

ΓΡΑΜΜΙΚΗ ΑΛΓΕΒΡΑ

Διάλεξη 10

Ορθογώνιες προβολές Βέλτιστη προσέγγιση Λύσεις ελαχίστων τετραγώνων

Ε. Στεφανόπουλος & Ε. Γαλλόπουλος

Τμήμα Μηχ. Η/Υ και Πληροφορικής, Πανεπιστήμιο Πατρών

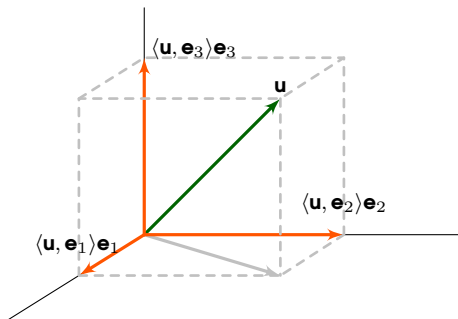
8 Δεκεμβρίου 2025

1. Προβολές

Η **προβολή** ενός σημείου (διανύσματος), έστω $\mathbf{x} \in X$ σε έναν υπόχωρο U είναι ένα αποτύπωμα του $\mathbf{x}$ στον υπόχωρο. Το αποτύπωμα αυτό κατασκευάζεται σύμφωνα με συγκεκριμένα κριτήρια. Για παράδειγμα η προβολή μπορεί να επιλέγεται έτσι ώστε το αποτύπωμα να είναι το πλησιέστερο (σύμφωνα με κάποια μετρική) στο αρχικό διάνυσμα.

- Διαισθητικά: είναι ο «ίσκιος» του διανύσματος πάνω σε μια ευθεία (αν επιλέξουμε μονοδιάστατο υπόχωρο διάστασης) ή επίπεδο (αν επιλέξουμε δισδιάστατο υπόχωρο διάστασης).
- Επικεντρώνοντας την προσοχή μας στις έννοιες, στις οντότητες και στις αφαιρέσεις (abstractions) της Γραμμικής Άλγεβρας που μας ενδιαφέρουν, στη γενική περίπτωση, η προβολή ενός σημείου (διανύσματος) σε έναν υπόχωρο μπορεί να θεωρηθεί ως το «αποτύπωμα» του στον υπόχωρο.
- Όπως ένα αντικείμενο μπορεί να δώσει διαφορετικές σκιές στο επίπεδο ανάλογα με τον τρόπο που το φωτίζουμε, έτσι και ένα διάνυσμα μπορεί να προβάλλεται σε έναν υπόχωρο με διαφορετικές προβολές, που εξαρτώνται από την κατεύθυνση ανάλογα με την κατεύθυνση ως προς την οποία γίνεται η προβολή. Η ορθογώνια προβολή αποτελεί απλώς μία συγκεκριμένη, ιδιαίτερα σημαντική επιλογή.
- Για να είναι πλήρης ο ορισμός της προβολής στη γενική περίπτωση, επιπλέον του υπόχωρου στον οποίο προβάλλουμε πρέπει να προσδιοριστεί και ο υπόχωρος παράλληλα προς τον οποίο επιτελείται η προβολή.

Παράδειγμα



Γράφοντας το διάνυσμα $\mathbf{u} \in \mathbb{R}^3$ ως

$$\mathbf{u} = \langle \mathbf{u}, \mathbf{e}_1 \rangle \mathbf{e}_1 + \langle \mathbf{u}, \mathbf{e}_2 \rangle \mathbf{e}_2 + \langle \mathbf{u}, \mathbf{e}_3 \rangle \mathbf{e}_3,$$

ονομάζουμε τα διανύσματα

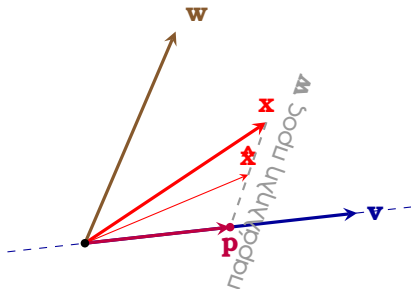
$$\mathbf{p}_1 = \langle \mathbf{u}, \mathbf{e}_1 \rangle \mathbf{e}_1, \quad \mathbf{p}_2 = \langle \mathbf{u}, \mathbf{e}_2 \rangle \mathbf{e}_2, \quad \mathbf{p}_3 = \langle \mathbf{u}, \mathbf{e}_3 \rangle \mathbf{e}_3$$

προβολές (projections) του $\mathbf{u}$, αντίστοιχα, στον x-άξονα, y-άξονα, z-άξονα, ή στους υποχώρους του $\mathbb{R}^3$, $\text{span}\{\mathbf{e}_1\}$, $\text{span}\{\mathbf{e}_2\}$, και $\text{span}\{\mathbf{e}_3\}$ αντίστοιχα.

Πλάγια Προβολή (oblique projection)

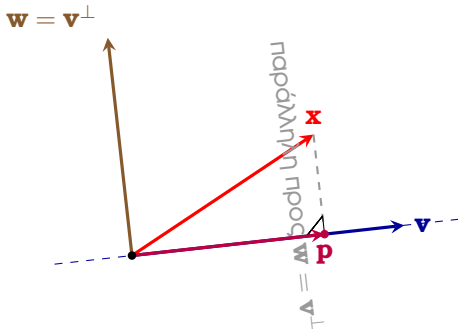
$$\mathbf{P}_{\mathbf{v},\mathbf{w}}(\mathbf{x})$$

- Η προβολή γίνεται επί του $\mathbf{v}$ κατά την κατεύθυνση του $\mathbf{w}$. Με γραμμοαλγεβρικούς όρους, η προβολή γίνεται επί του υπόχωρου $\text{span}\{\mathbf{v}\}$ παράλληλα στον υπόχωρο $\text{span}\{\mathbf{w}\}$



Ορθογώνια Προβολή $\mathbf{P}_{\mathbf{v}}(\mathbf{x})$

- Η προβολή γίνεται επί του $\mathbf{v}$ παράλληλα προς τον $\text{span}\{\mathbf{w}\} = \text{span}\{\mathbf{v}\}^\perp$.
- Ειδική περίπτωση της πλάγιας προβολής. Στο παράδειγμα στις 2 διαστάσεις $\mathbf{v} \perp \mathbf{w}$.



Στη συνέχεια, θα παραλείψουμε τις παρενθέσεις και θα γράφουμε π.χ. $\mathbf{P}_{\mathbf{v},\mathbf{w}}\mathbf{x}$ ή απλά $\mathbf{P}\mathbf{x}$.

Προβολή: ιδιότητες και παραδείγματα από τη ζωή

Στηριζόμαστε στη διαίσθηση και λαμβάνουμε ως δεδομένο ότι η απεικόνιση της προβολής μπορεί να αναπαρασταθεί με τετραγωνικό μητρώο, έστω P .

1. Απλοποίηση της πραγματικότητας Μαθηματικά: Προβολή από «μεγάλο» χώρο σε μικρότερο υπόχωρο $\sim$ Ζωή: Φέρνουμε μια πολύπλοκη κατάσταση σε πιο απλή μορφή, κρατώντας τα ουσιώδη.

2. Χάνεται πληροφορία Μαθηματικά: Δεν μπορούμε γενικά να ανακτήσουμε το αρχικό διάνυσμα από την προβολή, δεν υπάρχει το P^{-1} . $\sim$ Ζωή: Κάθε οπτική γωνία είναι μερική και κάτι πάντα αφήνουμε έξω.

3. Ιδιομορφία - Αυτοδυναμία Μαθηματικά: Αν ξαναπροβάλεις κάτι ήδη προβεβλημένο, δεν αλλάζει, $P^2 = P \sim$ Ζωή: Αν κοιτάς συνεχώς μέσα από το ίδιο φίλτρο/πλαίσιο, η εικόνα μένει ίδια μένεις «κολλημένος» στην ίδια ερμηνεία.

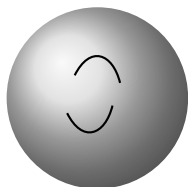
4. Ορθογώνια προβολή = πλησιέστερο σημείο Μαθηματικά: Η ορθογώνια προβολή δίνει το σημείο στον υπόχωρο με ελάχιστη απόσταση από το αρχικό διάνυσμα. $\sim$ Ζωή: Η «απλοποιημένη εκδοχή» μιας κατάστασης είναι η πιο κοντινή στην αλήθεια, μέσα στο δεδομένο πλαίσιο.

5. Εξάρτηση από τον υπόχωρο Μαθηματικά: Διαφορετικός υπόχωρος $\Rightarrow$ διαφορετική προβολή. $\sim$ Ζωή: Διαφορετικό πλαίσιο (αξίες, προτεραιότητες, ρόλος) $\Rightarrow$ διαφορετική εικόνα της ίδιας πραγματικότητας.

6. Κάθετες συνιστώσες «εξαφανίζονται» Μαθηματικά: Αν x είναι κάθετο στον υπόχωρο, τότε η ορθογώνια προβολή δίνει $P(x) = 0$. $\sim$ Ζωή: Ό,τι δεν «χωράει» στο πλαίσιο που χρησιμοποιούμε, συχνά το αντιμετωπίζουμε σαν να μην υπάρχει.

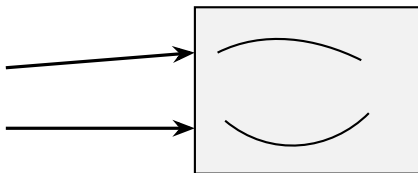
Προβολή και απώλεια πληροφορίας: $\Gamma_{\eta} \rightarrow \text{Χάρτης}$

Πραγματικότητα (3D)



Πλούσια πληροφορία

Προβολή (2D χάρτης)



Απλοποιημένη αναπαράσταση

Όλη η δομή της Γ_{η} : καμπυλότητα
πραγματικές αποστάσεις κ.λπ.

Κρατάμε μόνο ό,τι χωράει σε επίπεδο: χάνουμε/
παραμορφώνουμε μέρος της πληροφορίας

- Η σφαίρα (Γ_{η}) ζει στον «πλήρη» χώρο ($\mathbb{R}^3$)
- Ο χάρτης είναι η προβολή σε έναν δισδιάστατο υπόχωρο του $\mathbb{R}^3$.
- Χρήσιμο, αλλά με αναπόφευκτη απώλεια πληροφορίας.

