ΓΡΑΜΜΙΚΗ ΑΛΓΕΒΡΑ

Διάλεξη 10

Βέλτιστη προσέγγιση
Λύσεις ελαχίστων τετραγώνων
Ορθογώνια μητρώα

Ε. Στεφανόπουλος & Ε. Γαλλόπουλος

Τμήμα Μηχανικών Η/Υ & Πληροφορικής, Πανεπιστήμιο Πατρών

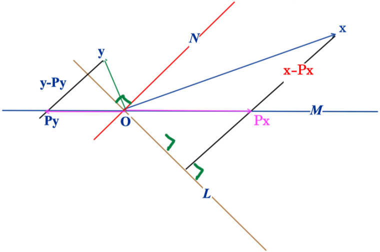
2 Δεκεμβρίου 2022

Προβολές: Ανασκόπηση θεωρίας & Παραδείγματα II

- Οι δύο περιπτώσεις είναι διαφορετικές, δηλ. γενικά $\left(\mathrm{Range}(P)\right)^{\perp} \neq \mathrm{Null}(P)$.
- Αν ισχύει ότι $P = P^{\top}$, τότε $\left(\mathrm{Range}(P)\right)^{\perp} = \mathrm{Null}(P^{\top}) = \mathrm{Null}(P)$, οι δύο περιπτώσεις συμπίπτουν και οι υπόχωροι $\mathrm{Range}(P)$ και $\mathrm{Null}(P)$ είναι ορθογώνιοι μεταξύ τους. Στην περίπτωση αυτή, ο P ονομάζεται τελεστής ορθογώνιας προβολής επί του M.
- Διαφορετικά ονομάζεται τελεστής πλάγιας προβολής (ή πλαγιογώνιας προβολής) επί του M.
- Προσέξτε γενικά ότι με κάθε τελεστή πλάγιας προβολής P, εμπλέκονται οι εξής επιπλέον υπόχωροι, οι οποίοι είναι ορθογώνιοι μεταξύ τους:
  - ο υπόχωρος, έστω $\mathcal{N}$, παράλληλα στον οποίο επιτελείται η προβολή, δηλ. παράλληλα στον οποίο κείται το υπόλοιπο $x - Px$.
  - είτει ο υπόχωρος, έστω $\mathcal{L}$, επί του οποίου το υπόλοιπο είναι κάθετο.
  - Οι υπόχωροι $\mathcal{N}$, $\mathcal{L}$ είναι ορθογώνιοι μεταξύ τους.

Στη συνέχεια θα αναφερόμαστε σχεδόν αποκλειστικά σε ορθογώνιες προβολές.

Αναπαράσταση πλάγιων προβολών Px, Py: επί του υπ. M και ορθογώνια στον υπ. L παράλληλα στον υπ. N



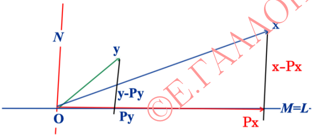
Παράδειγμα για τα σημεία x, y:

Τα y-Py και x-Px είναι κάθετα στον L,
και παράλληλα στον N.

Τα Px, Py ανήκουν στον M.

Επίσης R^n = range(P) + range(I-P) = range(P) + null(P) = M+N (ευθύ άθροισμα)

Αναπαράσταση ορθογώνιων προβολών Px, Py: επί του υπ. M (παράλληλα στον N=ορθογώνιο συμπλήρωμα M)



Αν ο L επιλέγει ίδιος με τον M τότε τα x-Px, y-Py θα είναι κάθετα στον M.

Τότε null(P) = ορθογώνιο συμπλήρωμα του M

Παραδείγματα

Ο δ.χ. είναι το $\mathbb{R}^3$ και μας ενδιαφέρουν οι υπόχωροι $\mathcal{S} := \operatorname{span}\{e_3\}$ και $\mathcal{T} := \operatorname{span}\{e_1, e_2\}$. Τότε αν

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

$$\Rightarrow \quad \text{προβολή επί του } \mathcal{S} \text{ είναι } x_{\mathcal{S}} = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}$$

$$\Rightarrow \quad \text{προβολή επί του } \mathcal{T} \text{ είναι } x_{\mathcal{T}} = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$$

Ο δ.χ. είναι το $\mathbb{R}^3$ και μας ενδιαφέρουν οι υπόχωροι $\mathcal{S} := \mathrm{span}\{e_3\}$ και $\mathcal{T} := \mathrm{span}\{e_1, e_2\}$. Τότε αν

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

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

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

Απλό παράδειγμα

Ο δ.χ. είναι το $\mathbb{R}^3$ και μας ενδιαφέρουν οι υπόχωροι $\mathcal{S} := \mathrm{span}\{e_3\}$ και $\mathcal{T} := \mathrm{span}\{e_1, e_2\}$. Τότε αν $x = (1, 2, -1)^\top$

$$\Rightarrow x_{\mathcal{S}} = P_{\mathcal{S}} x, \text{ όπου } P_{\mathcal{S}} := \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = e_3 (e_3^\top e_3)^{-1} e_3^\top$$

$$\Rightarrow x_{\mathcal{T}} = P_{\mathcal{T}} x, \text{ όπου } P_{\mathcal{T}} := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} e_1 & e_2 \end{pmatrix} \left(\begin{pmatrix} e_1^\top \\ e_2^\top \end{pmatrix} \begin{pmatrix} e_1 & e_2 \end{pmatrix} \right)^{-1} \begin{pmatrix} e_1^\top \\ e_2^\top \end{pmatrix}$$

Αναπαράσταση ορθογώνιας προβολής με μητρώα

Προβολή σε 1 διάσταση

- Κάθε μονοδιάστατος υποχώρος του $\mathbb{R}^n$ αποτελείται από διανύσματα που είναι συγγραμμικά με κάποιο διάνυσμα, έστω $u \in \mathbb{R}^n$. Δοθέντος του u, το μητρώο

$$P := \frac{uu^\top}{u^\top u}$$

είναι ο τελεστής ορθογώνιας προβολής επί του υπόχωρου $\mathcal{M} = \mathrm{span}\{u\}$.

- Εξετάζουμε την επίδραση του P σε τυχαίο $x \in \mathbb{R}^n$:

$$\begin{aligned} \frac{uu^\top}{u^\top u}x &= \frac{u}{\|u\|}\frac{u^\top x}{\|u\|} \\ &= \hat{u}\|x\|\cos(x,u), \quad \textit{όπου } \hat{u} = \frac{u}{\|u\|} \end{aligned}$$

Το Px είναι η ορθογώνια προβολή του x επί του $\mathcal{M}$.

- Προσέξτε ότι

$$(Px)^\top(x - Px) = 0 \Rightarrow x - Px \perp Px$$

που επιβεβαιώνει την ορθογωνιότητα

Παράδειγμα Αν $u = [1, 2, 3]^\top$, τότε

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

Για τυχαίο $x = [\xi_1, \xi_2, \xi_3]^\top$

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

Παράδειγμα

Δίνονται $b, u \neq 0 \in \mathbb{R}^m$. Να βρείτε το $\tilde{\alpha} \in \mathbb{R}$ τ.ώ.

$$\alpha = \arg\min_{\alpha} \|b - \alpha u\|_2^2.$$

Στη συνέχεια (εκτός αν λέγται κάτι διαφορετικό) θα χρησιμοποιούμε τη νόρμα-2 χωρίς να γράφουμε τον υποδείκτη. Χρησιμοποιώντας τον ορισμό της νόρμας και το ότι $b^\top u = u^\top b$,

$$\phi(\alpha) = (b - \alpha u)^\top (b - \alpha u) = \|b\|^2 + \alpha^2 \|u\|^2 - 2\alpha u^\top b$$

επομένως, από γνωστή θεωρία διαφορικού λογισμού, αν κάποιο α είναι ελαχιστοποιητής, θα πρέπει $\frac{d\phi}{d\alpha} = 0$ και $\frac{d^2\phi}{d\alpha^2} > 0$. Παραγωγίζοντας:

$$\frac{d\phi}{d\alpha} = 2\alpha \|u\|^2 - 2u^\top b \Rightarrow \alpha = \frac{u^\top b}{u^\top u}$$

Επίσης, $\frac{d^2\phi}{d\alpha^2} = 2\|u\|^2 > 0$. Επομένως, ο ελαχιστοποιητής είναι $\frac{u^\top b}{u^\top u}$, άρα το διάνυσμα που προσεγγίζει καλύτερα το b ως προς τη νόρμα-2 θα είναι το $\frac{u^\top b}{u^\top u} u$. Προσέξτε ότι είναι ακριβώς η ορθογώνια προβολή $P_u b$ του b στον υπόχωρο $\text{span}\{u\}$. Δηλ. no surprises!!!

Μητρώα προβολής

Ορισμός

Ένα μητρώο $P \in \mathbb{R}^{n\times n}$ αποκαλείται μητρώο προβολής αν είναι αυτοπαθές[α΄], δηλ. $PP = P$.
Ένα μητρώο προβολής P αποκαλείται μητρώο ορθογώνιας προβολής αν επιπλέον είναι συμμετρικό

[α΄] idempotent

Κατασκευή μητρώων ορθογώνιας προβολής

Έστω ο υπόχωρος $\mathcal{V} = \mathrm{span}\{v_1, \ldots, v_n\}$ όπου τα v_i είναι γραμμικά ανεξάρτητα, άρα αποτελούν βάση. Αν $V = (v_1, \ldots, v_n)$, τότε το μητρώο

$$P = V(V^\top V)^{-1} V^\top$$

είναι το μητρώο ορθογώνιας προβολής επί του $\mathcal{V}$. Προσεξτε ότι το μητρώο δεν εξαρτάται από τη βάση.

