ΓΡΑΜΜΙΚΗ ΑΛΓΕΒΡΑ

Διάλεξη (X) 19-20

Ορθογώνιες προβολές
Βέλτιστη προσέγγιση

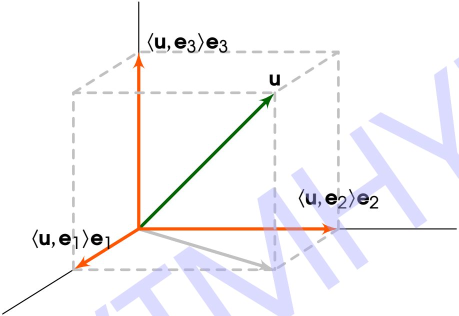
Λύσεις ελαχίστων τετραγώνων (εισαγωγή)

Ε. Γαλλόπουλος & Ε. Στεφανόπουλος

Τμήμα Μηχανικών Η/Υ & Πληροφορικής, Πανεπιστήμιο Πατρών

1 & 4 Δεκεμβρίου 2022

1. Προβολές



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

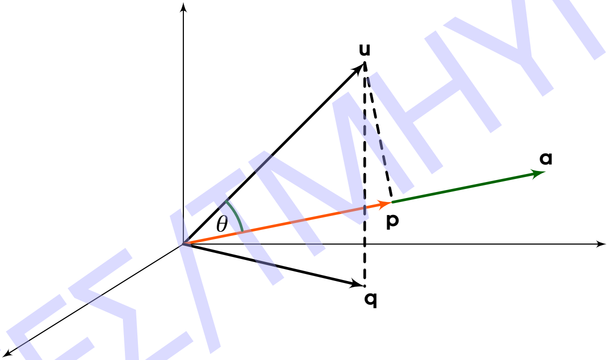
$$\mathbf{u} = \langle \mathbf{u}, \mathbf{e}_1 \rangle \mathbf{e}_1 + \langle \mathbf{u}, \mathbf{e}_2 \rangle \mathbf{e}_2 + \langle \mathbf{u}, \mathbf{e}_3 \rangle \mathbf{e}_3,$$

ονομάζουμε τα διανύσματα

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

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

Αν **a** είναι ένα σταθερό μη μηδενικό διάνυσμα στον $\mathbb{R}^3$ και **u** είναι ένα τυχαίο διάνυσμα στον $\mathbb{R}^3$ μας ενδιαφέρει να βρούμε την προβολή του **u** επί της ευθείας η οποία περιέχει το **a**.



Αν θ είναι η γωνία μεταξύ των **a** και **u** και **p** είναι η προβολή του **u** επί της ευθείας δια του **a**, τότε

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

κατά συνέπεια

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

επομένως από τον ορισμό του συνήθους εσωτερικού γινομένου προκύπτει ότι

$$\mathbf{p} = \frac{\mathbf{a} \cdot \mathbf{u}}{\mathbf{a} \cdot \mathbf{a}} \mathbf{a} = \frac{\mathbf{a}^{\mathrm{T}} \mathbf{u}}{\mathbf{a}^{\mathrm{T}} \mathbf{a}} \mathbf{a} \tag{1}$$

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

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

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

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

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

Το μητρώο της ορθογώνιας προβολής

Η εξοικείωσή μας με τα στοιχειώδη μητρώα επιτρέπει να δούμε, αναφερόμαστε στο αρχικό παράδειγμα, όπου $\mathbf{u} = \begin{pmatrix} u_1 & u_2 & u_3 \end{pmatrix}^{\mathrm{T}}$, ότι

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

ή

$$\mathbf{p}_1 = \begin{pmatrix} \mathbf{e}_1 & \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{u} = P_1\mathbf{u}, \qquad \mathbf{p}_2 = \begin{pmatrix} \mathbf{0} & \mathbf{e}_2 & \mathbf{0} \end{pmatrix}\mathbf{u} = P_2\mathbf{u}, \qquad \mathbf{p}_3 = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{e}_3 \end{pmatrix}\mathbf{u} = P_3\mathbf{u}$$