Μαθηματικές ιδιότητες προβολών: Σύνοψη

- Υπενθυμίζουμε (Διάλεξη 7) την έννοια του ευθέως άθροισματος υπόχωρων καθώς και το ότι αν $X = V \oplus W$ και $\mathbf{x} \in X$ τότε υπάρχουν μοναδικά $\mathbf{v} \in V, \mathbf{w} \in W$ τέτοια ώστε $\mathbf{x} = \mathbf{v} + \mathbf{w}$.
- Η **απεικόνιση** $P_{V,W} : X \rightarrow V \subseteq X$ που για κάθε $\mathbf{x} \in X$, επιστρέφει το $\mathbf{v} \in V$ του αντίστοιχου ευθέως αθροίσματος, δηλ. $P_{V,W}\mathbf{x} = \mathbf{v}$ καλείται **πλάγια (ή πλαγιογώνια) προβολή επί του V παράλληλα στο W** .
- Στην ειδική περίπτωση που $W = V^\perp$ η $P_{V,V^\perp}$ ονομάζεται **απεικόνιση ορθογώνιας προβολής** και γράφεται P_V .
- Στην περίπτωση που $X = \mathbb{R}^{n \times n}$, οι απεικονίσεις πλάγιας και ορθογώνιας προβολής μπορούν να αναπαρασταθούν ως $n \times n$ μητρώα που αποκαλούνται **μητρώα προβολής** με πολύ συγκεκριμένες ιδιότητες:
 - Όλα τα μητρώα προβολής είναι **αυτοδύναμα** (idempotent), δηλ. ισχύει ότι $P^2 = P$,
 - Ένα μητρώο $P \in \mathbb{R}^{n \times n}$ είναι μητρώο ορθογώνιας προβολής αν και μόνον αν ισχύουν ότι $P^2 = P$ και $P^T = P$.
- Εννοείται ότι στην πράξη δεν είναι ανάγκη να χρησιμοποιούμε τους παραπάνω συμβολισμούς και ότι οι υποδείκτες είναι προαιρετικοί αν προκύπτουν από τα συμφραζόμενα.

Θέμα: Μας δίνεται ένα διάνυσμα x από τον $\delta.x.$ και αναζητούμε :

- 1 Την προβολή, δηλ. το διάνυσμα που προκύπτει από την **προβολή** του διανύσματος σε κάποιον υπόχωρο.
- 2 Θα δούμε ότι μπορούμε να εκφράσουμε την προβολής ως ένα ειδικό μητρώο **προβολής** επί το $\mathbf{x}$.

Θεωρούμε εδώ ότι η προβολή ενός διανύσματος

- σε ευθεία είναι το μέρος του διανύσματος κατά μήκος της ευθείας.
- σε επίπεδο είναι το μέρος του διανύσματος επί του επιπέδου.
- παρόμοια γενικεύουμε την έννοια της προβολής διανύσματος επί μεγαλύτερου υπόχωρου.

Γεωμετρική σημασία της ορθογώνιας προβολής

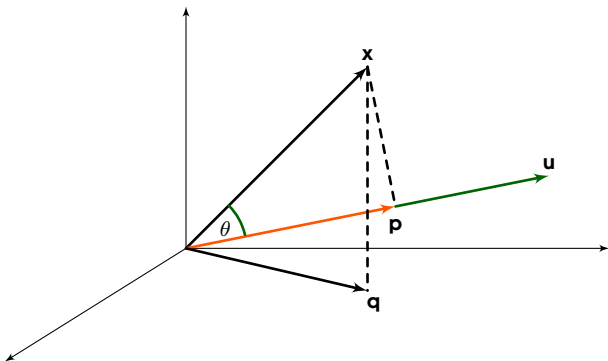
- Δοθέντος υποχώρου V , η **ορθογώνια προβολή** ενός διανύσματος $\mathbf{x}$ στο V είναι το μοναδικό $\mathbf{p} \in V$ τέτοιο ώστε:

$$\mathbf{x} - \mathbf{p} \in V^\perp.$$

- Δηλαδή, το «πιο κοντινό» σημείο του V στο $\mathbf{x}$.
- Ισοδύναμα, το $\mathbf{p}$ ελαχιστοποιεί τη νόρμα-2:

$$\|\mathbf{b} - \mathbf{v}\|_2 \quad \text{για } \mathbf{v} \in W.$$

Αν $\mathbf{u} \neq \mathbf{0}$ είναι ένα σταθερό διάνυσμα στον $\mathbb{R}^3$ και $\mathbf{x}$ είναι ένα τυχαίο διάνυσμα στον $\mathbb{R}^3$, μας ενδιαφέρει να βρούμε την προβολή του $\mathbf{x}$ επί της ευθείας η οποία περιέχει το $\mathbf{u}$.



Αν θ είναι η γωνία μεταξύ των $\mathbf{x}$ και $\mathbf{u}$ και $\mathbf{p}$ είναι η προβολή του $\mathbf{x}$ επί της ευθείας διαμέσου του $\mathbf{u}$, τότε

$$\|\mathbf{p}\| = \|\mathbf{x}\| \cos \theta$$

Κατά συνέπεια

$$\mathbf{p} = \|\mathbf{x}\| \cos \theta \frac{\mathbf{u}}{\|\mathbf{u}\|} = \frac{\|\mathbf{x}\| \|\mathbf{u}\| \cos \theta}{\|\mathbf{u}\|^2} \mathbf{u}$$

Επομένως, από τον ορισμό του συνήθους εσωτερικού γινομένου προκύπτει ότι

$$\mathbf{p} = \frac{\mathbf{x} \cdot \mathbf{u}}{\mathbf{x} \cdot \mathbf{u}} \mathbf{u} = \frac{\mathbf{u}^T \mathbf{x}}{\mathbf{u}^T \mathbf{u}} \mathbf{u} \quad (1)$$

Στο Σχήμα βλέπουμε, επίσης, την προβολή $\mathbf{q}$ του διανύσματος $\mathbf{x}$ επί του xy -επιπέδου, δηλαδή επί του υποχώρου $\text{span}\{\mathbf{e}_1, \mathbf{e}_2\}$ του $\mathbb{R}^3$, η οποία είναι

$$\mathbf{q} = \mathbf{p}_1 + \mathbf{p}_2 = \frac{\mathbf{e}_1 \cdot \mathbf{x}}{\mathbf{e}_1 \cdot \mathbf{e}_1} \mathbf{e}_1 + \frac{\mathbf{e}_2 \cdot \mathbf{x}}{\mathbf{e}_2 \cdot \mathbf{e}_2} \mathbf{e}_2$$

ισοδύναμα, αφού $\mathbf{e}_1 \cdot \mathbf{e}_1 = \mathbf{e}_2 \cdot \mathbf{e}_2 = 1$,

$$\mathbf{q} = \langle \mathbf{x}, \mathbf{e}_1 \rangle \mathbf{e}_1 + \langle \mathbf{x}, \mathbf{e}_2 \rangle \mathbf{e}_2$$

όπου $\langle \cdot, \cdot \rangle$ είναι το σύνηθες εσωτερικό γινόμενο στον $\mathbb{R}^3$.

Το μήτρωο της ορθογώνιας προβολής

Η εξοικείωσή μας με τα στοιχειώδη μήτρωα επιτρέπει να δούμε, αναφερόμενοι στο αρχικό παράδειγμα, όπου $\mathbf{u} = (\xi_1 \ \xi_2 \ \xi_3)^T$, ότι

$$\mathbf{p}_1 = \begin{pmatrix} u_1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{x}, \quad \mathbf{p}_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{x}, \quad \mathbf{p}_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{x}$$

ή

$$\mathbf{p}_1 = (\mathbf{e}_1 \ \mathbf{0} \ \mathbf{0})\mathbf{x} = P_1\mathbf{x}, \quad \mathbf{p}_2 = (\mathbf{0} \ \mathbf{e}_2 \ \mathbf{0})\mathbf{x} = P_2\mathbf{x}, \quad \mathbf{p}_3 = (\mathbf{0} \ \mathbf{0} \ \mathbf{e}_3)\mathbf{x} = P_3\mathbf{x}$$

και

$$\mathbf{q} = (\mathbf{e}_1 \ \mathbf{e}_2 \ \mathbf{0})\mathbf{x} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{x} = Q\mathbf{x}.$$

Ερώτημα: Ποιο είναι το μήτρωο P της προβολής ενός διανύσματος $\mathbf{x}$ του $\mathbb{R}^m$ πρώτον επί ενός διανύσματος $\mathbf{u}$, και δεύτερον επί ενός υποχώρου V του $\mathbb{R}^m$;

- 1 Αναζητούμε μήτρωο P ώστε $P\mathbf{x} = \mathbf{p} = \text{proj}_{\mathbf{u}} \mathbf{x}$, ισοδύναμα για ποιο P είναι

$$P\mathbf{x} = \frac{\mathbf{x}^T \mathbf{u}}{\mathbf{u}^T \mathbf{u}} \mathbf{u};$$

επειδή $(\mathbf{x}^T \mathbf{u}) \mathbf{u} = \mathbf{u}(\mathbf{u}^T \mathbf{x}) = (\mathbf{u} \mathbf{u}^T) \mathbf{x}$ (γιατί:), τελικά βρίσκουμε

$$\text{proj}_{\mathbf{u}} \mathbf{x} = P\mathbf{x} = \frac{\mathbf{u} \mathbf{u}^T}{\mathbf{u}^T \mathbf{u}} \mathbf{x}. \quad (2)$$

- 2 Έστω ότι τα $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ είναι γραμμικά ανεξάρτητα διανύσματα στο $\mathbb{R}^m$ και έστω $V = \text{span}\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\}$. Αν A είναι το μήτρωο με στήλες τα $\mathbf{a}_j$, τότε $\text{range } A = V$, δηλαδή ο υποχώρος V είναι ο χώρος στηλών του A . Έστω $\mathbf{c} \in \mathbb{R}^m$ τέτοιο ώστε $\text{proj}_V \mathbf{x} = A\mathbf{c}$. Επειδή το $\mathbf{x} - A\mathbf{c}$ είναι ορθογώνιο στον V έπεται ότι $\mathbf{a}_j^T (\mathbf{x} - A\mathbf{c}) = 0$ για $j = 1, 2, \dots, n$. Κατά συνέπεια

$$A^T (\mathbf{x} - A\mathbf{c}) = \mathbf{0} \Rightarrow A^T \mathbf{x} = A^T A \mathbf{c}$$

Το μήτρωο $A^T A$ είναι αντιστρέψιμο, βλέπε επόμενο Θεώρημα, επομένως

$$(A^T A)^{-1} A^T \mathbf{x} = \mathbf{c} \Rightarrow A(A^T A)^{-1} A^T \mathbf{x} = A\mathbf{c}$$

κατά συνέπεια

$$\text{proj}_V \mathbf{x} = P\mathbf{x} = A(A^T A)^{-1} A^T \mathbf{x}. \quad (3)$$

Θεώρημα 1

Εάν το $m \times n$ μήτρωο A έχει γραμμικά ανεξάρτητες στήλες, τότε το μήτρωο $A^T A$ είναι αντιστρέψιμο.