Παράδειγμα

Έστω $\operatorname{span}\{e_1, e_3\}$ στον $\mathbb{R}^4$ τότε $V = [e_1, e_3]$ και

$$
\begin{aligned}
P_{\mathcal{V}} &= \begin{pmatrix} e_1 & e_3 \end{pmatrix} \left(\begin{pmatrix} e_1^\top \\ e_3^\top \end{pmatrix} \begin{pmatrix} e_1 & e_3 \end{pmatrix} \right)^{-1} \begin{pmatrix} e_1^\top \\ e_3^\top \end{pmatrix} \\
&= \begin{pmatrix} e_1 & e_3 \end{pmatrix} \begin{pmatrix} e_1^\top e_1 & e_1^\top e_3 \\ e_3^\top e_1 & e_3^\top e_3 \end{pmatrix}^{-1} \begin{pmatrix} e_1^\top \\ e_3^\top \end{pmatrix} \\
&= \begin{pmatrix} e_1 & e_3 \end{pmatrix} \begin{pmatrix} e_1^\top \\ e_3^\top \end{pmatrix} \text{ προσέξτε ότι ο μεσαίος όρος παραπάνω είναι } I_2 \\
&= e_1 e_1^\top + e_3 e_3^\top = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}
\end{aligned}
$$

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

Παράδειγμα

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

$$
\begin{aligned}
P_{\mathcal{V}} &= \begin{pmatrix} a & b \end{pmatrix} \left(\begin{pmatrix} a^\top \\ b^\top \end{pmatrix} \begin{pmatrix} a & b \end{pmatrix} \right)^{-1} \begin{pmatrix} a^\top \\ b^\top \end{pmatrix} = \begin{pmatrix} a & b \end{pmatrix} \begin{pmatrix} a^\top a & a^\top b \\ b^\top a & b^\top b \end{pmatrix}^{-1} \begin{pmatrix} a^\top \\ b^\top \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & -1 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 4 & 3 \\ 3 & 11 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & -1 & 3 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & -1 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} \frac{11}{35} & -\frac{3}{35} \\ -\frac{3}{35} & \frac{4}{35} \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & -1 & 3 \end{pmatrix} \\
&= \begin{pmatrix} \frac{11}{35} & \frac{8}{35} & \frac{14}{35} & \frac{2}{35} \\ \frac{8}{35} & \frac{9}{35} & \frac{7}{35} & \frac{11}{35} \\ \frac{14}{35} & \frac{7}{35} & \frac{21}{35} & -\frac{7}{35} \\ \frac{2}{35} & \frac{11}{35} & -\frac{7}{35} & \frac{29}{35} \end{pmatrix}
\end{aligned}
$$

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

Αν μας έδιναν το παραπάνω $P_{\mathcal{V}}$ και ένα τυχαίο διάνυσμα $x = [\xi_1, \xi_2, \xi_3, \xi_4]^\top$ τότε

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

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

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

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

Παρατηρήσεις

Στη γενική περίπτωση, το (τετραγωνικό) μητρώο προβολής P δεν είναι αντιστρέψιμο.

Αυτό μπορεί να αναδειχθεί με διάφορους τρόπους.
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 ποιό διάνυσμα του χώρου να επιλέξει). Π.χ. είδαμε ότι αν

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

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

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

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.

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

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

- Πολλές εφαρμογές οδηγούν σε μετρήσεις, π.χ. ζεύγη τιμών $(\xi_1, \psi_1), \ldots, (\xi_m, \psi_m)$ (τα x_j διαφορετικά) και αναζητούμε απλή συνάρτηση f που να προσαρμόζεται στις τιμές αυτές (να τις μοντελοποιεί) όσο γίνεται καλύτερα. Εννοούμε ότι θα θέλαμε οι τιμές $f(\xi_1), \ldots, f(\xi_m)$ να είναι κοντά στις μετρήσεις.
- Πιο γενικά, χρειάζεται να προσεγγίσουμε μία δύσκολη ή άγνωστη συνάρτηση ως γραμμικό συνδυασμό γνωστών, απλούστερων συναρτήσεων.
- Η επίλυση των προβλημάτων αυτών αφορά στη γενική θεωρία προσεγγίσεων και ελαχιστοποίησης/αριστοποίησης.
- Η Γραμμική Άλγεβρα συνεισφέρει καθοριστικά στην επίλυσή τους με προσεγγίσεις Ελαχίστων Τετραγώνων και άλλες τεχνικές.

Παράδειγμα

Έστω τα παραπάνω ζεύγη τιμών και ότι αποφασίζουμε να χρησιμοποιήσουμε πολυώνυμο π.χ. με $\phi_i(\xi) = \xi^{i-1}$

$$f(\xi; c) = \gamma_0 + \gamma_1 x + \cdots + \gamma_{n-1}\xi^{n-1}.$$

Αν υπολογίσουμε τις τιμές $[\phi_1(\xi_j), \phi_2(\xi_j), \ldots, \phi_n(\xi_j)]$ για κάθε ξ_j, τότε προκύπτει το παρακάτω γραμμικό σύστημα που πρέπει να λυθεί ως προς τους συντελεστές (ή παραμέτρους) $c = [\gamma_0, \ldots, \gamma_{n-1}]^\top$.

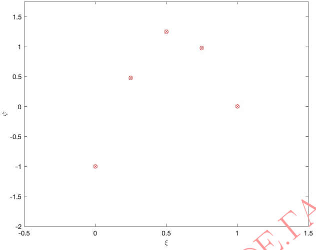
$$A = \begin{pmatrix} 1 & \xi_1 & \cdots & \xi_1^{n-1} \\ 1 & \xi_2 & & \xi_2^{n-1} \\ \vdots & & & \vdots \\ \vdots & & \vdots & \\ 1 & \xi_m & \cdots & \xi_m^{n-1} \end{pmatrix}, c = \begin{pmatrix} \gamma_0 \\ \vdots \\ \gamma_{n-1} \end{pmatrix}. \text{ και } \quad y = \begin{pmatrix} \psi_1 \\ \vdots \\ \vdots \\ \psi_m \end{pmatrix}$$

- Αν $m = n$, τότε $A \in \mathbb{R}^{n\times n}$ είναι τετραγωνικό και αν είναι αντιστρέψιμο, τότε υπάρχει μοναδική λύση $c = A^{-1}y$.
- Αν $m > n$, τότε το $A \in \mathbb{R}^{m\times n}$ και εκτός αν $y \in \text{Range}(A)$, δεν υπάρχει $c \in \mathbb{R}^n$ τέτοιο ώστε $y = Ac$.
- Αν χρησιμοποιήσουμε μητρώα, το πρόβλημα γράφεται ισοδύναμα ως $\arg\min_{c\in\mathbb{R}^n} \|y - Ac\|$

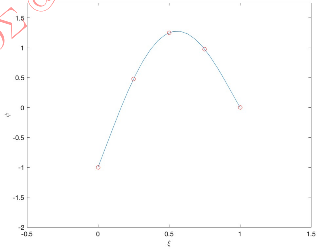
Παράδειγμα

Δίνονται

$$x = \begin{pmatrix} 0.0000 \\ 0.2500 \\ 0.5000 \\ 0.7500 \\ 1.0000 \end{pmatrix}, \; y = \begin{pmatrix} -1.0000 \\ 0.4767 \\ 1.2500 \\ 0.9767 \\ 0.0000 \end{pmatrix}, \text{ άρα } A = \begin{pmatrix} 1.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 1.0000 & 0.2500 & 0.0625 & 0.0156 & 0.0039 \\ 1.0000 & 0.5000 & 0.2500 & 0.1250 & 0.0625 \\ 1.0000 & 0.7500 & 0.5625 & 0.4219 & 0.3164 \\ 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 \end{pmatrix}$$



$$p(\xi) = -1 + 6.1699\xi + 2.1506\xi^2 - 14.6409\xi^3 + 7.3204\xi^4$$

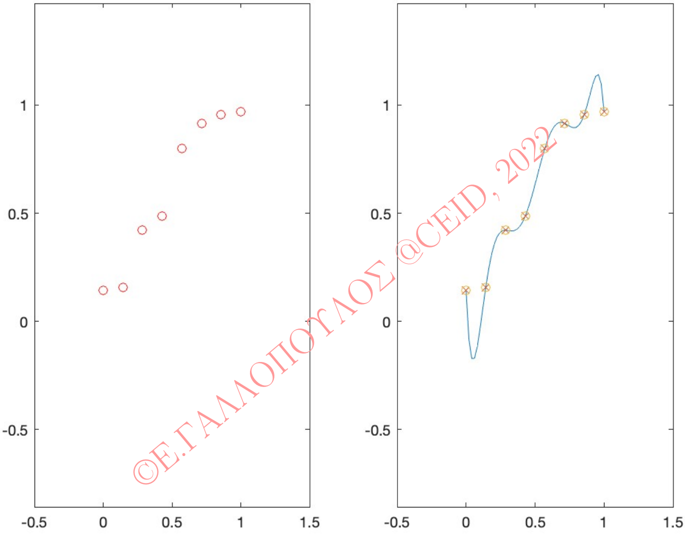


$$A^{-1}y = \begin{pmatrix} -1.0000 \\ 6.1699 \\ 2.1506 \\ -14.6409 \\ 7.3204 \end{pmatrix}$$

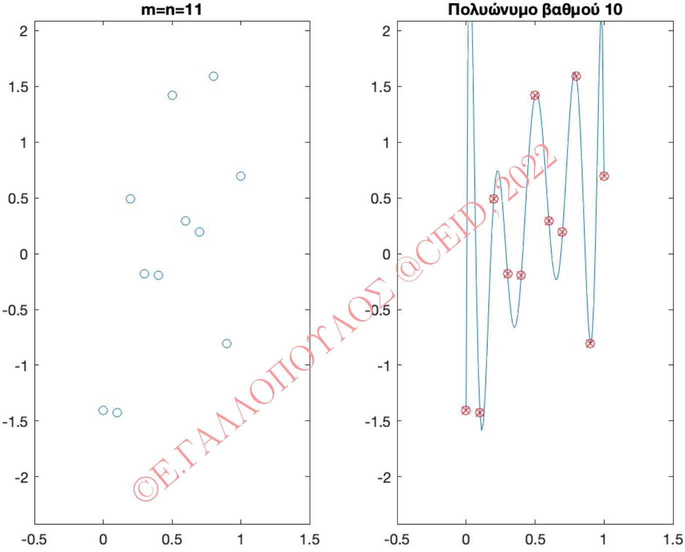
Από τα σημεία $x, y \in \mathbb{R}^5$, κατασκευάσαμε 4-βάθμιο πολυώνυμο p για το οποίο $p(\xi_i) = \psi_i, i = 1, \ldots, 5$.

