ΓΡΑΜΜΙΚΗ ΑΛΓΕΒΡΑ

Ε. Γαλλόπουλος & Ε. Στεφανόπουλος

CEID

Διάλεξη 8

17 Απριλίου 2018

Έννοιες, λέξεις κλειδιά, υπενθύμιση I

Υπενθύμιση της 7ης δ.

- Συνδυασμοί και πράξεις μεταξύ υπόχωρων. Ευθύ άθροισμα και συμπλήρωμα.
- Ορθογωνιότητα υπόχωρων.
- Προβολές, ορθογώνια προβολή και αναπαράστασή τους με μητρώα.
- Ιδιότητες των μητρώων προβολής.

Έννοιες, λέξεις κλειδιά, υπενθύμιση II

Σήμερα θα δούμε (όροι 8ης δ.):

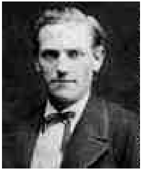
- Ορθοκανικοποίηση διανυσμάτων και διαδικασία Gram-Schmidt.
- Παραγοντοποίηση QR μέσω Gram-Schmidt
- Θεώρημα προβολής
- Γενική θεώρηση προβλημάτων ελαχιστοποίησης
- Προβλήματα και προσεγγίσεις ελαχίστων τετραγώνων
- Ελάχιστα τετράγωνα σε διανυσματικούς χώρους συναρτήσεων.
- Gram-Schmidt σε χώρους πολυωνύμων και βέλτιστη προσέγγιση. συναρτήσεων.

Υπό συζήτηση ενότητες

Ιστορικές πληροφορίες[1]

GRAM-SCHMIDT ORTHOGONALIZATION. When Erhard Schmidt presented the formulae on p. 442 of his ``Zur Theorie der linearen und nichtlinearen Integralgleichungen. I. Teil: Entwicklung willkürlicher Funktionen nach Systemen vorgeschriebener,'' Math. Ann., 63 (1907), 433-476 he said that essentially the same formulae were in J. P. Gram's ``Ueber die Entwickelung reeler Funtionen in Reihen mittelst der Methode der kleinsten Quadrate,'' Jrnl. für die reine und angewandte Math. (1883), 94, 71-73. Modern writers, however, distinguish the two procedures, sometimes using the term ``Gram-Schmidt'' for the Schmidt form and ``modified Gram-Schmidt'' for the Gram version.

Jorgen Pedersen Gram (1850-1916)



Erhard Schmidt (1876 - 1959)

[1] http://jeff560.tripod.com/g.html

Ορθοκανονικοποίηση διανυσμάτων: Διαδικασία Gram-Schmidt

Πρόβλημα: Δίνονται γραμμικά ανεξάρτητα διανύσματα $\{a_1, \ldots, a_n\}$ και θέλουμε να κατασκευάσουμε ορθοκανονικά (ΟΚ) διανύσματα $\{q_1, \cdots, q_n\}$ τ.ώ.

$$\mathrm{span}\{q_1, \cdots, q_n\} = \mathrm{span}\{a_1, \ldots, a_n\}$$

Ορθοκανονικοποίηση διανυσμάτων: Διαδικασία Gram-Schmidt

Πρόβλημα: Δίνονται γραμμικά ανεξάρτητα διανύσματα $\{a_1, \ldots, a_n\}$ και θέλουμε να κατασκευάσουμε ορθοκανονικά (ΟΚ) διανύσματα $\{q_1, \cdots, q_n\}$ τ.ώ.

$$\mathrm{span}\{q_1, \cdots, q_n\} = \mathrm{span}\{a_1, \ldots, a_n\}$$

Ιδέα: Αν αφαιρέσουμε από διάνυσμα, x, την ορθογώνια προβολή του, $P_{\mathcal{S}}x$, επί ενός υπόχωρου, $\mathcal{S}$, το διάνυσμα διαφοράς, $x - P_{\mathcal{S}}x$, είναι κάθετο στον υπόχωρο, δηλ. $x - P_{\mathcal{S}}x \perp \mathcal{S}$.

Ορθοκανονικοποίηση διανυσμάτων: Διαδικασία Gram-Schmidt

Πρόβλημα: Δίνονται γραμμικά ανεξάρτητα διανύσματα $\{a_1, \dots, a_n\}$ και θέλουμε να κατασκευάσουμε ορθοκανονικά (ΟΚ) διανύσματα $\{q_1, \cdots, q_n\}$ τ.ώ.

$$\mathrm{span}\{q_1, \cdots, q_n\} = \mathrm{span}\{a_1, \dots, a_n\}$$

Ιδέα: Αν αφαιρέσουμε από διάνυσμα, x, την ορθογώνια προβολή του, $P_{\mathcal{S}}x$, επί ενός υπόχωρου, $\mathcal{S}$, το διάνυσμα διαφοράς, $x - P_{\mathcal{S}}x$, είναι κάθετο στον υπόχωρο, δηλ. $x - P_{\mathcal{S}}x \perp \mathcal{S}$.

Συστηματική εφαρμογή, διαδοχικά:

I Υπολογίζουμε $q_1 = a_1/\|a_1\|$.

II Για $k = 2, \dots, n$ εκτελούμε:

1. Αφαιρούμε από το a_k την ορθογώνια προβολή του επί του υπόχωρου $\mathcal{S} := \mathrm{span}\{q_1, .., q_{k-1}\}$. Το παραγόμενο διάνυσμα, έστω $\tilde{q}_k$, θα είναι κάθετο στο $\mathcal{S}$.
2. Υπολογίζουμε $q_k = \tilde{q}_k/\|\tilde{q}_k\|$.