και

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

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

1. Αναζητάμε μητρώο P ώστε $P\mathbf{u} = \mathbf{p} = \mathrm{proj}_{\mathbf{a}}\mathbf{u}$, ισοδύναμα για ποιο P είναι
$$P\mathbf{u} = \frac{\mathbf{a}^{\mathrm{T}}\mathbf{u}}{\mathbf{a}^{\mathrm{T}}\mathbf{a}}\mathbf{a}?$$
Επειδή $(\mathbf{a}^{\mathrm{T}}\mathbf{u})\mathbf{a} = \mathbf{a}(\mathbf{a}^{\mathrm{T}}\mathbf{u}) = (\mathbf{a}\mathbf{a}^{\mathrm{T}})\mathbf{u}$ (γιατί;), τελικά βρίσκουμε
$$\mathrm{proj}_{\mathbf{a}}\mathbf{u} = P\mathbf{u} = \frac{\mathbf{a}\mathbf{a}^{\mathrm{T}}}{\mathbf{a}^{\mathrm{T}}\mathbf{a}}\mathbf{u}. \qquad (2)$$

2. Έστω $\mathbf{a}_1, \mathbf{a}_2 \ldots, \mathbf{a}_n$ γραμμικά ανεξάρτητα διανύσματα στο $\mathbb{R}^m$ και έστω $W = \mathrm{span}\{\mathbf{a}_1, \mathbf{a}_2 \ldots, \mathbf{a}_n\}$. Αν A είναι το μητρώο με στήλες τα $\mathbf{a}_j$, τότε $\mathrm{range}\, A = W$, δηλαδή ο υπόχωρος W είναι ο χώρος στηλών του A. Έστω $\mathbf{c} \in \mathbb{R}^m$ τέτοιο ώστε $\mathrm{proj}_W \mathbf{u} = A\mathbf{c}$. Επειδή το $\mathbf{u} - A\mathbf{c}$ είναι ορθογώνιο στον W έπεται ότι $\mathbf{a}_j^{\mathrm{T}}(\mathbf{u} - A\mathbf{c}) = 0$ για $j = 1, 2, \ldots, n$. Κατά συνέπεια
$$A^{\mathrm{T}}(\mathbf{u} - A\mathbf{c}) = 0 \Rightarrow A^{\mathrm{T}}\mathbf{u} = A^{\mathrm{T}}A\mathbf{c}$$
Το μητρώο $A^{\mathrm{T}}A$ είναι αντιστρέψιμο, βλέπε επόμενο Θεώρημα, επομένως
$$(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\mathbf{u} = \mathbf{c} \Rightarrow A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\mathbf{u} = A\mathbf{c}$$
κατά συνέπεια
$$\mathrm{proj}_W \mathbf{u} = P\mathbf{u} = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\mathbf{u}. \qquad (3)$$

Θεώρημα 1

Εάν το $m \times n$ μητρώο A έχει γραμμικά ανεξάρτητες στήλες, τότε το μητρώο $A^{\mathrm{T}}A$ είναι αντιστρέψιμο.

Απόδειξη.

Το μητρώο $A^{\mathrm{T}}A$ είναι τετραγωνικό ($n \times n$), οπότε αρκεί να δείξουμε ότι $\operatorname{null} A^{\mathrm{T}}A = \{\mathbf{0}\}$. Πρώτα δείχνουμε ότι τα μητρώα A και $A^{\mathrm{T}}A$ έχουν τον ίδιο μηδενόχωρο, δηλαδή $\operatorname{null} A = \operatorname{null} A^{\mathrm{T}}A$. Πράγματι από τη σχέση