- Θεωρητικά, αυτό μπορούμε να το κάνουμε για αυθαίρετα μεγάλες τιμές $m = n$.
- Μάλιστα αποδεικνύεται ότι αν η άγνωστη συνάρτηση που παρήγαγε τις τιμές ψ_i ήταν ένα πολυώνυμο βαθμού $n - 1$ και είχαμε τις τιμές του στα $m = n$ διακριτά σημεία $\xi_1, ..., \xi_m$, τότε θα μπορούσαμε να υπολογίσουμε ακριβώς τους συντελεστές $\gamma_0, ..., \gamma_{n-1}$ και επομένως να κατασκευάσουμε το πολυώνυμο (σε μορφή αθροίσματος δυνάμεων του ξ, λέγεται δυναμομορφή).
- Αυτό το ζήτημα είναι κεντρικό σε ένα κεφάλαιο του αντικειμένου της Αριθμητικής Ανάλυσης και της Θεωρίας Προσεγγίσεων (Πολυωνυμική Παρεμβολή).
- Δυστυχώς όμως, για λόγους που αφορούν την (μη ακριβή) αριθμητική που επιτελείται στην αριθμητική μονάδα των Η/Υ και άλλους παράγοντες, η εύρεση του p καθίσταται δυσχερής καθώς μεγαλώνει το n (υπενθυμίζουμε ότι στο παράδειγμα $m = n$.)
- Τα επόμενα σχήματα, αναδεικνύουν κάπως τις δυσκολίες : Παρατηρήστε ότι καθώς μεγαλώνει το πλήθος των σημείων n, η πολυωνυμική συνάρτηση έχει όλο και πιο έντονες διακυμάνσεις, σε βαθμό που σύντομα (π.χ. αν $m > 10$) καθίσταται μη πρακτική.
- Όπως θα δείτε, προσθέτοντας στοιχεία, φαίνεται να δυσκολεύει η διαδικασία (και δεν έχουμε πει τίποτα για το κόστος !) Προσοχή: δεν θέλουμε να υποδείξουμε ότι είναι κακό να έχεις πολλά δεδομένα (θα ήταν ομολογουμένως περίεργο) αλλά οτι ενδεχομένωςμ θα ήταν σκόπιμο να μειώσουμε τον αριθμό των παραμέτρων, δηλαδή το n[1] . Εξάλλου, η συνάρτηση που δίνει τις τιμές είναι άγνωστή και δεν υπάρχει κανένας λόγος να ισχυει ότι $m = n$. Εξάλλου, και οι τιμές ακόμα μπορεί να μην είναι ακριβείς.

[1] Υπερπαράμετρος !!!



Προσέξτε την έντονη διακύμανση στις πιο ακραίες τιμές.



Φαίνεται ότι η διακύμανση γίνεται εντονότερη.

Το παραπάνω είναι ειδική περίπτωση του εξής γενικότερου προβλήματος:

- Έστω ξ μια ανεξάρτητη μεταβλητή και ότι $Y(\xi)$ είναι η συνάρτηση που θέλουμε να προσεγγίσουμε. Μας δίνονται m παρατηρήσεις, δηλ. οι τιμές της Y που έχουν μετρηθεί σε συγκεκριμένες τιμές του ξ.

$$\psi_i = Y(\xi_i), i = 1, \ldots, m$$

- Θέλουμε να μοντελοποιήσουμε τη συνάρτηση $Y(\xi)$ μέσω μιας συνάρτησης $f(\xi; c)$ που είναι γραμμικός συνδυασμός n (απλών) συναρτήσεων $\phi_j(t)$, δηλ.

$$\begin{aligned} f(\xi; c) &= \gamma_1\phi_1(t) + \cdots + \gamma_n\phi_n(t) \\ Y(\xi) &\approx f(\xi; c) \end{aligned}$$

- Το μητρώο $A = [\phi_j(\xi_i)]$ (design matrix) είναι $m \times n$.
- Θέτουμε $x = [\xi_1, \ldots, \xi_m]^\top$, $y = [\psi_1, \ldots, \psi_m]^\top$ και $A = [\phi_j(\xi_i)]$ (design matrix).
- Οι διαστάσεις είναι $y \in \mathbb{R}^m, x \in \mathbb{R}^n, A \in \mathbb{R}^{m \times n}$.
- Επομένως θέλουμε να υπολογίσουμε τους αγνώστους $c = [\gamma_0, \ldots, \gamma_{n-1}]^\top$ έτσι ώστε

$$y \approx Ac$$

Με βάση την προηγούμενη συζήτηση, θα θέλαμε να διερευνήσουμε την επιλογή λιγότερων παραμέτρων, δηλ. $n < m$.

- Εμπόδιο: Το γραμμικό σύστημα θα έχει περισσότερες εξισώσεις από αγνώστους (αποκαλείται υπερπροσδιορισμένο).
- $y \in \mathbb{R}^m$ αλλά rank (range)$(A) \leq n \leq m$.
- Γενικά, δεν υπάρχει λύση! $\boxed{\text{ΤΙ ΚΑΝΟΥΜΕ;}}$

Χαλαρώνουμε ... και αρκούμαστε σε προσέγγιση (με κάποια κριτήρια). Για παράδειγμα:

$$\arg\min_{c \in \mathbb{R}^n} \|y - Ac\|_2 \text{ (κριτήριο ελαχίστων τετραγώνων)}$$

Δηλαδή, ζητούμε να υπολογίσουμε το διάνυσμα $\hat{c} \in \mathbb{R}^n$ που ελαχιστοποιεί τη 2-νόρμα της διαφοράς (υπολοίπου) $\|y - Ac\|_2$. Το διάνυσμα λέγεται και ελαχιστοποιητής.
Ζητούμε δηλαδή να προσεγγίσουμε το $y \in \mathbb{R}^n$ όσο γίνεται καλύτερα από το διανυσματικό υπόχωρο Range(A).
Αν είμαστε πολύ τυχεροί (που δεν συμβαίνει στη γενική περίπτωση), μπορεί $y \in$ Range(A), οπότε υπάρχει ακριβής λύση, δηλ. η σχέση $y = A\hat{c}$ ισχύει ακριβώς για τον ελαχιστοποιητή $\hat{c}$.
Για διευκόλυνση, θα θεωρήσουμε ότι Range$(A) = n$. Τότε, από τη θεωρία των προβολών, γνωρίζουμε ότι η βέλτιστη προσέγγιση θα είναι η ορθογώνια προβολή του y επί του Range(A): Δηλαδή

$$P_{\text{Range}(A)}\, y = A(A^\top A)^{-1} A^\top y$$

επομένως $\hat{c} = (A^\top A)^{-1} A^\top y$. (Στη συνέχεια, θα χρησιμοποιούμε περισσότερο το x για το διάνυσμα των παραμέτρων.)

Περιγραφή

Δίνονται $A \in \mathbb{R}^{m\times n}$, $y \in \mathbb{R}^m$, $m \geq n$. Το **γραμμικό πρόβλημα ελαχίστων τετραγώνων** συνίσταται στην εύρεση ενός διανύσματος από το σύνολο

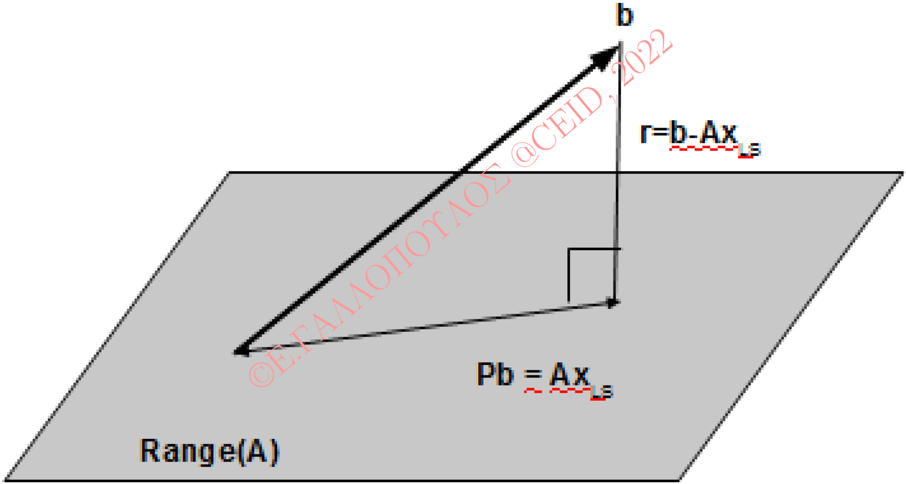
$$\mathcal{X} = \{x \in \mathbb{R}^n : \text{ ελαχιστοποιεί το } \rho(x) = \|Ax - y\|_2\}.$$

- Η ελαχιστοποίηση του $\|Ax - y\|_2$ ισοδυναμεί με την εύρεση διανύσματος $x_{LS} \in \mathbb{R}^n$ τέτοιου ώστε, μεταξύ όλων των διανυσμάτων που παράγονται από γραμμικό συνδυασμό των στηλών του A, το $p = Ax_{LS}$ να είναι το πλησιέστερο στο b (ως προς την ευκλείδεια νόρμα).
- Για να συμβαίνει αυτό, το $r = y - Ax_{LS}$ είναι κάθετο στον υπόχωρο $\text{Range}(A)$ άρα $r \in \text{Null}(A^\top)$. Επομένως, για οποιοδήποτε αυθαίρετο διάνυσμα Az του $\text{Range}(A)$

$$0 = (Az)^\top (y - Ax_{LS}) = z^\top (A^\top y - A^\top Ax_{LS}).$$

Επομένως το x_{LS} είναι η λύση του συστήματος $\boxed{A^\top Ax_{LS} = A^\top b}$ που αποκαλούνται κανονικές εξισώσεις του προβλήματος των ελαχίστων τετραγώνων. Το μητρώο $A^\top A$ αποκαλείται συχνά μητρώο Gram.