Ειδικότερα

Έστω $\mathcal{S} := \mathrm{span}\{q_1, .., q_{k-1}\}$ και ότι Θέτουμε

$$Q_{k-1} := [q_1, ..., q_{k-1}]$$

$$P_{k-1} = Q_{k-1}(Q_{k-1}^\top Q_{k-1})^{-1} Q_{k-1}^\top$$

Ειδικότερα

Έστω $\mathcal{S} := \mathrm{span}\{q_1, .., q_{k-1}\}$ και ότι Θέτουμε

$$Q_{k-1} := [q_1, ..., q_{k-1}]$$

$$\begin{aligned} P_{k-1} &= Q_{k-1}(Q_{k-1}^\top Q_{k-1})^{-1} Q_{k-1}^\top \\ &= Q_{k-1}\cancel{(Q_{k-1}^\top Q_{k-1})^{-1}} Q_{k-1}^\top \end{aligned}$$

Ειδικότερα

Έστω $\mathcal{S} := \mathrm{span}\{q_1, .., q_{k-1}\}$ και ότι Θέτουμε

$$Q_{k-1} := [q_1, ..., q_{k-1}]$$

$$\begin{aligned} P_{k-1} &= Q_{k-1}(Q_{k-1}^\top Q_{k-1})^{-1} Q_{k-1}^\top \\ &= Q_{k-1}\cancel{(Q_{k-1}^\top Q_{k-1})^{-1}} Q_{k-1}^\top \\ &= Q_{k-1} Q_{k-1}^\top \end{aligned}$$

Ειδικότερα

Έστω $\mathcal{S} := \mathrm{span}\{q_1, .., q_{k-1}\}$ και ότι Θέτουμε

$$Q_{k-1} := [q_1, ..., q_{k-1}]$$

$$\begin{aligned}
P_{k-1} &= Q_{k-1}(Q_{k-1}^\top Q_{k-1})^{-1} Q_{k-1}^\top \\
&= Q_{k-1}\cancel{(Q_{k-1}^\top Q_{k-1})^{-1}} Q_{k-1}^\top \\
&= Q_{k-1} Q_{k-1}^\top \\
&= q_1 q_1^\top + \cdots + q_{k-1} q_{k-1}^\top
\end{aligned}$$

Ειδικότερα

Έστω $\mathcal{S} := \mathrm{span}\{q_1, .., q_{k-1}\}$ και ότι Θέτουμε

$$Q_{k-1} := [q_1, ..., q_{k-1}]$$

$$\begin{aligned} P_{k-1} &= Q_{k-1}(Q_{k-1}^\top Q_{k-1})^{-1} Q_{k-1}^\top \\ &= Q_{k-1}\cancel{(Q_{k-1}^\top Q_{k-1})^{-1}} Q_{k-1}^\top \\ &= Q_{k-1} Q_{k-1}^\top \\ &= q_1 q_1^\top + \cdots + q_{k-1} q_{k-1}^\top \end{aligned}$$

Άρα ισοδύναμα τα απαραίτητα βήματα στην επανάληψη k είναι:

$$\begin{aligned} \mathbf{1}.\ \tilde{q}_k &= a_k - q_1(q_1^\top a_k) - \ldots - q_{k-1}(q_{k-1}^\top a_k) \\ \mathbf{2}.\ q_k &= \tilde{q}_k / \|\tilde{q}_k\|. \end{aligned}$$

Παράδειγμα (εφαρμογή)

Θα υπολογίσουμε βάσεις για τους υπόχωρους μητρώου (όπως κεφ. 3) και θα κατασκευάσουμε ΟΚ βάσεις γι' αυτές.

$$A = \begin{pmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 1 & 3 & 1 & 6 \end{pmatrix}$$

Παράδειγμα (εφαρμογή)

Θα υπολογίσουμε βάσεις για τους υπόχωρους μητρώου (όπως κεφ. 3) και θα κατασκευάσουμε ΟΚ βάσεις γι' αυτές.

$$A = \begin{pmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 1 & 3 & 1 & 6 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{pmatrix} A = \begin{pmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Παράδειγμα (εφαρμογή)

Θα υπολογίσουμε βάσεις για τους υπόχωρους μητρώου (όπως κεφ. 3) και θα κατασκευάσουμε ΟΚ βάσεις γι' αυτές.

$$A = \begin{pmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 1 & 3 & 1 & 6 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{pmatrix} A = \begin{pmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{pmatrix} \leftarrow \text{ΑΓΚΜ}$$

Παράδειγμα (εφαρμογή)

Θα υπολογίσουμε βάσεις για τους υπόχωρους μητρώου (όπως κεφ. 3) και θα κατασκευάσουμε ΟΚ βάσεις γι΄ αυτές.

$$A = \begin{pmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 1 & 3 & 1 & 6 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{pmatrix} A = \begin{pmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{pmatrix} \leftarrow \text{ΑΓΚΜ}$$

Επομένως

$$R = \begin{pmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Παράδειγμα (εφαρμογή)

Θα υπολογίσουμε βάσεις για τους υπόχωρους μητρώου (όπως κεφ. 3) και θα κατασκευάσουμε ΟΚ βάσεις γι' αυτές.

$$A = \begin{pmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 1 & 3 & 1 & 6 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{pmatrix} A = \begin{pmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{pmatrix} \leftarrow \text{ΑΓΚΜ}$$

Επομένως

$$R = \begin{pmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{pmatrix} \text{ και το μητρώο που το ανάγει } M = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{pmatrix}.$$