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

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

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

έπεται ότι $\operatorname{null} A^{\mathrm{T}}A \subseteq \operatorname{null} A$. Στη συνέχεια δείχνουμε ότι $\operatorname{null} A = \{\mathbf{0}\}$. Πράγματι αν $A\mathbf{x} = \mathbf{0}$ το διάνυσμα $A\mathbf{x}$ ως γραμμικός συνδυασμός των γραμμικά ανεξαρτήτων στηλών του A θα είναι το μηδενικό διάνυσμα, συνεπώς $\mathbf{x} = \mathbf{0}$. Έτσι τελικά έχουμε ότι $\operatorname{null} A^{\mathrm{T}}A = \operatorname{null} A = \{\mathbf{0}\}$ που είναι το ζητούμενο. $\square$

Παράδειγμα 1

Το μητρώο

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

έχει γραμμικά ανεξάρτητες στήλες. Υπολογίζοντας

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

βλέπουμε, όπως εξασφαλίζει το Θεώρημα 1, ότι το $A^{\mathrm{T}} A$ είναι αντιστρέψιμο και

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

Αντίθετα το 3×3 μητρώο AA^{T} δεν αντιστρέφεται αφού $\operatorname{rank}(AA^{\mathrm{T}}) \leq \operatorname{rank} A = 2$. Έτσι το μητρώο

παράδειγμα συνέχεια

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

είναι η προβολή επί του χώρου στηλών του A, ισοδύναμα για κάθε $\mathbf{u} \in \mathbb{R}^3$, έστω $\mathbf{u} = \begin{pmatrix} x & y & z \end{pmatrix}^{\mathrm{T}}$, 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

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

Παρατήρηση 1

Εάν P είναι μητρώο ορθογώνιας προβολής στο $\mathbb{R}^m$ επί υπόχωρου W, τότε από την (2) ή την (3), ανάλογα την περίπτωση, προκύπτει ότι το P είναι συμμετρικό, δηλαδή $P^{\mathrm{T}} = P$.

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 μητρώο, είναι επίσης μητρώο προβολής. Προβάλει κάθε διάνυσμα επί του ορθογωνίου 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 συνέχεια

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

οπότε $\langle P\mathbf{u}, (I-P)\mathbf{v}\rangle = 0$. Έτσι μπορούμε να γράψουμε (γιατί;)

1. $\operatorname{range} P = W$ και
2. $\operatorname{null} P = \operatorname{range}(I-P)$.
3. $\mathbb{R}^m = \operatorname{range} P \oplus \operatorname{null} P$ και
4. $\operatorname{range} P \perp \operatorname{null} P$.

Σημείωση:

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

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

ικανοποιεί την σχέση $P_\lambda^2 = P_\lambda$ κατά συνέπεια είναι μητρώο προβολής, αλλά $P_\lambda^{\mathrm{T}} \neq P_\lambda$ 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

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

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

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

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

αν και μόνο αν $\lambda = 0$. Συνεπώς ο μηδενόχωρος $\operatorname{null} P$ δεν είναι το ορθογώνιο συμπλήρωμα του $\operatorname{range} P$ αν $\lambda \neq 0$.

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

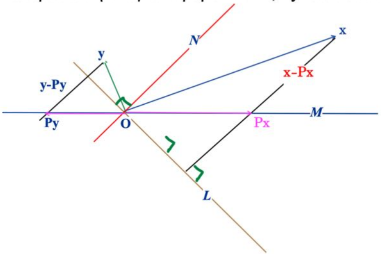
Θεώρημα 2

Μια προβολή P, ένα μητρώο δηλαδή με την ιδιότητα $P^2 = P$, είναι ορθογώνια αν και μόνο αν $P^{\mathrm{T}} = P$.

Αν το μητρώο δεν είναι συμμετρικό, η προβολή λέγεται πλάγια. Στις πλάγιες προβολές, πρέπει να οριστεί και ένας επιπλέον υπόχωρος.

Συγκεκριμένα, η πλάγια προβολή σε έναν υπόχωρο M θα είναι παράλληλη σε έναν υπόχωρο N και ορθογώνια επί ενός υπόχωρου L.

