη προκύπτουν από τις συναρτήσεις συμμετοχής. Η έξοδος του συστήματος TSK μηδενικής τάξης είναι

$$f(X) = \frac{\sum_{i=1}^{R} \mu_i(X) \cdot b_i}{\sum_{i=1}^{R} \mu_i(X)} = \sum_{j=1}^{R} w_j \cdot b_j, \quad \text{όπου} \quad w_j = \frac{\mu_j(X_i)}{\sum_{i=1}^{R} \mu_i(X_i)} \tag{A.2}$$

ενώ του TSK πρώτης τάξης είναι

$$f(X) = \frac{\sum_{i=1}^{R} \mu_i(X) \cdot (A_i^T X + b_i)}{\sum_{i=1}^{R} \mu_i(X)} = \sum_{j=1}^{R} w_j \{A_j^T X_i + b_j\}, \quad \text{όπου} \quad w_j = \frac{\mu_j(X_i)}{\sum_{i=1}^{R} \mu_i(X_i)} \tag{A.3}$$

Με βάση τις σχέσεις αυτές γίνεται μία ομαλή μετάβαση από τη μία περιοχή του χώρου εισόδου στην άλλη, και μπορούμε να προσεγγίζουμε ικανοποιητικά περίπλοκες μη-γραμμικές συναρτήσεις, χρησιμοποιώντας πολύ απλά μοντέλα, με μικρό αριθμό ελεύθερων παραμέτρων. Αντιθέτως, αν θέσουμε αυστηρά όρια μεταξύ των περιοχών του χώρου εισόδου, (όπως γίνεται όταν χρησιμοποιούμε π.χ. τοπικά πολυώνυμα) αφ'ενός δεν παίρνουμε μία συνεχή συνάρτηση, ενώ μπορεί να χρειαστεί πολύ μεγαλύτερος αριθμός παραμέτρων για να πάρουμε παρόμοια αποτελέσματα.

Κατάτμηση του χώρου εισόδου

Για την κατάτμηση του χώρου εισόδου υπάρχουν δύο επιλογές:

(1) Μπορούμε να χρησιμοποιήσουμε μονοδιάστατες συναρτήσεις συμμετοχής και να κατατμήσουμε κάθε διάσταση ξεχωριστά, οπότε η συνάρτηση συμμετοχής ενός στοιχείου καθορίζεται από το γινόμενο των ποσοστών συμμετοχής κάθε συνιστώσας του στα αντίστοιχα ασαφή σύνολα, δηλαδή έχουμε

$$\mu_i(X) = \prod_{i=1}^{D_e} \mu_{R_i}(X_i) \tag{A.4}$$

όπου η μ μπορεί να είναι οποιαδήποτε μονοδιάστατη συνάρτηση συμμετοχής, όπως π.χ. η Gaussian , η τραπεζοειδής ή η τριγωνική. Η μέθοδος αυτή προτιμάται όταν η σημασία των δεδομένων διαφοροποιείται από διάσταση σε διάσταση, π.χ. για έναν κλιματιστικό, όπου η έξοδος είναι η παρεχόμενη ισχύς στο σύστημα ψύξης, όταν η πρώτη διάσταση είναι η εξωτερική θερμοκρασία, η δεύτερη είναι η θερμοκρασία του εσωτερικού χώρου η τρίτη η προσδιορισμένη από το χρήστη θερμοκρασία κ.ο.κ.. Στην περίπτωση αυτή έχει νόημα να κατατμήσουμε χωριστά την κάθε διάσταση.

(2) Μπορούμε να χρησιμοποιήσουμε πολυδιάστατες συναρτήσεις συμμετοχής, όπως π.χ. μία πολυδιάστατη Gaussian κέντρου C_i και διασποράς σ_i:

$$\mu_i(X) = e^{\frac{|X-C_i|^2}{2\sigma_i^2}} = e^{\left\{\frac{\sum_{j=1}^{D} e(X_j - C_{ji})^2}{2\sigma_i^2}\right\}} = \prod_{j=1}^{D_e} e^{\left\{\frac{(X_j - C_{ji})}{2\sigma_i^2}\right\}} \tag{A.5}$$

Οι πολυδιάστατες συναρτήσεις συμμετοχής γενικότερα προτιμώνται στην περίπτωση που έχουμε δεδομένα τα οποία είναι 'καθαρά' από σημασία, όπως π.χ. αυτά τα οποία προκύπτουν από την εμβύθιση ενός παρατηρούμενου σήματος, οπότε τα δεδομένα σε όλες τις διαστάσεις είναι ίδιας φύσης. Στην περίπτωση αυτή είναι προτιμότερο να δημιουργήσουμε τοπικά μοντέλα στις περιοχές εκείνες του χώρου εισόδου όπου είναι πυκνότερα τα δεδομένα (συστάδες δεδομένων), οπότε εξασφαλίζουμε μεγαλύτερη ακρίβεια και μικρότερο αριθμό παραμέτρων. Π.χ. για ένα απλό σύστημα το οποίο έχει εμβυθιστεί σε 5 διαστάσεις αν κατατμήσουμε την κάθε διάσταση χωριστά σε δύο περιοχές θα κατασκεύασουμε $2^6 = 64$ τοπικά μοντέλα, ενώ μπορεί και 3 ή 4 να αρκούν. Για το λόγο αυτό χρησιμοποιήσαμε πολυδιάστατες συναρτήσεις συμμετοχής και συγκεκριμένα Gaussians που έχουν το ιδαίτερα χρήσιμο χαρακτηριστικό ότι η παράγωγός τους είναι συνεχής (σε αντίθεση π.χ. με τις τριγωνικές ή τις τραπεζοειδείς συναρτήσεις συμμετοχής).

Έχοντας προσδιορίσει τις συναρτήσεις σμμετοχής οι παράμετροι A_i, b_i υπολογίζονται χρησιμοποιώντας τις σχέσεις που ισχύουν και για τα συστήματα WCP,WLP.

Καθορισμός του πλήθους των συναρτήσεων συμμετοχής -SVD

Η τακτική που ακολουθείται για τον καθορισμό των σημάτων ενεργοποιήσης που τελικά συμμετέχουν στην έκφραση για την έξοδο του συστήματος είναι να έχουμε αρχικά ένα μεγάλο αριθμό συναρτήσεων συμμετοχής και να επιλέξουμε τα πλέον χρήσιμα σήματα ενεργοποίησης.

Η μέθοδος που χρησιμοποιείται για την επιλογή ενός υποσυνόλου των σημάτων ενεργοποίησης βασίζεται στην ανάλυση του πίνακα ενεργοποίησης P χρησιμοποιώντας την τεχνική SVD (Singular Value Decomposition)[40], με την οποία μπορούμε να απαλλαχθούμε από εκείνα τα σήματα ενεργοποίησης τα οποία είτε επικαλύπτονται με άλλα σήματα ενεργοποίησης ή που παίζουν μικρό ρόλο στον καθορισμό της έξοδου του προβλέπτη. Αυτό μπορεί να γίνει γράφοντας τον $N \times M$ πίνακα P ως

$$P = U\Sigma V^T, \quad \text{όπου} \quad U : N \times M, \quad \Sigma : M \times M \quad V : M \times M \tag{A.6}$$

Ο πίνακας Σ είναι διαγώνιος και οι πίνακες V,U έχουν ορθογώνιες στύλες. Με βάση την παραπάνω σχέση ο πίνακας P αποσυντίθεται σε μία *ορθογώνια* βάση V στην οποία μπορούν να εκφράστούν τα σήματα ενεργοποιήσεων, ένα διαγώνιο πίνακα Σ του οποίου το στοιχείο Σ_{ii} δείχνει πόσο σημαντική είναι η στύλη V_i του πίνακα V στο σχηματισμό των σημάτων ενεργοποίησης και έναν πίνακα U ο οποίος δείχνει με ποιο συνδυασμό από τα 'ζυγισμένα' με τον Σ διανύσματα του πίνακα V παίρνουμε τα σήματα ενεργοποίησης για το κάθε σημείο. Σημειώνεται εδώ ότι οι όποιες συσχετίσεις των αρχικών σημάτων ενεργοποίησης έχουν αναιρεθεί, καθώς η βάση V είναι ορθογώνια.

Αυτές οι στύλες του V στις οποίες αντιστοιχεί μικρό Σ_{ii} είναι εκείνες οι κατευθύνσεις του χώρου R που παίζουν το μικρότερο ρόλο στον καθορισμό των σημάτων ενεργοποίησης και το αντίστροφο. Αφαιρώντας τις κατευθύνσεις αυτές θα μπορούσαμε να έχουμε μικρή αύξηση του σφάλματος, ενώ θα μειώναμε το χρησιμοποιούμενο πλήθος σημάτων ενεργοποίησης. Συγκεκριμένα, αν τα στοιχεία της διαγωνίου του Σ είναι (σε φθίνουσα σειρά)

$$\sigma_1, \sigma_2, \ldots, \sigma_M, \quad \text{όπου} \quad \sigma_i = \Sigma_{ii} \tag{A.7}$$

και εμείς κρατήσουμε μόνο τις r πρώτες από αυτές, παίρνουμε μία προσέγγιση της LS λύσης του προβλήματος η οποία ισούται με

$$\tilde{\theta}'_{LS} = \sum_{i=1}^{r} \frac{U_i^T \vec{y}}{\sigma_i} V_i \qquad \& \qquad \rho_{LS} = \|\vec{y} - P\vec{\theta}\|_2^2 = \sum_{i=r+1}^{M} (U_i^T \vec{y})^2 \tag{A.8}$$

όπου ρ_{LS} είναι το ελάχιστο υπόλοιπο λάθους. Το $\vec{\theta}'_{LS}$ είναι η LS λύση του προβλήματος

$$\vec{y} = \vec{\theta}' P', \quad P' = \sum_{i=1}^{r} \sigma_i U_i V_i^T \tag{A.9}$$

Πρέπει να σημειωθεί ότι πολλές φορές μπορεί να επιτευχθεί και ελάττωση του σφάλματος καθώς οι στύλες που αντιστοιχούν σε μικρό Σ_{ii} μπορούν να οδηγήσουν και σε πίνακες P κακής κατάστασης, με αποτέλεσμα να πάρουμε ασταθείς λύσεις.

Καθώς για καινούρια δεδομένα δεν έχουμε τις αντι΄στοιχες τιμές του U για να κάνουμε πρόβλεψη μέσω των πινάκων U, Σ, V πρέπει να κάνουμε πρόβλεψη χρησιμοποιώντας

στύλες του πίνακα P οπότε χρειάζεται να βρούμε πόσες στύλες θα χρησιμοποιήσουμε για την πρόβλεψη και ποιες στύλες είναι οι περισσότερο ανεξάρτητες μεταξύ τους.

Ο αριθμός των στυλών του P που θα συμμετάσχουν στον τελικό πρόβλέπτη καθορίζεται από την effective rank του πίνακα P, που μπορεί να την υπολογιστεί από την SVD αποσύνθεση του P αν μηδενίσουμε όλες τις τιμές του πίνακα Σ που είναι μικρότερες από ένα όριο, π.χ. $\Sigma_{ii} < 10^{-2}\Sigma_{11}$. Η τάξη r του πίνακα Σ μετά από το μηδενισμό των μικρότερων ιδιοτιμών του δίνει μία εκτίμηση της effective rank του πίνακα P που ονομάζεται numerical rank (αριθμητική τάξη). Έχοντας βρεί πόσες στύλες του πίνακα P χρησιμεύουν τελικά για την πρόβλεψη, απομένει να εντοπίσουμε ποιές θα χρησιμοποιήσουμε (προς το παρόν έχουμε δουλέψει μόνο με τους πίνακες U, Σ, V). Αποδεικνύεται [40] ότι αν ο Π είναι ένας $N \times M$ πίνακας αναδιάταξης και έχουμε

$$\begin{array}{rcccc} P\Pi = & \underset{r}{[B_1} & \underset{n-r}{B_2]} & \& & P^TV = \end{array} \begin{array}{c} \left[\begin{array}{cc} \tilde{V}_{11} & \tilde{V}_{12} \\ \tilde{V}_{21} & \tilde{V}_{22} \end{array}\right] \\ \begin{array}{cc} r & n-r \end{array} \end{array} \begin{array}{c} r \\ n-r \end{array} \tag{A.10}$$

τότε έχουμε

$$\rightarrow \sigma_r(B_1) \geq \frac{\sigma_r(P)}{\|\tilde{V}_{11}^{-1}\|_2} \tag{A.11}$$

όπου $\sigma_r(A)$ είναι η μικρότερη ιδιοτιμή του πίνακα A. Αυτό το οποίο μπορούμε να χρησιμοποιήσουμε από αυτό το θεώρημα είναι το συμπέρασμα ότι επιλέγοντας μία τέτοια αναδιάταξη Π των στοιχείων του πίνακα P ώστε ο αντίστοιχος υποπίνακας $\tilde{V}_{11}$ που προκύπτει εφαρμόζοντας την ίδια αναδιάταξη στον πίνακα V να είναι καλής κατάστασης (που σημαίνει σχετικά μικρό $\|V_{11}\|^{-1}$), εξασφαλίζουμε ότι ο πίνακας B_1 θα είναι και αυτός καλής κατάστασης (που προκύπτει από το σχετικά μεγάλο κάτω όριο για το $\sigma_r(B_1)$). Οι στύλες λοιπόν του B_1 είναι σχετικά ανεξάρτητες και είναι οι στύλες του P που θα επιλεγούν για να χρησιμοποιηθούν στον προβλέπτη.

Για την επιλογή ενός συγκεκριμένου πίνακα μετάθεσης Π χρησιμοποιείται ευρεστικά η τεχνική QR αποσύνθεσης, η οποία εφαρμόζεται στον πίνακα $[V_{11}^T, V_{21}^T]$, όπου

$$V = \begin{array}{c} \left[\begin{array}{cc} V_{11} & V_{12} \\ V_{21} & V_{22} \end{array}\right] \\ \begin{array}{cc} r & n-r \end{array} \end{array} \begin{array}{c} r \\ n-r \end{array} \tag{A.12}$$

Συγκεκριμένα εφαρμόζοντας την QR αποσύνθεση έχουμε

$$Q^T[V_{11}^TV_{21}^T]\Pi = [R_{11}R_{12}] \rightarrow \left[\begin{array}{c} \tilde{V}_{11} \\ \tilde{V}_{21} \end{array}\right] = \Pi^T \left[\begin{array}{c} V_{11} \\ V_{21} \end{array}\right] = \left[\begin{array}{c} R_{11}^TQ^T \\ R_{12}^TQ^t \end{array}\right] \tag{A.13}$$

Συνεπώς έχουμε $\|\tilde{V}_{11}^{-1}\|_2 = \|R_{11}^{-1}\|_2$ οπότε καθώς ο πίνακας R_{11} είναι καλής κατάστασης (από την τεχνική QR) ευριστικά έχουμε έναν καλό πίνακα Π. Η τεχνική που παρουσιάστηκε ονομάζεται τεχνική SVD -QR και είναι από τις ευρύτερα διαδεδομένες στο χώρο των ασαφών συστημάτων για το σκοπό επιλογή υποσυνόλου κανόνων˙ έχει δεχτεί διάφορες βελτιώσεις, αλλά ουσιαστικά ο πυρήνας είναι ο ίδιος.