Οπτικοποίηση



Επιλυσιμότητα των κανονικών εξισώσεων $A^\top Ax = A^\top b$

Υπάρχει λύση? Είναι μοναδική?

- Επιλυσιμότητα Είναι εξασφαλισμένο ότι υπάρχει τουλάχιστον μία λύση: Γιατί? Η τάξη $r = \mathbf{rank}(A^\top A) \leq n$. Αν $r = n$ τότε υπάρχει μοναδική λύση. Αν $r < n$, τότε ο μηδενόχωρος του $A^\top A$ έχει διάσταση $n - r$ και υπάρχουν άπειρες λύσεις. Στην περίπτωση αυτή, επιλέγουμε από τις λύσεις με κάποιο επιπλέον κριτήριο, π.χ. το διάνυσμα που έχει το μικρότερο μήκος (διπλή βελτιστοποίηση).
- Παρατήρηση: Το μητρώο $A^\top A$ είναι αντιστρέψιμο αν και μόνον αν οι στήλες του είναι γραμμικά ανεξάρτητες.

Συμμετρικά θετικά (ημι)ορισμένα μητρώα

Ενδιαφέρον χαρακτηριστικό π.χ. στο μητρώο των κανονικών εξισώσεων

$$(A^\top A)x = A^\top y.$$

Το $B = A^\top A$:

- είναι συμμετρικό
- ισχύει πάντα ότι

$$x^\top A^\top A x \geq 0$$

- όταν οι στήλες του A είναι γραμμικά ανεξάρτητες, $x^\top A^\top A x > 0$, είναι αυστηρά θετικό.

Ορισμός

Κάθε $A \in \mathbb{R}^{n\times n}$ για το οποίο ισχύει ότι $^\top = A$ και $f(x) := x^\top A x \geq 0$ για κάθε $x \in \mathbb{R}^n$ και $f(x) = 0$ ανν $x = 0$ αποκαλείται **συμμετρικό θετικά ορισμένο μητρώο (ΣΘΟ)**. Αν υπάρχει $x \neq 0$ ώστε $f(x) = 0$ τότε αποκαλείται **συμμετρικό θετικά ημιορισμένο μητρώο**. Αν λαμβάνει και αρνητικές τιμές, αποκαλείται **αόριστο**.

Συμμετρικά θετικά (ημι)ορισμένα μητρώα

Ενδιαφέρον χαρακτηριστικό π.χ. στο μητρώο των κανονικών εξισώσεων

$$(A^\top A)x = A^\top y.$$

Το $B = A^\top A$:

- είναι συμμετρικό
- ισχύει πάντα ότι

$$x^\top A^\top A x \geq 0$$

- όταν οι στήλες του A είναι γραμμικά ανεξάρτητες, $x^\top A^\top A x > 0$, είναι αυστηρά θετικό.

Ορισμός

Κάθε $A \in \mathbb{R}^{n\times n}$ για το οποίο ισχύει ότι $^\top = A$ και $f(x) := x^\top A x \geq 0$ για κάθε $x \in \mathbb{R}^n$ και $f(x) = 0$ ανν $x = 0$ αποκαλείται **συμμετρικό θετικά ορισμένο μητρώο (ΣΘΟ)**. Αν υπάρχει $x \neq 0$ ώστε $f(x) = 0$ τότε αποκαλείται **συμμετρικό θετικά ημιορισμένο μητρώο**. Αν λαμβάνει και αρνητικές τιμές, αποκαλείται **αόριστο**.

Επομένως, το $A^\top A$ είναι ΣΘΟ ανν το έχει γραμμικά ανεξάρτητες στήλες, διαφορετικά είναι θετικά ημιορισμένο. Δεν μπορεί όμως να είναι αόριστο.

Γενική μορφή γραμμικού προβλήματος ελαχίστων τετραγώνων

$$f(\xi; c) = \gamma_1\phi_1(\xi) + \cdots + \gamma_n\phi_n(\xi).$$

Υπολογίζουμε τις τιμές $[\phi_1(\xi_j), \phi_2(\xi_j), \ldots, \phi_n(\xi_j)]$ για κάθε ξ_j, όπου $\phi_1, \ldots, \phi_n$ είναι γνωστές (συνήθως απλές) συναρτήσεις, ενώ οι συντελεστές $\gamma_1, \ldots, \gamma_n$ είναι σταθερές που πρέπει να καθοριστούν έτσι ώστε να ελαχιστοποιείται το άθροισμα των τετραγώνων του σφάλματος σε κάθε σημείο. Αν χρησιμοποιήσουμε μητρώα, το πρόβλημα γράφεται ισοδύναμα ως $\arg\min_{c\in\mathbb{R}^n} \|y - Ac\|$ όπου

$$A = \begin{pmatrix} \phi_1(\xi_1) & \cdots & \phi_n(\xi_1) \\ \vdots & & \vdots \\ \vdots & & \vdots \\ \phi_1(\xi_m) & \cdots & \phi_n(\xi_m) \end{pmatrix}, c = \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_n \end{pmatrix}. \text{ και } \quad y = \begin{pmatrix} \psi_1 \\ \vdots \\ \vdots \\ \psi_m \end{pmatrix}$$

Γραμμικό πρόβλημα ελαχίστων τετραγώνων[2]

- Στο γραμμικό πρόβλημα ελαχίστων τετραγώνων, η συνάρτηση μοντέλο $f(t;x)$ εξαρτάται γραμμικά από τις παραμέτρους:

$$f(t;x) = \xi_1 f_1(t) + \cdots + \xi_n f_n(t)$$

- Στο γραμμικό πρόβλημα ελαχίστων τετραγώνων, οι συναρτήσεις $f_j(t)$ μπορεί να είναι μη γραμμικές!
- π.χ.

$$f(t;x) = \xi_1 + \xi_2 e^{-t^2} + \xi_3 \sin^2(t\pi)$$

[2] Ονομάζεται και ((πρόβλημα γραμμικών ελαχίστων τετραγώνων))

Παράδειγμα

Πρόβλημα: Λαμβάνουμε μέτρησεις $\psi_1, \psi_2, \ldots, \psi_m$ (π.χ. ένα σήμα) στις χρονικές στιγμές $t = t_1, t_2, \ldots, t_m$. Παρατηρούμε ότι το σήμα έχει ταλαντούμενη συμπεριφορά που φαίνεται να μειώνεται με εκθετικά. Κρίνουμε σκόπιμο να το μοντελοποιήσουμε με μια συνάρτηση

$$f(t;x) \quad = \quad \xi_1 + \xi_2 e^{-(\xi_3 - t)^2} \sin(\xi_4 t)$$

όπου τα στοιχεία του $x = (\xi_1, \xi_2, \ldots, \xi_4)^\top$ είναι οι παράμετροι του μοντέλου. Για να ορίσουμε τη συνάρτηση, θέλουμε να επιλέξουμε το διάνυσμα των παραμέτρων x ώστε να επιτυγχάνεται η βέλτιστη προσαρμογή τιμών $f(t_j;x)$ στις μετρήσεις.

Διατύπωση ως πρόβλημα ελαχίστων τετραγώνων: Να υπολογιστεί $x \in \mathbb{R}^4$ τέτοιο ώστε να ελαχιστοποιείται το άθροισμα των τετραγώνων

$$\min_{x \in \mathbb{R}^6} \sum_{j=1}^{m} (\psi_j - f(t_j;x))^2(x) = \min_{x \in \mathbb{R}^6} \sum_{j=1}^{m} r_j^2(x)$$

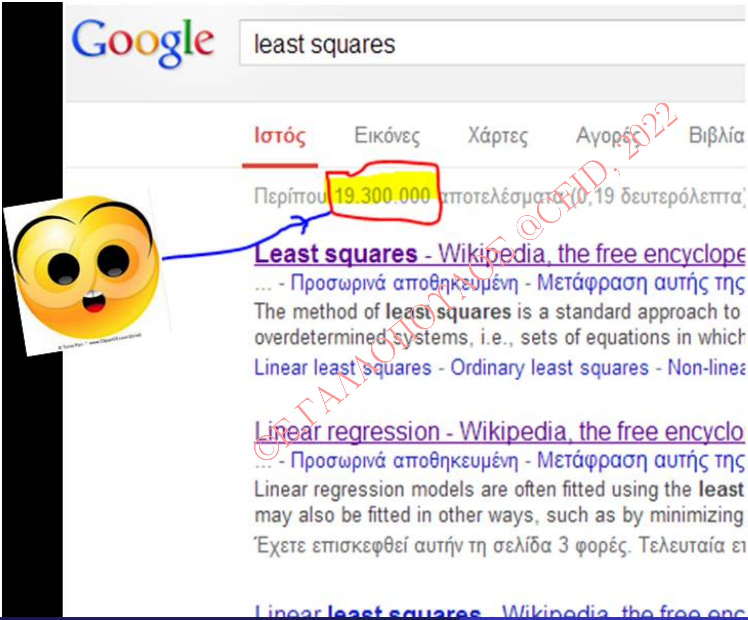
όπου $r_j(x) = \psi_j - f(t_j;x)$, $j = 1, \ldots, m$ είναι οι τιμές των υπολοίπων στα σημεία $t_1, \ldots, t_m$.