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



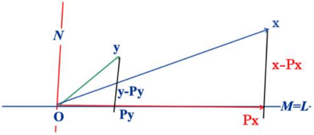
Παράδειγμα για τα σημεία x, y:

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 $\mathcal{S} := \mathrm{span}\{e_3\}$ και $\mathcal{T} := \mathrm{span}\{e_1, e_2\}$. Τότε αν

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

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 $x = (1, 2, -1)^\top$

$$\begin{aligned}
\Rightarrow x_{\mathcal{S}} = P_{\mathcal{S}} x, \text{ όπου } P_{\mathcal{S}} &:= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = e_3 (e_3^\top e_3)^{-1} e_3^\top \\
\Rightarrow x_{\mathcal{T}} = P_{\mathcal{T}} x, \text{ όπου } P_{\mathcal{T}} &:= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \\
&= \begin{pmatrix} e_1 & e_2 \end{pmatrix} \left(\begin{pmatrix} e_1^\top \\ e_2^\top \end{pmatrix} \begin{pmatrix} e_1 & e_2 \end{pmatrix} \right)^{-1} \begin{pmatrix} e_1^\top \\ e_2^\top \end{pmatrix}
\end{aligned}$$

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 $\mathcal{M} = \text{span}\{u\}$.

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

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

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

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

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

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

Παρατηρήσεις: 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

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

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

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

$\mathcal{V} = \mathrm{span}\{e_1, e_3\}$ στον $\mathbb{R}^4$ 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 ορθογώνιο μητρώο προβολής επί ενός υπόχωρου 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 ίδιου υπόχωρου). Επομένως υπάρχει αντιστρέψιμο μητρώο M τέτοιο ώστε

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

Άρα αν γράψουμε

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

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 λέγεται 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!!!

2. Βέλτιστη προσέγγιση

Ορισμός 1

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

$$\mathrm{proj}_W\,\mathbf{u} := \langle\mathbf{u},\mathbf{w}_1\rangle\mathbf{w}_1 + \langle\mathbf{u},\mathbf{w}_2\rangle\mathbf{w}_2 + \cdots + \langle\mathbf{u},\mathbf{w}_n\rangle\mathbf{w}_n. \tag{4}$$

Το διάνυσμα

$$\mathbf{u} - \mathrm{proj}_W\,\mathbf{u} = \mathbf{u} - \langle\mathbf{u},\mathbf{w}_1\rangle\mathbf{w}_1 - \langle\mathbf{u},\mathbf{w}_2\rangle\mathbf{w}_2 - \cdots - \langle\mathbf{u},\mathbf{w}_n\rangle\mathbf{w}_n \tag{5}$$

θα το λέμε **ορθογώνια στο** W **συνιστώσα του u**.

Ο ορισμός του $\mathbf{u} - \mathrm{proj}_W\,\mathbf{u}$ ως ορθογώνια στο W συνιστώσα του $\mathbf{u}$ διακαιολογείται από το γεγονός ότι γράφοντας

$$\mathbf{u} = \mathrm{proj}_W\,\mathbf{u} + \mathbf{u} - \mathrm{proj}_W\,\mathbf{u}$$

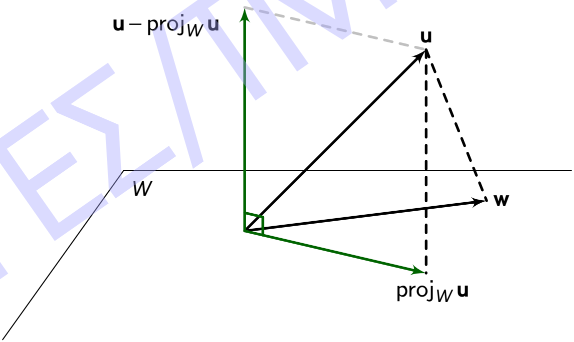
ισχύει

Θεώρημα 3 (Βέλτιστη προσέγγιση)

Εάν W είναι ένας πεπερασμένης διάστασης υπόχωρος 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 $\mathbf{w} \neq \operatorname{proj}_W \mathbf{u}$ ισχύει

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

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



Απόδειξη.