Απόδειξη.

Το μήτρωο $A^T A$ είναι τετραγωνικό, οπότε αρκεί να δείξουμε ότι $\text{null } A^T A = \{\mathbf{0}\}$. Πρώτα δείχνουμε ότι τα μήτρωα A και $A^T A$ έχουν τον ίδιο μηδενοχώρο, δηλαδή $\text{null } A = \text{null } A^T A$. Πράγματι, από τη σχέση

$$A\mathbf{x} = \mathbf{0} \Rightarrow A^T A\mathbf{x} = A^T \mathbf{0} = \mathbf{0}$$

έπεται ότι $\text{null } A \subseteq \text{null } A^T A$, ενώ από την

$$A^T A\mathbf{x} = \mathbf{0} \Rightarrow \mathbf{x}^T A^T A\mathbf{x} = \mathbf{x}^T \mathbf{0} \Rightarrow (A\mathbf{x})^T A\mathbf{x} = 0 \Rightarrow \|A\mathbf{x}\|^2 = 0$$

έπεται ότι $\text{null } A^T A \subseteq \text{null } A$. Στη συνέχεια δείχνουμε ότι $\text{null } A = \{\mathbf{0}\}$.

Πράγματι, αν $A\mathbf{x} = \mathbf{0}$, το διάνυσμα $A\mathbf{x}$ ως γραμμικός συνδυασμός των γραμμικά ανεξάρτητων στηλών του A για να είναι το μηδενικό διάνυσμα, συνεπώς $\mathbf{x} = \mathbf{0}$. Έτσι τελικά έχουμε ότι $\text{null } A^T A = \text{null } A = \{\mathbf{0}\}$, που είναι το ζητούμενο. $\square$

Παράδειγμα 1

Το μήτρωο

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

έχει γραμμικά ανεξάρτητες στήλες. Υπολογίζοντας

$$A^T A = \begin{pmatrix} 1 & 2 & -1 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2 & 0 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 6 & 0 \\ 0 & 2 \end{pmatrix}$$

βλέπουμε, όπως εξασφαλίζει το Θεώρημα 1, ότι το $A^T A$ είναι αντιστρέψιμο και

$$(A^T A)^{-1} = \frac{1}{12} \begin{pmatrix} 2 & 0 \\ 0 & 6 \end{pmatrix} = \frac{1}{6} \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}.$$

Αντίθετα, το 3×3 μήτρωο AA^T δεν αντιστρέφεται αφού $\text{rank}(AA^T) \leq \text{rank } A = 2$. Έτσι το μήτρωο

Παράδειγμα 1 (συνέχεια)

$$P = A(A^T A)^{-1} A^T = \frac{1}{6} \begin{pmatrix} 1 & 1 \\ 2 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 1 & 2 & -1 \\ 1 & 0 & 1 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & -1 \\ 1 & -1 & 2 \end{pmatrix}$$

είναι η προβολή επί του χώρου στηλών του A , ισοδύναμα για κάθε $\mathbf{u} \in \mathbb{R}^3$, έστω $\mathbf{u} = (x \ y \ z)^T$, υπάρχουν σταθερές λ και μ ώστε

$$P\mathbf{u} = A \begin{pmatrix} \lambda \\ \mu \end{pmatrix} \Leftrightarrow \frac{1}{3} \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & -1 \\ 1 & -1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \lambda \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} + \mu \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \lambda + \mu \\ 2\lambda \\ -\lambda + \mu \end{pmatrix}$$

απ' όπου επιλύοντας βρίσκουμε

$$\lambda = \frac{1}{6}(x + 2y - z), \quad \mu = \frac{1}{2}(x + z).$$

Παρατήρηση 1

Εάν P είναι μήτρωο ορθογώνιας προβολής στο $\mathbb{R}^m$ επί υποχώρου W , τότε από την (2) ή την (3), ανάλογα με την περίπτωση, προκύπτει ότι το P είναι συμμετρικό, δηλαδή

$$P^T = P.$$

Επιπλέον ισχύει ότι $P^2 = P$. Πράγματι, για κάθε διάνυσμα $\mathbf{u}$ έχουμε ότι $P\mathbf{u} \in W$, κατά συνέπεια

$$P^2\mathbf{u} = P(P\mathbf{u}) = P\mathbf{u}.$$

Το $I - P$, όπου I είναι το ταυτοτικό μήτρωο, είναι επίσης μήτρωο προβολής. Προβάλλει κάθε διάνυσμα επί του ορθογώνιου συμπληρώματος $W^\perp$ του W . Πράγματι, για κάθε διάνυσμα $\mathbf{u}$ είναι

$$\mathbf{u} = P\mathbf{u} + \mathbf{u} - P\mathbf{u} = P\mathbf{u} + (I - P)\mathbf{u}$$

και $P\mathbf{u} \perp (I - P)\mathbf{u}$, αφού για κάθε διάνυσμα $\mathbf{v}$, άρα και για $\mathbf{v} = \mathbf{u}$ είναι

Παρατήρηση 1 (συνέχεια)

$$\begin{aligned}\langle \mathbf{P}\mathbf{u}, (\mathbf{I} - \mathbf{P})\mathbf{v} \rangle &= \langle \mathbf{P}^2\mathbf{u}, (\mathbf{I} - \mathbf{P})\mathbf{v} \rangle & (\mathbf{P}^2 = \mathbf{P}) \\ &= \langle \mathbf{P}\mathbf{u}, \mathbf{P}^T(\mathbf{I} - \mathbf{P})\mathbf{v} \rangle \\ &= \langle \mathbf{P}\mathbf{u}, \mathbf{P}(\mathbf{I} - \mathbf{P})\mathbf{v} \rangle & (\mathbf{P}^T = \mathbf{P}) \\ &= \langle \mathbf{P}\mathbf{u}, (\mathbf{P} - \mathbf{P}^2)\mathbf{v} \rangle \\ &= \langle \mathbf{P}\mathbf{u}, (\mathbf{P} - \mathbf{P})\mathbf{v} \rangle \\ &= \langle \mathbf{P}\mathbf{u}, \mathbf{P}\mathbf{v} \rangle - \langle \mathbf{P}\mathbf{u}, \mathbf{P}\mathbf{v} \rangle \\ &= 0,\end{aligned}$$

οπότε $\mathbf{P}\mathbf{u} \perp (\mathbf{I} - \mathbf{P})\mathbf{v}$.

Έτσι μπορούμε να γράψουμε (γιατί:)

- ❶ $\text{range } \mathbf{P} = \mathbf{W}$ και
- ❷ $\text{null } \mathbf{P} = \text{range}(\mathbf{I} - \mathbf{P})$.
- ❸ $\mathbb{R}^m = \text{range } \mathbf{P} \oplus \text{null } \mathbf{P}$ και
- ❹ $\text{range } \mathbf{P} \perp \text{null } \mathbf{P}$.

Σημείωση (Προαιρετικό):

Για ένα μήτρωο προβολής P η ιδιότητα $P^2 = P$ είναι χαρακτηριστική μάλιστα μπορεί να αποδειχθεί ότι ένα μήτρωο P είναι μήτρωο προβολής επί ενός υποχώρου, του $\text{range } P$, αν και μόνον αν $P^2 = P$. Η ιδιότητα $P^T = P$ χαρακτηρίζει την ορθογωνιότητα.

Για παράδειγμα το μήτρωο

$$P_\lambda = \begin{pmatrix} 0 & 0 \\ \lambda & 1 \end{pmatrix}, \quad \lambda \in \mathbb{R}$$

ικανοποιεί τη σχέση $P_\lambda^2 = P_\lambda$, κατά συνέπεια είναι μήτρωο προβολής, αλλά $P_\lambda^T \neq P_\lambda$ αν $\lambda \neq 0$. Παρατηρούμε ότι

$$\begin{pmatrix} 0 & 0 \\ \lambda & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ \lambda x + y \end{pmatrix},$$

επομένως

$$\text{range } P_\lambda = \left\{ \begin{pmatrix} 0 \\ x \end{pmatrix} : x \in \mathbb{R} \right\} = \text{span} \left\{ \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\},$$

$$\text{null } P_\lambda = \left\{ \begin{pmatrix} x \\ -\lambda x \end{pmatrix} : x \in \mathbb{R} \right\} = \text{span} \left\{ \begin{pmatrix} 1 \\ -\lambda \end{pmatrix} \right\}.$$

Άρα, όπως περιμένουμε άλλωστε, $\mathbb{R}^2 = (\text{range } P_\lambda) \oplus (\text{null } P_\lambda)$. Ελέγχουμε αν $\text{range } P_\lambda \perp \text{null } P_\lambda$, όπως συμβαίνει στην περίπτωση της ορθογώνιας προβολής. Έχουμε λοιπόν

$$\begin{pmatrix} 0 \\ x \end{pmatrix} \cdot \begin{pmatrix} y \\ -\lambda y \end{pmatrix} = -\lambda xy = 0, \quad \forall x \quad \forall y$$

αν και μόνον αν $\lambda = 0$. Συνεπώς ο μηδενοχώρος $\text{null } P$ δεν είναι το ορθογώνιο συμπλήρωμα του $\text{range } P$ αν $\lambda \neq 0$.

Ο χαρακτηρισμός ορθογωνιότητας για μια προβολή P εκφράζει ακριβώς το γεγονός ότι ο χώρος στηλών του P , $\text{range } P$, και ο μηδενοχώρος του P , $\text{null } P = \text{range}(I - P)$, είναι ορθογώνια συμπληρώματα ο ένας του άλλου.

Θεώρημα 2

Μια προβολή P , ένα μήτρωο δηλαδή με την ιδιότητα $P^2 = P$, είναι ορθογώνια αν και μόνον αν $P^T = P$.

Προσοχή: Υπενθυμίζουμε ότι υπάρχει και προβολή που ΔΕΝ είναι ορθογώνια. Πρόκειται για την **πλάγια προβολή** (oblique projection). Το αντίστοιχο μητρώο πληροί τη σχέση $P^2 = P$ χωρίς να είναι συμμετρικό.