Διάφορες άλλες τεχνικές έχουν προταθεί για την επιλογή ενός υποσυνόλου κανόνων, που χρησιμοποιούν μεθόδους της θεωρίας πληροφορίας, της στατιστικής ακόμα και γενετικούς αλγόριθμους, αλλά καθώς η SVD-QR έδωσε καλά αποτελέσματα, δεν κρίθηκε

σκόπιμο να εξετασθούν όλες οι δυνατές τεχνικές. Εξάλλου αρκετά από τα λάθη που ενδεχομένως να γίνουν μπορούν να διορθωθούν με τον αλόριθμο back propagation που παρουσιάζεται στη συνέχεια.

Είναι χρήσιμο να αναφερθούμε εδώ στο ρόλο που παίζει η παράμετρος σ στην κατάσταση (λόγος μεγαλύτερης προς μικρότερη ιδιοτιμή) του πίνακα P: Μία μεγάλη σ σημαίνει μία απλωμένη συνάρτηση συμμετοχής, που μπορεί να έχει σημαντική επικάλυψη με μία άλλη συνάρτηση συμμετοχής. Συνεπώς τα σήματα ενεργοποίησής τους θα έχουν σημαντική ομοιότητα, οπότε ο πίνακας σημάτων ενεργοποίησης θα έχει τουλάχιστον δύο στύλες του παρόμοιες. Αυτό σημαίνει, όμως, τουλάχιστον μία πολύ μικρή ιδιοτιμή και συνεπώς κακή κατάσταση του πίνακα P. Από την άλλη, αν μία συνάρτηση συμμετοχής έχει μικρή σ και οι υπόλοιπες μεγάλη, επόμενο είναι το σήμα ή τα σήματα ενεργοποίησης που αντιστοιχούν στη συνάρτηση αυτή να είναι πολύ ασθενέστερα από ότι αυτά των υπόλοιπων συναρτήσεων συμμετοχής (η συνάρτηση αυτή θα ενεργοποιείται μόνο όταν ένα διάνυσμα εισόδου βρίσκεται πολύ κοντά στο κέντρο της, ενώ οι υπόλοιπες θα έχουν σχετικά ισχυρή έξοδο για μία ευρύτερη περιοχή γύρω από το κέντρο τους. Συνεπώς, αυτή η συνάρτηση συμμετοχής αν και μπορεί να μην δίνει σήματα ενεργοποίησης που να επικαλύπτονται με άλλα σήματα ενεργοποίησης είναι πολύ πιθανό να θεωρηθεί πλεονάζουσα, λόγω του μικρού, συνολικά, ρόλου που παίζει στον καθορισμό της εξόδου του προβλέπτη.

Με βάση τις παραπάνω παρατηρήσεις ορίσαμε τις αρχικές διασπορές των συναρτήσεων συμμετοχής να είναι ίσες με ένα ποσοστό της διασποράς των δεδομένων στο χώρο εισόδου R, έτσι ώστε να μην είναι ούτε υπερβολικά απλωμένες ούτε πολύ συγκεντρωμένες. Για το μοντέλο TSK-0 που έχει και μικρότερο αριθμό σημάτων ενεργοποίησης (M έναντι $M + Md$ του TSK-1) αφήσαμε μία σχετικά μεγαλύτερη διασπορά των συναρτήσεων συμμετοχής, ίση με το $0.6 \cdot R$, ενώ για το TSK-1 τέθηκε αρχικά ίση με $0.5 \cdot R$.

Επίσης, κατά την ενημέρωση των διασπορών με βάση τον αλγόριθμο back propagation που θα παρουσιαστεί παρακάτω, είναι δυνατό μία διασπορά να οδηγηθεί σε αρκετά μεγάλη τιμή, που ενδεχομένως να μας δώσει και μικρότερο MSE. Παρ'αυτα, προκειμένου να έχουμε πίνακα P καλύτερης κατάστασης, προτιμήσαμε να περιορίσουμε τις διασπορές των συναρτήσεων συμμετοχής σ_i με βάση τη διασπορά των δεδομένων στο χώρο εισόδου R ως εξής: $\quad 0.1 \cdot R < \sigma_i < R, \quad \forall i$

Νευροασαφή συστήματα

Μία εναλλακτική προσέγγιση στον υπολογισμό των παραμέτρων των συστημάτων TSK-0,1 είναι αντί να ελαχιστοποιείται το MSE της πρόβλεψης, θεωρώντας δεδομένες τις τιμές των κέντρων και των διασπορών των Gaussians που χρησιμοποιούνται ως συναρτήσεις συμμετοχής, να προσπαθούμε παράλληλα να μάθουμε τις κατάλληλες τιμές των κέντρων C_i και των διασπορών σ_i καθώς και τις τιμές των παραμέτρων A_i, b_i. Για το σκοπό αυτό χρησιμοποιείται ο αλγόριθμός back-propagation για την ενημέρωση των παραμέτρων που επηρεάζουν τη συνάρτηση λάθους με βάση την αναλυτική της έκφραση [25]. Εδώ υπάρχουν (τουλάχιστον) δύο διαφορερικές προσεγγίσεις:

1. Σύμφωνα με την πρώτη ενημερώνονται όλες οι παράμετροι με βάση τον αλγόριθο back-propagation, οπότε δεν εκμεταλλευόμαστε τη γραμμική σχέση που συνδέει τα σήματα ενεργοποίησης με τις τιμές της εξόδου του συστήματος.

2. Σύμφωνα με τη δεύτερη, στο πρώτο στάδιο κάθε βήματος , ενημερώνονται μόνο οι

τιμές των κέντρων και των διασπορών των συμαρτήσεων συμμετοχής και οι τιμές των παραμέτρων A_i, b_i θεωρούνται, προσωρινά σταθερές. Στο δεύτερο στάδιο θεωρούμε τα σήματα ενεργοποίησης σταθερά και εφαρμόζουμε τις κανονικές εξισώσεις για να εξάγουμε τις νέες τιμές των παραμέτρων A_i, b_i.

Ο δεύτερος αλγόριθμος προτιμάται γενικότερα, γιατί από τις κανονικές εξισώσεις καταλήγουμε απευθείας στο ελάχιστο της συνάρτησης κόστους (στο μέρος της που αφορά τις παραμέτρους A, b), ενώ ο αλγόριθμος back-propagation είναι στοχαστικής φύσης και μπορεί να αντιμετωπίσει διάφορα προβλήματα (να παγιδευτεί σε τοπικά ελάχιστα κ.α.)

Παρακάτω παρουσιάζουμε τις σχέσεις με τις οποίες υπολογίζεται η μερική παράγωγος του συνολικού MSE ως προς τις παραμέτρους C_i, σ_i ενός συστήματος TSK-1 (η περίπτωση του TSK-0 προκύπτει αν απλά αγνοήσουμε τους όρους που περιέχουν το διάνυσμα A). Ως το σημείο αυτό δεν έχει υπάρξει κάποια διαφοροποίηση στο σύστημα TSK στο οποίο καταλήγουμε ανάλογα με τη διάσταση του χώρου εξόδου του συστήματος. Τόσο η τεχνική SVD όσο και οι κανονικές εξισώσεις δίνουν τα ίδια αποτελέσματα, είτε θεωρήσουμε ότι κάνουμε πρόβλεψη σε μία διάσταση ή σε D_e. Εδώ όμως προκύπτει ένα διαφορετικό σύστημα καθώς ο όρος λάθους προκύπτει από το MSE στο σύνολο των εξόδων του συστήματος, οπότε αν χρησιμοποιήσουμε ένα σύστημα TSK για κάθε έξοδο εν γένει θα καταλήξουμε σε διαφορετικές συναρτήσεις συμμετοχής για κάθε έξοδο από ότι αν χρησιμοποιήσουμε ένα σύστημα TSK για όλες τις εξόδους. Εμείς θεωρούμε την περίπτωση όπου γίνεται πρόβλεψη και στις D_e διαστάσεις, ενώ οι εξισώσεις στην περίπτωση όπου γίνεται πρόβλεψη σε μία διάσταση προκύπτουν απλά αφαιρώντας τον όρο αθροίσματος, ($\sum_{j=1}^{D} {}_e$). Στους τύπους που ακολουθούν γράφοντας π.χ. A_{ji} εννοούμε το διάνυσμα A που χρησιμοποιείται από τον i-οστό κανόνα για την πρόβλεψη της j-οστής εξόδου (ομοίως και για το b_{ji}).

Ο αλγόριθμος back-propagation για το σύστημα TSK

$$
\begin{aligned}
E(X) &= \sum_{j=1}^{D} {}_e \frac{1}{2}(y_j - f_j(X))^2 \rightarrow \\
\frac{\partial E}{\partial a_k} &= \sum_{j=1}^{D} {}_e (y_j - f_j(X)) \frac{\partial f_j(X)}{\partial a_k} == \sum_{j=1}^{D} {}_e (y_j - f_j(X)) \cdot \frac{\partial \left\{ \frac{\sum_{i=1}^{R} \mu_i(X)(A_{ji}^T X + b_{ji})}{\sum_{i=1}^{R} \mu_i(X)} \right\}}{\partial a_k} = \\
&= \sum_{j=1}^{D} {}_e (y_j - f_j(X)) \cdot \frac{\frac{\partial \sum_{i=1}^{R} \mu_i(X)(A_{ji}^T X + b_{ji})}{\partial a_k} \cdot \sum_{i=1}^{R} \mu_i(X) - \left\{ \sum_{i=1}^{R} \mu_i(X)(A_{ji}^T X + b_{ji}) \right\} \frac{\partial \sum_{i=1}^{R} \mu_i(X)}{\partial a_k}}{\left[\sum_{i=1}^{R} \mu_i(X) \right]^2} \\
&= \sum_{j=1}^{D} {}_e (y_j - f_j(X)) \cdot \frac{\sum_{i=1}^{R} \frac{\partial \mu_i(X)}{\partial a_k} (A_{ji}^T X + b_{ji}) \cdot \sum_{i=1}^{R} \mu_i(X) - \left\{ \sum_{i=1}^{R} \mu_i(X)(A_{ji}^T X + b_{ji}) \right\} \sum_{i=1}^{R} \frac{\partial \mu_i(X)}{\partial a_k}}{\left[\sum_{i=1}^{R} \mu_i(X) \right]^2}
\end{aligned}
$$

Στην περίπτωση όπου η συνάρτηση συμμετοχής είναι Gaussian έχουμε

$$\mu_i(X) = e^{\left\{-\frac{\sum_{i=1}^{D} e(X_i - C_i)^2}{\sigma_i^2}\right\}} \Rightarrow$$

$$a_k = \sigma_k \rightarrow \delta_{ik} = \frac{\partial \mu_i(X)}{\partial \sigma_k} = \begin{cases} \frac{2\sum_{j=1}^{D} e(X_j - C_j)^2}{\sigma_k^3} \cdot e^{\left\{-\frac{\sum_{j=1}^{D} e(X_j - C_j)^2}{\sigma_k^2}\right\}} & i = k \\ 0 & i \neq k \end{cases}$$

$$a_k = C_k \rightarrow \Delta_{ik} = \frac{\partial \mu_i(X)}{\partial C_k} = \begin{cases} -\frac{2(X - C_k)}{\sigma_k^2} \cdot e^{\left\{-\frac{\sum_{j=1}^{D} e(X_j - C_j)^2}{\sigma_k^2}\right\}} & i = k \quad (C_k : \text{Διάνυσμα!}) \\ 0 & i \neq k \end{cases}$$

Η τελευταία σχέση μπορεί να γραφεί για κάθε συνιστώσα του διανύσματος C_i ξεχωριστα, αλλά γράφηκε έτσι για λόγους απλότητας (το Δ είναι διανυσματική ποσότητα). Οι εξισώσεις του αλγορίθμου back-propagation γράφονται μς βάση τις παραπάνω σχέσεις ως εξής:

- Έξάρτηση του λάθους από την παράμετρο σ_k (δεδομένου ότι για $i \neq k \rightarrow \delta_{ik} = 0$):

$$\begin{aligned} \frac{\partial E(X)}{\partial \sigma_k} &= \sum_{j=1}^{D} e \left\{ (y_j - f_j(X)) \cdot \frac{\delta_{kk}(X)(A_{jk}^T X + b_{jk}) \cdot \sum_{i=1}^{R} \mu_i(X) - \sum_{i=1}^{R} \left\{ \mu_i(X)(A_{ji}^T X + b_{ji}) \right\} \cdot \delta_{kk}(X)}{\left[\sum_{i=1}^{R} \mu_i(X) \right]^2} \right\} \\ &= \delta_{kk} \cdot \frac{\sum_{j=1}^{R} (y_j - f_j(X)) \cdot \left\{ (A_{jk}^T X + b_{jk}) \cdot \sum_{i=1}^{R} \mu_i(X) - \sum_{i=1}^{R} \left\{ \mu_i(X)(A_{ji}^T X + b_{ji}) \right\} \right\}}{\left[\sum_{i=1}^{R} \mu_i(X) \right]^2} \end{aligned} \quad (A.14)$$

- Εξάρτηση λάθους από την παράμετρο C_{ki} (k-οστή συνιστώσα i-οστού κέντρου): Η σχέση που προκύπτει είναι πανομοιότυπη με την προηγούμενη, με τη διαφορά ότι αντί της ποσότητας d_{kk} χρησιμοποιούμε τη διανυσματική ποσότητα Δ_{kk}

Χρησιμοποιώντας τις παραπάνω σχέσεις μπορούμε να γράψουμε τους κανόνα ενημέρωσης των παραμέτρων του συστήματος ως εξής:

$$\begin{aligned} \sigma_k(n+1) &= \sigma_k(n) - \nu_1 \frac{\partial J}{\partial \sigma_k(n)} \\ C_i(n+1) &= C_i(n) - \nu_2 \frac{\partial J}{\partial C_i(n)} \end{aligned} \quad (A.15)$$

όπου ν_1, ν_2 οι ρυθμοί μάθησης για τις δύο παραμέτρους του συστήματος. Υπό τη συνθήκη οι ν_1, ν_2 να είναι αρκετά μικροί , ο αλγόριθμος συγκλίνει σε ένα ελάχιστο - αν και όχι πάντα στο καθολικό ελάχιστο. Μία επιλογή που γίνεται συχνά είναι να επιλεγεί ένας αρχικά σχετικά μεγάλος ρυθμός μάθησης, ο οποίος μετά από ένα πλήθος επαναλήψεων να μειώνεται. Ετσι αρχικά γίνονται οι μεγάλες αλλαγές που πρέπει να γίνουν και κατόπιν γίνεται ένα fine-tuning αποφεύγοντας τις 'ταλαντώσεις' που μπορεί να προκληθούν από μία μεγάλη τιμή των ρυθμών μάθησης. Επίσης για να αποφευχθούν οι ταλαντώσεις οι οποίες μπορεί να οφείλονται στην αλλαγή από το ένα βήμα στο άλλο του προσήμου της μεταβολής μίας παραμέτρου, χρησιμοποιείται και μία παράμετρος 'ορμής' m η οποία συμμετέχει στην ενημέρωση των παραμέτρων ως εξής:

$$\delta w_k(n) = m \cdot \delta w_k(n-1) + \frac{\partial J}{\partial w_k(n)} \tag{A.16}$$

Οι ευρεστικές αυτές δεν είναι πάντα τόσο αποτελεσματικές όσο θα έπρεπε να είναι , οπότε χρησιμοποιήθηκε τελικά ένας λίγο διαφορετικός αλγόριθμος ο οποίος παρουσιάζεται παρακάτω.