Γράφοντας

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

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

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

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

$$\|\mathbf{u} - \mathbf{w}\|^2 > \|\mathbf{u} - \mathrm{proj}_W \mathbf{u}\|^2.$$

απ' όπου έπεται το ζητούμενο. $\square$

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 $\{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ είναι ένα ορθοκανονικό σύνολο διανυσμάτων του X, τότε για κάθε διάνυσμα $\mathbf{u}$ του X ισχύει

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

Πράγματι υπολογίζοντας

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

παρατήρηση 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

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

και αυτό συμβαίνει όταν

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

Συνεπώς

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

Παράδειγμα 2

Στο χώρο των συνεχών συναρτήσεων $\mathscr{C}[-1,1]$, με εσωτερικό γινόμενο

$$\langle f, g\rangle = \int_{-1}^{1} f(x)g(x)\,\mathrm{d}x \tag{6}$$

να βρεθεί η βέλτιστη προσέγγιση της $f(x) = x^3$ από πολυώνυμα βαθμού το πολύ δύο.

Η βέλτιστη προσέγγιση της f είναι η προβολή της f επί του υπόχωρου $\mathbb{P}_2[-1,1]$. Μια βάση για τον $\mathbb{P}_2$ είναι η συνήθης, $\mathscr{B} = \{1, x, x^2\}$, αλλά αυτή δεν είναι ορθοκανονική.

Βήμα 1. Βρίσκουμε μια ορθοκανονική, ως προς το εσωτερικό γινόμενο (6), βάση στον $\mathbb{P}_2$ εφαρμόζοντας τον αλγόριθμο Gram-Schmidt στη συνήθη βάση $\mathscr{B} = \{1, x, x^2\}$.

$$q_0 = 1$$

$$\|q_0\|^2 = \int_{-1}^{1} 1\,\mathrm{d}x = 2 \qquad p_0(x) = \frac{1}{\sqrt{2}}$$

παράδειγμα συνέχεια

$$q_1(x) = x - \langle x, p_0 \rangle p_0(x),$$

$$\langle x, p_0 \rangle = \int_{-1}^{1} x \frac{1}{\sqrt{2}} \, \mathrm{d}x = 0$$

$$\|q_1\|^2 = \int_{-1}^{1} x^2 \, \mathrm{d}x = \frac{2}{3} \qquad p_1(x) = \sqrt{\frac{3}{2}} x$$

$$q_2(x) = x^2 - \langle x^2, p_0 \rangle p_0(x) - \langle x^2, p_1 \rangle p_1(x)$$

$$\langle x^2, p_0 \rangle = \int_{-1}^{1} x^2 \frac{1}{\sqrt{2}} \, \mathrm{d}x = \frac{\sqrt{2}}{3}$$

$$\langle x^2, p_1 \rangle = \int_{-1}^{1} x^2 \sqrt{\frac{3}{2}} x \, \mathrm{d}x = 0$$

$$\|q_2\|^2 = \int_{-1}^{1} \left(x^2 - \frac{1}{3}\right)^2 \mathrm{d}x = \frac{8}{45} \qquad p_2(x) = \sqrt{\frac{45}{8}} \left(x^2 - \frac{1}{3}\right).$$

Έτσι μια ορθοκανονική βάση για το $\mathbb{P}_2[-1,1]$ αποτελείται από τα πολυώνυμα p_0, p_1 και p_2 τα οποία είναι τα τρία πρώτα πολυώνυμα Legendre.