Απλό παράδειγμα I

Ο δ.χ. είναι το $\mathbb{R}^3$ και μας ενδιαφέρουν οι υπόχωροι $\mathcal{S} := \text{span}\{\mathbf{e}_3\}$ και $\mathcal{T} := \text{span}\{\mathbf{e}_1, \mathbf{e}_2\}$. Τότε αν

$$\mathbf{x} = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}$$

$$\Rightarrow \text{προβολή επί του } \mathcal{S} \text{ είναι } \mathbf{x}_{\mathcal{S}} = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}$$

$$\Rightarrow \text{προβολή επί του } \mathcal{T} \text{ είναι } \mathbf{x}_{\mathcal{T}} = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$$

Απλό παράδειγμα II

Ο δ.χ. είναι το $\mathbb{R}^3$ και μας ενδιαφέρουν οι υπόχωροι $\mathcal{S} := \text{span}\{\mathbf{e}_3\}$ και $\mathcal{T} := \text{span}\{\mathbf{e}_1, \mathbf{e}_2\}$. Τότε αν

$$\mathbf{x} = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}$$

$$\Rightarrow \text{προβολή επί του } \mathcal{S} \text{ είναι } \mathbf{x}_{\mathcal{S}} = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}$$

$$\Rightarrow \text{προβολή επί του } \mathcal{T} \text{ είναι } \mathbf{x}_{\mathcal{T}} = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}$$

Απλό παράδειγμα III

Ο δ.χ. είναι το $\mathbb{R}^3$ και μας ενδιαφέρουν οι υπόχωροι $\mathcal{S} := \text{span}\{\mathbf{e}_3\}$ και $\mathcal{T} := \text{span}\{\mathbf{e}_1, \mathbf{e}_2\}$. Τότε αν $\mathbf{x} = (1, 2, -1)^\top$

$$\Rightarrow \mathbf{x}_{\mathcal{S}} = \mathbf{P}_{\mathcal{S}}\mathbf{x}, \text{ όπου } \mathbf{P}_{\mathcal{S}} := \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \mathbf{e}_3(\mathbf{e}_3^\top \mathbf{e}_3)^{-1}\mathbf{e}_3^\top$$

$$\begin{aligned} \Rightarrow \mathbf{x}_{\mathcal{T}} = \mathbf{P}_{\mathcal{T}}\mathbf{x}, \text{ όπου } \mathbf{P}_{\mathcal{T}} &:= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \\ &= (\mathbf{e}_1 \quad \mathbf{e}_2) \left(\begin{pmatrix} \mathbf{e}_1^\top \\ \mathbf{e}_2^\top \end{pmatrix} (\mathbf{e}_1 \quad \mathbf{e}_2) \right)^{-1} \begin{pmatrix} \mathbf{e}_1^\top \\ \mathbf{e}_2^\top \end{pmatrix} \end{aligned}$$

Παράδειγμα 2

Έστω ότι ενδιαφέρει ο υπόχωρος που παράγεται από το διάνυσμα $u = [1, 2, 3]^T$. Τότε

$$P = \frac{1}{14} \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{pmatrix}$$

Για τυχαίο διάνυσμα x έχουμε

$$\begin{aligned} Px &= \frac{1}{14} \begin{pmatrix} \xi_1 + 2\xi_2 + 3\xi_3 \\ 2\xi_1 + 4\xi_2 + 6\xi_3 \\ 3\xi_1 + 6\xi_2 + 9\xi_3 \end{pmatrix} \\ &= \frac{\xi_1 + 2\xi_2 + 3\xi_3}{14} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}. \end{aligned}$$

Παράδειγμα 3

Έστω $\text{span}\{\mathbf{e}_1, \mathbf{e}_3\}$ στον $\mathbb{R}^4$ τότε $V = [\mathbf{e}_1, \mathbf{e}_3]$ και

$$\begin{aligned}P_V &= (\mathbf{e}_1 \quad \mathbf{e}_3) \left(\begin{pmatrix} \mathbf{e}_1^\top \\ \mathbf{e}_3^\top \end{pmatrix} (\mathbf{e}_1 \quad \mathbf{e}_3) \right)^{-1} \begin{pmatrix} \mathbf{e}_1^\top \\ \mathbf{e}_3^\top \end{pmatrix} \\&= (\mathbf{e}_1 \quad \mathbf{e}_3) \begin{pmatrix} \mathbf{e}_1^\top \mathbf{e}_1 & \mathbf{e}_1^\top \mathbf{e}_3 \\ \mathbf{e}_3^\top \mathbf{e}_1 & \mathbf{e}_3^\top \mathbf{e}_3 \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{e}_1^\top \\ \mathbf{e}_3^\top \end{pmatrix} \\&= (\mathbf{e}_1 \quad \mathbf{e}_3) \begin{pmatrix} \mathbf{e}_1^\top \\ \mathbf{e}_3^\top \end{pmatrix} \text{ προσέξτε ότι ο μεσαίος όρος παραπάνω είναι } I_2 \\&= \mathbf{e}_1 \mathbf{e}_1^\top + \mathbf{e}_3 \mathbf{e}_3^\top = \begin{pmatrix} \color{red}{1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \color{red}{1} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}\end{aligned}$$

Παρατηρήσεις: Προφανώς $P_V \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \end{pmatrix} = \begin{pmatrix} \xi_1 \\ 0 \\ \xi_3 \\ 0 \end{pmatrix}$ που θα μπορούσαμε να το ((μαντέψουμε)) από την αρχή. Δείτε όμως το επόμενο παράδειγμα που αυτό δεν είναι εύκολο.

Παράδειγμα: Κατασκευή μητρώου προβολής

Παράδειγμα 4

Έστω $\mathcal{V} = \text{span}\{\mathbf{a}, \mathbf{b}\}$ στον $\mathbb{R}^4$ όπου $\mathbf{a} = [1, 1, 1, 1]^T$, $\mathbf{b} = [0, 1, -1, 3]^T$ οπότε $V = [\mathbf{a}, \mathbf{b}]$. και

$$\begin{aligned} P_{\mathcal{V}} &= (\mathbf{a} \ \mathbf{b}) \left(\begin{pmatrix} \mathbf{a}^T \\ \mathbf{b}^T \end{pmatrix} (\mathbf{a} \ \mathbf{b}) \right)^{-1} \begin{pmatrix} \mathbf{a}^T \\ \mathbf{b}^T \end{pmatrix} = (\mathbf{a} \ \mathbf{b}) \begin{pmatrix} \mathbf{a}^T \mathbf{a} & \mathbf{a}^T \mathbf{b} \\ \mathbf{b}^T \mathbf{a} & \mathbf{b}^T \mathbf{b} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{a}^T \\ \mathbf{b}^T \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & -1 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 4 & 3 \\ 3 & 11 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & -1 & 3 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & -1 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} \frac{11}{35} & -\frac{3}{35} \\ -\frac{3}{35} & \frac{4}{35} \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & -1 & 3 \end{pmatrix} \\ &= \begin{pmatrix} \frac{11}{35} & \frac{8}{35} & \frac{14}{35} & \frac{2}{35} \\ \frac{35}{35} & \frac{35}{35} & \frac{35}{35} & \frac{35}{35} \\ \frac{35}{35} & \frac{35}{35} & \frac{35}{35} & \frac{35}{35} \\ \frac{2}{35} & \frac{11}{35} & -\frac{7}{35} & \frac{29}{35} \end{pmatrix} \end{aligned}$$

Παρατηρήσεις: $P_{\mathcal{V}}^2 = P_{\mathcal{V}} = P_{\mathcal{V}}^T$, $P_{\mathcal{V}}\mathbf{a} = \mathbf{a}$, $P_{\mathcal{V}}\mathbf{b} = \mathbf{b}$, $P_{\mathcal{V}}(\gamma\mathbf{a} + \delta\mathbf{b}) = \gamma\mathbf{a} + \delta\mathbf{b}$
(αναμενόμενο: $P_{\mathcal{V}}\mathbf{v} = \mathbf{v}$ για κάθε $\mathbf{v} \in \mathcal{V}$.)

Παράδειγμα 5

(συνέχεια του προηγούμενου παραδείγματος) Αν μας έδιναν το παραπάνω P_V και ένα τυχαίο διάνυσμα $x = [\xi_1, \xi_2, \xi_3, \xi_4]^T$ τότε

$$P_V x = \frac{1}{35} \begin{pmatrix} 11\xi_1 + 8\xi_2 + 14\xi_3 + 2\xi_4 \\ 8\xi_1 + 9\xi_2 + 7\xi_3 + 11\xi_4 \\ 14\xi_1 + 7\xi_2 + 21\xi_3 - 7\xi_4 \\ 2\xi_1 + 11\xi_2 - 7\xi_3 + 29\xi_4 \end{pmatrix}$$

Για να επαληθεύσουμε ότι για κάθε $x \in \mathbb{R}^4$, $P_V x \in V$, αρκεί να δείξουμε ότι υπάρχει πάντα $z \in \mathbb{R}^2$ που ικανοποιεί το σύστημα $Vz = P_V x$. Εξετάζουμε το επαυξημένο σύστημα

$$\left(\begin{array}{cc|c} 1 & 0 & 11\xi_1 + 8\xi_2 + 14\xi_3 + 2\xi_4 \\ 1 & 1 & 8\xi_1 + 9\xi_2 + 7\xi_3 + 11\xi_4 \\ 1 & -1 & 14\xi_1 + 7\xi_2 + 21\xi_3 - 7\xi_4 \\ 1 & 3 & 2\xi_1 + 11\xi_2 - 7\xi_3 + 29\xi_4 \end{array} \right)$$

και το φέρνουμε σε ΑΓΚΜ,

$$\left(\begin{array}{cc|c} 1 & 0 & 11\xi_1 + 8\xi_2 + 14\xi_3 + 2\xi_4 \\ 0 & 1 & -3\xi_1 + \xi_2 - 7\xi_3 + 9\xi_4 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{array} \right)$$

Προσέξτε ότι $n = \text{rank}(A) = \text{rank}([A, b])$ επομένως υπάρχει μοναδική λύση.

Σχετικά με τη (μη) αντιστρεψιμότητα του μητρώου προβολής

Παρατήρηση

Στη γενική περίπτωση για **γνήσιο υπόχωρο** το (τετραγωνικό) μητρώο προβολής P δεν είναι αντιστρέψιμο.