Παράδειγμα (εφαρμογή)

Θα υπολογίσουμε βάσεις για τους υπόχωρους μητρώου (όπως κεφ. 3) και θα κατασκευάσουμε ΟΚ βάσεις γι' αυτές.

$$A = \begin{pmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 1 & 3 & 1 & 6 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{pmatrix} A = \begin{pmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{pmatrix} \leftarrow \text{ΑΓΚΜ}$$

Επομένως

$$R = \begin{pmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{pmatrix} \text{ και το μητρώο που το ανάγει } M = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{pmatrix}.$$

Από τα παραπάνω: $m = 3, n = 4, r = 2, \dim(\mathrm{null}(A)) = n - r = 2, \dim(\mathrm{null}(A^\top)) = m - r = 1.$

Παράδειγμα (εφαρμογή)

Θα υπολογίσουμε βάσεις για τους υπόχωρους μητρώου (όπως κεφ. 3) και θα κατασκευάσουμε ΟΚ βάσεις γι' αυτές.

$$A = \begin{pmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 1 & 3 & 1 & 6 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{pmatrix} A = \begin{pmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{pmatrix} \leftarrow \text{ΑΓΚΜ}$$

Επομένως

$$R = \begin{pmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{pmatrix} \text{ και το μητρώο που το ανάγει } M = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{pmatrix}.$$

Από τα παραπάνω: $m = 3, n = 4, r = 2$, $\dim(\mathrm{null}(A)) = n - r = 2$, $\dim(\mathrm{null}(A^\top)) = m - r = 1$.
Βάσεις:

$$\mathrm{range}(A) = \mathrm{span}\{a_1, a_3\} = \left\{\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}\right\}, \text{ οι στήλες οδηγών του } A \text{ (επίσης του } R\text{)}$$

Παράδειγμα (εφαρμογή)

Θα υπολογίσουμε βάσεις για τους υπόχωρους μητρώου (όπως κεφ. 3) και θα κατασκευάσουμε ΟΚ βάσεις γι' αυτές.

$$A = \begin{pmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 1 & 3 & 1 & 6 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{pmatrix} A = \begin{pmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{pmatrix} \leftarrow \text{ΑΓΚΜ}$$

Επομένως

$$R = \begin{pmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{pmatrix} \text{ και το μητρώο που το ανάγει } M = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{pmatrix}.$$

Από τα παραπάνω: $m = 3, n = 4, r = 2, \dim(\mathrm{null}(A)) = n - r = 2, \dim(\mathrm{null}(A^\top)) = m - r = 1.$

Βάσεις:

$$\mathrm{range}(A) = \mathrm{span}\{a_1, a_3\} = \{\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}\}, \text{ οι στήλες οδηγών του } A \text{ (επίσης του } R\text{)}$$

$$\mathrm{null}(A^\top) = \mathrm{span}\{\begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}\}, \ m - r = 1 \text{ άρα παράγεται από την τελευταία γραμμή του } M$$

Προσοχή: Τα διανύσματα βάσης επαληθεύουν ότι $\mathbb{R}^m = \mathrm{range}(A) \oplus \mathrm{null}(A^\top)$.

Παράδειγμα (μέρος 2)

Όπως πριν

$$A = \begin{pmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 1 & 3 & 1 & 6 \end{pmatrix}, R = \begin{pmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{pmatrix}, M = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{pmatrix}$$

Από τα παραπάνω: $m = 3, n = 4$, $r = 2$, $\dim(\mathrm{null}(A)) = n - r = 2$, $\dim(\mathrm{null}(A^\top)) = m - r = 1$.

Παράδειγμα (μέρος 2)

Όπως πριν

$$A = \begin{pmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 1 & 3 & 1 & 6 \end{pmatrix}, R = \begin{pmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{pmatrix}, M = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{pmatrix}$$

Από τα παραπάνω: $m = 3, n = 4$, $r = 2$, $\dim(\mathrm{null}(A)) = n - r = 2$, $\dim(\mathrm{null}(A^\top)) = m - r = 1$.
Βάσεις:

$$\mathrm{range}(A^\top) = \mathrm{span}\{r^{(1)}, r^{(2)}\} = \{ \begin{pmatrix} 1 \\ 3 \\ 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 4 \end{pmatrix} \},$$ πρώτες r γραμμές του R

Παράδειγμα (μέρος 2)

Όπως πριν

$$A = \begin{pmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 1 & 3 & 1 & 6 \end{pmatrix}, R = \begin{pmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{pmatrix}, M = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{pmatrix}$$

Από τα παραπάνω: $m = 3, n = 4$, $r = 2$, $\dim(\mathrm{null}(A)) = n - r = 2$, $\dim(\mathrm{null}(A^\top)) = m - r = 1$.
Βάσεις:

$$\mathrm{range}(A^\top) = \mathrm{span}\{r^{(1)}, r^{(2)}\} = \{\begin{pmatrix} 1 \\ 3 \\ 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 4 \end{pmatrix}\},$$ πρώτες r γραμμές του R

$$\mathrm{null}(A) = \mathrm{span}\{\begin{pmatrix} -3 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -2 \\ 0 \\ -4 \\ 1 \end{pmatrix}\}$$

από τις $n - r = 2$ ελεύθερες στήλες του R προσαυξημένες με στήλες του I_r.