παράδειγμα συνέχεια

Βήμα 2. Η βέλτιστη προσέγγιση της f από πολυώνυμο βαθμού το πολύ δύο είναι

$$\operatorname{proj}_{\mathbb{P}_2} f = \langle f, p_0\rangle p_0 + \langle f, p_1\rangle p_1 + \langle f, p_2\rangle p_2.$$

Υπολογίζουμε λοιπόν

$$\langle f, p_0\rangle = \int_{-1}^{1} x^3 \frac{1}{\sqrt{2}}\,\mathrm{d}x = 0$$

$$\langle f, p_1\rangle = \int_{-1}^{1} x^3 \sqrt{\frac{3}{2}}x\,\mathrm{d}x = \sqrt{\frac{3}{2}}\int_{-1}^{1} x^4\,\mathrm{d}x = \frac{\sqrt{6}}{5}$$

$$\langle f, p_2\rangle = \int_{-1}^{1} x^3 \sqrt{\frac{45}{8}}\left(x^2 - \frac{1}{3}\right)\mathrm{d}x = \sqrt{\frac{45}{8}}\int_{-1}^{1}\left(x^5 - \frac{1}{3}x^3\right)\mathrm{d}x = 0$$

επομένως

$$\operatorname{proj}_{\mathbb{P}_2} f(x) = \frac{\sqrt{6}}{5}\sqrt{\frac{3}{2}}x = \frac{3}{5}x,$$

είναι η βέλτιστη προσέγγιση της $f(x) = x^3$ από πολυώνυμα του $\mathbb{P}_2[-1, 1]$.

Προαιρετικό:

Στο τελευταίο Παράδειγμα δείξαμε ότι

$$\min_{a,b,c} \|x^3 - a - bx - cx^2\| = \left\| x^3 - \frac{3}{5}x \right\|.$$

Δείχνουμε ότι το αποτέλεσμα αυτό μπορεί να προκύψει εφαρμόζοντας μεθόδους του Απειροστικού Λογισμού. Θεωρούμε την συνάρτηση

$$E(a,b,c) = \|x^3 - a - bx - cx^2\|^2 = \int_{-1}^{1} (x^3 - a - bx - cx^2)^2 \,\mathrm{d}x.$$

Για την εύρεση του ελαχίστου της E βρίσκουμε τα κρίσιμα σημεία της E, εδώ τα σημεία (a,b,c) για τα οποία

$$\frac{\partial E}{\partial a} = \frac{\partial E}{\partial b} = \frac{\partial E}{\partial c} = 0.$$

Χρησιμοποιώντας το αποτέλεσμα της Ανάλυσης(*)

$$\frac{\partial}{\partial \lambda} \int_{-1}^{1} (x^3 - a - bx - cx^2)^2 \,\mathrm{d}x = \int_{-1}^{1} \frac{\partial}{\partial \lambda} (x^3 - a - bx - cx^2)^2 \,\mathrm{d}x \tag{7}$$

όπου $\lambda \in \{a, b, c\}$, και επιλύοντας το σύστημα

$$\frac{\partial E}{\partial a} = \int_{-1}^{1} -2(x^3 - a - bx - cx^2)\,\mathrm{d}x = 2a\int_{-1}^{1} -2a\,\mathrm{d}x + 2c\int_{-1}^{1} x^2\,\mathrm{d}x = 4a + \frac{4}{3}c = 0$$
$$\frac{\partial E}{\partial b} = \int_{-1}^{1} -2x(x^3 - a - bx - cx^2)\,\mathrm{d}x = -2\int_{-1}^{1} x^4\,\mathrm{d}x + 2b\int_{-1}^{1} x^2\,\mathrm{d}x = -\frac{4}{5} + \frac{4}{5}b = 0$$
$$\frac{\partial E}{\partial c} = \int_{-1}^{1} -2x^2(x^3 - a - bx - cx^2)\,\mathrm{d}x = 2a\int_{-1}^{1} x^2\,\mathrm{d}x + 2c\int_{-1}^{1} x^4\,\mathrm{d}x = \frac{4}{3}a + \frac{4}{5}c = 0$$