Σκεπτικό: Έστω ότι ο αναφερόμαστε στον $\mathbb{R}^m$. Αν υπήρχε αντίστροφο P^{-1} , για κάθε διάνυσμα του υπόχωρου $v \in \mathcal{V}$ θα επέστρεφε ένα μοναδικό διάνυσμα του χώρου $P^{-1}v \in \mathbb{R}^m$. Επειδή ένας γνήσιος υπόχωρος είναι ((μικρότερος)), θα υπάρχουν περισσότερα από ένα διανύσματα του χώρου των οποίων η προβολή στον υπόχωρο θα είναι το ίδιο διάνυσμα, επομένως ο αντίστροφος δεν μπορεί να οριστεί (θα προβληματίζεται ποιο διάνυσμα του χώρου να επιλέξει). Π.χ. είδαμε ότι αν

$$\mathcal{V} = \text{span}\{e_1, e_3\} \text{ στον } \mathbb{R}^4 \text{ τότε } P_{\mathcal{V}} \begin{pmatrix} 1 \\ 5 \\ 2 \\ -1 \end{pmatrix} = P_{\mathcal{V}} \begin{pmatrix} 1 \\ 1000 \\ 2 \\ -100 \end{pmatrix}.$$

Μία απόδειξη Για να υπάρχει αντίστροφο, θα πρέπει το μητρώο P να είναι αντιστρέψιμο. Όμως η βάση $V \in \mathbb{R}^{m \times k}$ περιέχει λιγότερα διανύσματα από τη διάσταση όλου του χώρου, δηλ. $k < m$, επομένως

$$\text{rank}(P) \leq \min\{\text{rank}(V), \text{rank}((V^T V)^{-1}), \text{rank}(V^T)\}.$$

Η τάξη των μητρώων στα δεξιά είναι k ή μικρότερη (αποδεικνύεται ότι είναι ακριβώς k) άρα το μητρώο δεν είναι πλήρους τάξης και δεν είναι αντιστρέψιμο.

Γιατί χρειάζεται γνήσιος υπόχωρος; επειδή το (μοναδικό) μητρώο ορθογώνιας προβολής στον $\mathbb{R}^n$ είναι το ταυτοτικό I , το οποίο είναι αντιστρέψιμο.

Παρατήρηση

Το (τετραγωνικό) μητρώο προβολής P **είναι μοναδικό** (δεν εξαρτάται από τη βάση).

Γιατί αν οι στήλες των V και U είναι δύο διαφορετικές βάσεις για τον υπόχωρο προβολής, τότε οι στήλες του ενός μπορούν να γραφτούν ως γραμμικοί συνδυασμοί των στηλών του άλλου (ως μέλη του ίδου υπόχωρου). Επομένως υπάρχει αντιστρέψιμο τέτοιο ώστε

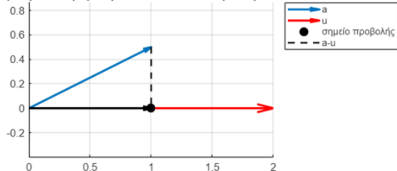
$$V = UM, \quad V, U \in \mathbb{R}^{m \times n}, M \in \mathbb{R}^{n \times n}.$$

Άρα αν γράψουμε

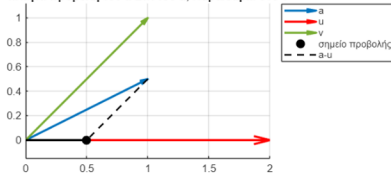
$$\begin{aligned} P &= V(V^T V)^{-1} V^T \\ &= (UM)((UM)^T UM)^{-1} (UM)^T \\ &= UM(M^T U^T UM)^{-1} M^T U \\ &= U(M^{-T} M^T U^T U M M^{-1})^{-1} U \\ &= U(U^T U)^{-1} U^T \end{aligned}$$

Παραδείγματα

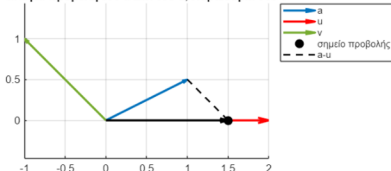
ορθογώνια προβολή του a Επί του u , Παράλληλα στο v



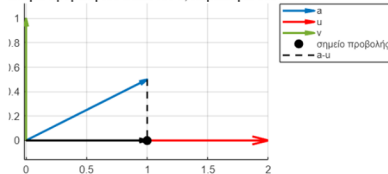
πλάγια προβολή του a Επί του u , Παράλληλα στο v



πλάγια προβολή του a Επί του u , Παράλληλα στο v



πλάγια προβολή του a Επί του u , Παράλληλα στο v



- Από το σύνδεσμο εδώ μπορείτε να κατεβάσετε πρόγραμμα MATLAB που υλοποιεί πλάγιες και ορθογώνιες προβολές στον $\mathbb{R}^2$. Οι παραπάνω οπτικοποιήσεις προήλθαν από το εν λόγω πρόγραμμα.

2. Βέλτιστη προσέγγιση

Ορισμός 1

Έστω X ένας διανυσματικός χώρος με εσωτερικό γινόμενο $\langle \cdot, \cdot \rangle$. Εάν W είναι ένας πεπερασμένης διάστασης υποχώρος του X και $\{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n\}$ είναι μια ορθοκανονική βάση του W και $\mathbf{u}$ είναι ένα διάνυσμα του X , ορίζουμε την **ορθογώνια προβολή του $\mathbf{u}$ επί του W** το διάνυσμα

$$\text{proj}_W \mathbf{u} := \langle \mathbf{u}, \mathbf{w}_1 \rangle \mathbf{w}_1 + \langle \mathbf{u}, \mathbf{w}_2 \rangle \mathbf{w}_2 + \dots + \langle \mathbf{u}, \mathbf{w}_n \rangle \mathbf{w}_n. \quad (4)$$

Το διάνυσμα

$$\mathbf{u} - \text{proj}_W \mathbf{u} = \mathbf{u} - \langle \mathbf{u}, \mathbf{w}_1 \rangle \mathbf{w}_1 - \langle \mathbf{u}, \mathbf{w}_2 \rangle \mathbf{w}_2 - \dots - \langle \mathbf{u}, \mathbf{w}_n \rangle \mathbf{w}_n \quad (5)$$

θα το λέμε **ορθογώνια στο W συνιστώσα του $\mathbf{u}$** .

Ο ορισμός του $\mathbf{u} - \text{proj}_W \mathbf{u}$ ως ορθογώνιας στο W συνιστώσας του $\mathbf{u}$ δικαιολογείται από το γεγονός ότι γράφοντας

$$\mathbf{u} = \text{proj}_W \mathbf{u} + \mathbf{u} - \text{proj}_W \mathbf{u}$$

ισχύει

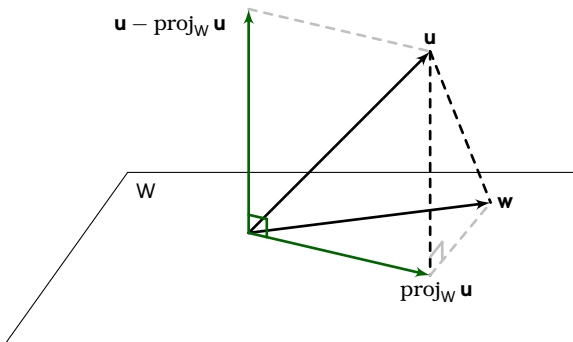
$$\langle \text{proj}_W \mathbf{u}, \mathbf{u} - \text{proj}_W \mathbf{u} \rangle = 0 \Rightarrow \mathbf{u} - \text{proj}_W \mathbf{u} \in W^\perp.$$

Θεώρημα 3 (Βέλτιστη προσέγγιση)

Εάν W είναι ένας πεπερασμένης διάστασης υποχώρος ενός διανυσματικού χώρου X με εσωτερικό γινόμενο και επαγόμενη νόρμα $\|\cdot\|$, και εάν $\mathbf{u} \in X$, τότε για κάθε $\mathbf{w} \in W$ με $\mathbf{w} \neq \text{proj}_W \mathbf{u}$ ισχύει

$$\|\mathbf{u} - \text{proj}_W \mathbf{u}\| < \|\mathbf{u} - \mathbf{w}\|$$

δηλαδή το διάνυσμα $\text{proj}_W \mathbf{u}$ είναι η βέλτιστη προσέγγιση του $\mathbf{u}$ από το W .



Απόδειξη.

Γράφοντας

$$\mathbf{u} - \mathbf{w} = (\text{proj}_W \mathbf{u} - \mathbf{w}) + (\mathbf{u} - \text{proj}_W \mathbf{u})$$

επειδή $\text{proj}_W \mathbf{u} - \mathbf{w}$ είναι, ως διαφορά διανυσμάτων του W , διάνυσμα του W και είναι ορθογώνιο στο $\mathbf{u} - \text{proj}_W \mathbf{u} \in W^\perp$, κατά συνέπεια από το Πυθαγόρειο Θεώρημα για νόρμες έχουμε ότι

$$\|\mathbf{u} - \mathbf{w}\|^2 = \|\text{proj}_W \mathbf{u} - \mathbf{w}\|^2 + \|\mathbf{u} - \text{proj}_W \mathbf{u}\|^2.$$

Αφού $\mathbf{w} \neq \text{proj}_W \mathbf{u}$, θα είναι $\|\text{proj}_W \mathbf{u} - \mathbf{w}\| > 0$, επομένως από την τελευταία ισότητα έπεται ότι

$$\|\mathbf{u} - \mathbf{w}\|^2 > \|\mathbf{u} - \text{proj}_W \mathbf{u}\|^2,$$

απ' όπου έπεται το ζητούμενο. □

Παρατήρηση 2

Το Θεώρημα της βέλτιστης προσέγγισης μπορεί να διατυπωθεί και ως ένα πρόβλημα βελτιστοποίησης, συγκεκριμένα αν X είναι ένας διανυσματικός χώρος με εσωτερικό γινόμενο και $\{\mathbf{w}_1, \dots, \mathbf{w}_n\}$ είναι ένα ορθοκανονικό σύνολο διανυσμάτων του X , τότε για κάθε διάνυσμα $\mathbf{u}$ του X ισχύει

$$\min_{\lambda_1, \dots, \lambda_n} \|\mathbf{u} - \lambda_1 \mathbf{w}_1 - \dots - \lambda_n \mathbf{w}_n\| = \|\mathbf{u} - \langle \mathbf{u}, \mathbf{w}_1 \rangle \mathbf{w}_1 - \dots - \langle \mathbf{u}, \mathbf{w}_n \rangle \mathbf{w}_n\|.$$

Πράγματι, υπολογίζοντας

$$\begin{aligned} \|\mathbf{u} - \lambda_1 \mathbf{w}_1 - \dots - \lambda_n \mathbf{w}_n\|^2 &= \langle \mathbf{u} - \sum_{i=1}^n \lambda_i \mathbf{w}_i, \mathbf{u} - \sum_{i=1}^n \lambda_i \mathbf{w}_i \rangle \\ &= \langle \mathbf{u}, \mathbf{u} \rangle - 2 \sum_{i=1}^n \lambda_i \langle \mathbf{u}, \mathbf{w}_i \rangle + \sum_{i,j=1}^n \lambda_i \lambda_j \langle \mathbf{w}_i, \mathbf{w}_j \rangle \\ &= \|\mathbf{u}\|^2 - 2 \sum_{i=1}^n \lambda_i \langle \mathbf{u}, \mathbf{w}_i \rangle + \sum_{i=1}^n \lambda_i^2 \end{aligned}$$