Μετά από κάθε ενημέρωση των τιμών των κέντρων και των διασπορών των Gaussian συναρτήσεων συμμετοχής πρέπει να επαναπροσδιοριστούν οι τιμές των παραμέτρων των τοπικών μοντέλων. Για να αποφεύγεται ο επαναπροδιορισμός των τοπικών μοντέλων σε κάθε βήμα προτιμάται να χρησιμοποιείται η batch μέθοδος ενημέρωσης των παραμέτρων των συναρήσεων συμμετοχής όπου όλοι οι όροι που αναφέρθηκαν παραπάνω προκύπτουν ως μέσοι σε όλο το training set.

Ενημέρωση βαρών- αλγόριμθος rprop

Ένα πρόβλημα το οποίο αντιμετωπίσαμε ήταν ο προσδιορισμός του ρυθμού μάθησης: έτσι όπως είναι διατυπωμένες οι εξισώσεις ενημέρωσης των βαρών η μεταβολή της κάθε παραμέτρου καθορίζεται από το μέτρο της μερικής παραγώγου της συνάρτησης σφάλματος το οποίο μπορεί να γίνει εξαιρετικά μικρό όταν υπάρχουν πολλές συναρτήσεις συμμετοχής (αφού η κάθε μία παίζει σχετικά μικρό ρόλο στο συνολικό σφάλμα) και αντιστρόφως. Διάφορες ευρεστικές που αναζητήθηκαν για τον καθορισμό της παραμέτρου μάθησης (με χρήση όρων ορμής, μεταβαλλόμενου ρυθμού μάθησης) δεν απέδωσαν τα αναμενόμενα, καθώς όλες εξαρτώνται από τον αρχικό ρυθμό μάθησης. Για το λόγο αυτό χρησιμοποιήθηκε μία παραλλαγή του αλγορίθμου back-propagation, που ονομάζεται rprop ($r$esilient $prop$agation) και παρουσιάζεται στο [41]. Η τεχνική αυτή έχει εφαρμοστεί σε neuro -fuzzy συστήματα (π.χ. [42]) με επιτυχία και έδωσε και στην περίπτωσή μας τα καλύτερα αποτελέσματα.

Ο αλγόριθμος RPROP χρησιμοποιεί μόνο το πρόσημο της μερικής παραγώγου του κόστους ως προς την εκάστοτε μεταβλητή για να καθορίσει τη μεταβολή της μεταβλητής αυτής, αποφεύγοντας να χρησιμοποιήσει και το μέτρο της. Το μέτρο της μεταβολής τίθεται αρχικά ίσο με μία σχετικά μικρή σταθερά και σε κάθε βήμα μεταβάλλεται, ανάλογα με το αν σημειώθηκε μεταβολή στο πρόσημο της παραγώγου ή όχι:

- Αν διατηρηθεί ίδιο το πρόσημο της μερικής παραγώγου σημαίνει ότι βρισκόμαστε στη σωστή κατεύθυνση αναζήτησης, οπότε η μεταβολή της παραμέτρου γίνεται λίγο μεγαλύτερη, ώστε να επιταχυνθεί η αναζήτηση προς την σωστή κατεύθυνση (πολλαπλασιασμός μέτρου μεταβολής παραμέτρου με συντελεστή $\eta^+ > 1$).

- Αλλαγή στο πρόσημο της παραγώγου σημαίνει ότι η μεταβολή στο προηγούμενο βήμα ήταν τόσο μεγάλη ώστε προσπεράστηκε ένα (τοπικό) ελάχιστο και πήγαμε από την άλλη πλευρά της κοιλάδας της συνάρτησης κόστους, οπότε πρέπει να οπισθοχωρήσουμε. Συνεπώς το μέτρο της μεταβολής πρέπει να μικρύνει ώστε να πλησιάσουμε το τοπικό ελάχιστο το οποίο βρίσκεται μεταξύ της τρέχουσας τιμής της παραμέτρου και της προηγούμενής της (πολλαπλασιασμός μέτρου μεταβολής παραμέτρου με συντελεστή $\eta^- < 1$).

Οι τιμές που χρησιμοποιήθηκαν είναι ελάχιστα διαφορετικές από αυτές πού έχουν προταθεί στο [41] και είναι $\eta^- = 0.5, \eta^+ = 1.05$. Ο αλγόριθμος ενημέρωσης των βαρών για τον αλγόριθμο RPROP είναι τελικά:

- Αρχικοποίηση των μέτρων της μεταβολής των παραμέτρων

$$\Delta_i^1 = 0.1, \quad i = 1 \cdots P \quad (= \#parameters) \tag{A.17}$$

- Επανάλαβε εως ότου να επιτευχθεί σύγκλιση (ή να ξεπεραστεί max πλήθος επαναλήψεων)
  $n = n + 1$,

$$\Delta_i^n = \begin{cases} \eta^+ \cdot \Delta_i^{n-1} & \text{αν} \quad \frac{\partial J}{\partial w_i}^n \cdot \frac{\partial J}{\partial w_i}^{n-1} > 0 \\ \eta^- \cdot \Delta_i^{n-1} & \text{αν} \quad \frac{\partial J}{\partial w_i}^n \cdot \frac{\partial J}{\partial w_i}^{n-1} < 0 \\ \Delta_i^{n-1} & \text{αν} \quad \frac{\partial J}{\partial w_i}^n \cdot \frac{\partial J}{\partial w_i}^{n-1} = 0 \end{cases} \tag{A.18}$$

$$w_i^n = w_i^{n-1} - \text{sign}\left\{ \frac{\partial J}{\partial w_i}^n \right\} \cdot \Delta_i^n \tag{A.19}$$

Μία μικρή τροποποίηση του αλγορίθμου οφείλεται στο γεγονός ότι έχει προταθεί για νευρωνικά δίκτυα, όπου έχει νόημα να θέσουμε την αρχική τιμή του μέτρου της μεταβολής ενός βάρους ίση με 0.1. Στην περίπτωσή μας μία αλλαγή της διασποράς ίση με 0.1 είναι πολύ μεγάλη για να γίνει στο πρώτο βήμα, και μπορεί να καθυστερήσει τον αλγόριθμο. Αυτό το οποίο πήραμε ως αρχική μεταβολή για τις διασοπορές των συναρτήσεων συμμετοχής είναι το $1/20$ της αρχικής διασποράς και για τα κέντρα των συναρτήσεων συμμετοχής το $1/20$ της απόστασης των κέντρων από το μέσο των δεδομένων στο χώρο εισόδου. Με αυτές τις επιλογές μπορέσαμε με 5 επαναλήψεις του αλγορίθμου rprop να επιτύχουμε σημαντική μείωση του σφάλματος πρόβλεψης, από 75% εως και 95%

Τελειώνοντας, δίνουμε συνοπτικά τον τελικό αλγόριθμο εκπαίδευσης ενός TSK

- Πρώτο στάδιο -καθορισμός αρχιτεκτονικής συστήματος
  - Ομαδοποίηση δεδομένων χρησιμοποιώντας ένα σχετικά μεγάλο πλήθος συστάδων (π.χ. $R = 30$).
  - Σχηματισμός πίνακα σημάτων ενεργοποίησης P

$$\begin{aligned} P &= [P_1, P_2, \cdots, P_N] \quad N \times M (= R + R \cdot D_e) \\ P_i &= [w_1, w_1 \cdot X_{1i}, \cdots, w_1 \cdot X_{D_e i}, w_2, w_2 \cdot X_{1i}, \text{-}, w_2 \cdots, X_{D_e i} \cdots w_R, \cdots, w_R \cdot X_{1i}, \cdots, w_R \cdot X_{D_e 1}]^T \end{aligned}$$

  - SVD ανάλυση του P -καθορισμός efficient rank του P:

$$\begin{aligned} P &= U \Sigma V^T, \quad U : N \times M, \quad \Sigma : M \times M \quad V : M \times M \\ \forall i : \quad \Sigma_{ii} &< q \cdot \Sigma_{11} \quad \Rightarrow \quad \Sigma'_{ii} \leftarrow 0 \\ r = \text{rank}(P) &= \#(\Sigma'_{ii} \neq 0) \end{aligned}$$

  - QR-αποσύνθεση του $r \times M$ πίνακα $[V_{11}, V_{21}$ - επιλογή $r \times N$ υποπίνακα B_1 του πίνακα σημάτων ενεργοποίησης του P για την επίλυση των κανονικών εξισώσεων:

$$Q \cdot [V_{11}^T V_{21}^T] \cdot \Pi = R \Rightarrow B_1 = P' \Pi$$

  - Επίλυση κανονικών εξισώσεων:

$$\vec{\theta_i}' = (B_1^T B_1)^{-1} (B_{1i}) i = 1 \ldots D_e$$

- Δεύτερο στάδιο - back-propagation:
  Εως ότου να έχουμε σύγκλιση (ή να ξεπεραστεί μέγιστο πλήθος βημάτων) επανάλαβε
  - Υπολογισμός διεύθυνσης μεταβολής βάρους στο βήμα n

$$\begin{aligned} \delta\sigma_k(n) &= \frac{\partial J}{\partial \sigma_k} \\ \delta C_k(n) &= \frac{\partial J}{\partial C_k} \qquad k = 1 \ldots R \end{aligned} \tag{A.20}$$

  - Ενημέρωση βαρών με αλγόριθμο RPROP για το n -οστό βήμα:

$$w_i^n = w_i^{n-1} - \text{sign} \left\{ \frac{\partial J}{\partial w_i}^n \right\} \cdot \Delta_i^n$$

$$\text{όπου} \quad \Delta_i^n = \begin{cases} \eta^+ \cdot \Delta_i^{n-1} & \text{αν} \quad \frac{\partial J}{\partial w_i}^n \cdot \frac{\partial J}{\partial w_i}^{n-1} > 0 \\ \eta^- \cdot \Delta_i^{n-1} & \text{αν} \quad \frac{\partial J}{\partial w_i}^n \cdot \frac{\partial J}{\partial w_i}^{n-1} < 0 \\ \Delta_i^{n-1} & \text{αν} \quad \frac{\partial J}{\partial w_i}^n \cdot \frac{\partial J}{\partial w_i}^{n-1} = 0 \end{cases} \tag{A.21}$$

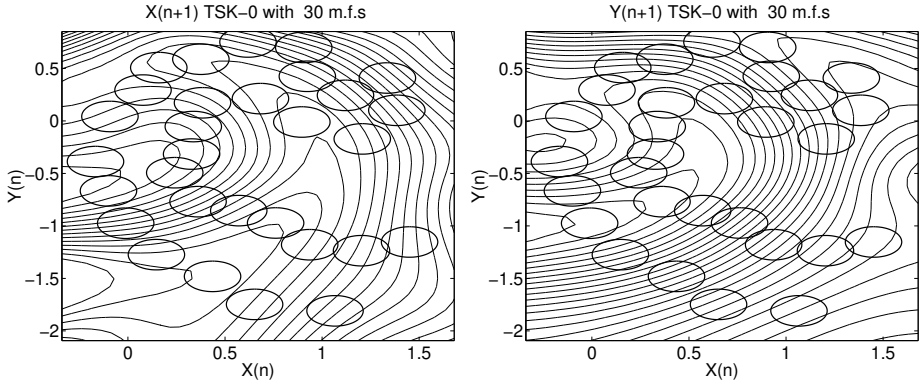
  - Υπολογισμός νέου πίνακα ενρεγοποίησης B_1, επίλυση εκ νέου των κανονικών εξισώσεων.

Αποτελέσματα

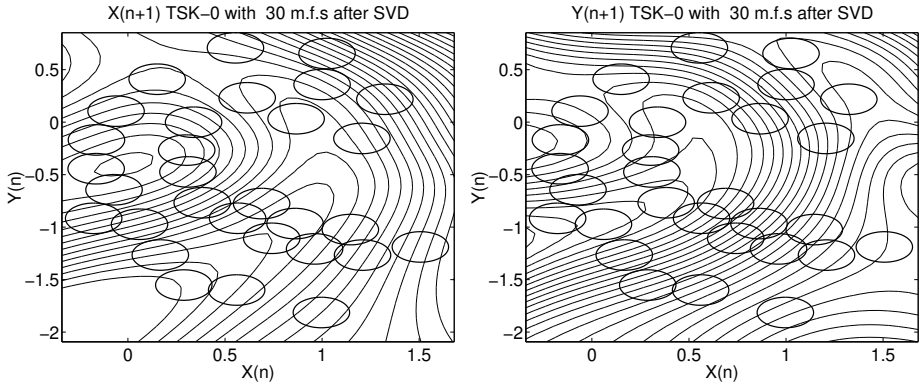
Παρακάτω βλέπουμε την ανακατασκευή της δυναμικής του συστήματος Ikeda αν χρησιμοποιηθούν

Συστήματα TSK-0

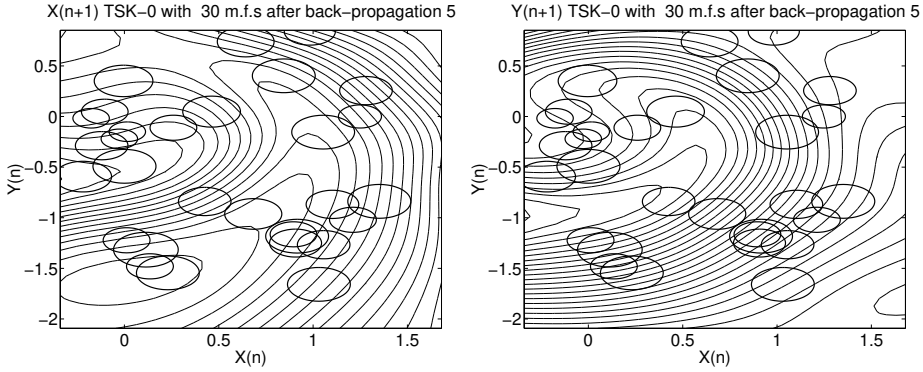
- Ένα σύστημα TSK-0 με 30 τοπικά μοντέλα (MSE=$6.1 \cdot 10^{-4}$, 30 σήματα ενεργοποίησης)