Ορολογία Πρόκειται για ένα γενικό πρόβλημα ελαχίστων τετραγώνων.

Προσοχή Είναι μη γραμμικό γιατί η συνάρτηση μοντέλο f εξαρτάται μη γραμμικά από τις παραμέτρους.

☺ Δεν θα ασχοληθούμε με μη γραμμικά προβλήματα!

Τι λέγεται για τα ελάχιστα τετράγωνα



Τι λέγεται για τα ελάχιστα τετράγωνα

least squares problems

All · Images · Videos · News · Shopping · More · Tools

About 96,300,000 results (0.64 seconds)

Least squares

Overview · Practice problems · Examples · Videos

https://en.wikipedia.org › wiki › Least_squares

Least squares - Wikipedia

Least squares problems fall into two categories: linear or ordinary least squares and nonlinear least squares, depending on whether or not the residuals are ...

Linear least · Least-squares function · Ordinary least · Generalized least

https://textbooks.math.gatech.edu › ila › least-squares

The Method of Least Squares

We will present two methods for finding **least-squares** solutions, and we will give several applications to best-fit **problems**.

Abo

The
in re
over
squa
indiv
in da

Τι λέγεται για τα ελάχιστα τετράγωνα

Τι λέγεται για τα ελάχιστα τετράγωνα

Least squares

From Wikipedia, the free encyclopedia

"Least squares approximation" redirects here. Not to be confused with Least-squares function approximation.

The method of **least squares** is a standard approach in regression analysis to approximate the solution of overdetermined systems (sets of equations in which there are more equations than unknowns) by minimizing the sum of the squares of the residuals (a residual being the difference between an observed value and the fitted value provided by a model) made in the results of each individual equation.

The most important application is in data fitting. When the problem has substantial uncertainties in the independent variable (the *x* variable), then simple regression and least-squares methods have problems; in such cases, the methodology required for fitting errors-in-variables models may be considered instead of that for least squares.

Least squares problems fall into two categories: linear or ordinary least squares and nonlinear least squares, depending on whether or not the residuals are linear in all unknowns. The linear least-squares problem occurs in statistical regression analysis; it has a closed-form solution. The nonlinear problem is usually solved by iterative refinement; at each iteration the system is approximated by a linear one, and thus the core calculation is similar in both cases.

Polynomial least squares describes the variance in a prediction of the dependent variable as a function of the independent variable and the deviations from the fitted curve.

When the observations come from an exponential family with identity as its natural sufficient statistics and mild-conditions are satisfied (e.g. for normal, exponential, Poisson and binomial distributions), standardized least-squares estimates and maximum-likelihood estimates are identical.[1] The method of least squares can also be derived as a method of moments estimator.

The following discussion is mostly presented in terms of linear functions but the use of least squares is valid and practical for more general families of functions. Also, by iteratively applying local quadratic approximation to the likelihood (through the Fisher information), the least-squares method may be used to fit a generalized linear model.

The least-squares method was officially discovered and published by Adrien-Marie Legendre (1805),[2] though it is usually also co-credited to Carl Friedrich Gauss (1795)[3][4] who contributed significant theoretical advances to the method and may have previously used it in his work.[5][6]

Contents [hide]

Part of a series on
Regression analysis

Models

Linear regression · Simple regression · Polynomial regression · General linear model · Proportional hazards model

Generalized linear model · Vector generalized linear model · Discrete choice · Binomial regression · Binary regression · Logistic regression · Multinomial logistic regression · Mixed logit · Probit · Multinomial probit · Ordered logit · Ordered probit · Poisson

Multilevel model · Fixed effects · Random effects · Linear mixed-effects model · Nonlinear mixed-effects model

Nonlinear regression · Support vector regression · Nonparametric · Semiparametric · Robust · Quantile · Isotonic · Principal components · Least angle · Local · Segmented

Errors-in-variables

Estimation

Least squares · Linear · Non-linear

Ordinary · Weighted · Generalized · Generalized estimating equation

Partial · Total · Non-negative · Ridge regression · Regularized

Least absolute deviations · Iteratively reweighted · Bayesian · Bayesian multivariate · Least-squares spectral analysis · Heteroscedasticity Consistent Regression Standard Errors · Heteroscedasticity and Autocorrelation Consistent Regression Standard Errors · Instrumental variables estimation

Background

Regression validation · Mean and predicted response · Errors and residuals · Goodness of fit · Studentized residual · Gauss–Markov theorem

Mathematics portal

Ορισμός (Wikipedia)

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

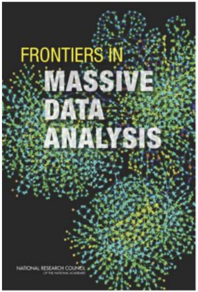
Η μηχανική μάθηση διερευνά τη μελέτη και την κατασκευή αλγορίθμων που μπορούν **να μαθαίνουν από τα δεδομένα** και **να κάνουν προβλέψεις σχετικά με αυτά**.

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**.

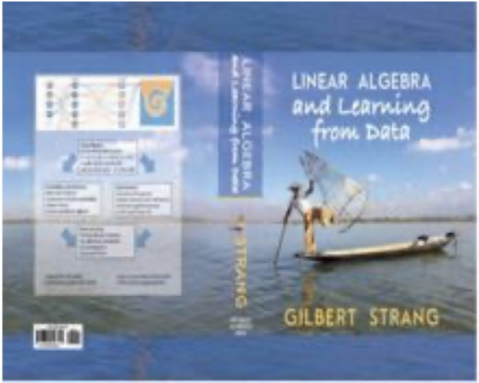
The 7 Computational Giants of Massive Data Analysis (National Research Council, 2013)

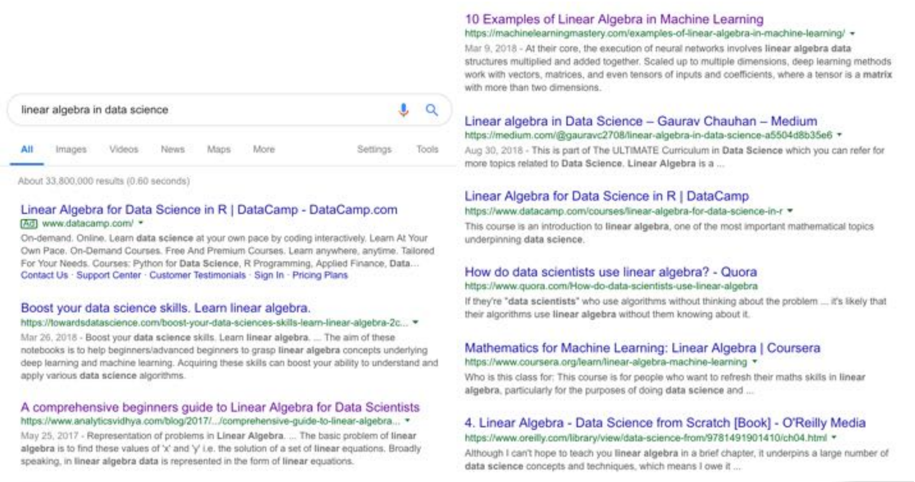


1. Basic statistics,
2. Generalized N-body problem,
3. Graph-theoretic computations,
4. **Linear algebraic computations**,
5. Optimization,
6. Integration, and
7. Alignment problems.

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**.




linear algebra in data science

All Images Videos News Maps More Settings Tools

About 33,800,000 results (0.60 seconds)

Linear Algebra for Data Science in R | DataCamp - DataCamp.com
Ad www.datacamp.com/
On-demand. Online. Learn data science at your own pace by coding interactively. Learn At Your Own Pace. On-Demand Courses. Free And Premium Courses. Learn anywhere, anytime. Tailored For Your Needs. Courses: Python for Data Science, R Programming, Applied Finance, Data...
Contact Us · Support Center · Customer Testimonials · Sign In · Pricing Plans

Boost your data science skills. Learn linear algebra.
https://towardsdatascience.com/boost-your-data-sciences-skills-learn-linear-algebra-2c...
Mar 26, 2018 - Boost your data science skills. Learn linear algebra. ... The aim of these notebooks is to help beginners/advanced beginners to grasp linear algebra concepts underlying deep learning and machine learning. Acquiring these skills can boost your ability to understand and apply various data science algorithms.

A comprehensive beginners guide to Linear Algebra for Data Scientists
https://www.analyticsvidhya.com/blog/2017/.../comprehensive-guide-to-linear-algebra...
May 25, 2017 - Representation of problems in Linear Algebra. ... The basic problem of linear algebra is to find these values of 'x' and 'y' i.e. the solution of a set of linear equations. Broadly speaking, in linear algebra data is represented in the form of linear equations.

10 Examples of Linear Algebra in Machine Learning
https://machinelearningmastery.com/examples-of-linear-algebra-in-machine-learning/
Mar 9, 2018 - At their core, the execution of neural networks involves linear algebra data structures multiplied and added together. Scaled up to multiple dimensions, deep learning methods work with vectors, matrices, and even tensors of inputs and coefficients, where a tensor is a matrix with more than two dimensions.

Linear algebra in Data Science – Gaurav Chauhan – Medium
https://medium.com/@gauravc2708/linear-algebra-in-data-science-a5504d8b35e6
Aug 30, 2018 - This is part of The ULTIMATE Curriculum in Data Science which you can refer for more topics related to Data Science. Linear Algebra is a ...

Linear Algebra for Data Science in R | DataCamp
https://www.datacamp.com/courses/linear-algebra-for-data-science-in-r
This course is an introduction to linear algebra, one of the most important mathematical topics underpinning data science.

How do data scientists use linear algebra? - Quora
https://www.quora.com/How-do-data-scientists-use-linear-algebra
If they're "data scientists" who use algorithms without thinking about the problem ... it's likely that their algorithms use linear algebra without them knowing about it.