Παρατήρηση 2 (συνέχεια)

οπότε

$$\begin{aligned}\|\mathbf{u} - \lambda_1 \mathbf{w}_1 - \cdots - \lambda_n \mathbf{w}_n\|^2 &= \|\mathbf{u}\|^2 + \sum_{i=1}^n (\lambda_i^2 - 2\lambda_i \langle \mathbf{u}, \mathbf{w}_i \rangle + \langle \mathbf{u}, \mathbf{w}_i \rangle^2 - \langle \mathbf{u}, \mathbf{w}_i \rangle^2) \\ &= \|\mathbf{u}\|^2 - \sum_{i=1}^n \langle \mathbf{u}, \mathbf{w}_i \rangle^2 + \sum_{i=1}^n (\lambda_i - \langle \mathbf{u}, \mathbf{w}_i \rangle)^2\end{aligned}$$

από την τελευταία ισότητα βλέπουμε ότι

$$\min_{\lambda_1, \dots, \lambda_n} \|\mathbf{u} - \lambda_1 \mathbf{w}_1 - \cdots - \lambda_n \mathbf{w}_n\|^2 = \|\mathbf{u}\|^2 - \sum_{i=1}^n \langle \mathbf{u}, \mathbf{w}_i \rangle^2$$

και αυτό συμβαίνει όταν

$$\lambda_i = \langle \mathbf{u}, \mathbf{w}_i \rangle \quad \text{για κάθε } i = 1, 2, \dots, n.$$

Συνεπώς

$$\|\mathbf{u} - \langle \mathbf{u}, \mathbf{w}_1 \rangle \mathbf{w}_1 - \cdots - \langle \mathbf{u}, \mathbf{w}_n \rangle \mathbf{w}_n\| \leq \|\mathbf{u} - \lambda_1 \mathbf{w}_1 - \cdots - \lambda_n \mathbf{w}_n\|.$$

3. Προσεγγίσεις ελαχίστων τετραγώνων

Για ένα σύστημα $A\mathbf{x} = \mathbf{b}$ ένα διάνυσμα $\mathbf{x}$ θα είναι προσεγγιστική λύση αν η ποσότητα $\|A\mathbf{x} - \mathbf{b}\|$ είναι «μικρή». Σημειώνουμε ότι

$$\|A\mathbf{x} - \mathbf{b}\| = 0 \Leftrightarrow A\mathbf{x} = \mathbf{b}$$

κατά συνέπεια αν το σύστημα δεν έχει λύση, θέλουμε να γνωρίζουμε εκείνο το διάνυσμα $\mathbf{x}$ για το οποίο η ποσότητα $\|A\mathbf{x} - \mathbf{b}\|$ είναι η ελάχιστη δυνατή. Αν

$$A\mathbf{x} - \mathbf{b} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

έχουμε ότι

$$\|A\mathbf{x} - \mathbf{b}\|^2 = \epsilon_1^2 + \epsilon_2^2 + \cdots + \epsilon_n^2,$$

κατά συνέπεια το διάνυσμα $\mathbf{x}$ που ελαχιστοποιεί την $\|A\mathbf{x} - \mathbf{b}\|$ ελαχιστοποιεί το άθροισμα $\epsilon_1^2 + \epsilon_2^2 + \cdots + \epsilon_n^2$. Ένα τέτοιο διάνυσμα θα το λέμε **λύση των ελαχίστων τετραγώνων** και την ποσότητα $\|A\mathbf{x} - \mathbf{b}\|$ θα την λέμε **σφάλμα των ελαχίστων τετραγώνων**.

Πρόβλημα:

Να βρεθεί η λύση ελαχίστων τετραγώνων για το σύστημα $A\mathbf{x} = \mathbf{b}$.

Αν $W = \text{range } A$ είναι ο χώρος στηλών του μήτρου A , τότε

$$\mathbf{b} = \text{proj}_W \mathbf{b} + (\mathbf{b} - \text{proj}_W \mathbf{b}), \quad \& \quad \text{proj}_W \mathbf{b} \perp \mathbf{b} - \text{proj}_W \mathbf{b},$$

έτσι

$$\mathbf{b} - A\mathbf{x} = \text{proj}_W \mathbf{b} - A\mathbf{x} + \mathbf{b} - \text{proj}_W \mathbf{b}, \quad \& \quad \text{proj}_W \mathbf{b} - A\mathbf{x} \perp \mathbf{b} - \text{proj}_W \mathbf{b}, \quad (6)$$

αφού $A\mathbf{x} \in \text{range } A = W$, επομένως

$$\|\mathbf{b} - A\mathbf{x}\|^2 = \|\text{proj}_W \mathbf{b} - A\mathbf{x}\|^2 + \|\mathbf{b} - \text{proj}_W \mathbf{b}\|^2.$$

Κατά συνέπεια το σφάλμα ελαχιστοποιείται όταν $A\mathbf{x} = \text{proj}_W \mathbf{b}$, ισοδύναμα η λύση των ελαχίστων τετραγώνων, η οποία υπάρχει πάντα (γιατί;), είναι η $\hat{\mathbf{x}}$ για την οποία

$$A\hat{\mathbf{x}} = \text{proj}_W \mathbf{b}. \quad (7)$$

Έτσι από την (6) παίρνουμε

$$\mathbf{b} - A\hat{\mathbf{x}} = \mathbf{b} - \text{proj}_W \mathbf{b}, \quad \text{και} \quad \mathbf{b} - \text{proj}_W \mathbf{b} \in W^\perp,$$

κατά συνέπεια, αφού $\mathbf{b} - \text{proj}_W \mathbf{b} \in \text{null } A^T$ (θεμελιώδες Θεώρημα)

$$A^T(\mathbf{b} - A\hat{\mathbf{x}}) = A^T(\mathbf{b} - \text{proj}_W \mathbf{b}) = \mathbf{0}.$$

επομένως

$$A^T A \hat{\mathbf{x}} = A^T \mathbf{b}. \quad (8)$$

Σημειώνουμε ότι για την εύρεση του $\hat{\mathbf{x}}$ δεν χρειάζεται να βρούμε την προβολή του $\mathbf{b}$ επί του W και στη συνέχεια να λύσουμε το σύστημα $A\mathbf{x} = \text{proj}_W \mathbf{b}$. Η λύση των ελαχίστων τετραγώνων προκύπτει από την (8). Την εξίσωση στην (8) την λέμε **κανονική εξίσωση**.

Παρατήρηση 3

Αν το μήτρωο $A^T A$ είναι αντιστρέψιμο, από την κανονική εξίσωση παίρνουμε

$$\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b},$$

απ' όπου πολλαπλασιάζοντας από αριστερά με A προκύπτει μέσω της (7) ότι

$$\text{proj}_W \mathbf{b} = A \hat{\mathbf{x}} = A(A^T A)^{-1} A^T \mathbf{b}.$$

Η τελευταία σχέση μας λέει ότι για το μήτρωο A το $A(A^T A)^{-1} A^T$ προβάλλει επί του χώρου στηλών του A .

Εφαρμογή

Με τη μέθοδο των ελαχίστων τετραγώνων να βρεθεί η ευθεία που προσεγγίζει καλύτερα τα σημεία $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ του επιπέδου.

Έστω $y = \alpha + \beta x$ η ζητούμενη ευθεία. Αν τα δοσμένα σημεία βρίσκονταν επάνω στην ευθεία θα είχαμε

$$\left. \begin{array}{l} y_1 = \alpha + \beta x_1 \\ y_2 = \alpha + \beta x_2 \\ \vdots \\ y_n = \alpha + \beta x_n \end{array} \right\} \Leftrightarrow \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \Leftrightarrow \mathbf{Ax} = \mathbf{b}.$$

Για τη λύση των ελαχίστων τετραγώνων υπολογίζουμε

$$\mathbf{A}^T \mathbf{A} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} = \begin{pmatrix} n & \sum x_j \\ \sum x_j & \sum x_j^2 \end{pmatrix}.$$

Εφαρμογή – συνέχεια

και

$$A^T \mathbf{b} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \sum y_j \\ \sum x_j y_j \end{pmatrix}.$$

Η κανονική εξίσωση (8) γίνεται

$$\begin{pmatrix} n & \sum x_j \\ \sum x_j & \sum x_j^2 \end{pmatrix} \begin{pmatrix} \alpha' \\ \beta' \end{pmatrix} = \begin{pmatrix} \sum y_j \\ \sum x_j y_j \end{pmatrix}. \quad (9)$$

Οι στήλες του A είναι γραμμικά ανεξάρτητες, εκτός αν $x_1 = x_2 = \dots = x_n$ (γιατί;) κατά συνέπεια, αν συμβολίσουμε με $\mathbf{a}_1$ και $\mathbf{a}_2$ τις στήλες, από την ανισότητα Cauchy–Bunyakovsky–Schwarz έχουμε

$$\langle \mathbf{a}_1, \mathbf{a}_2 \rangle^2 < \langle \mathbf{a}_1, \mathbf{a}_1 \rangle \langle \mathbf{a}_2, \mathbf{a}_2 \rangle \Leftrightarrow \left(\sum_{j=1}^n x_j \right)^2 < \left(\sum_{j=1}^n 1^2 \right) \left(\sum_{j=1}^n x_j^2 \right) = n \sum_{j=1}^n x_j^2.$$

Εφαρμογή – συνέχεια

επομένως η ορίζουσα του μήτρωου των συντελεστών στη κανονική εξίσωση είναι διάφορη του μηδενός (Διαφορετικά, η ορίζουσα του μήτρωου των συντελεστών, με $\mathbf{x} = (x_1 \ x_2 \ \cdots \ x_n)^T$, είναι

$$n\|\mathbf{x}\|^2 - \left(\sum_{j=1}^n x_j\right)^2 = \|\mathbf{x}\|^2 \left(n - \left(\sum_{j=1}^n \frac{x_j}{\|\mathbf{x}\|}\right)^2\right) > 0$$

αφού $x_j \leq |x_j| = (x_j^2)^{1/2} < (x_1^2 + x_2^2 + \cdots + x_n^2)^{1/2} = \|\mathbf{x}\|$, για κάθε j , όταν τουλάχιστον δύο από τα x_j είναι διάφορα του μηδενός), συνεπώς

$$\begin{pmatrix} \alpha' \\ \beta' \end{pmatrix} = \frac{1}{n \sum x_j^2 - (\sum x_j)^2} \begin{pmatrix} \sum x_j^2 & -\sum x_j \\ -\sum x_j & n \end{pmatrix} \begin{pmatrix} \sum y_j \\ \sum x_j y_j \end{pmatrix}.$$