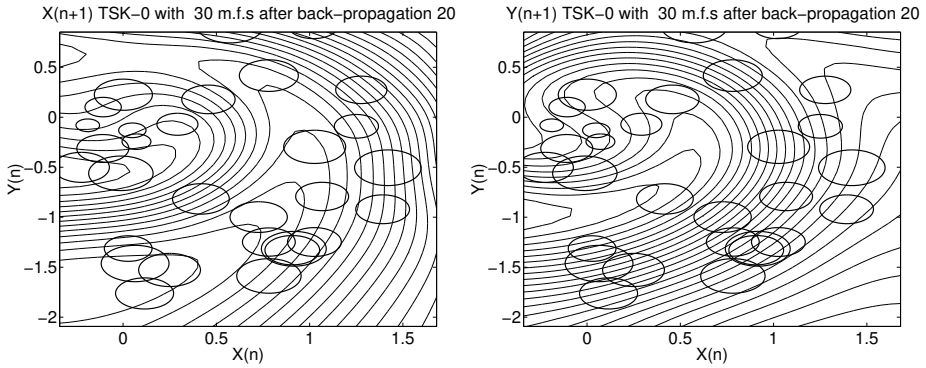
- Ένα σύστημα TSK-0 με 30 τοπικά μοντέλα μετά από την SVD των σημάτων ενεργοποίησης, όπου απόρριπεται μία ιδιοτιμή σ_i του πίνακα Σ αν $\sigma_i < 10^{-2}\sigma_1$. (Ποσοστό απορριμμένων ιδιοτιμών:6.7%, MSE=$1.1 \cdot 10^{-3}$,29 σήματα ενεργοποίησης)



- Ένα σύστημα TSK-0 με 30 τοπικά μοντέλα μετά από SVD μετά από 5 βήματα του αλγόριθμου rprop, (MSE=$1.67 \cdot 10^{-4}$ βελτίωση του MSE κατά 84.4%)

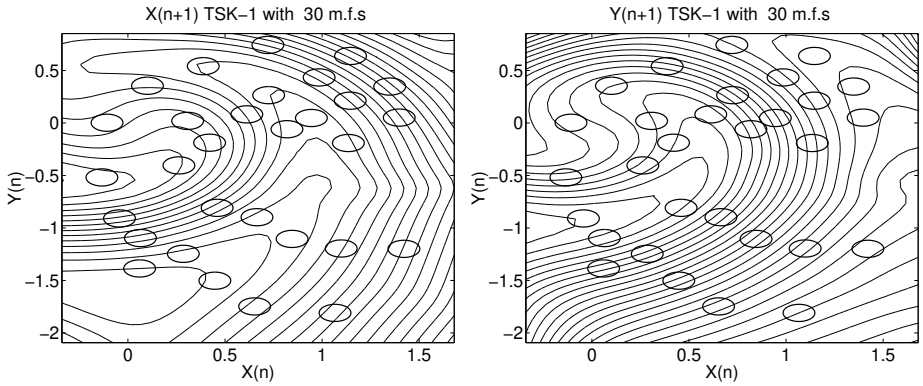


- Ένα σύστημα TSK-0 με 30 τοπικά μοντέλα μετά από 20 βήματα του αλγόριθμου rprop, (MSE=$2.83 \cdot 10^{-5}$ βελτίωση του MSE κατά 97.34%)

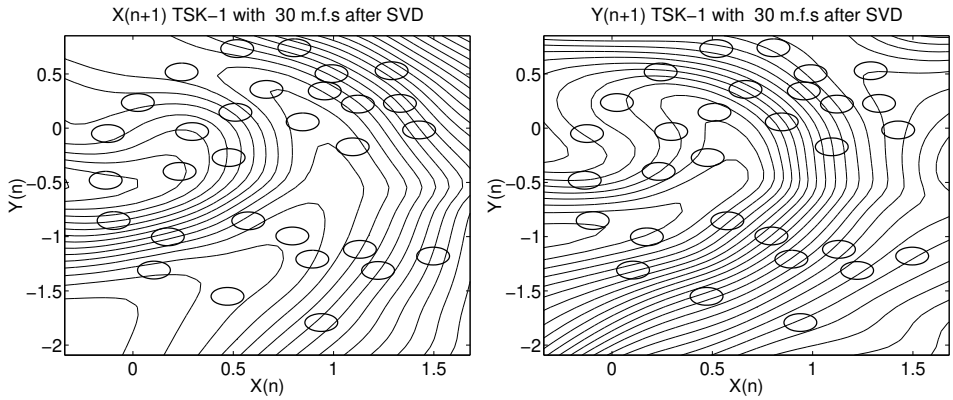


Συστήματα TSK-1

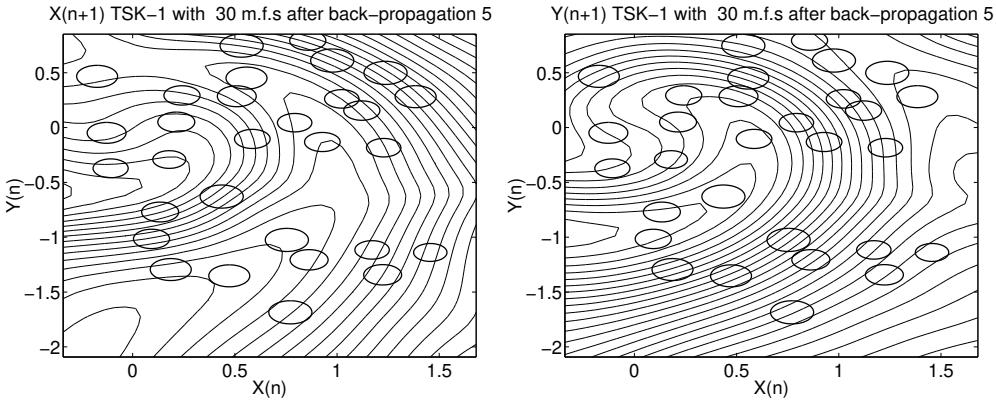
- Ένα σύστημα TSK-1 με 30 τοπικα μοντέλα (MSE=$1.7 \cdot 10^{-6}$, 90 σήματα ενεργοποίησης)



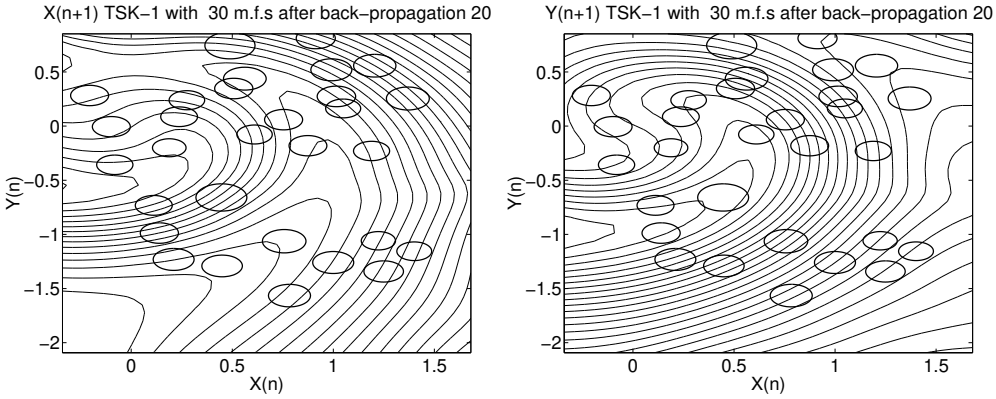
- Ένα σύστημα TSK-1 με 30 τοπικα μοντέλα μετά από την SVD των σημάτων ενεργοποίησης. Ποσοστό απορριμμένων ιδιοτιμών:56.7% (MSE=$2.78 \cdot 10^{-4}$ - 39 σήματα ενεργοποίησης)



- Ένα σύστημα TSK-1 με 30 τοπικα μοντέλα μετά από την SVD των σημάτων ενεργοποίησης, μετά από 5 βήματα του αλγόριθμου rprop (MSE=$3.61 \cdot 10^{-5}$, βελτίωση του MSE κατά 87.1%)



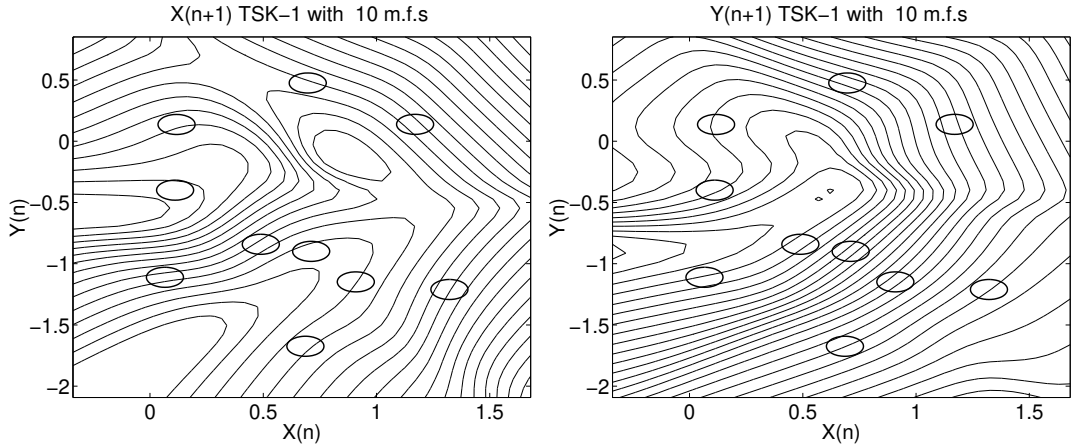
- Ένα σύστημα TSK-1 με 30 τοπικα μοντέλα, μετά από 20 βήματα του αλγόριθμου rprop (MSE=$3.61 \cdot 10^{-6}$, βελτίωση του MSE κατά 98.6%)



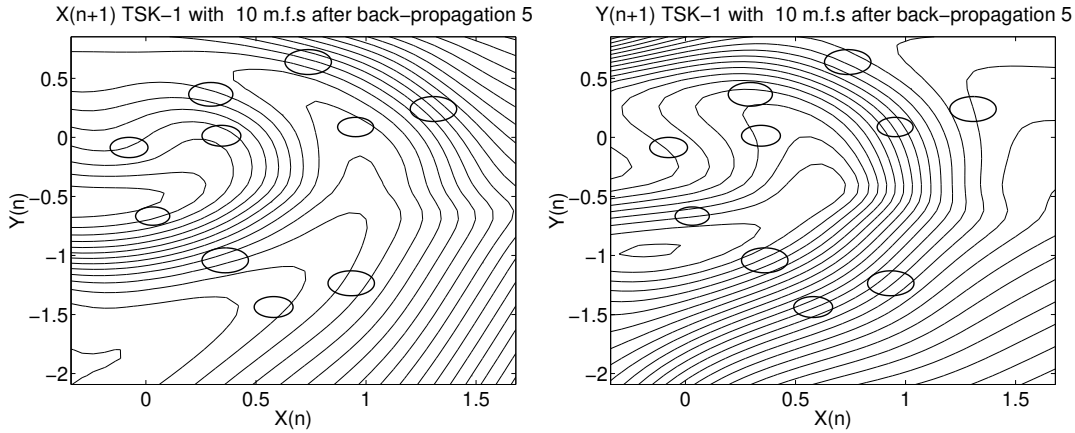
Είναι ξεκάθαρο ότι χρησιμοποιώντας τον αλγόριθμο rprop μπορούμε να πετύχουμε χωρίς κάποιο ιδιαίτερα μεγάλο υπολογιστικό κόστος σημαντική ελάττωση του MSE, καθώς και πολύ ακριβέστερη ανακατασκευή της δυναμικής ενός συστήματος. Αυτό φαίνεται τόσο από το μικρότερο MSE όσο και από την πολύ ακριβέστερη (οπτικά) ανακατασκευή της δυναμικής του συστήματος Ikeda. Εντυπωσιακά είναι τα αποτελέσματα αν χρησιμοποιηθεί ένας σχετικά μικρός αριθμός από τοπικά μοντέλα αρχικά και εκτελέσουμε λίγες επαναλήψεις του αλγορίθμου back-propagation.

Με μόνο 10 τοπικά μοντέλα μπορέσαμε να πετύχουμε μία πολύ ακριβή ανακατασκευή της δυναμικής του συστήματος, παρόμοια με αυτή που επιτεύχηκε ξεκινώντας από 30 τοπικά μοντέλα (βέβαια το 57% των σημάτων ενεργοποίησης απορρίφθηκε χρησιμοποιώντας την SVD-QR τεχνικη). Τελικά πάντως προτιμήθηκε να χρησιμοποιήσουμε ένα σχετικά μεγάλο αριθμό από συναρτήσεις συμμετοχής και σημάτων ενεργοποίησης, και να χρησιμοποιήσουμε την τεχνική SVD-QR για να κρατήσουμε όσα σήματα ενεργοποίησης χρειαζόμαστε έτσι ώστε να έχουμε ένα πιο ευέλικτο δίκτυο το οποίο να μπορεί, αν χρειαστεί, να κάνει και μία δυσκολότερη απεικόνιση.