Mathematics for Machine Learning: Linear Algebra | Coursera
https://www.coursera.org/learn/linear-algebra-machine-learning
Who is this class for: This course is for people who want to refresh their maths skills in linear algebra, particularly for the purposes of doing data science and ...

4. Linear Algebra - Data Science from Scratch [Book] - O'Reilly Media
https://www.oreilly.com/library/view/data-science-from-scratch/9781491901410/ch04.html
Although I can't hope to teach you linear algebra in a brief chapter, it underpins a large number of data science concepts and techniques, which means I owe it ...

MIT MATHEMATICS

18.065 - Matrix Methods In Data Analysis, Signal Processing, And Machine Learning (Spring 2018)

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**.

Κατηγορία: News
Τίτλος: Machine Learning and Machine Unlearning for Data Systems
Ομιλητής: Peter Triantafillou, Professor of Data Systems at the Department of Computer Science at the University of Warwick
Ημερομηνία-χώρος: Παρασκευή 2 Δεκεμβρίου, 3μμ, ΤΜΗΥΠ, αμφιθέατρο Γ.

Περίληψη: Learned Data Systems (that is, data systems with machine learning components) bear the promise of increased performance, especially for resource-hungry analytics tasks over large datasets. As such, they are enjoying large attention by researchers. However, DB systems differ from other domains where machine learning (ML) plays a key role in that DBs are continuously updated. How can we ensure then that previously trained neural-network ML models continue to be accurate in the face of DB updates, such as data insert and/or delete operations? New data insertions may carry out-of-distribution (OOD) data for which models may be highly inaccurate. Likewise, for data deletions, which additionally introduce an additional challenge, namely that of unlearning. How can we then surgically unlearn what was previously learned and now deleted without erasing knowledge about relevant retained data? And how can we ensure the above efficiently, i.e., without retraining the models from scratch (which is a time-consuming operation)? In this talk, I will highlight our research results for the above problems. To our knowledge this is the first research results achieving the above goals.

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.

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)

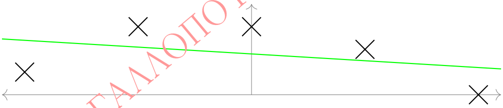
Πρόβλημα: Να υπολογίσουμε την ευθεία γραμμή $\gamma_1 + \gamma_2\xi$ που προσαρμόζεται σε ένα σύνολο γνωστών τιμών $\{(\xi_j, \psi_j)\}_{j=1}^{m}$ με το ελάχιστο δυνατό σφάλμα.

Ορολογία: Στη Στατιστική λέγεται γραμμική παλινδρόμηση (linear regression) Περί τίνος πρόκειται; Είναι η προσπαθεια να κατασκευάσουμε ένα απλό (γραμμικό) μοντέλο για τις παρατηρήσεις που αντιστοιχούν στα m ζεύγη των παραπάνω τιμών.

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.

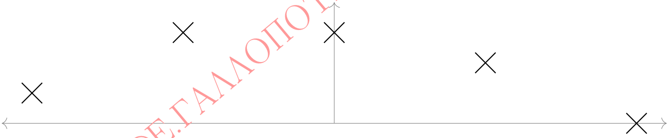
Ορολογία: Στη Στατιστική λέγεται γραμμική παλινδρόμηση (linear regression) Περί τίνος πρόκειται; Είναι η προσπαθεια να κατασκευάσουμε ένα απλό (γραμμικό) μοντέλο για τις παρατηρήσεις που αντιστοιχούν στα m ζεύγη των παραπάνω τιμών. Πώς μετράμε το σφάλμα? Το ορίζουμε ως το άθροισμα των τετραγώνων των 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 $\begin{pmatrix}1 & \xi\end{pmatrix}\begin{pmatrix}\gamma_1\\ \gamma_2\end{pmatrix} = \psi$

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

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

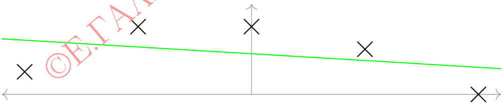
$$A^\top A = \begin{pmatrix} 5.0 & 0.0 \\ 0.0 & 2.5 \end{pmatrix}, A^\top b = \begin{pmatrix} 0.9 \\ -0.15 \end{pmatrix}$$

Αν κάνουμε τους υπολογισμούς

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

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

σφάλμα $\|b - Ax_{\text{LS}}\|_2 = 0.2429$



Προσαρμογή με καμπύλη γραμμή 2ου βαθμού

$\phi_0(x) = 1, \phi_1(\xi) = \xi, \phi_2(\xi) = \xi^2$ 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

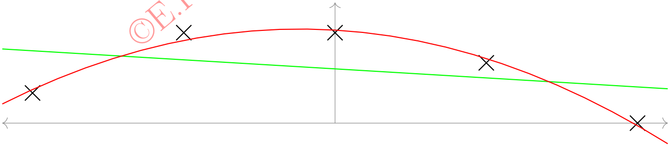
$$x_{LS} = (A^\top A)^{-1}A^\top b = \begin{pmatrix} 0.3086 \\ -0.0600 \\ -0.2571 \end{pmatrix}$$

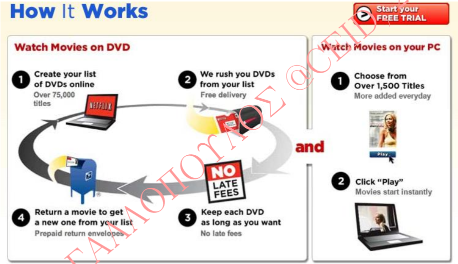
$$A^\top A = \begin{pmatrix} 5.0 & 0.0 & 2.5 \\ 0.0 & 2.5 & 0.0 \\ 2.5 & 0.0 & 2.125 \end{pmatrix},$$

επομένως $\boxed{\phi(\xi) = 0.3086 - 0.0600\xi - 0.2571\xi^2}$

σφάλμα $\|b - Ax_{LS}\|_2 = 0.0338$

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





With Netflix you can rent as many DVDs as you want and watch movies instantly on your PC for one low price (...) no late fees (...) no due dates, and DVD shipping free both ways. Plans start from only $4.99 (...) With our most popular plan, you can rent as many DVDs as you want (3 at-a-time) (...) all for just $17.99 a month +tax. ...

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 <user, movie, date of grade, grade>. 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".

Επιχειρηματικό Κίνητρο: Βραβείο Netflix $1,000,000 III

The BigChaos Solution to the Netflix (

Andreas Töscher and Michael Jahrer

commendo research & consulting

Neuer Weg 23, A-8580 Köflach, Austr

{*andreas.toescher,michael.jahrer*}*@comme*

Robert M. Bell*

AT&T Labs - Research

Florham Park, NJ

Input: A matrix **P** with all previous probe predictions. **P** always includes a constant predictor (a column with ones).

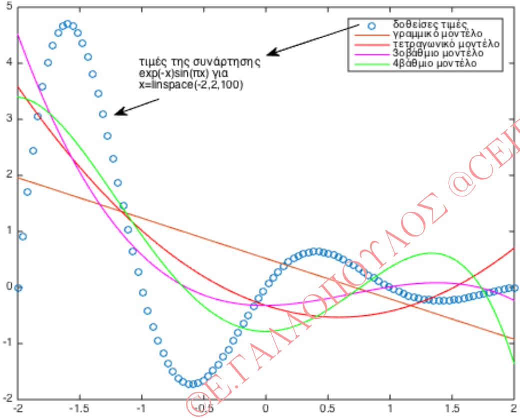
Exclude the Probe ratings **r** from the training set.
Initialize the weights.
$RMSE_{best} = \infty$
$RMSE_{epoch} = 1000$
$epochs = 0$
while $RMSE_{epoch} \leq RMSE_{best}$ **do**
  Train one epoch.
  if $RMSE_{epoch} \leq RMSE_{best}$ **then**
    $RMSE_{best} = RMSE_{epoch}$
    Save the current weights.
  end
  Predict the probe set $\hat{\mathbf{p}}$.
  Merge current probe prediction $\hat{\mathbf{p}}$ and previous predictions: $\mathbf{X} = [\mathbf{P}\ \hat{\mathbf{p}}]$
  Calculate blending weights: $\mathbf{w} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{r}$
  Calculate prediction of the current blend: $\mathbf{p} = \mathbf{X} \cdot \mathbf{w}$
  Calculate the RMSE of the blend: $RMSE_{epoch} = \sqrt{\frac{1}{R}\sum_{i=1}^{R}(p_i - r_i)^2}$; r_i is probe rating i, R is #ratings in the probe set
  $epochs = epochs + 1$
end
Load the weights.
Generate predictions for the probe set.
Insert the probe set into the training set.