Παράδειγμα (μέρος 2)

Όπως πριν

$$A = \begin{pmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 1 & 3 & 1 & 6 \end{pmatrix}, R = \begin{pmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{pmatrix}, M = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{pmatrix}$$

Από τα παραπάνω: $m = 3, n = 4$, $r = 2$, $\dim(\text{null}(A)) = n - r = 2$, $\dim(\text{null}(A^\top)) = m - r = 1$.
Βάσεις:

$$\text{range}(A^\top) = \text{span}\{r^{(1)}, r^{(2)}\} = \{\begin{pmatrix} 1 \\ 3 \\ 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 4 \end{pmatrix}\}, \quad \text{πρώτες } r \text{ γραμμές του } R$$

$$\text{null}(A) = \text{span}\{\begin{pmatrix} -3 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -2 \\ 0 \\ -4 \\ 1 \end{pmatrix}\}$$

από τις $n - r = 2$ ελεύθερες στήλες του R προσαυξημένες με στήλες του I_r.
Προσοχή: Τα διανύσματα βάσης επαληθεύουν ότι $\mathbb{R}^n = \text{range}(A^\top) \oplus \text{null}(A)$.

Ορθοκανονικές βάσεις

Χρησιμοποιούμε Gram-Schmidt για να κατασκευάσουμε ΟΚ βάσεις για τους υπόχωρους (εκτός αν είναι ήδη.)

ΟΚ βάση για $\mathrm{range}(A)$ από τις στήλες οδηγών του A

$$\mathrm{range}(A) = \mathrm{span}\{a_1, a_3\} = \{\begin{pmatrix}1\\0\\1\end{pmatrix}, \begin{pmatrix}0\\1\\1\end{pmatrix}\}$$

επομένως $q_1 = a_1/\|a_1\| = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 0 & 1\end{pmatrix}^\top$

$$\tilde{q}_2 = a_2 - q_1(q_1^\top a_2) = \frac{1}{2}\begin{pmatrix}-1 & 2 & 1\end{pmatrix}^\top \Rightarrow q_2 = \frac{1}{\sqrt{6}}\begin{pmatrix}-1 & 2 & 1\end{pmatrix}^\top$$

Άρα $\mathrm{range}(A) = \mathrm{span}\{\frac{1}{\sqrt{2}}\begin{pmatrix}1\\0\\1\end{pmatrix}, \frac{1}{\sqrt{6}}\begin{pmatrix}-1\\2\\1\end{pmatrix}\}$

ΟΚ βάση για $\mathrm{null}(A^\top)$ $\{\frac{1}{\sqrt{3}}\begin{pmatrix}-1\\-1\\1\end{pmatrix}\}$ (διάσταση 1, αρκεί κανονικοποίηση)

ΟΚ βάση για $\mathrm{range}(A^\top)$ $q_1 = \frac{1}{\|a_1\|}a_1 = \frac{1}{\sqrt{14}}\begin{pmatrix}1\\3\\0\\2\end{pmatrix}$, $\tilde{q}_2 = a_2 - q_1(q_1^\top a_2) = \begin{pmatrix}0\\0\\1\\4\end{pmatrix} - \frac{8}{14}\begin{pmatrix}1\\3\\0\\2\end{pmatrix} = \frac{1}{7}\begin{pmatrix}-4\\-12\\7\\20\end{pmatrix}$

άρα μία ΟΚ βάση για το $\mathrm{range}(A^\top)$ είναι $\{q_1, q_2\} = \{\frac{1}{\sqrt{14}}\begin{pmatrix}1\\3\\0\\2\end{pmatrix}, \frac{1}{\sqrt{609}}\begin{pmatrix}-4\\-12\\7\\20\end{pmatrix}\}$

ΟΚ βάση για $\mathrm{null}(A)$ Η βάση που υπολογίσαμε αποτελείται από τα διανύσματα $b_1 = \begin{pmatrix}-3\\1\\0\\0\end{pmatrix}, b_2 = \begin{pmatrix}-2\\0\\-4\\1\end{pmatrix}$.

Άρα $p_1 = \frac{1}{\|b_1\|}b_1 = \frac{1}{\sqrt{10}}\begin{pmatrix}-3\\1\\0\\0\end{pmatrix}$,

$\tilde{p}_2 = b_2 - p_1(p_1^\top b_2) = \begin{pmatrix}-2\\0\\-4\\1\end{pmatrix} - \frac{6}{10}\begin{pmatrix}-3\\1\\0\\0\end{pmatrix} = \frac{1}{5}\begin{pmatrix}-1\\-3\\-20\\5\end{pmatrix}$

άρα μια ΟΚ βάση για το $\mathrm{null}(A)$ είναι $\{p_1, p_2\} = \{\frac{1}{\sqrt{10}}\begin{pmatrix}-3\\1\\0\\0\end{pmatrix}, \frac{1}{\sqrt{435}}\begin{pmatrix}-1\\-3\\-20\\5\end{pmatrix}\}$

Gram-Schmidt και παραγοντοποίηση QR I

Η διαδικασία GS παράγει τις στήλες του Q με γραμμικούς συνδυασμούς των στηλών του A. Ειδικότερα, μπορούμε να πούμε ότι:

Παρατήρηση 1: Οι στήλες του $Q = (q_1, q_2, \ldots, q_n)$ είναι ΟΚ και είναι γραμμικός συνδυασμός των στηλών του $A = (a_1, a_2, .., a_n)$.

Παρατήρηση 1: Οι στήλες του $A = (a_1, a_2, .., a_n)$ είναι γραμμικός συνδυασμός των στηλών του $Q = (q_1, q_2, \ldots, q_n)$ που είναι ΟΚ.