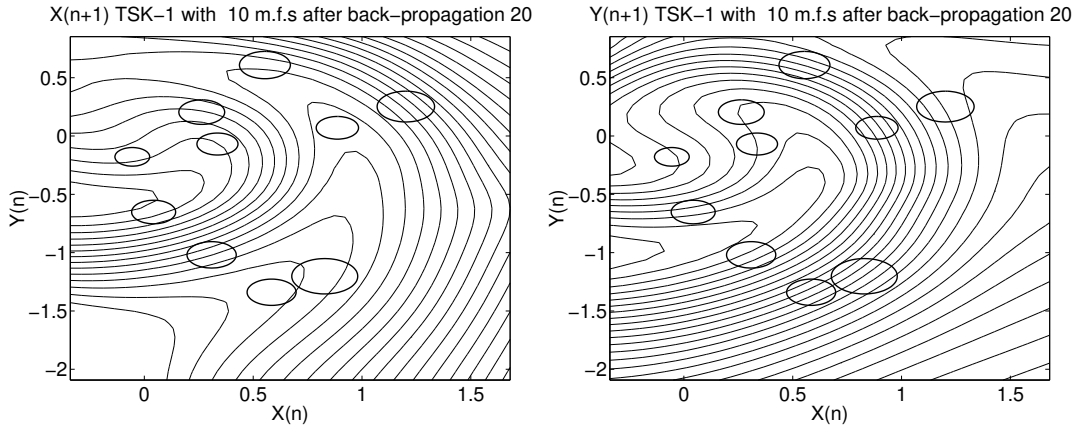
- Ανακτασκευή της δυναμικής του συστήματος Ikeda με ένα σύστημα TSK-1 με 10 συναρτήσεις συμμετοχής (MSE=$1.2 \cdot 10^{-3}$)



- Ανακατασκευή της δυναμικής του συστήματος Ikeda με ένα σύστημα TSK-1 με 10 συναρτήσεις συμμετοχής μετά από 5 βήματα του αλγορίθμου rprop (MSE=$2.24 \cdot 10^{-4}$)



- Ανακατασκευή της δυναμικής του συστήματος Ikeda με ένα σύστημα TSK-1 με 10 συναρτήσεις συμμετοχής μετά από 20 βήματα του αλγορίθμου rprop (MSE=$5.66 \cdot 10^{-5}$)



Παράρτημα Β

Δίκτυα Support Vector Machine

'Ενα από τα βασικά ζητούμενα κατά την εκπαίδευση ενός νευρωνικού δικτύου είναι να έχει τη δυνατότητα καλής γενίκευσης, δηλαδή να μπορεί να έχει καλές επιδόσεις σε δεδομένα τα οποία δεν έχει συναντήσει κατά τη διάρκεια της εκπαίδευσής του. Η ιδιότητα αυτή είναι άρρηκτα συνδεδεμένη με την αρχιτεκρονική του η οποία στις περισσότερες περιπτώσεις είναι δύσκολο να καθοριστεί εκ των προτέρων.

Στην περίπτωση των νευρωνικών δικτύων ακτινικών συναρτήσεων βάσης (radial basis function networks) και των πολυεπίπεδων perceptrons, για παράδειγμα, δεν είναι εκ των προτέρων γνωστό πόσοι εσωτερικοί νευρώνες χρειάζονται για να δώσουν ικανοποιητικά αποτελέσματα πλέον. Στην περίπτωση που χρησιμοποιηθούν πολλοί εσωτερικοί νευρώνες το δίκτυο έιναι σε θέση να μάθει με πολύ μεγάλη ακρίβεια τα δεδομένα με τα οποία εκπαιδεύτηκε, γεγονός το οποίο μπορεί να οδηγήσει σε μέτρια επίδοση με καινούρια δεδομένα στην περίπτωση όπου το σύνολο εκπαίδευσης περιείχε θόρυβο ή δεν ήταν αντιπροσωπευτικό του συνόλου των δεδομένων. Από την άλλη χρησιμοποιώντας μικρό αριθμό νευρώνων οδηγούμαστε σε δίκτυα με περιορισμένη δυνατότητα μάθησης, τα οποία δεν μπορούν να εκμεταλλευτούν όλη την προσφερόμενη σε αυτά πληροφορία κατά τη διάρκεια της εκπαίδευσης (πρόκειται για το γνωστό δίλημμα bias-variance).

Συνοψίζοντας μπορούμε να πούμε ότι επιδιώκουμε το δίκτυο να μαθαίνει μόνο τις κανονικότητες ˙που υπάρχουν στα δεδομένα με τα οποία το εκπαιδεύουμε και όχι τα δεδομένα αυτα καθ'αυτά.

Διάφορες τεχνικές που έχουν προταθεί για την επίλυση αυτού του προβλήματος είναι

- Κατασκευή ενό μικρού αρχικά δικτύου και σταδιακή αύξηση του αριθμού των νευρώνων του εως ότου οι επιδόσεις του δικτύου γίνουν ικανοποιητικές
- Δημιουργία ένός μεγάλου δικτύου και σταδιακή αφαίρεση νευρώνων εως ότου η επίδοση του δικτύου να πέσει κάτω από ένα όριο.
- Αφαίρεση των πλεονάζοντων η σχετικά ασήμαντων νευρώνων με βάση τεχνικές SVD
- Πρόωρος τερματισμός της εκπαίδευσης του δικτύου (ώστε να μην προλάβει να μάθει πλήρως το σύνολο εκπαίδευσης
- Εισαγωγή ενός όρου ποινής στη συνάρτηση κόστους του δικτύου, ο οποίος να 'τιμωρεί' την πολυπλοκότητα του δικτύου. Η τεχνική αυτή ονομάζεται κανονικοποίηση και είναι μάλλονη πλέον ευρέως διαδεδομένη από τις παραπάνω . Αν το λάθος πάνω

στο σύνολο εκπαίδευσης έχει βρεθεί ίσο με το κανονικοποιημένο κόστος γράφεται

$$\tilde{E} = E + v\Omega$$

όπου το Ω είναι ένας όρος ποινής για την πολυπλοκότητα του δικτύου (π.χ. $\Omega = \sum_i w_i^2$,όπου w_i τα βάρη του δικτύου) και v η σταθερά κανονικοποίησης, η οποία εκφράζει πόσο σημαντικό θεωρούμε να έχουμε ένα απλό δίκτυο σε σχέση με το να έχουμε καλές προβλέψεις στο σύνολο εκπαίδευσης. Βασικό πλεονέκτημα αυτής της μεθόδου είναι ότο μπορούμε να τροποιήσουμε τον αλγόριθμο Back-Propagation ώστε να λαμβάνει υπ'οψιν και τον όρο της κανονικοποίησης, οδηγώντας σε ένα αποδοτικό αλγόριμο εκπαίδευσης.

Τα δίκτυα SVM τα οποία εισήχθησαν από τον Vapnik το 1992 [43] ,προτείνουν μία νέα τεχνική σχεδιασμού των δικτύων που βασίζεται στα αποτελέσματα της στατιστικής θεωρίας μάθησης [44]. Αποτελούν έναν απο τους ταχύτερα αναπτυσσόμενους τομείς έρευνας στο χώρο των νευρωνικών δικτύων και θεωρούνται ήδη ως η ενδεδειγμένη μέθοδος για πολλούς σκοπούς, π.χ. αναγνώριση χαρακτήρων, όραση υπολογιστών, βιοπληροφορική, κατηγοριοποίηση κειμένου , ανάλυση χρονοσειρών κ.α. [45]. Αυτό το οποίο τα διαφοροποιεί σε σχέση με τις άλλες μεθόδους είναι ότι προσφέρουν, όπως θα δούμε, έναν πολύ πιο πειθαρχημένο τρόπο στην επιλογή του αριθμού των 'ενεργών' νευρώνων.

B.1 Δίκτυα SVM για ταξινόμιση

Αν έχουμε δύο κλάσεις l γραμμικά διαχωρίσιμων d-διάστατων προτύπων x και θέλουμε να τα διαχωρίσουμε με βάση μία γραμμική συνάρτηση $f(x_i) = \sum_{i=1}^{d} w_i x_i + b$ χρησιμοποιούμε τον παρακάτω αλγόριθμο εκμάθησης του percpetron για να εξάγουμε το διάνυσμα w από τα δεδομένα, όπου $y_i \in \{-1, 1\}$ είναι η γνωστή εκ των προτέρων κλάση στην οποία ανήκει το x_i)

Αλγόριθμος μάθησης perceptron

$$\mathsf{w}_{k+1} = \begin{cases} w_k + \nu\delta w_k = w_k + \nu f(x_i) y_i, & f(x_i) y_i < 0 \\ w_k + \nu\delta w_k = w_k & f(x_i) y_i > 0 \end{cases} \Rightarrow$$

$\mathsf{w}=\sum_{i=1}^{l} a_i y_i x_i$ Αξίζει να προσέξουμε τη σχέση B.1 καθώς μας δείχνει ότι μπορεί το τελικό διάνυσμα w να εκφραστεί ως ένας γραμμικός συνδυασμός των δεδομένων στην είσοδο του ταξινομητή. Από αυτά, μεγαλύτερο βάρος (το οποίο εκφράζεται μέσω του a_i) θα έχουν αυτά για τα οποία το λάθος στην ταξινόμιση ήταν μεγαλύτερο, ενώ όπως προκύπτει και από τον αλγόριθμο ενημέρωσης του βάρους w το a_i για τα διανύσματα εκείνα στα οποία η ταξινόμιση ήταν επιτυχημένη είναι ίσο με μηδέν. Συνεπώς μπορούμε να πούμε ότι στον καθορισμό του w παίζουν ρόλο μόνο τα σημεία αυτά τα οποία δίνουν λάθος αποτέλεσμα και σκοπός μας είναι να ανατρέψουμε το λάθος αυτό.

Η σχέση B.1 μας δίνει τη δυνατότητα να ξαναγράψουμε τη διακριτική συνάρτηση f και τον κανόνα ενημέρωσης των βαρών ως εξής

$$f(x_i) = \sum_{i=1}^{d} w_i x_i + b = \sum_{i=1}^{l} a_i y_i x_i x_i = \sum_{i=1}^{l} a_i y_i < x_i, x_i >$$

$$\text{Αν} \quad y_i(\sum_{j=1}^{l} a_j y_j < x_j, x_j > + b) < 0 \rightarrow a_i \leftarrow a_i + \nu \tag{B.1}$$

Συνεπώς βλέπουμε ότι με βάση τη δυική αυτή έκφραση μπορούμε να επιλύσουμε το πρόβλημα του καθορισμού των βαρών του δικτύου χρησιμοποιώντας μόνο τους πολλαπλασιαστές a_i. Τα δεδομένα εμφανίζονται μόνο με εσωτερικά γινόμενα, τα οποία μπορούν να έχουν υπολογιστεί εκ των προτέρων , χωρίς να αλλάζουν κατά τη διάρκεια της εκπαίδευσης. Η σημασία των παραπάνω θα φανεί όταν καταλήξουμε στις εξισώσεις για τα δίκτυα SVM

Πρωτού προχωρήσουμε χρήσιμο είναι να αναφερθούμε στο περιθώριο διαχωρισμού της διακρτικής συνάρτησης $f(x_i) = w^T x_i + b$: όπως είναι γνωστό, η απόσταση ενός σημείου x_i από το υπερεπίπεδο που ορίζει η διακριτική συνάρτηση ισούται με $\frac{f(x_i)}{|w|}$. Σύμφωνα, τώρα, με την υπόθεσή μας ότι τα δεδομένα είναι γραμμικά διαχωρίσιμα έχουμε ότι

$$\begin{aligned} w^T x_i + b &> 0, \quad \text{αν } y_i = 1 \\ w^T x + b &< 0, \quad \text{αν } y_i = -1 \end{aligned} \tag{B.2}$$

Οι παραπάνω σχέσεις μπορούν να ξαναγραφούν ,αν θεωρήσουμε ότι η ελάχιστη απόλυτη τιμή της συνάρτησης διαχωρισμού στο σύνολο των δεδομένων είναι ϵ, μετά από την κανονικοποίηση $\cdot 1/\epsilon$ οπότε θα έχουμε

$$\begin{aligned} w^T x_i + b &> 1, \quad \text{αν } y_i = 1 \\ w^T x_i + b &< -1, \quad \text{αν } y_i = -1 \end{aligned} \tag{B.3}$$

Οι δύο σχέσεις αυτές μπορούν να γραφούν συμπυκνωμένα ως εξής:

$$y_i(w^T x_i + b) > 1 \tag{B.4}$$

Τα σημεία τα οποία απέχουν ελάχιστα από το διαχωριστικό επίπεδο είναι αυτά για τα οποία η $w^T x + b$ παίρνει τιμή ίση με 1, οπότε θα απέχουν $1/|w|$ από το διαχωριστικό υπερεπίπεδο. Συνεπώς θα υπάρχει μία 'νεκρή ζώνη' έυρους $1/2|w|$ μέσα στην οποία δεν θα υπάρχουν σημεία από τις δύο τάξεις. Η ζώνη αυτή λέγεται διαχωριστικό διάστημα (margin of seperation) και αποτελεί ένα μέτρο της διαχωρισιμότητας των δεδομένων.

Δομική ελαχιστοποίηση του κινδύνου (Structural Risk Minimization) και Support Vector Machines

Όπως είπαμε στην αρχη, το ζητούμενο κατά την έκπαίδευση ενός νευρωνικού δικτύου είναι να μπορεί να αντιμετωπίσει άγνωστα δεδομένα ˙ η ιδιότητα αυτή μπορεί να εκτιμηθεί με κριτήρια όπως αυτό του ελάχιστου μήκους περιγραφής (Minimum Description Length, την Probably Approximately Correct (PAC) μέθοδο και με βάση τη Στατιστική Θεωρία Μάθησης -(Statistical Learning Theory) η οποία θα παρουσιαστεί με πολύ σύντομο τρόπο παρακάτω.

Κεντρικής σημασίας για τη στατιστική θεωρία μάθησης είναι ή έννοια της διάστασης Vapnik-Chervonenkis (VC-dimension) ενός δικτύου. Η διάσταση αυτή ισούται με την πληθυκότητα του μεγαλύτερου συνόλου το οποίο μπορεί να ταξινομίσει με οποιονδήποτε δυνατό τρόπο σε δύο τάξεις, δηλαδή να το χωρίσει σε δύο τάξεις με όλους τους δυνατούς συβδυσμούς. Συνεπώς ένα δίκτυο με μεγάλη διάσταση VC έχει και μεγάλη δυνατότητα μάθησης (αφού είναι αρκετά ευέλικτο ώστε να κάνει πολλές διαφορετικές διχοτομήσεις του χώρου εισόδου του) και το αντίστροφο.

Αυτό το οποίο αποτελεί το κριτήριο με το οποίο καταλήγουμε σε ένα συκεκριμένο δίκτυο είναι ένα θεώρημα της στατιστικής θεωρίας μάθησης (Vapnik 1995) σύμφωνα με το οποίο

,αν ένα δίκτυο το οποίο εκτελεί την απεικόνιση $x \rightarrow f(x,a)$,όπου a οι παράμετροι του δικτύου, έχει διάσταση VC , και μέσο λάθος R_{emp} (Empirical Risk) στο σύνολο εκπαίδευσης τότε το λάθος γενίκευσης είναι με πιθανότητα $1-\nu$ άνω φραγμένο από την ποσότητα

$$R_{guaranteed}(a) = R_{emp}(a) + \sqrt{\left(\frac{h(\log(2l/h)+1) - log(\nu/4)}{l}\right)} \tag{B.5}$$

όπου l το πλήθος των στοιχείων με βάση τα οποία έγινε ο υπολογισμός του R_{emp}.

Το παραπάνω άθροισμα (Guaranteed Risk) περιλαμβάνει δύο αντικρουόμενες ποσότητες: το $R_{emp}(a)$ είναι φθίνουσα συνάρτηση του h καθώς ένα δίκτυο με μεγάλη διάσταση VC είναι σε θέση να μάθει πολύ καλά το σύνολο εκπαίδευσής του (οπότε θα έχει μικρό R_{emp}) ενώ η δεύτερη ποσότητα (VC -διάστημα εμπιστοσύνης) είναι άυξουσα συνάρτηση του h. Σύμφωνα με την αρχή της δομικής ελαχιστοποίησης του σφάλματος (Structural Risk Minimization) , αν γνωρίζουμε την ποσότητα h για ένα συγκεκριμένο τύπο δικτύου μπορούμε ελαχιστοποιώντας το $R_{guaranteed}$ να επιλέξουμε το συγκεκριμένο δίκτυο το οποίο μας δίνει το μικρότερο (αναμενόμενο) σφάλμα γενίκευσης.