Παραδείγματα προσέγγισης ελαχίστων τετραγώνων

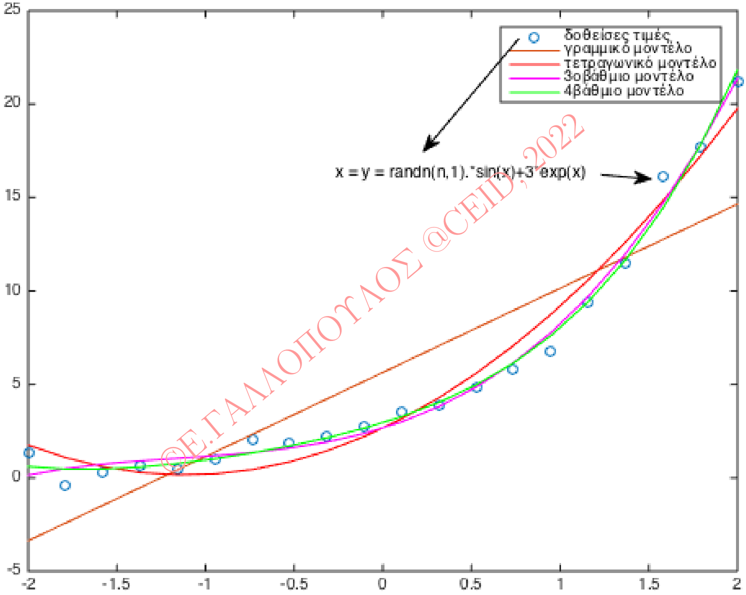
Χρησιμοποιούμε τον παρακάτω κώδικα MATLAB για διάφορες επιλογές $x, \ y$

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

n=100; x=linspace(-2,2,n)'; y = exp(-x).*sin(pi*x)



Σφάλματα: nerr(1:4) = 14.4664 12.3761 11.7957 11.0530 n=20; x = linspace(-2,2,n)'; y = randn(n,1).*sin(x)+3*exp(x)



Δοθέντων m ζευγών τιμών $\{(\xi_j, \psi_j)\}_{j=1}^{m}$ (θεωρούμε ότι τα ξ δεν είναι όλα ίσα μεταξύ τους) ενδιαφερόμαστε να υπολογίσουμε τις τιμές των παραμέτρων γ_1, γ_2 ώστε να ελαχιστοποιείται η συνάρτηση

$$E(\gamma_1, \gamma_2) = \sum_{j=1}^{m} (\psi_j - \gamma_2 \xi_j - \gamma_1)^2$$

Από βασική θεωρία, για να υπάρχει τοπικό ακρότατο, πρέπει να ικανοποιούνται οι ακόλουθες σχέσεις (κριτήριο 1ης παραγώγου):

$$\begin{aligned} 0 &= \frac{\partial E}{\partial \gamma_1} = 2\sum_{j=1}^{m} (\psi_j - \gamma 2\xi_j - \gamma_1)(-1) \\ 0 &= \frac{\partial E}{\partial \gamma_2} = 2\sum_{j=1}^{m} (\psi_j - \gamma_2\xi_j - \gamma_1)(-\xi_j) \end{aligned}$$

Τα τοπικά ακρότατα είναι στις τιμές $\tilde{\gamma}_1, \tilde{\gamma}_2$ που ικανοποιούν το σύστημα

$$\begin{aligned} \sum_{j=1}^{m} \psi_j &= m\gamma_1 + (\sum_{j=1}^{m} \xi_j)\gamma_2 \\ \sum^{m} \xi_j\psi_j &= (\sum^{m} \xi_j)\gamma_1 + (\sum^{m} \xi_j^2)\gamma_2 \end{aligned}$$

Αν η λύση είναι $(\tilde{\gamma}_1, \tilde{\gamma}_2)$, για να ελαχιστοποιεί το $E(\tilde{\gamma}_1, \tilde{\gamma}_2)$, θα πρέπει (κριτήριο 2ης παραγώγου)

$$\frac{\partial^2 E}{\partial \gamma_1^2} > 0 \text{ και } \Delta = \frac{\partial^2 E}{\partial \gamma_1^2}\frac{\partial^2 E}{\partial \gamma_2^2} - \frac{\partial^2 E}{\partial \gamma_1 \partial \gamma_2} > 0 \frac{\partial^2 E}{\partial \gamma_1^2} > 0.$$

Η πρώτη ανισότητα προφανώς ικανοποιείται:

$$\frac{\partial^2 E}{\partial \gamma_1^2} = 2m > 0$$

Σχετικά με τη δεύτερη: Υπολογίζουμε

$$\frac{\partial^2 E}{\partial \gamma_2^2} = 2\sum_{j=1}^{m} \xi_j^2, \quad \frac{\partial^2 E}{\partial \gamma_1 \partial \gamma_2} = 2\sum_{j=1}^{m} \xi_j.$$

άρα

$$\Delta > 0 \Leftrightarrow m\sum_{j=1}^{m} \xi_j^2 > (\sum_{j=1}^{m} \xi_j)^2.$$

Αλλά $m = e^\top e$, $x^\top x = \sum_{j=1}^m \xi_j^2$ και $e^\top x = \sum_{j=1}^m \xi$. Επομένως η ανισότητα

$$\sum_{j=1}^{m} \xi_j^2 > (\sum_{j=1}^{m} \xi_j)^2$$

μπορεί να γραφτεί ως

$$\|e\|^2\|x\|^2 > \langle e, x\rangle|^2.$$

Θυμηθείτε τώρα την ανισότητα Cauchy-Schwarz. Για κάθε $z, x, \in \mathbb{R}^m$

$$\langle z, x\rangle| \leq \|z\|\|x\|$$

όπου ισότητα ισχύει μόνον αν τα z, x είναι συγγραμμικά, δηλ. $z = \alpha x$ για κάποιο $\alpha \in \mathbb{R}$. Όμως στην περιπτωσή μας αυτό δεν μπορεί να συμβεί γιατί τότε θα ίσχυε ότι $x = \alpha e$, άρα $\xi_1 = \xi_2 = \cdots = \xi_m = \alpha$. Αυτό αποκλείεται αφού θεωρούμε ότι δεν είναι όλα τα ξ_j ίδια.
Είδαμε ότι η λύση του παρακάτω συστήματος ελαχιστοποιεί το $E(\gamma_1, \gamma_2)$:

$$\begin{pmatrix} m & \sum_{j=1}^m \xi_j \\ \sum_{j=1}^m \xi_j & \sum_{j=1}^m \xi_j^2 \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^m \psi_j \\ \sum_{j=1}^m \xi_j \psi_j \end{pmatrix}$$

Ενδιαφέρον: Αν θέσουμε

$$A = \begin{pmatrix} 1 & \xi_1 \\ \vdots & \vdots \\ 1 & \xi_m \end{pmatrix}, \quad y = \begin{pmatrix} \psi_1 \\ \vdots \\ \psi_m \end{pmatrix}$$

το παραπάνω είναι ίδιο με το σύστημα των κανονικών εξισώσεων

$$A^\top A c = A^\top y$$

που θα προέκυπτε απευθείας αν επιχειρούσαμε να λύσουμε το $Ax = y$ πολλαπλασιάζοντας πρώτα από τα αριστερά με το $A^\top$ ή αν χρησιμοποιούσαμε την ιδέα της προβολής του y στον υπόχωρο στηλών του A.

Προσοχή: Η τάξη $\operatorname{rank}(A^\top A) \leq 2 = \min(m, n)$ επομένως το σύστημα θα έχει μοναδική λύση ανν υπάρχουν τουλάχιστον δύο i, j τ.ώ. $\xi_i \neq \xi_j$.

Πρακτική επίλυση του προβλήματος ελαχίστων τετραγώνων

- Στο μάθημα αυτού του εξαμήνου, η συζήτηση σχετικά με τον τρόπο επίλυσης θα είναι πολύ συνοπτική και εισαγωγική.
- Η επίλυση μπορεί να επιτευχθεί με διάφορους τρόπους που θεωρητικά παράγουν το ίδιο αποτέλεσμα αλλά πρακτικά μπορεί να έχουν πολύ διαφορετικά χαρακτηριστικά όσον αφορά το κόστος και την ακρίβεια, δεδομένου ότι υλοποιούνται στην αριθμητική του υπολογιστή.
- Επιγραμματικά, μπορούμε να χρησιμοποιήσουμε:
  - Σχηματισμός και επίλυση των κανονικών εξισώσεων, π.χ. με παραγοντοποίηση (Cholesky, δηλ. ειδική εκδοχή της LU για μητρώα που είναι ΣΘΟ).
  - Χρησιμοποιούμε παραγοντοποίηση QR του , π.χ. από τη διαδικασία Gram-Schmidt επί των στηλών του A.
    Πρόκειται για την παραγοντοποίηση του $A = QR$ όπου το μητρώο Q είναι ορθογώνιο ή έχει ορθογώνιες στήλες και το μητρώο R είναι άνω τριγωνικό.
  - Οι επόμενες διαφάνειες αναφέρονται στα ορθογώνια μητρώα. Στην επόμενη διάλεξη θα μιλήσουμε για την QR και την υλοποίηση μέσω Gram-Schmidt.

3. Ορθογώνια μητρώα

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

Αν $\mathbf{c}_1$ και $\mathbf{c}_2$ είναι οι στήλες του A, τότε

$$\mathbf{c}_1 \cdot \mathbf{c}_2 = 0(-1) + 1(0) = 0, \qquad \text{και} \qquad \|\mathbf{c}_1\| = \|\mathbf{c}_2\| = 1.$$

Αν $\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_n$ είναι ορθοκανονικά διανύσματα στον $\mathbb{R}^n$, τότε $\mathbf{q}_i \cdot \mathbf{q}_j = \delta_{ij}$, κατά συνέπεια αν $Q = \begin{pmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \end{pmatrix}$, τότε $Q^{\mathrm{T}} Q = I$. Επειδή το μητρώο Q είναι τετραγωνικό έπεται ότι $Q^{\mathrm{T}} = Q^{-1}$.

Ορισμός (2)

Ένα τετραγωνικό μητρώο Q λέγεται **ορθογώνιο** (orthogonal) αν $Q^{-1} = Q^{\mathrm{T}}$.

Σημειώνουμε ότι κάθε ορθογώνιο 2×2 μητρώο είναι της μορφής

$$Q_\omega^+ = \begin{pmatrix} \cos\omega & -\sin\omega \\ \sin\omega & \cos\omega \end{pmatrix}, \qquad \text{ή} \qquad Q_\omega^- = \begin{pmatrix} \cos\omega & \sin\omega \\ \sin\omega & -\cos\omega \end{pmatrix}, \tag{10}$$

για κατάλληλο $\omega \in \mathbb{R}$.