Παραγοντοποίηση QR

Το (1) εκφράζεται και ως παραγοντοποίηση:

$$\mathbf{A} = \mathbf{QR}$$

Χαρακτηρίζεται από το ότι: α) το Q έχει ορθογώνιες στήλες (εκ κατασκευής) και αν το A είναι τετραγωνικό, τότε το Q είναι ορθογώνιο. β) το R είναι άνω τριγωνικό (θα το δούμε στη συνέχεια).

Gram-Schmidt και παραγοντοποίηση QR II

Κάθε στήλη του A παράγεται από γραμμικό συνδυασμό των στηλών του Q, π.χ. αν με a_j, r_j συμβολίσουμε τη στήλη j των Q, R αντίστοιχα:

$$a_j = Qr_j, \text{ για } j = 1, \ldots, n \text{ όπου } r_j = (\rho_{1,j}, \rho_{2,j}, \ldots, \rho_{n,j})^\top.$$

Επειδή οι στήλες του Q είναι ΟΚ, ισχύει ότι $Q^\top A = R$, επομένως

$$R = \begin{pmatrix} q_1^\top a_1 & q_1^\top a_2 & \cdots & & q_1^\top a_n \\ q_2^\top a_1 & q_2^\top a_2 & \cdots & & q_2^\top a_n \\ q_3^\top a_1 & q_3^\top a_2 & q_3^\top a_3 & \cdots & \\ \cdots & \cdots & & & \cdots \\ q_n^\top a_1 & q_n^\top a_2 & q_n^\top a_3 & \cdots & q_n^\top a_n \end{pmatrix}$$

Εκ κατασκευής, $q_i \perp \{a_1, a_2, \ldots, a_{i-1}\}$, άρα το R είναι άνω τριγωνικό.

Παρατηρήσεις και συμπεράσματα I

$$\begin{aligned}
a_1 &= q_1\rho_{11}, \text{ όπου } \|a_1\| = \rho_{11}\\
a_2 &= q_1(q_1^\top a_2) + q_2\rho_{22}, \text{ όπου } \rho_{22} = \|a_2 - q_1(q_1^\top a_2)\|\\
a_3 &= q_1(q_1^\top a_3) + q_2(q_2^\top a_3) + q_3\rho_{33}
\end{aligned}$$

δηλ.

$$(a_1, a_2, a_3) = (q_1, q_2, q_3)\begin{pmatrix} \rho_{11} & \rho_{12} & \rho_{13} \\ 0 & \rho_{22} & \rho_{23} \\ 0 & 0 & \rho_{33} \end{pmatrix}$$

και γενικά $A = QR$ όπου το Q έχει ορθογώνιες στήλες και το R είναι άνω τριγωνικό.

Παρατηρήσεις και συμπεράσματα II

Παραγοντοποίηση QR

Για κάθε μητρώο $A \in \mathbb{R}^{m\times n}$ με γραμμικά ανεξάρτητες στήλες μπορούμε να γράψουμε $A = QR$ όπου το $Q \in \mathbb{R}^{m\times n}$ έχει ορθογώνιες στήλες και το R είναι άνω τριγωνικό με μη μηδενικά στη διαγώνιο.

- Οι στήλες $q_1, \ldots, q_i$ αποτελούν ΟΚ βάση για τον υπόχωρο $\mathrm{span}\{a_1, \cdots, a_i\}$ για κάθε $i = 1, 2, \ldots, n$
- Η παραγοντοποίηση επεκτείνεται και σε $A = QR$ όπου $Q \in \mathbb{R}^{m\times m}$ και $R \in \mathbb{R}^{m\times n}$. Τότε το Q περιέχει ΟΚ βάση για όλον τον $\mathbb{R}^m$.
- Προσοχή: Η παραγοντοποίηση αστοχεί όταν οι στήλες του A είναι γραμμικά εξαρτημένες. Όπως και με την LU μπορεί να χρησιμοποιηθεί οδήγηση αλλά με εναλλαγές στηλών για να παρακάμψουμε το πρόβλημα. Ειδικότερα, τότε, η παραγοντοποίηση έχει τη μορφή

$$AP = QR,$$ όπου P είναι μητρώο μετάθεσης.

Απόσταση από υπόχωρο και το πρόβλημα της προσέγγισης

Απόσταση από υπόχωρο

Έστω $\mathcal{S}$ ένα μη κενό υποσύνολο ενός διανυσματικού χώρου $\mathcal{V}$, που διαθέτει τη νόρμα $\|\cdot\|$.

Για κάθε $v \in \mathcal{V}$, η απόσταση του v από το $\mathcal{S}$ ορίζεται ως

$$d(v, \mathcal{S}) = \inf_{s \in \mathcal{S}} \|v - s\|.$$

Ένα στοιχείο $\hat{s} \in \mathcal{S}$ αποκαλείται βέλτιστη προσέγγιση (ή πλησιέστερο στοιχείο) στο v από τον $\mathcal{S}$ αν

$$\|v - \hat{s}\| = d(v, \mathcal{S}).$$

Απόσταση από υπόχωρο και το πρόβλημα της προσέγγισης

Απόσταση από υπόχωρο

Έστω $\mathcal{S}$ ένα μη κενό υποσύνολο ενός διανυσματικού χώρου $\mathcal{V}$, που διαθέτει τη νόρμα $\|\cdot\|$.
Για κάθε $v \in \mathcal{V}$, η απόσταση του v από το $\mathcal{S}$ ορίζεται ως

$$d(v, \mathcal{S}) = \inf_{s \in \mathcal{S}} \|v - s\|.$$

Ένα στοιχείο $\hat{s} \in \mathcal{S}$ αποκαλείται βέλτιστη προσέγγιση (ή πλησιέστερο στοιχείο) στο v από τον $\mathcal{S}$ αν

$$\|v - \hat{s}\| = d(v, \mathcal{S}).$$

Η αναζήτηση βέλτιστης προσέγγισης και τα συναφή προβλήματα (ύπαρξη, μοναδικότητα, κατασκευή) είναι κεντρικό ζήτημα των Μαθηματικών με πάρα πολλές εφαρμογές. Η αποτελεσματική επίλυσή του απαιτεί συνδυασμό μαθηματικών τεχνικών και υπολογιστικών τεχνολογιών και γνώσεις της εκάστοτε εφαρμογής.