Έχοντας τα παραπάνω υπ'οψιν μπρούμε να εκμεταλλευτούμε το παρακάτω θεώρημα:

[VC-διάσταση σφαιρικού ταξινομητή]

Έστω D η διάμετρος του ελάχιστου κύκλου που περιέχει όλα τα δεδομένα που θέλουμε να κατατάξουμε. Αν η κατατάξη βασίζεται σε μία διακριτική συνάρτηση

$$\sum_{i=1}^{D} w_i x_i + b = w^T x_i + b \tag{B.6}$$

και το διάστημα διαχωρισμού γύρω από το διαχωριστικό επίπεδο $w^T x + b$ είναι ρ τότε η VC διάσταση των ταξινομητών που χρησιμοποιούν τη διακριτική αυτή συνάρτηση είναι άνω φραγμένη από την ποσότητα

$$VC \leq \min\{\lceil \frac{D^2}{\rho^2} \rceil, D\} + 1 \tag{B.7}$$

Η ποσότητα ρ αποδειχθηκε προηγουμένως ότι είναι ίση με $2/|w|$. Η μείωση της ποσότητας $2/|w|$ σημαίνει, όπως είδαμε μεγαλύτερο ρ, οπότε, βάσει του τύπου B.7 και μικρότερη διάσταση VC. Συνεπώς μπορούμε να πούμε ότι ακολουθώντας την αρχή της δομικής ελαχιστοποίησης του κινδύνου ζητάμε, ανάμεσα στο σύνολο των υπερεπιπέδων που κάνουν το διαχωρισμό των δύο κλάσεων το συγκεκριμένο υπερεπίπεδο για το οποίο ελαχιστοποιείται η ποσότητα $\frac{2}{|w|}$. Επιζητούμε δηλαδή να ελαχιστοποιήσουμε το VC διάστημα εμπιστοσύνης που υπεισέρχεται στον τύπο B.5 έτσι ώστε να πάρουμε ένα προβλέπτη με καλή δυνατότητα γενίκευσης.

Αυτή είναι και η αρχή με την οποία λειτουργούν τα SVM , αντί να χρησιμοποιούν κάποιον έμμεσο τρόπο, όπως οι μέθοδοι που αναφέρθηκαν στην εισαγωγή, για τον έλεγχο της πολυπλοκότητας του δικτύου.

Το πρόβλημα που προαναφέρθηκε μπορεί να διατυπωθεί ως εξής:

Ελαχιστοποίησε την ποσότητα	$\frac{1}{2}w^T, w$	(Μέγιστο)
Υπό τη συνθήκη	$y_i(w^T x_i + b) \geq 1$	(Μηδενικό σφάλμα)

Η Langrangian του προβλήματος αυτού είναι :

$$\begin{aligned} L(w) &= \frac{1}{2}w^T w - \sum_{i=1}^{l} \alpha_i y_i \left[\left(w^T x + b\right) - 1\right] \\ a_i &= 0, \qquad \forall i \end{aligned}$$

Θέτοντας $\frac{\partial L}{\partial b} = 0, \frac{\partial L}{\partial w} = 0$ και μετά από κάποιες αντικαταστάσεις (βλ. π.χ. [45]) παίρνουμε το δυικό πρόβλημα:

$$\text{Μεγιστοποίησε την ποσότητα} \qquad w(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} a_i a_j y_i y_j < x_i, x_j >$$

$$\text{Υπό τις συνθήκες} \qquad \begin{cases} \alpha_i \geq 0 \\ \sum_{i=1}^{l} \alpha_i y_i = 0 \end{cases}$$

Το πρόβλημα αυτό είναι ένα πρόβλημα τετραγωνικού προγραμματισμού, το οποίο μπορεί να επιλυθεί με χρήση οποιασδήποτε βιβλιοθήκης βελτιστοποίησης. Σημασία έχει να τονιστεί ότι λόγω της φύσης του προβλήματος, σε αντίθεση με την εκπαίδευση νευρωνικών δικτύων, δεν υπάρχει κίνδυνος να παγιδευτούμε σε τοπικά ελάχιστα, ενώ η λύση μπορεί να βρεθεί σε πολυωνυμικό χρόνο. Μία ιδιαίτερα ενδιαφέρουσα παρατήρηση, η οποία εξηγεί και την ονομασία SVM είναι ότι με βάση τις συνθήκες Kuhn–Karush–Tucker (KKT) για να είναι μία λύση του παραπάνω προβλήματος η βέλτιστη θα πρέπει να ισχύει:

$$\alpha_i \left[w^T x_i + b - 1\right] = 0 \qquad \forall i \tag{B.8}$$

το οποίο σημαίνει ότι οι πολλαπλασιαστές Lagrange α_i είναι μη μηδενικοί μόνο για τα σημεία εκείνα τα οποία βρίσκονται ακριβώς πάνω στο χάσμα μεταξύ των δύο τάξεων. Τα σημεία αυτά ονομάζονται Support Vectors και είναι αυτά τα οποία 'οδηγούν' τη διαδικασία μάθησης: είναι τα πιο δύσκολα στην ταξινόμιση και είναι αυτά τα οποία εμποδίζουν το χάσμα να μεγαλώσει πέρα από ένα όριο. Επίσης, με δεδομένο ότι η ελαχιστοποίηση της Langrangian ως προς το διάνυσμα α μας δίνει

$$w = \sum_{i=1}^{l} a_i y_i x_i \tag{B.9}$$

βλέπουμε ότι το τελικό διάνυσμα βάρους μπορεί να εκφραστεί μόνο με βάση τα Support Vectors. Αντιπαραβάλλοντας την τελευταία παρατήρηση με την παρατήρηση που έγινε για την εξίσωση B.1 μπορούμε να πούμε ότι ο αλγόριθμος εκπαίδευσης των SVM ωθεί στα άκρα αυτό που παρατηρήσαμε για τον αλγόριθμο εκμάθησης του perceptron, δηλαδή αφήνει μόνο τα πολύ δύσκολα στην ταξινόμιση σημεία να συμμετάσχουν στον καθορισμό της διακριτικής συνάρτησης.

Μη–διαχωρίσιμα δεδομένα

Προς το παρόν θεωρήσαμε ότι τα δεδομένα μας είναι δυνατόν να διαχωριστούν με βάση τη διακριτική συνάρτηση χωρίς την παρουσία λαθών. Αυτό, προφανώς, δεν είναι να δυνατό να ισχύει πάντοτε, οπότε πρέπει να αφήσουμε περιθώριο για την ύπαρξη λαθών, αν θέλουμε το πρόβλημα βελτιστοποίησης να είναι επιλύσιμο. Για το σκοπό αυτό χαλαρώνουμε

τις συνθήκες B.3 με αντάλλαγμα μία αύξηση στη συνάρτηση κόστους, εισάγοντας τις slack μεταβλητες ξ_i με βάση τις οποίες το πρόβλημα βελτιστοποίησης γράφεται ως εξής:

$$\begin{array}{rlll} \text{Ελαχιστοποίησε την ποσότητα} & w^T w + & 2C\sum_{i=1}^{l}\xi_i & \\ \text{Υπό τη συνθήκη} & w^T x_i + b & \geq y_i - \xi_i & , y_i = 1 \\ & w^T x_i + b & \leq y_i + \xi_i & , y_i = -1 \\ & \xi_i & \geq 0 & \end{array}$$

Παρατηρούμε ότι η συνάρτηση κόστους έχει αλλάξει ώστε να περιλαμβάνει και τους όρους ξ_i. Η παράμετρος C δείχνει πόσο σημαντικό θεωρούμε ένα λάθος αποτέλεσμα της μηχανής ταξινόμησης: ένα μεγάλο C μπορεί να οδηγήσει σε μία λύση με μεγάλο σχετικά $|w|$ αρκεί να έχουμε μικρό λάθος και αντίστροφα. Μπορούμε να βρούμε μία αναλογία με το πρόβλημα του bias-variance όπου πρέπει να διαλέξουμε μεταξύ ενός δικτύου το οποίο μαθαίνει σωστά τα δεδομένα ή ενός το οποίο είναι απλούστερο και του επιτρέπονται ορισμένα λάθη.

Το δυικό πρόβλημα είναι:

$$\begin{array}{rll} \text{Μεγιστοποίησε την ποσότητα} & w(\alpha) = & \sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{l} a_i a_j y_i y_j x_i^T x_j \\ \text{Υπό τις συνθήκες} & & \begin{cases} 0 \leq \alpha_i \leq C & \forall i \\ \sum_{i=1}^{l}\alpha_i y_i = 0 \end{cases} \end{array}$$

Η λύση του προβλήματος έχει πάλι τη μορφή $w = \sum_{i=1}^{l}\alpha_i y_i x_i$. Η μόνη διαφορά σε σχέση με την προηγούμενη περίπτωση είναι ότι έχει τεθεί ένα άνω όριο στην τιμή του α, ώστε τα διανύσματα τα οποία δεν μπορούν εκ φύσεως να καταταχθούν σωστά σε μία κατηγορία να μην οδηγούν σε μη-φραγμένες τίμες των α_i.

Από τις συνθήκες KKT για το τροποποιημένο αυτό πρόβλημα έχουμε

$$\begin{aligned} \alpha_i\{y_i(w^T x_i + b) - 1 + \xi_i\} &= 0 \\ C - \alpha_i - {}_i &= 0 \\ {}_i\xi_i &= 0 \end{aligned}$$

Από τις δύο τελευταίες σχέσεις προκύπτει ότι όταν $\alpha_i < C$,έχουμε $\xi_i = 0$. Προκύπτει, λοιπόν ότι τα διανύσματα x_i τα οποία βρίσκονται ακριβώς πάνω στο όριο του χάσματος της διαχωριστικής επιφάνειας ($y_i(w^T x_i + b) = 1$) θα έχουν τον αντίστοιχο συντελεστή α_i στο διάστημα $(0, C)$ τα σημεία που βρίσκονται στο λάθος μέρος της διαχωριστικής επιφάνειας συντελεστή α_i ίσο με C και τα υπόλοιπα δεν θα συμμετέχουν στον καθορισμό του διανύσματος w.

Πάλι δηλαδή θα είναι ένα υποσύνολο από τα δεδομένα που θα καθορίζει την απόφαση του ταξινομητη (αυτά τα x_i για τα οποία $\alpha_I \neq 0$), το οποίο θα αποτελεί τους support vectors.

Μη-γραμμικοί πυρήνες εσωτερικού γινομένου

Αν επιθυμούμε να ταξινομήσουμε δεδομένα τα οποία να ανήκουν σε μη-γραμμικά διαχωρίσιμες τάξεις, αυτό το οποίο γίνεται, γενικότερα, στην αναγνώριση προτύπων είναι μέσω ενός μη γραμμικού μετασχηματισμού να απεικονίζουμε τα δεδομένα σε ένα χώρο συνήθως

μεγαλύτερης διαστάσεως (χώρος χαρακτηριστικών -feature space) , όπου να είναι δυνατός ο γραμμικός διαχωρισμός των δεδομένων.

Ο ορισμός των SVMs μπορεί να επεκταθεί άμεσα από τη γραμμική περίπτωση στη μη-γραμμική ,αν αντί του εσωτερικού γινομένου χρησιμοποιηθεί ένας πυρήνας εσωτερικού γινομένου˙ οι πυρήνες εσωτερικού γινομένου έχουν το ιδιαίτερα ελκυστικό χαρακτηριστικό ότι αντικαθιστούν την προβολή στο χώρο προτύπων και την εκτέλεση του εσωτερικού γινομένου στον χώρο προτύπων (τον οποίο απαιτεί ο γραμμικός διαχωρισμός των δεδομένων) με μία απλούστερη πράξη, όπως φαίνεται από το παρακάτω απλό παράδειγμα:

$$\begin{array}{rrcl} \text{Αρχικά δεδομένα:} & (x_1,x_2),(y_1,y_2) & \overset{\text{Μ/μος } \Phi:(x_1,x_2)\to(x_1^2,\sqrt{2}x_1x_2,x_2^2)}{\Rightarrow} & \\ & <\Phi(x_1,x_2),\Phi(y_1,y_2)> & = & x_1^2y_1^2+2x_1x_2y_1y_2+x_2^2y_2^2= \\ = & <(x_1,x_2),(y_1,y_2)> & = & K((x_1,x_2),(y_1,y_2)) \end{array} \tag{B.10}$$

Με τη χρήση πυρήνων εσωτερικού γινομένου μπορούμε με μικρό αριθμό πράξεων να υπολογίζουμε εσωτερικά γινόμενα σε διανυσματικούς χώρους άπειρων διαστάσεων. Οι συνθήκες ώστε μία συνάρτηση να αποτελεί πυρήνα εσωτερικού γινομένου είναι γνωστές ως συνθήκες του Mercer (βλ. π.χ. [46]) και ορισμένες από τις συναρτήσεις που τις ικανοποιούν είναι οι ακόλουθες:

$K(x,x_i)=$	$e^{-\frac{\lvert x-x_i\rvert}{\sigma^2}}$	Gaussian
$K(x,x_i)=$	$\tanh(\beta_0 x^T x_i+\beta_1)$	Σιγμοειδής
$K(x,x_i)=$	$(x^T x_i+1)^p$	Πολυωνυμική
$K(x,x_i)=$	$B_n^d(\lvert x-x_i\rvert)$	n-τάξης d-διάστατη spline

Σημειώνεται ότι Ο σιγμοειδής πυρήνας αποτελεί πυρήνα εσωτερικού γινομένου μόνο για ορισμένες τιμές των β_0,β_1.

Στη δυική παράσταση (B.1) αντικαθιστώντας το εσωτερικό γινομένο με έναν πυρήνα προκύπτει η έκφραση:

$$f(x)=\sum_{i=1}^{l} a_i y_i K(x,x_i)+b \tag{B.11}$$

Όλες οι σχέσεις που γράφηκαν προηγουμένως και χρησιμοποιούσαν το εσωτερικό γινόμενο μπορούν να ξαναγραφούν έχοντας στη θέση του εσωτερικού γινομένου έναν πυρήνα.