Παράδειγμα (2 Περιστροφή στο επίπεδο)

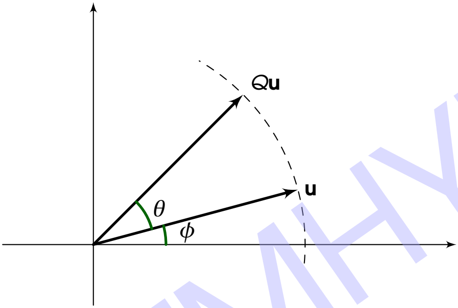
Το μητρώο

$$Q = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

είναι ορθογώνιο μητρώο. Ας δούμε το αποτέλεσμα της δράσης του Q επί ενός μοναδιαίου διανύσματος του $\mathbb{R}^2$, $\mathbf{u} = \begin{pmatrix} \cos\phi & \sin\phi \end{pmatrix}^{\mathrm{T}}$. Υπολογίζοντας

$$\begin{aligned} Q\begin{pmatrix} \cos\phi \\ \sin\phi \end{pmatrix} &= \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} \cos\phi \\ \sin\phi \end{pmatrix} \\ &= \begin{pmatrix} \cos\theta\cos\phi - \sin\theta\sin\phi \\ \sin\theta\cos\phi + \cos\theta\sin\phi \end{pmatrix} \\ &= \begin{pmatrix} \cos(\phi+\theta) \\ \sin(\phi+\theta) \end{pmatrix} \end{aligned}$$

βλέπουμε ότι ο πολλαπλασιασμός επί Q έχει σαν αποτέλεσμα την περιστροφή του μοναδιαίου διανύσματος, άρα και κάθε διανύσματος, κατά γωνία θ. Το αποτέλεσμα αποτυπώνεται στο Σχήμα που ακολουθεί



Άσκηση 1.

Δείξτε ότι το μητρώο

$$Q = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

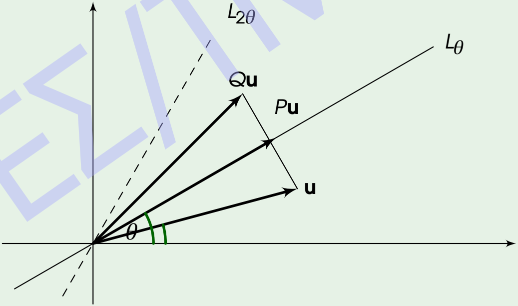
(α΄) Είναι ορθογώνιο.

(β΄) Περιστρέφει κάθε διάνυσμα στο $\mathbb{R}^3$ γύρω από τον z-άξονα κατά γωνία θ.

Παράδειγμα (3 Ανάκλαση στο επίπεδο)

Αν $\mathbf{q} = \begin{pmatrix}\cos\theta & \sin\theta\end{pmatrix}^{\mathrm{T}}$ είναι ένα μοναδιαίο διάνυσμα και L_θ η ευθεία που το περιέχει θέλουμε να βρούμε το μητρώο Q το οποίο υλοποιεί την ανάκλαση οποιουδήποτε διανύσματος ως προς την ευθεία.

Αν $\mathbf{u}$ είναι τυχαίο διάνυσμα στο επίπεδο, $Q\mathbf{u}$ είναι η ανάκλασή του ως προς την ευθεία L_θ, και $P\mathbf{u}$ η προβολή του επί της ευθείας L_θ, τότε, "μεταφράζοντας" μετρικές σχέσεις του ισοσκελούς τριγώνου με "πλευρές" τα διανύσματα $\mathbf{u}$ και $Q\mathbf{u}$ στο Σχήμα παίρνουμε



Παράδειγμα (Ανάκλαση στο επίπεδο (συνέχεια))

$$P\mathbf{u} - \mathbf{u} = \frac{1}{2}(Q\mathbf{u} - \mathbf{u}) \Rightarrow Q\mathbf{u} = 2P\mathbf{u} - \mathbf{u}.$$

άρα $Q = 2P - I$, ή μέσω της $P = \mathbf{q}\mathbf{q}^{\mathrm{T}}/\mathbf{q}^{\mathrm{T}}\mathbf{q}$, με $\|\mathbf{q}\| = 1$

$$Q = 2\mathbf{q}\mathbf{q}^{\mathrm{T}} - I. \tag{11}$$

Αναλυτικά το μητρώο Q είναι

$$\begin{aligned} Q &= 2\begin{pmatrix} \cos^2\theta & \sin\theta\cos\theta \\ \sin\theta\cos\theta & \sin^2\theta \end{pmatrix} - I \\ &= \begin{pmatrix} 2\cos^2\theta - 1 & 2\sin\theta\cos\theta \\ 2\sin\theta\cos\theta & 2\sin^2\theta - 1 \end{pmatrix} \\ &= \begin{pmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{pmatrix} \end{aligned}$$

του οποίου οι στήλες είναι ορθοκανονικά διανύσματα στο $\mathbb{R}^2$, επομένως είναι ένα ορθογώνιο μητρώο.

Παράδειγμα (Ανάκλαση στο επίπεδο (συνέχεια))

Η δράση του Q επί ενός μοναδιαίου διανύσματος $\begin{pmatrix}\cos\phi & \sin\phi\end{pmatrix}^{\mathrm{T}}$ έχει σαν αποτέλεσμα το διάνυσμα

$$Q\begin{pmatrix}\cos\phi\\ \sin\phi\end{pmatrix} = \begin{pmatrix}\cos 2\theta & \sin 2\theta\\ \sin 2\theta & -\cos 2\theta\end{pmatrix}\begin{pmatrix}\cos\phi\\ \sin\phi\end{pmatrix}$$
$$= \begin{pmatrix}\cos 2\theta\cos\phi + \sin 2\theta\sin\phi\\ \sin 2\theta\cos\phi - \cos 2\theta\sin\phi\end{pmatrix}$$
$$= \begin{pmatrix}\cos(2\theta-\phi)\\ \sin(2\theta-\phi)\end{pmatrix}$$

το οποίο συμφωνεί με την εικόνα του σχήματος ότι το αρχικό διάνυσμα, ας πούμε, $\mathbf{u} = \begin{pmatrix}\cos\phi & \sin\phi\end{pmatrix}^{\mathrm{T}}$ περιστρέφεται κατά 2θ αρχικά και στη συνέχεια κατά $-\phi$, δηλαδή κατά ϕ με αντίθετη φορά, προκειμένου να προκύψει τελικά το $Q\mathbf{u}$.

Παράδειγμα (4 Μετάθεση)

Κάθε μητρώο μετάθεσης προκύπτει από το ταυτοτικό μητρώο μεταθέτοντας γραμμές του, κατά συνέπεια είναι ορθογώνιο. Στο $\mathbb{R}^2$ τα δύο (γιατί;) μητρώα μετάθεσης είναι

$$Q_0^+ = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \text{και} \qquad Q_{\pi/2}^- = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

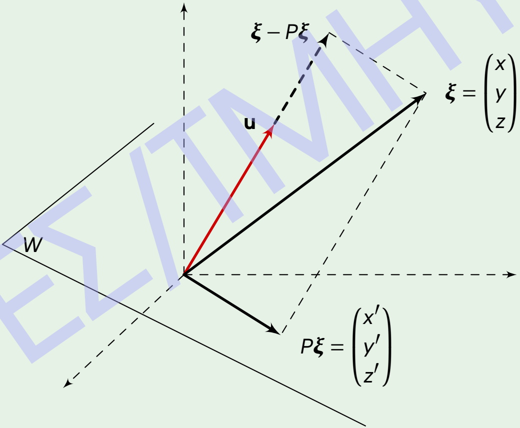
Θεώρημα (2)

Αν Q είναι ένα $n \times n$ μητρώο, οι παρακάτω προτάσεις είναι ισοδύναμες

1. *Το Q είναι ορθογώνιο.*
2. *Το Q διατηρεί το μήκος, δηλαδή για κάθε $\mathbf{x} \in \mathbb{R}^n$ ισχύει $\|Q\mathbf{x}\| = \|\mathbf{x}\|$.*
3. *Το Q διατηρεί το εσωτερικό γινόμενο, δηλαδή για $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ ισχύει $\langle Q\mathbf{x}, Q\mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle$.*

Παράδειγμα (5)

Θεωρούμε το επίπεδο W με εξίσωση $ax + by + cz = 0$ στο $\mathbb{R}^3$. Να βρεθεί η προβολή του τυχαίου διανύσματος $\xi \in \mathbb{R}^3$ επί του W, καθώς και το μητρώο προβολής επί του W.



Παράδειγμα (συνέχεια)

Το $\mathbf{u} = \begin{pmatrix} a & b & c \end{pmatrix}^{\mathrm{T}}$ είναι ορθογώνιο στο W, επιπλέον διαιρώντας την εξίσωση με $(a^2+b^2+c^2)^{1/2}$ μπορούμε να υποθέσουμε ότι το $\mathbf{u}$ είναι μοναδιαίο. Αν $P\boldsymbol{\xi}$ είναι η προβολή του ξ επί του W, βλέπε Σχήμα το διάνυσμα $\boldsymbol{\xi} - P\boldsymbol{\xi}$ είναι η προβολή του ξ επί της ευθείας δια του $\mathbf{u}$, κατά συνέπεια

$$\begin{aligned} \boldsymbol{\xi} - P\boldsymbol{\xi} &= \langle \boldsymbol{\xi}, \mathbf{u} \rangle \mathbf{u} \\ \Rightarrow \; P\boldsymbol{\xi} &= \boldsymbol{\xi} - \langle \boldsymbol{\xi}, \mathbf{u} \rangle \mathbf{u}. \end{aligned}$$

Για δε το μητρώο της προβολής έχουμε

$$\begin{aligned} P\boldsymbol{\xi} &= I\boldsymbol{\xi} - \mathbf{u}\langle \boldsymbol{\xi}, \mathbf{u} \rangle \\ &= I\boldsymbol{\xi} - \mathbf{u}\mathbf{u}^{\mathrm{T}}\boldsymbol{\xi} \\ &= (I - \mathbf{u}\mathbf{u}^{\mathrm{T}})\boldsymbol{\xi}, \end{aligned}$$

ισοδύναμα

$$P = I - \mathbf{u}\mathbf{u}^{\mathrm{T}}.$$