Παράδειγμα

Περιγραφή: Λαμβάνουμε μέτρησεις $\psi_1, \psi_2, \ldots, \psi_m$ στις χρονικές στιγμές $\xi_1, \xi_2, \ldots, \xi_m$. Επειδή παρατηρούμε ότι το σήμα συνδυάζει εκθετική συμπεριφορά μαζί με ταλαντώσεις, αποφασίζουμε να το μοντελοποιήσουμε με συνάρτηση που έχει την εξής μορφή:

$$f(\xi; c) \quad = \quad \gamma_1 + \gamma_2 e^{-(\gamma_3 - \xi)^2/\gamma_4} + \gamma_5 \sin(\gamma_6 \xi)$$

όπου οι (άγνωστες προς το παρόν) τιμές $c = (\gamma_1, \gamma_2, \ldots, \gamma_6)^\top$ είναι οι παράμετροι του μοντέλου. Για να οριστεί πλήρως η συνάρτηση, πρέπει να επιλεγεί το c. Αυτό γίνεται έτσι ώστε να επιτευχθεί η βέλτιστη προσαρμογή των τιμών $f(\xi_j; c)$ στις (γνωστές) μετρήσεις.

Διατύπωση ως πρόβλημα ελαχίστων τετραγώνων Να υπολογιστεί $c \in \mathbb{R}^6$ τέτοιο ώστε να ελαχιστοποιείται το άθροισμα των τετραγώνων

$$\min_{c \in \mathbb{R}^6} \sum_{j=1}^{m} r_j^2(c)$$

όπου $r_j(c) = \psi_j - f(\xi_j; c)$, $j = 1, \ldots, m$ είναι τα κατάλοιπα.

Ορολογία Πρόκειται για μη γραμμικό πρόβλημα ελαχίστων τετραγώνων. Είναι μη γραμμικό γιατί η συνάρτηση μοντέλο f εξαρτάται μη γραμμικά από τις παραμέτρους.

☺ Δεν θα ασχοληθούμε με μη γραμμικά προβλήματα!

Γραμμικό πρόβλημα ελαχίστων τετραγώνων

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

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

- Στο γραμμικό πρόβλημα ελαχίστων τετραγώνων, οι συναρτήσεις $\phi_j(\xi)$ μπορεί να είναι μη γραμμικές! Για παράδειγμα στην

$$f(\xi;c) \quad = \quad \gamma_1 + \gamma_2 e^{-\xi^2} + \gamma_3 \sin^2(\xi\pi)$$

έχουμε επιλέξει τις συναρτήσεις

$$\phi_1(\xi) = 1, \phi_2(\xi) = e^{-\xi^2}, \phi_3(\xi) = \sin^2(\xi p)$$

όπου οι ϕ_2, ϕ_3 είναι μη γραμμικές.

Υπενθύμιση

Γνωρίζουμε ότι για κάθε διανυσματικό υπόχωρο $\mathcal{S}$ ενός δ.χ. $\mathcal{V}$, ισχύουν τα εξής:

1. (θεώρημα προβολής) $\mathcal{V} = \mathcal{S} \oplus \mathcal{S}^{\perp}$ όπου
$$\mathcal{S}^{\perp} = \{y | y \perp x, \forall x \in \mathcal{S}\}$$
2. Κάθε $x \in \mathcal{V}$ γράφεται με μοναδικό τρόπο ως $x = x_{\mathcal{S}} + x_{\mathcal{S}^{\perp}}$, όπου $x_{\mathcal{S}} \in \mathcal{S}$ και $x_{\mathcal{S}^{\perp}} \perp \mathcal{S}$.
3. Ισχύουν ότι $x_{\mathcal{S}} = Px$ και $x_{\mathcal{S}^{\perp}} = (I - P)x$, όπου P είναι ο τελεστής ορθογώνιας προβολής επί του $\mathcal{S}$.
4. $\|x\|^2 = \|x_{\mathcal{S}} + x_{\mathcal{S}^{\perp}}\|^2 = \|x_{\mathcal{S}}\|^2 + \|x_{\mathcal{S}^{\perp}}\|^2$