Έτσι τελικά οι συντελεστές στην ευθεία $y = \alpha' + \beta'x$ των ελαχίστων τετραγώνων είναι

$$\alpha' = \frac{\sum x_j^2 \sum y_j - \sum x_j y_j \sum x_j}{n \sum x_j^2 - (\sum x_j)^2}, \quad \beta' = \frac{-\sum x_j \sum y_j + n \sum x_j y_j}{n \sum x_j^2 - (\sum x_j)^2}. \quad (10)$$

Γενικότερη περίπτωση

- Έχουμε ζεύγη τιμών από μετρήσεις (π.χ. χρόνους εκτέλεσης στον υπολογιστή μας ενός προγράμματος για την επίλυση προβλήματος για

$$(t_i, \psi_i), \quad i = 1, \dots, m \text{ όπου το } m \text{ μπορεί να είναι αρκετά μεγάλο.}$$

- Θέλουμε να κατασκευάσουμε μία “απλή” συνάρτηση $f(t)$ που προσαρμόζεται στις μετρήσεις ώστε να μοντελοποιεί το φαινόμενο όσο γίνεται καλύτερα (π.χ. την αγνωστη συνάρτηση χρονοεπίδοσης του αλγορίθμου στην πλατφόρμα)

$$\mathbf{y} \approx \mathbf{f}(\mathbf{t});$$

- Υπάρχει το εξής θεωρητικό αποτέλεσμα (π.χ. μάθημα Αριθμητικής Ανάλυσης 2ου έτους): Αν τα t_i είναι διακριτά, υπάρχει μοναδικό **πολυώνυμο** $p(t)$ βαθμού το πολύ $m - 1$ για το οποίο ισχύει ότι $p(t_i) = y_i$ για $i = 1, \dots, m$. Οι συντελεστές του πολυωνύμου προκύπτουν επιλύοντας ένα $m \times m$ γραμμικό σύστημα.
- Συχνότερα όμως, προτιμάμε να χρησιμοποιήσουμε πολυώνυμο μικρού βαθμού, π.χ. $n = 1, n = 2$ οπότε προκύπτει σύστημα m εξισώσεων με $n + 1$ αγνώστους (τους συντελεστές της δυναμομορφής του, π.χ. τα α_0, α_1 αν επιλέξουμε $p(t) = \alpha_0 + \alpha_1 t$).
- Ακόμα γενικότερα, το μοντέλο που αναζητούμε μπορεί να έχει τη μορφή

$$f(t) = \xi_1 \phi_1(t) + \xi_2 \phi_2(t) + \dots + \xi_n \phi_n(t)$$

όπου οι συναρτήσεις $\phi_i(t)$ επιλέγονται κατάλληλα (ό,τι κρίνεται καλύτερο για τα δεδομένα) π.χ. $1, t, t^2, \sin(t), \exp(t), \dots$

- Ορίζουμε τη μήτρα σχεδιασμού (design matrix):

$$\mathbf{A} = \begin{bmatrix} \phi_1(\mathbf{t}_1) & \phi_2(\mathbf{t}_1) & \cdots & \phi_n(\mathbf{t}_1) \\ \phi_1(\mathbf{t}_2) & \phi_2(\mathbf{t}_2) & \cdots & \phi_n(\mathbf{t}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(\mathbf{t}_m) & \phi_2(\mathbf{t}_m) & \cdots & \phi_n(\mathbf{t}_m) \end{bmatrix}.$$

- Διάνυσμα παραμέτρων:

$$\mathbf{x} = \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_n \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} \psi_1 \\ \vdots \\ \psi_m \end{bmatrix}.$$

- Μοντέλο σε μορφή μήτρας:

$$\mathbf{Ax} \approx \mathbf{y}.$$

Πρόβλημα Ελαχίστων Τετραγώνων

- Θέλουμε το $A\mathbf{x}$ να είναι όσο πιο κοντά γίνεται στο $\mathbf{y}$:

$$\min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{y}\|^2.$$

- Η λύση προκύπτει από τις **κανονικές εξισώσεις**:

$$A^T A \mathbf{x} = A^T \mathbf{y}.$$

- Αν η μήτρα $A^T A$ είναι αντιστρέψιμη:

$$\mathbf{x}_{\min} = (A^T A)^{-1} A^T \mathbf{y}.$$

- Αυτό ισχύει για **οποιοδήποτε** μοντέλο γραμμικό ως προς τις παραμέτρους (όχι μόνο ευθεία).

Μοντέλο Πολυωνύμου 2ου Βαθμού

- Θέλουμε να ταιριάξουμε καμπύλη:

$$y \approx \xi_2 t^2 + \xi_1 t + \xi_0.$$

- Εδώ:

$$\phi_1(t) = t^2, \quad \phi_2(t) = t, \quad \phi_3(t) = 1.$$

- Το μητρώο A γίνεται:

$$A = \begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ \vdots & \vdots & \vdots \\ 1 & t_m & t_m^2 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} \xi_0 \\ \xi_1 \\ \xi_2 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} \psi_1 \\ \vdots \\ \psi_m \end{bmatrix}.$$

Λύση Ελαχίστων Τετραγώνων για Πολυώνυμο

- Διατυπώνουμε το πρόβλημα:

$$\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{y}\|^2.$$

- Κανονικές εξισώσεις:

$$\mathbf{A}^T \mathbf{Ax} = \mathbf{A}^T \mathbf{y}.$$

- Λύνοντας για $\mathbf{x}$ βρίσκουμε:

$$\hat{\xi}_0, \hat{\xi}_1, \hat{\xi}_2$$

που ορίζουν την καμπύλη

$$\hat{y}(t) = \hat{\xi}_2 t^2 + \hat{\xi}_1 t + \hat{\xi}_0.$$

Ελάχιστα Τετράγωνα σε 3Δ: Προσέγγιση με Επίπεδο

- Έχουμε μετρήσεις σε 3 διαστάσεις:

$$(x_i, y_i, z_i), \quad i = 1, \dots, m.$$

- Θέλουμε να βρούμε ένα επίπεδο:

$$z \approx ax + by + c$$

που να «περνάει όσο γίνεται πιο κοντά» από τα σημεία.

- Γράφουμε:

$$z_i \approx ax_i + by_i + c.$$

- Στόχος:

$$\min_{a,b,c} \sum_{i=1}^m (z_i - (ax_i + by_i + c))^2.$$

- Ορίζουμε τη μήτρα:

$$A = \begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ \vdots & \vdots & \vdots \\ x_m & y_m & 1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}, \quad \mathbf{z} = \begin{bmatrix} z_1 \\ \vdots \\ z_m \end{bmatrix}.$$

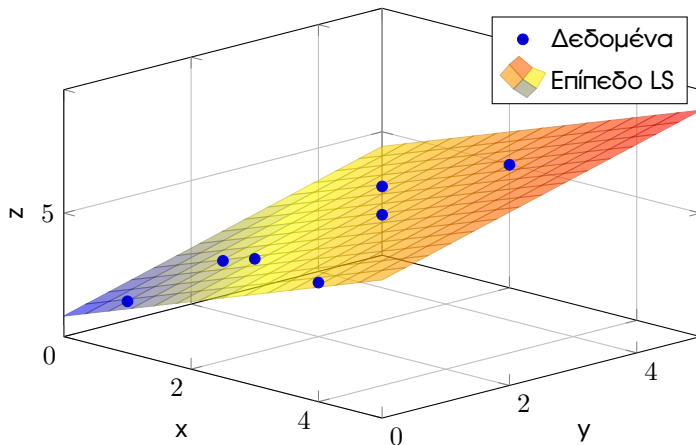
- Σύστημα:

$$A\mathbf{x} \approx \mathbf{z}.$$

- Επίλυση μέσω ελαχίστων τετραγώνων:

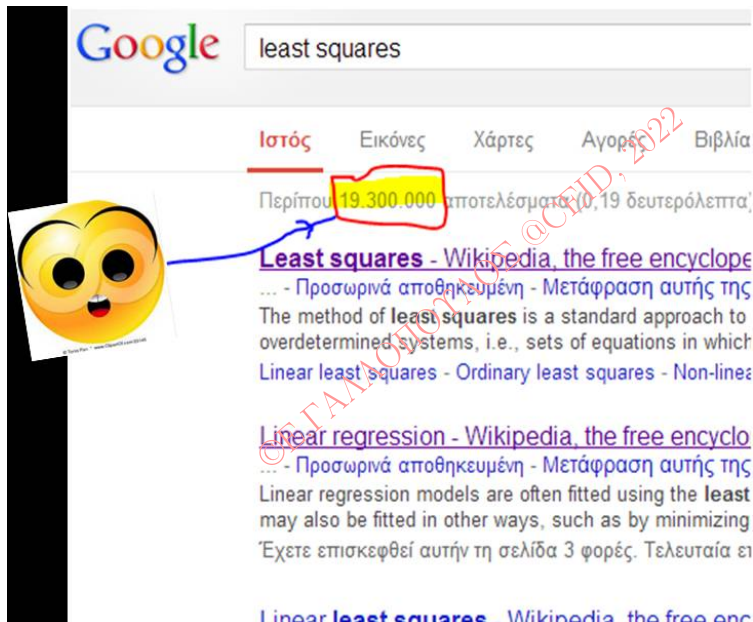
$$\min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{z}\|^2 \Rightarrow \mathbf{x}_{\min} = (A^T A)^{-1} A^T \mathbf{z}.$$

Fit Επιπέδου σε 3 Διαστάσεις (Οπτικοποίηση)



- Τα μπλε σημεία είναι τα δεδομένα (x_i, y_i, z_i) .
- Το ημιδιαφανές επίπεδο είναι η λύση ελαχίστων τετραγώνων.
- Η μέθοδος είναι απολύτως ανάλογη με το 2D fit, αλλά τώρα σε 3Δ.

Τι λέγεται για τα ελάχιστα τετράγωνα



A screenshot of a Google search for "least squares". The search bar shows "least squares". Below the search bar, there are tabs for "Ιστός", "Εικόνες", "Χάρτες", "Αγορές", and "Βιβλία". The "Ιστός" tab is selected. The search results show "Περίπου 19.300.000 αποτελέσματα (0,19 δευτερόλεπτα)". A yellow emoji with a wide-eyed, open-mouthed expression is pointing to the search results. A red box highlights the number "19.300.000". The search results include links to Wikipedia articles: "Least squares - Wikipedia, the free encyclopedia", "Linear regression - Wikipedia, the free encyclopedia", and "Linear least squares - Wikipedia, the free encyclopedia".