βρίσκουμε $a = 0$, $b = 3/5$, και $c = 0$. Υπάρχει λοιπόν ένα κρίσιμο σημείο και επειδή

$$E(0,0,0) = \int_{-1}^{1} x^6\,\mathrm{d}x = \frac{2}{7}$$
$$E(0,3/5,0) = \int_{-1}^{1}\left(x^3 - \frac{3}{5}x\right)^2\mathrm{d}x = \int_{-1}^{1}\left(x^6 - \frac{6}{5}x^4 + \frac{9}{25}x^2\right)\mathrm{d}x = \frac{2}{7} - \frac{6}{25}$$

έπεται ότι στο $(0,3/5,0)$ η συνάρτηση E παίρνει την ελάχιστη τιμή της, όπως εξάλλου είδαμε στο Παράδειγμα.

Σημειώσεις:

- (*) **Θεώρημα.** *Έστω ότι το ολοκλήρωμα*

$$F(t) = \int_a^b \phi(x,t)\,\mathrm{d}x$$

υπάρχει για κάθε $t \in [p,q]$. *Αν η μερική παράγωγος* $\partial\phi/\partial t$ *είναι συνεχής στο* $Q = \{(x,t) : a \le x \le b,\ p \le t \le q\}$, *τότε η* $F'(t)$ *υπάρχει για κάθε* $t \in (p,q)$ *επιπλέον*

$$F'(t) = \int_a^b \frac{\partial}{\partial t}\phi(x,t)\,\mathrm{d}x.$$

- Αν δυσπιστείτε για την ισχύ της (7) υπολογίστε

$$\int_{-1}^{1} (x^3 - a - bx - cx^2)^2\,\mathrm{d}x = \cdots$$

$$= \frac{2}{7} + 2a^2 + \frac{2}{3}b^2 + \frac{2}{5}c^2 + \frac{4}{3}ac - \frac{4}{5}b.$$

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". Έχουμε

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

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

$$A\mathbf{x} - \mathbf{b} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

έχουμε ότι

$$\|A\mathbf{x} - \mathbf{b}\|^2 = e_1^2 + e_2^2 + \cdots + e_n^2,$$

κατά συνέπεια το διάνυσμα που ελαχιστοποιεί την $\|A\mathbf{x} - \mathbf{b}\|$ ελαχιστοποιεί το άθροισμα $e_1^2 + e_2^2 + \cdot + e_n^2$. Ένα τέτοιο διάνυσμα θα το λέμε **λύση των ελαχίστων τετραγώνων** και την ποσότητα $\|A\mathbf{x} - \mathbf{b}\|$ θα τη λέμε **σφάλμα των ελαχίστων τετραγώνων**.

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

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

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

έτσι

$$\mathbf{b} - A\mathbf{x} = \operatorname{proj}_W \mathbf{b} - A\mathbf{x} + \mathbf{b} - \operatorname{proj}_W \mathbf{b}, \qquad \& \qquad \operatorname{proj}_W \mathbf{b} - A\mathbf{x} \perp \mathbf{b} - \operatorname{proj}_W \mathbf{b}, \quad (8)$$

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

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

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

$$A\widehat{\mathbf{x}} = \operatorname{proj}_W \mathbf{b}. \quad (9)$$

Έτσι από την (8) παίρνουμε

$$\mathbf{b} - A\widehat{\mathbf{x}} = \mathbf{b} - \operatorname{proj}_W \mathbf{b}, \qquad \text{και} \qquad \mathbf{b} - \operatorname{proj}_W \mathbf{b} \in W^{\perp},$$

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

$$A^{\mathrm{T}}(\mathbf{b} - A\widehat{\mathbf{x}}) = A^{\mathrm{T}}(\mathbf{b} - \operatorname{proj}_W \mathbf{b}) = \mathbf{0}$$

επομένως

$$A^{\mathrm{T}}A\hat{\mathbf{x}} = A^{\mathrm{T}}\mathbf{b}. \tag{10}$$

Σημειώνουμε ότι για την εύρεση του $\hat{\mathbf{x}}$ δεν χρειάζεται να βρούμε την προβολή του b επί του W και στη συνέχεια να λύσουμε το σύστημα $A\mathbf{x} = \mathrm{proj}_W\,\mathbf{b}$. Η λύση των ελαχίστων τετραγώνων προκύπτει από την (10). Την εξίσωση στην (10) λέμε **κανονική εξίσωση**.