Προσέγγιση ελαχίστων τετραγώνων

Γραμμικό πρόβλημα ελαχίστων τετραγώνων

Περιγραφή

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

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

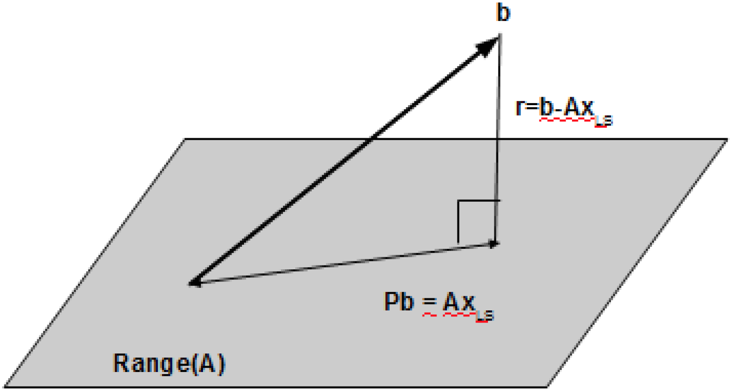
- Η ελαχιστοποίηση του $\|Ax-b\|_2$ ισοδυναμεί με την εύρεση διανύσματος $x_{LS} \in \mathbb{R}^n$ τέτοιου ώστε, μεταξύ όλων των διανυσμάτων που παράγονται από γραμμικό συνδυασμό των στηλών του A, το $p = Ax_{LS}$ να είναι το πλησιέστερο στο b (ως προς την ευκλείδεια νόρμα).
- Για να συμβαίνει αυτό,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

$$0 \quad = \quad (Ay)^\top(b - Ax_{LS}) = y^\top(A^\top b - A^\top Ax_{LS}).$$

Επομένως το x_{LS} είναι η λύση του συστήματος $\boxed{A^\top Ax_{\mathrm{LS}} = A^\top b}$ που αποκαλούνται κανονικές εξισώσεις του προβλήματος των ελαχίστων τετραγώνων[2]

[2]Το μητρώο $A^\top A$ αποκαλείται συχνά μητρώο Gram ή Γραμμιανό.

Οπτικοποίηση



Γεωμετρική υπενθύμιση: Στις 2 διαστάσεις, αποδεικνύεται μέσω πυθαγορείου θεωρήματος ότι η απόσταση σημείου από ευθεία είναι ίση με το μήκος της καθέτου από το σημείο ως την τομή της με την ευθεία.

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 $r = \mathrm{rank}(A^\top A) \leq n$. 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: Ικανή και αναγκαία συνθήκη για να μην είναι αντιστρέψιμο το $A^\top A$ είναι να είναι γραμμικά εξαρτημένες οι στήλες του A (αποδείξτε το!)

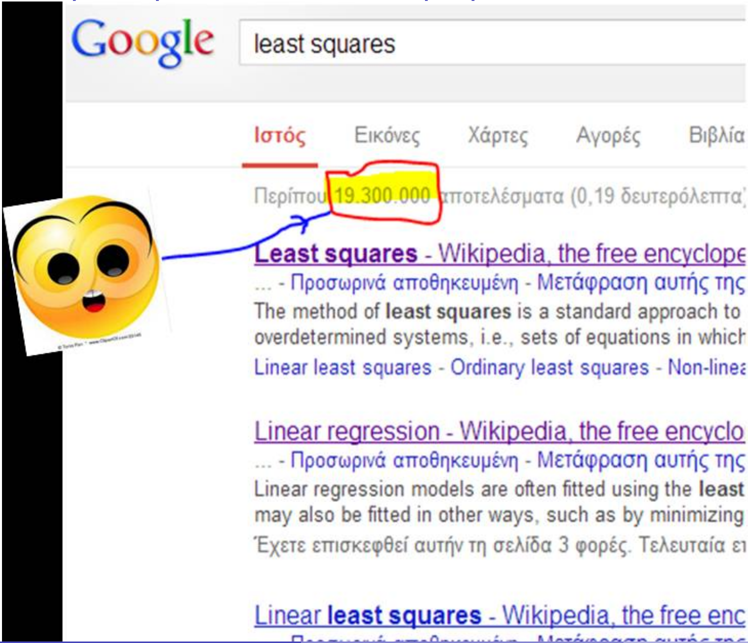
Γενική μορφή γραμμικού προβλήματος ελαχίστων τετραγώνων

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

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

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

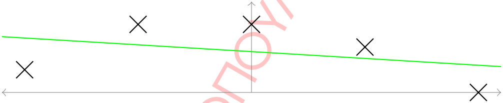
Τι λέγεται για τα ελάχιστα τετράγωνα



Τι λέγεται για τα ελάχιστα τετράγωνα

Προσαρμογή ελαχίστων τετραγώνων με ευθεία γραμμή

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».
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 $\begin{pmatrix} 1 & \xi \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} = \psi$

'Αλλες μετρικές σφάλματος Ανάλογα με την εφαρμογή, μπορεί οι απαιτήσεις να είναι διαφορετικές (και πιο απαιτητικές). Για παράδειγμα στα ολικά ελάχιστα τετράγωνα (total least squares) ελαχιστοποιούμε το άθροισμα τετραγώνων της απόστασης κάθε σημείου (ξ_j, ψ_j) από την ευθεία $\gamma_0 + \gamma_1\xi = \psi$ (δηλ. το μήκος της καθέτου από το κάθε σημείο στην ευθεία).