Google

least squares

Ιστός Εικόνες Χάρτες Αγορές Βιβλία

Περίπου 19.300.000 αποτελέσματα (0,19 δευτερόλεπτα)

[Least squares - Wikipedia, the free encyclopedia](#)
... - Προσωρινά αποθηκευμένη - Μετάφραση αυτής της
The method of **least squares** is a standard approach to
overdetermined systems, i.e., sets of equations in which
[Linear least squares](#) - [Ordinary least squares](#) - [Non-linear](#)

[Linear regression - Wikipedia, the free encyclopedia](#)
... - Προσωρινά αποθηκευμένη - Μετάφραση αυτής της
Linear regression models are often fitted using the **least**
may also be fitted in other ways, such as by minimizing
Έχετε επισκεφθεί αυτήν τη σελίδα 3 φορές. Τελευταία ει

[Linear least squares - Wikipedia, the free encyclopedia](#)

Τι λέγεται για τα ελάχιστα τετράγωνα



Μηχανική μάθηση και Μοντελοποίηση

Ορισμός (Wikipedia)

Μηχανική μάθηση είναι υποπεδίο της επιστήμης των υπολογιστών, που δίνει στους υπολογιστές την ικανότητα να μαθαίνουν, χωρίς να έχουν ρητά προγραμματιστεί.

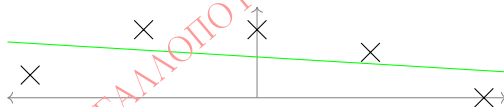
Η μηχανική μάθηση διερευνά τη μελέτη και την κατασκευή αλγορίθμων που μπορούν **να μαθαίνουν από τα δεδομένα** και **να κάνουν προβλέψεις σχετικά με αυτά**.

©Ε.ΓΑΛΛΟΠΟΥΛΟΣ @CEID, 2022

Προσαρμογή ελαχίστων τετραγώνων με ευθεία (least squares fit)

Πρόβλημα: Να υπολογίσουμε την ευθεία γραμμή $\gamma_1 + \gamma_2 \xi$ που προσαρμόζεται σε ένα σύνολο γνωστών τιμών $\{(\xi_j, \psi_j)\}_{j=1}^m$ με το ελάχιστο δυνατό σφάλμα.

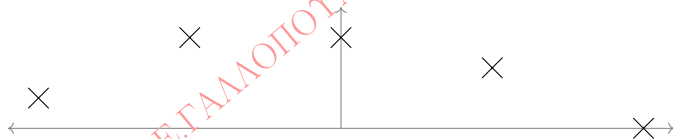
Ορολογία: Στη Στατιστική λέγεται γραμμική παλινδρόμηση (linear regression) Περί τίνος πρόκειται; Είναι η προσπάθεια να κατασκευάσουμε ένα απλό (γραμμικό) μοντέλο για τις παρατηρήσεις που αντιστοιχούν στα m ζεύγη των παραπάνω τιμών. Πώς μετράμε το σφάλμα? Το ορίζουμε ως το άθροισμα των τετραγώνων των αποκλίσεων των τιμών της (γραμμικής) συνάρτησης, $\{\gamma_1 + \gamma_2 \xi_j\}_{j=1}^m$ από τις τιμές των μετρήσεων ψ_j .



Με γραμμοαλγεβρική γραφή: $\gamma_1 + \gamma_2 \xi = \psi$ γράφεται και ως $(1 \quad \xi) \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} = \psi$

Παράδειγμα

Σημεία:	$\xi_1 = -1$	$\xi_2 = -1/2$	$\xi_3 = 0$	$\xi_4 = 1/2$	$\xi_5 = 1$
Τιμές:	$\psi_1 = 0.1$	$\psi_2 = 0.3$	$\psi_3 = 0.3$	$\psi_4 = 0.2$	$\psi_5 = 0.0$



$\phi_0(\xi) = 1, \phi_1(\xi) = \xi$ τότε $\phi(\xi) = \gamma_0\phi_0(\xi) + \gamma_1\phi_1(\xi)$ οπότε

$$A = \begin{pmatrix} 1.0 & -1.0 \\ 1.0 & -0.5 \\ 1.0 & 0.0 \\ 1.0 & 0.5 \\ 1.0 & 1.0 \end{pmatrix}$$

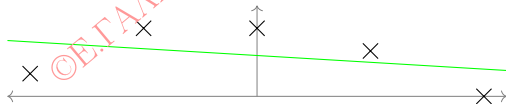
Αν κάνουμε τους υπολογισμούς

$$(A^T A)^{-1} A^T b = \begin{pmatrix} 0.18 \\ -0.06 \end{pmatrix}$$

$$\Rightarrow \boxed{\phi(\xi) = 0.18 - 0.06\xi}$$

$$A^T A = \begin{pmatrix} 5.0 & 0.0 \\ 0.0 & 2.5 \end{pmatrix}, A^T b = \begin{pmatrix} 0.9 \\ -0.15 \end{pmatrix}$$

σφάλμα $\|b - Ax_{LS}\|_2 = 0.2429$



Προσαρμογή με καμπύλη γραμμή 2ου βαθμού

$$\phi_0(x) = 1, \phi_1(\xi) = \xi, \phi_2(\xi) = \xi^2 \text{ και } \phi(\xi) = \gamma_0 \phi_0(\xi) + \gamma_1 \phi_1(\xi) + \gamma_2 \phi_2(\xi)$$

$$A = \begin{pmatrix} 1.0 & -1.0 & 1.0 \\ 1.0 & -0.5 & 0.25 \\ 1.0 & 0.0 & 0.0 \\ 1.0 & 0.5 & 0.25 \\ 1.0 & 1.0 & 1.0 \end{pmatrix}$$

Αν κάνουμε τους υπολογισμούς

$$x_{LS} = (A^T A)^{-1} A^T b = \begin{pmatrix} 0.3086 \\ -0.0600 \\ -0.2571 \end{pmatrix}$$

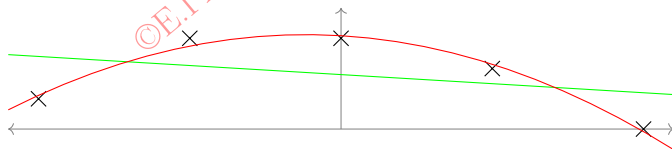
$$A^T A = \begin{pmatrix} 5.0 & 0.0 & 2.5 \\ 0.0 & 2.5 & 0.0 \\ 2.5 & 0.0 & 2.125 \end{pmatrix},$$

επομένως

$$\phi(\xi) = 0.3086 - 0.0600\xi - 0.2571\xi^2$$

$$\text{σφάλμα } \|b - Ax_{LS}\|_2 = 0.0338$$

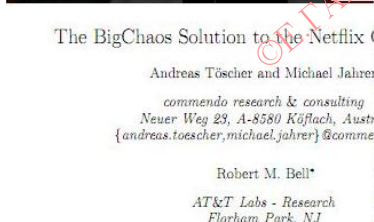
$$A^T b = \begin{pmatrix} 0.9 \\ -0.15 \\ 0.225 \end{pmatrix}$$



Επιχειρηματικό Κίνητρο : Βραβείο Netflix \$1,000,000 II

Περιγραφή (Wikipedia)

- The Netflix Prize was an open competition for the best collaborative filtering algorithm to predict user ratings for films, based on previous ratings without any other information about the users or films, i.e. without the users or the films being identified ...
- Netflix provided a training data set of 100,480,507 ratings that 480,189 users gave to 17,770 movies. Each training rating is a quadruplet of the form $\langle \text{user}, \text{movie}, \text{date of grade}, \text{grade} \rangle$. The user and movie fields are integer IDs, while grades are from 1 to 5 (stars).
- Prizes were based on **improvement over Netflix's own algorithm, called Cinematch**, or the previous year's score if a team has made improvement beyond a certain threshold. **A trivial algorithm that predicts for each movie in the quiz set its average grade from the training data** produces an RMSE of 1.0540. Cinematch uses "straightforward **statistical linear models** with a lot of data conditioning".



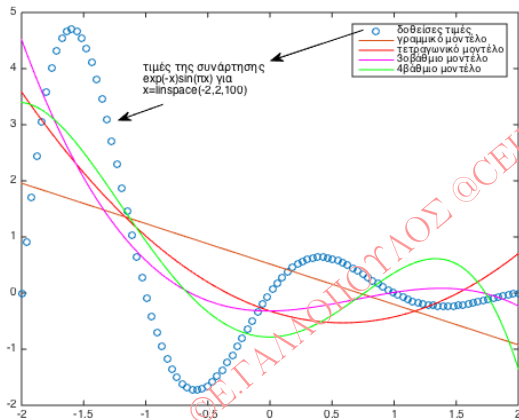
21. Insert the probe set into the training set.

Παραδείγματα προσέγγισης ελαχίστων τετραγώνων

Χρησιμοποιούμε τον παρακάτω κώδικα MATLAB για διάφορες επιλογές x , y

```
1      % κατασκευώνμητρώων
2      A1 = ( ones(n,1), x ); % γραμμικόμοντέλο
3      A2 = ( A1, x.^2 ); % τετραγωνικόμοντέλο
4      A3 = ( A2, x.^3 ); % βάθμιο3 μοντέλο
5      A4 = ( A3, x.^4 ); % βάθμιο4 μοντέλο
6      % επίλυσηκανονικώνεξισώσεων
7      z1 = ( A1' * A1 ) \ ( A1' * y ); z2 = ( A2' * A2 ) \ ( A2' * y );
8      z3 = ( A3' * A3 ) \ ( A3' * y ); z4 = ( A4' * A4 ) \ ( A4' * y );
9      % οπτικοποίηση
10     plot(x, y, 'o', x, A1*z1, x, A2*z2, 'r', x, A3*z3, 'm', x, A4*z4, 'g')
11     % σφάλματα
12     nerr(1) = norm(y - A1*z1); nerr(2) = norm(y - A2*z2);
13     nerr(3) = norm(y - A3*z3); nerr(4) = norm(y - A4*z4);
```

$n=100$; $x=\text{linspace}(-2,2,n)'$; $y = \exp(-x).\sin(\pi*x)$



Σφάλματα: $\text{nerr}(1:4) = 14.4664 \ 12.3761 \ 11.7957 \ 11.0530$

$n=20$; $x = \text{linspace}(-2,2,n)'$; $y = \text{randn}(n,1) \cdot \sin(x) + 3 \cdot \exp(x)$