Παρατήρηση 3

Αν το μητρώο $A^{\mathrm{T}}A$ είναι αντιστρέψιμο από την κανονική εξίσωση παίρνουμε

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

απ' όπου πολλαπλασιάζοντας από αριστερά με A προκύπτει μέσω της (9) ότι

$$\mathrm{proj}_W\,\mathbf{b} = A\hat{\mathbf{x}} = A\left(A^{\mathrm{T}}A\right)^{-1}A^{\mathrm{T}}\mathbf{b}.$$

Η τελευταία σχέση μας λέει ότι για το μητρώο A το $A\left(A^{\mathrm{T}}A\right)^{-1}A^{\mathrm{T}}$ προβάλει επί του χώρου στηλών του A, όπως είδαμε στη Διάλεξη 9.

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

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

Για την λύση των ελαχίστων τετραγώνων υπολογίζουμε

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

εφαρμογή συνέχεια

και

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

η κανονική εξίσωση (10) γίνεται

$$\begin{pmatrix} n & \sum x_j \\ \sum x_j & \sum x_j^2 \end{pmatrix} \begin{pmatrix} \alpha' \\ \beta' \end{pmatrix} = \begin{pmatrix} \sum y_j \\ \sum x_j y_j \end{pmatrix}. \tag{11}$$

Οι στήλες του A είναι γραμμικά ανεξάρτητες, εκτός αν $x_1 = x_2 = \cdots = x_n$ (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 μητρώου των συντελεστών στη κανονική εξίσωση είναι διάφορη του μηδενός[α΄], 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

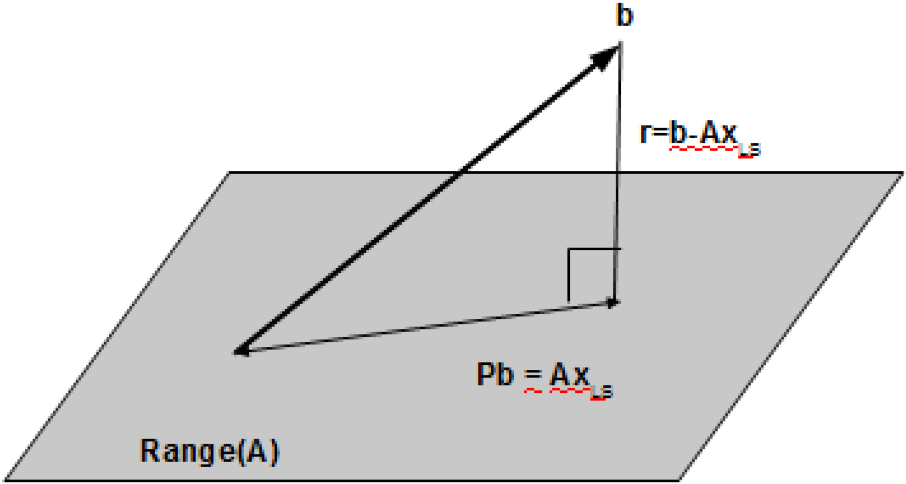
$$\alpha' = \frac{\sum x_j^2 \sum y_j - \sum x_j y_j \sum x_j}{n\sum x_j^2 - (\sum x_j)^2}, \qquad \beta' = \frac{-\sum x_j \sum y_j + n\sum x_j y_j \sum x_j}{n\sum x_j^2 - (\sum x_j)^2}. \tag{12}$$

[α΄] Διαφορετικά, η ορίζουσα του μητρώου των συντελεστών με $\mathbf{x} = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix}^{\mathrm{T}}$ είναι

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

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

Οπτικοποίηση



Εισαγωγή

- 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.