Παρατηρούμε ότι μπορούμε. θέτοντας τους κατάλληλους πυρήνες. να πάρουμε μία γραμμική μηχανή (πυρήνας εσωτερικού γινομένου), ένα δίκτυο rbf (Gaussian πυρήνας), ένα perceptron 2επιπέδων (Σιγμοειδής πυρήνας) καθώς και άλλους, καινούργιους τύπους δικτύων. Συνεπώς με βάση τον αλγόριθμο μάθησης του SVM μπορούμε να εκπαιδεύσουμε τα προαναφερθέντα δίκτυα, έτσι ώστε να μεγιστοποιείται το χάσμα μεταξύ κλάσεων στον χώρο χαρακτηριστικών, που όπως είδαμε μας εξασφαλίζει (θεωρητικά τουλάχιστον) μικρότερο σφάλμα γενίκευσης.

Το πρόβλημα της βελτιστοποίησης-SMO

Για την επίλυση του προβλήματος της βελτιστοποίησης μπορεί να χρησιμοποιηθεί μία οπιαδήποτε διαθέσιμη βιβλιοθήκη τετραγωνικού προγραμματισμού, της οποίας η πολυπλοκότητα, όμως, είναι τετραγωνική ως προς τον αριθμό των δεδομένων. Για μικρά

σύνολα εκπαίδευσης (500-1000 δεδομένα) είναι δυνατό να χρησιμοποίησουμε έναν standard αλγόριθμο τετραγωνικού προγραμματισμού όπως ο LOQO του Vanderbrei ([?],port σε Matlab: [?]) αλλά για μεγαλύτερα σύνολα εκπαίδευσης ο χρόνος εκπαίδευσης γίνεται απαγορευτικά μεγάλος. Εκτός αυτού, χρείαζεται να ληφθεί υπ'οψιν και το γεγονός ότι μπορεί να μην χωράνε όλα τα απαραίτητα δεδομένα ($K(x_i, x_j)$) στην μνήμη του υπολογιστή ενώ είναι εξαιρετικά ασύμφορο να υπολογίζονται εκ νέου τα γινόμενα πυρήνα όποτε χρειάζονται.

Μία πρώτη λύση που προτάθηκε για να αντιμετωπιστεί αυτό το πρόβλημα είναι η επιλογή ενός υποσυνόλου πολλαπλασιαστών Lagrange πάνω στο οποίο γίνεται η προσπάθεια ελαχιστοποίησης του κόστους. Το σκεπτικό είναι ότι αν για έναν πολλαπλασιαστή Lagrange α_i ικανοποιούνται για αρκετά μεγάλο χρονικό διάστημα οι συνθήκες ΚΚΤ (που, όπως είπαμε, ισχύουν όταν έχει επιτευχθεί η βέλτιση λύση) μπορούμε να θεωρήσουμε ότι η συνάρτηση κόστους δεν μπορεί να αλλάξει αλλάζοντας την τιμή του πολλαπλασιαστή α_i. Αντιθέτως, προσπαθώντας να ελαχιστοποιήσουμε τη συνάρτηση κόστους στον χώρο που ορίζεται από τους πολ/τες Lagrange στους οποίους παραβιάζονται οι συνθήκες ΚΚΤ είναι δυνατό να επιτύχουμε μείωση της συνάρτησης κόστους.

Οι αλγόριθμοι οι οποίοι βασίζονται στην αποκοπή των πολλαπλασιαστών Lagrange που δεν είναι 'πολλά υποσχόμενοι' για την ελάττωση της συνάρτησης κόστους ονομάζονται αλγόριθμοι chunking, ενώ αν επιπλέον χρησιμοποιούν ένα σταθερού μεγέθους υποσύνολο των πολλαπλασιαστών Lagrange ονομάζονται αλγόριθμοι διάσπασης. Αναφέρεται ότι έχει αποδειχθεί ότι τελικά συγκλίνουν στη λύση του προβλήματος. Δεν θα αναπτύξουμε κανέναν από τους αλγόριθμους αυτούς ούτε για την περίπτωση της ταξινόμησης ούτε για την πρόβλεψη για να παρουσιάσουμε απ'ευθείας την τεχνική που προτάθηκε το '99 από τον John Platt για την πρόβλημα της ταξινόμησης και έχει έκτοτε μεταφερθεί και στην πρόβλεψη.

Η τεχνική αυτή ονομάζεται Sequential Minimization Optimization (SMO) [48] και έχει καθιερωθεί ως η ταχύτερη τεχνική για ταξινόμηση, ενώ διάφορες παραλλαγές της συναγωνίζονται για την περίπτωση της πρόβλεψης. Αυτό το οποίο γίνεται με την τεχνική αυτή είναν να εξωθείται στο όριά του ο αλγόριθμος της επιλογής υποσυνόλου, επιλέγοντας κάθε φορά 2 πολλαπλασιαστές Lagrange οπότε υπάρχει και αναλυτική λύση για το πρόβλημα της βελτιστοποίησης.

Αλγόριθμος SMO

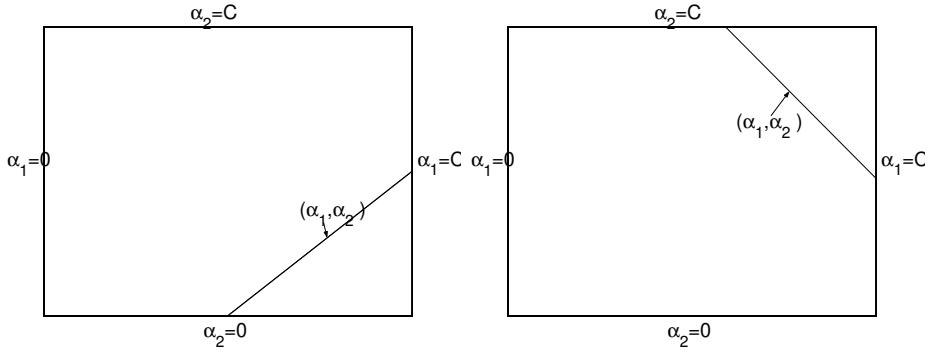
Σε κάθε βήμα του αλγορίθμου SMO επιλέγονται δύο πολλαπλασιαστές Lagrange α_1, α_2 που παραβιάζουν τις συνθήκες ΚΚΤ και με βάση αυτούς προσπαθούμε να ελαχιστοποιήσουμε τη συνάρτηση κόστους, κρατώντας τους υπόλοιπους σταθερούς στην προηγούμενη τιμή τους. Δεδομένου ότι οι συνθήκες του δυικού προβλήματος βελτιστοποίησης απαιτούν να ισχύει

$$\sum_{i=1}^{l} \alpha_i y_i = 0$$
$$0_i \leq C \tag{B.12}$$

συμπεραίνουμε ότι εφόσον οι υπόλοιποι πολλαπλασιαστές μένουν σταθεροί, πρέπει για τους νέους α_i', α_j' να ισχύει

$$\alpha_1' y_1 + \alpha_2' y_2 = \alpha_1 y_1 + \alpha_2 y_2 \tag{B.13}$$

ενώ είναι φραγμένοι από τις ίδιες συνθήκες ($0 \leq a_i \leq C, i = 1, 2$) στην περιοχή $[0, C] \times [0, C]$ (βλ.σχ. Β.1) Ο αλγόριθμος υπολογίζει πρώτα τον α_2' θεωρώντας ότι δεν είναι φραγμένος,



Β.1: Όρια πολλαπλασιαστών Lagrange αν (a)$y_1 y_2 = 1$,(b)$y_1 y_2 = 1$

και μετά με βάση την απαίτηση να είναι εφικτές όλες οι συνθήκες συγχρόνως, κάνει ένα clipping με βάση τις τιμές (L, H)–άνω και κάτω όριο . Οι τιμές (L, H) βρίσκονται

- αν $y_1 = y_2$,

$$L = \max(0, \alpha_2 - \alpha_1), H = \min(C, C + \alpha_2 - \alpha_1) \tag{B.14}$$

- αν $y_1 \neq y_2$,

$$L = \max(0, \alpha_1 + \alpha_2 - C), H = \min(C, \alpha_1 + \alpha_2) \tag{B.15}$$

Η δεύτερη παράγωγος της συνάρτησης κόστους κατά μήκος της ευθείας (Β.12) αποδεικνύεται εύκολα ότι είναι ίση με

$$\nu = 2k(x_1, x_2) - k(x_1, x_1) - k(x_2, x_2) \tag{B.16}$$

οπότε μπορούμε να βρούμε την τιμή του α_2 για την οποία ελαχιστοποιείται η συνάρτηση κόστους (για μία απόδειξη βλ. [48])

$$\alpha_2^{max} = \alpha_2 - \frac{y_2(E_1 - E_2)}{\nu} \tag{B.17}$$

όπου E_1, E_2 το σφάλμα της συνάρησης f στα δείγματα $1, 2$. Η νέα τιμή του α_2' μετά από το clipping είναι

$$\alpha_2' = \begin{cases} H, & \alpha_2^{max} \geq H \\ \alpha_2^{max}, & L < \alpha_2^{max} \leq 0 \\ L & \alpha_2^{max} \leq L \end{cases} \tag{B.18}$$

Με βάση την τιμή του α_2' και με τους περιορισμούς που έχουμε, παίρνουμε

$$\alpha_1' = \alpha_1 + y_1 y_2(\alpha_2 - \alpha_2') \tag{B.19}$$

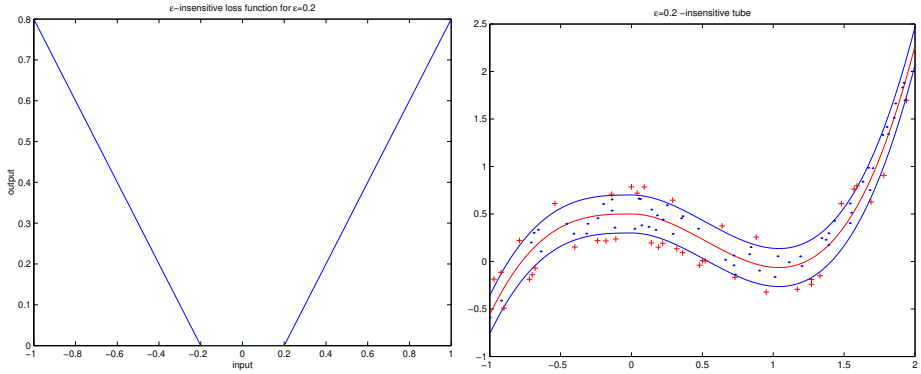
Δεν επιχειρήθηκε να δοθεί πλήρης απόδειξη όλων των τύπων που υπεισέχονται στην απόδειξη για λόγους συντομίας' η απόδειξή τους μπορεί να βρεθεί στο ([48]). Εκτός της

ύπαρξης αναλυτικής λύσης για την επίλυση του προβλήματος βελτιστοποίησης στην ίδια αναφορά υποδεικνύονται διάφορες ευριστικές για την αποθήκευση των πιο χρήσιμων εσωτερικών γινομένων καθώς και για την επιλογή των πολλαπλασιαστών Lagrange που επιλέγονται σε κάθε βήμα.

Ο αλγόριθμος αυτός έχει καθιερωθεί ως ο καταλληλότερος για την εκπαίδευση δικτύων SVM για ταξινόμηση και έχει δώσει νέα ώθηση στο πεδίο των SVM που τώρα μπορούν να εκπαιδευτούν σε χρονικό διάστημα περίπου 3-4 τάξεων μεγέθους μικρότερο από ότι αν χρησιμοποιηθούν οι 'έτοιμες' interior point τεχνικές βελτιστοποίησης.

B.2 Support Vector Machines για πρόβλεψη

Ως τώρα δεν έχουμε αναφερθεί καθόλου στο πρόβλημα της πρόβλεψης. Για το σκοπό αυτό χρησιμοποιείται συνήθως στα πλάισια των SVM η ϵ-αδιάφορη (ϵ-insnesitive) συνάρτηση λάθους (σχ. B.2) αν και δεν είναι η μοναδική που επιτρέπεται (βλ. π.χ. [46] σχετικά με διαφορετικές συναρτήσεις κόστους). Τα σημεία που φαίνονται στο σχήμα B.2 με



B.2: ϵ-αδιάφορη συνάρτηση κόστους

κόκκινους σταυρούς τιμωρούνται ενώ τα υπόλοιπα δεν συμβάλουν στο συνολικό σφάλμα. Τα σημεία, δηλαδή που βρίσκονται γύρω από μία περιοχή της πραγματικής τιμής της συνάρτησης που καθορίζεται από την τιμή του ϵ χαίρουν της ανοχής της συνάρτησης σφάλματος αυτής. Έτσι, σε αναλογία με την περίπτωση της ταξινόμησης, μπορούμε να πούμε ότι δεν ασχολούμαστε τόσο με τα εύκολα σημεία και επικεντρώνουμε στα δύσκολα.

Το πρόβλημα της έκπαιδευσης ενός δικτύου SVM αλλάζει ως εξής

$$\text{Ελαχιστοποίησε την ποσότητα} \quad \frac{1}{2}w^T w$$
$$\text{Υπό τη συνθήκη} \quad \begin{cases} y_i - w^T x_i - b & \leq \epsilon \\ w^T x_i + b - y_i & \leq \epsilon \end{cases}$$

Παρατηρούμε ότι οι συνθήκες του προβλήματος βελτιστοποίησης δεν μπορούν να εκφραστούν μέσω μίας ενιαίας σχέσης, όπως στην περίπτωση της ταξινόμησης.

Στην περίπτωση που θέλουμε να επιβάλλουμε ασθενέστερες συνθήκες για το σφάλμα της πρόβλεεψης, δηλαδή επιτρέψουμε να υπάρχει και ένα σφάλμα ξ μεγαλύτερο απο ϵ, καταλήγουμε στο παρακάτω πρόβλημα βελτιστοποίησης:

Το πρόβλημα της έκπαιδευσης ενός δικτύου SVM αλλάζει ως εξής

$$\text{Ελαχιστοποίησε την ποσότητα} \quad \frac{1}{2}w^T w + C\sum_{i=1}^{l}(\xi_i + \xi_i^*)$$

$$\text{Υπό τη συνθήκη} \quad \begin{cases} y_i - w^T x_i - b & \leq \epsilon + \xi_i \\ w^T x_i + b - y_i & \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* & \geq 0 \end{cases}$$

Η Lagrangian αυτού του προβλήματος ισούται με

$$\begin{aligned} L &= \frac{1}{2} w^T w + C \sum_{i=1}^{l} (\xi_i + \xi_i^*) - \sum_{i=1}^{l} \alpha_i (\epsilon + \xi_i - y_i + w^T x_i + b) - \\ &- \sum_{i=1}^{l} \alpha_i^* (\epsilon + \xi_i^* + y_i - w^T x_i - b) - \sum_{i=1}^{l} (\nu_i \xi_i^* + \nu_i^* \xi_i^*) \\ \text{όπου} \quad & \alpha_i^*, \nu_i^* \geq 0 \end{aligned}$$

Από τη Lagrangian μπορούμε να πάρουμε το δυικό πρόβλημα (βλ. π.χ. [46])

$$\begin{aligned} \text{Μεγιστοποίησε την ποσότητα} \quad & - \frac{1}{2} \sum_{i,j=1}^{l} \left[(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) x_i^T x_j \right] - \epsilon \sum_{i=1}^{l} (\alpha_i + \alpha_i^*) + \\ & + \sum_{i=1}^{l} y_i (\alpha_i - \alpha_i^*) \\ \text{Υπό τις συνθήκες} \quad & \begin{cases} \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) & = 0 \\ \alpha_i, \alpha_i^* & \in [0, C] \end{cases} \end{aligned} \tag{B.20}$$

Και σε αυτήν την περίπτωση το διάνυσμα w μπορεί να γραφεί ως συνδυασμός των δεδομένων εκπαίδευσης, όπως προκύπτει από την ελαχιστοποίηση της Langrangian ως προς το w:

$$w = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) x_i \rightarrow f(x) = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) x^T x_i + b \tag{B.21}$$

Από τη σχέση αυτή βλέπουμε, πάλι ότι τα δεδομένα τα οποία επηρεάζουν περισσότερο την συνάρτηση f είναι αυτά τα οποία έχουν πολλαπλασιαστή με μεγάλο μέτρο (σημειώνεται ότι αν ένας από τους δύο πολλαπλασιαστές α_i, α_i^* είναι μη μηδενικός ο άλλος πρέπει να είναι ίσος με μηδέν, καθώς μόνο μία από τις δύο ανισότητες B.23 είναι δυνατόν να ισχύει για ένα σημείο). Τα δεδομένα που έχουν το μεγαλύτερο πολλαπλασιαστή (C) είναι αυτά τα οποία βρίσκονται εκτός της περιοχής ανοχής της συνάρτησης λάθους ενώ αυτά τα οποία βρίσκονται στο εσωτερικό της έχουν

Το b μπορεί να υπολογιστεί με βάση τις συνθήκες KKT, οι οποίες μας λένε ότι στη λύση του προβλήματος βελτιστοποίησης πρέπει να ικανοποιούνται οι σχέσεις

$$\begin{aligned} \alpha_i (\epsilon + \xi_i - y_i + < w, x_i > + b) = 0 \qquad & \& \qquad (C - \alpha_i) \xi_i = 0 \\ \alpha_i^* (\epsilon + \xi_i^* + y_i - < w, x_i > - b) = 0 \qquad & \& \qquad (C - \alpha_i^*) \xi_i^* = 0 \end{aligned} \tag{B.22}$$

με τις οποίες μπορούμε να υπολογίσουμε το b από τις τιμές των y_i, w, x_i, ξ_i από τους πολλαπλασιαστές Lagrange $\alpha_i \in (0, C)$.

Η χρήση πυρήνων μπορεί άμεσα να εφαρμοστεί και στην περίπτωση της πρόβλεψης, αφού παντού τα δεδομένα εμφανίζονται με τη μορφή εσωτερικών γινομένων οπότε μπορούν να αντικαταστηθούν με πυρήνες εσωτερικού γινομένου. Δηλαδή μπορούμε άμεσα να χρησιμοποιήσουμε radial basis functions ,πολυεπίπεδα perceptrons κ.α. για το σκοπό της πρόβλεψης, με τον αλγόριθμο εκμάθησης του SVM να αναλαμβάνει την εκμάθηση του δικτύου.

ν–SVM για πρόβλεψη

Μία καινούρια τεχνική για εκπαίδευση δικτύων SVM έχει προταθεί για το πρόβλημα της πρόβλεψης ([47]) και έχει εφαρμοστεί και στο πρόβλημα της ταξινόμησης που μας επιτρέπει αντί να δίνουμε εκ των προτέρων ένα συκεκριμένο διάστημα άνοχης ϵ της συνάρτησης, το αφήνουμε να μεταβάλλεται και αντί αυτού προσδιορίζουμε ένα συντελεστή σημασίας ν του ϵ αυτού καθ'αυτού. Έτσι προσπαθούμε παράλληλα με την υπόλοιπη συνάρτηση κόστους να ελαχισοποιήσουμε και το ϵ όπου η ελαχιστοποίησή του παίζει ένα ρόλο στη συνολική συνάρτηση που καθορίζεται από την παράμετρο ν. Το νέο πρόβλημα βελτιστοποίησης γράφεται τώρα (ο όρος $\frac{1}{l}$ υπάρχει αποκλειστικά για να είμαστε σύμφωνοι με την αναφορά ([47]) και θα μπορούσε να μην υπάρχει, αν προηγηθεί ορισμός $C^* = \frac{C}{l}$) :

$$\begin{array}{rl} \text{Ελαχιστοποίησε την ποσότητα} & \frac{1}{2}w^T w + C(\nu\epsilon + \frac{1}{l}\sum_{i=1}^{l}(\xi_i + \xi_i^*)) \\ \text{Υπό τη συνθήκη} & \left\{ \begin{array}{ll} y_i - < w, x_i > -b & \leq \epsilon + \xi_i \\ < w, x_i > +b - y_i & \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* & \geq 0 \end{array} \right. \end{array}$$

Το δυικό αυτού του προβλήματος είναι :

$$\begin{array}{rl} \text{Μεγιστοποίησε την ποσότητα} & -\ \frac{1}{2}\sum_{i,j=1}^{l}\left[(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)x_i^T x_j\right] - \epsilon\sum_{i=1}^{l}(\alpha_i + \alpha_i^*) + \\ & +\ \sum_{i=1}^{l} y_i(\alpha_i - \alpha_i^*) \\ \text{Υπό τις συνθήκες} & \left\{ \begin{array}{ll} \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) & = 0 \\ \alpha_i, \alpha_i^* & \in [0, C] \end{array} \right. \end{array} \quad \text{(B.23)}$$

Στο ([47]) αποδεικνύεται ότι το ν είναι ασυμπτωτικά ίσο με το ποσοστό των Support Vectors που τελικά χρησιμοποιούνται για την πρόβλεψη, οπότε μπορούμε να πούμε ότι έχουμε τη δυνατότητα να ελέγχουμε την πολυπλοκότητα του προβλέπτη, χωρίς να καταφεύγουμε σε συμβιβασμούς σχετικά με το επιτρεπόμενο σφάλμα. Στην περίπτωση, βέβαια, που γνωρίζουμε εκ των προτέρων το αναμενόμενο σφάλμα (π.χ. από κβαντισμό ενός σήματος) είναι προτιμότερο να χρησιμοποιούμε το ϵ. (Για μία αναλυτική παρουσιάση σχετικά, βλ. ([46]).

Αλγόριθμος SMO για πρόβλεψη

Ο αλγόριθμος SMO μπορεί να εφαρμοστεί και στην περίπτωση του προβλήματος της παλινδρόμισης, αλλά με κάποιες άλλαγες που οφείλονται κυρίως στο γεγονός ότι για κάθε σημείο δεν αντιστοιχεί ένας μόνο πολλαπλασιαστής Lagrange αλλά δύο (οι α_i, α_i^* του προβλήματος βελτιστοποίηση).

Κάθώς η απόδειξη είναι μάλλον τεχνική και δεν παρουσιάζει κάποιο ιδιαίτερο ενδιαφέρον, μπορούμε να παραπέμψουμε στο [46], ενώ έχουν προταθεί διάφορες βελτιώσεις τόσο στον πυρήνα του αλγορίθμου βελτιστοποίησης, όσο και στις ευρεστικές που χρησιμοποιούνται για την επιλογή των πολλαπλασιαστών Lagrange με βάση τους οποίους γίνεται η βελτιστοποίηση σε κάθε βήμα[50],[49].

Βιβλιογραφία

[1] Abarbanel H.D.I. *Analysis of Observed Chaotic Data* Springer 1996

[2] Editors:Casdagli M.,Eubank S. *Nonlinear modeling and forecasting* Proceedings volume in the Santa Fe Institute Studies in the Sciences of Complexity

[3] Editors:Weigend A.,Gerschenfeld N. *Time-series prediction*Proceedings volume in the Santa Fe Institute Studies in the Sciences of Complexity

[4] Mandelbrot, B. *The fractal geometry of nature* W.H. Freeman and co. New York,1982

[5] Peitgen,Jurgens,Saupe *Chaos and Fractals, New frontiers of Science* Springer Verlag 1992

[6] http://www.cis.hut.fi/projects/somtoolbox/

[7] Flanagan,J *Speech Analysis Synthesis and Perception* Springer -Verlag 1972

[8] Rabiner,R Schafer,W *Digital Processing of Speech Signals* Prentice Hall 1978

[9] Tishby Naftali *A dynamical systems approach to speech processing* 1990 IEEE

[10] Townshed Brent *Nonlinear Prediction of Speech Signals* 1992 in [2]

[11] Teager H.M,Teager S.M. *Evidence for nonlinear Sound Production mechansims in the vocal tract*

[12] Eckmann,Ruelle *Ergodic theory of chaos and strange attractors* Reviews of Modern Physics 1985

[13] Πιτσικάλης Βασίλης Ανάλυση χαοτικών χρονοσειρών με εφαρμογές σε ήχους φωνής Φεβρουάριος 2001 Διπλωματική εργασία στο τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Εθνικού Μετσοβίου Πολυτεχνείου

[14] Abarbanel H.D.I.,Carrol T.A.,Pecora L.M.,Sidorowich J.J.,Tsimring L.S, *Prediciting Physical variables in time-delay embedding* Physical Review E,Volume 49,Number 3,March 1994

[15] Giona M,Lentini F.,Cimagalli V.,*Functional reconstruction and local prediction of chaotic time series* ,Physical Reveiw A,Volume 44,Number 6,September 1991

[16] Kugiumtzis D.,Lillekjlie,Christophersen N.,*Chaotic time series Part II:System identiication and prediction* February 1995

[17] Kugiumtzis D.,*State Space Local Linear Prediction* in 'Nonlinear Deterministic Modelling and Forecasting of Economic and Financial Time Series' by A.Soofi and L.Cao

[18] Gouesbet *Global Vector Field Reconstruction by using a multivariate polynomial L_2 approximation on nets*

[19] Haykin,S. *Neural Networks:A comprehensive foundation* Prentice Hall 1999

[20] Λύκας Αριστείδης Εισαγωγή στην υπολογιστική νοημοσύνη Πανεπιστήμιο Ιωαννίνων 1999

[21] Lapedes A. Farber R .*How Neural Nets Work* Theoretical Division, Los Alamos National Laboratory

[22] Kuo J.-M. Principe J. *Reconstructed Dynamics and Chaotic Signal Modeling* Computational NeuroEngineering Laboratory, University of Florida

[23] Principe J., Wang L. Motter M. *Local Dynamic Modeling with Self-Organizing Maps and Applications to Nonlinear System Identification and Control*

[24] Takagi T. Sugeno M. *Fuzzy identification of systems and its applications to modeling and control* IEEE transactions on Systems, Man and Cybenetics Vol 15 No 1 1985

[25] Jang J. *ANFIS:Adaptive Network-Based Fuzzy Inference System* IEEE transactions on Systems, Man and Cybenetics Vol 23 pp.665-685 1993

[26] Stokbro L.,Umberger D.K.,*Forecasting with weighted maps* in [2] Computational NeuroEngineering Laboratory, University of Florida

[27] Davis M., Sigal R. Weyker J. *Computability, complexity, and languages : fundamentals of theoretical computer science* Academic Press, 1994

[28] Abarbanel H.D.I.,Kennel Matthew *Local false nearest neighbors and dynamical dimensions of observed chaotic data* Physical Review E,Volume 47,Number 5, May 1993

[29] Abarbanel H.D.I. ,Suschik M.M. *Local or dynamical dimensions of nonlinear systems inferred from observations* International Journal on Bifurcations and Chaos Vol.3. ,No 3,1993

[30] Abarbanel H.D.I. *Local and Global Lyapunov Exponents on a Strange Attractor* in [2]

[31] Bryant Paul,Brown Reggie,Abarbanel H.D.I.,*Lyapunov Exponents from Observed Time Series* Physical Review Letters ,Volume 65,September 1990,Number 13

[32] Brown Reggie *Calculating Lyapunov exponents for short and/or noisy data sets* Physical Review E,Volume 47,Number 6,June 1993

[33] Brown Reggie,Paul Bryant,Abarbanel Henry D.I.,*Computing the Lyapunov spectrum of a dynamical system from an observed time series* Physical Review A ,Volume 43 Number 6 March 1991

[34] Eckmann J.-P., Oliffson Kamphorst,Ruelle D.,Ciliberto S *Lyapunov exponents from time series* Physical Review A,Volume 34,Number 6, December 1986

[35] Kennel Matthew,Abarbanel H.D.I.,Sidorowich J.J.,*Predicition Errors and Local Lyapunov Exponents*

[36] Kugiumtzis D.,Lillekjlie,Christophersen N.,*Chaotic time series Part I:Estimation of some invariant properties in state space* February 1995

[37] Wolf,A. Swift,J. Swinney,H. Vastano,J. *Determining Lyapunov exponents from a time series* Physica 16D 285-317

[38] Taylor W. *Application of Nonlinear Prediction on Signal Separation* in [2]

[39] Farmer J.,Sidorowich J. *Exploiting Chaos to Predict the Future and Reduce Noise* Theoretical Division and Center for Nonlinear Studies, Los Alamos National Laboratory

[40] Golub G., Van Loan C.: Matrix Computations The John Hopkins University Press, 1989

[41] Riedmiller M., Braun H.: A Direct Adaptive Method for Faster Backpropagation Learning: The RPROP Algorithm *in* Proceedings of the IEEE International Conference on Neural Networks San Francisco, CA, March 28- April 1, 1993

[42] Manfred Mannle: FTSM- Fast Takagi-Sugeno Fuzzy Modeling

[43] Boser B., Guyon I.,Vapnik V., *A training algorithm for optimal margin classifiers* Fifth Annual Workshop on Computational Learning Theory ,Pittsburg, 1992 ACM

[44] Vapnik V. *Estimation of dependencies based on empirical data* Springer Verlag 1982

[45] Burges C. *A Tutorial on Support Vector Machines for Pattern Recognition* 1998 www.support-vector.net

[46] Smola Alexander *Learning with Kernels* Διδακτορική διατριβή ,1998 G.M.D. Berlin

[47] Scholkopf B.,Bartlett P.,Smola A.,Williamson R.*Shrinking the Tube:A New Support Vector Regression Algorithm* Advances in NIPS 1998

[48] Platt J.,*Fast Training of Support Vector Machines using Sequential Minimal Optimization* Learning with Kernels ,MIT Press 1999

[49] Flake G,Lawrence S.,*Efficient SVM Regression Training with SMO* Machine Learning 2001

[50] Shevade S., Keerthi S., Bhattacharrya C., Murthy K., *Improvemens to SMO Algorithm for SVM Regression* Technical Report CD 99 16

[51] Munkherjee S.,Osuna E., Girosi F.,*Nonlinear Prediction of Chaotic Time Series Using Support Vector Machines* IEEE NNSP'97