Ελάχιστα τετράγωνα και ολικά ελάχιστα τετράγωνα (least squares vs. total least squares (MH'07))

Available online at www.sciencedirect.com
ScienceDirect

ELSEVIER

Signal Processing 87 (2007) 2283–2302

SIGNAL PROCESSING

www.elsevier.com/locate/sigpro

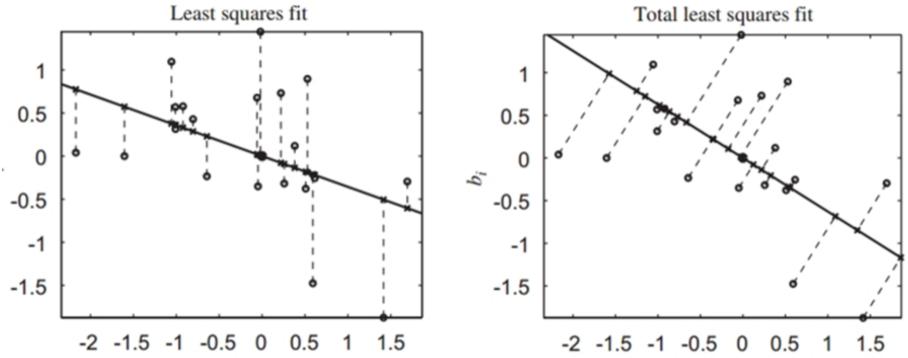
Overview of total least-squares methods

Ivan Markovsky[a,*], Sabine Van Huffel[b]

[a]*School of Electronics and Computer Science, University of Southampton, SO17 1BJ, UK*
[b]*Katholieke Universiteit Leuven, ESAT-SCD/SISTA, Kasteelpark Arenberg 10 B–3001 Leuven, Belgium*

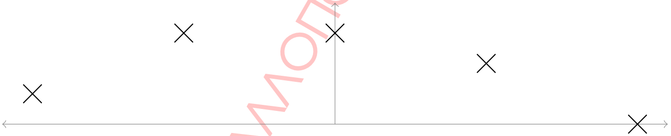
Received 28 September 2006; received in revised form 30 March 2007; accepted 3 April 2007
Available online 14 April 2007

Ελάχιστα τετράγωνα και ολικά ελάχιστα τετράγωνα least squares vs. total least squares



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

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

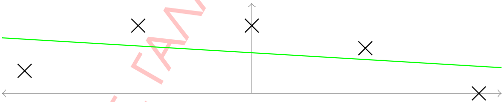
$$A^\top A = \begin{pmatrix} 5.0 & 0.0 \\ 0.0 & 2.5 \end{pmatrix}, A^\top y = \begin{pmatrix} 0.9 \\ -0.15 \end{pmatrix}$$

Αν κάνουμε τους υπολογισμούς

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

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

σφάλμα $\|y - Ac_{LS}\|_2 = 0.2429$



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

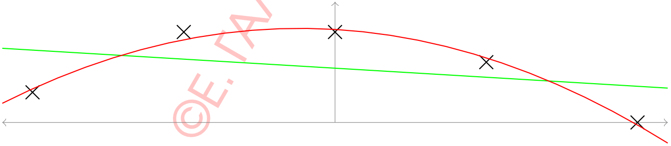
$$c_{LS} = (A^\top A)^{-1} A^\top y = \begin{pmatrix} 0.3086 \\ -0.0600 \\ -0.2571 \end{pmatrix}$$

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

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

σφάλμα $\|y - Ac_{LS}\|_2 = 0.0338$

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



Συμμετρικά θετικά (ημι)ορισμένα μητρώα

Ενδιαφέρον χαρακτηριστικό π.χ. στο μητρώο των κανονικών εξισώσεων

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

Συμμετρικά θετικά (ημι)ορισμένα μητρώα

Ενδιαφέρον χαρακτηριστικό π.χ. στο μητρώο των κανονικών εξισώσεων

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

Το $B = A^\top A$:

- είναι συμμετρικό
- ισχύει πάντα ότι

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

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

Συμμετρικά θετικά (ημι)ορισμένα μητρώα

Ενδιαφέρον χαρακτηριστικό π.χ. στο μητρώο των κανονικών εξισώσεων

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

Το $B = A^\top A$:

- είναι συμμετρικό
- ισχύει πάντα ότι

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

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

Ορισμός

Κάθε $A \in \mathbb{R}^{n\times n}$ για το οποίο ισχύει ότι $A^\top = A$ και $f(x) := x^\top Ax \geq 0$ 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

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

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 $A^\top = A$ 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**.

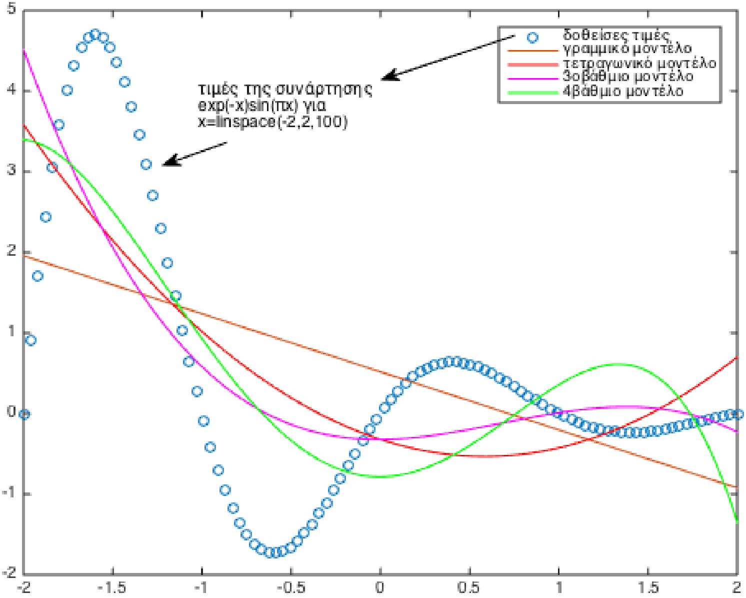
Επομένως, το $A^\top A$ είναι ΣΘΟ ανν το A έχει γραμμικά ανεξάρτητες στήλες, διαφορετικά είναι θετικά ημιορισμένο. Δεν μπορεί όμως να είναι αόριστο.

Παραδείγματα προσέγγισης ελαχίστων τετραγώνων

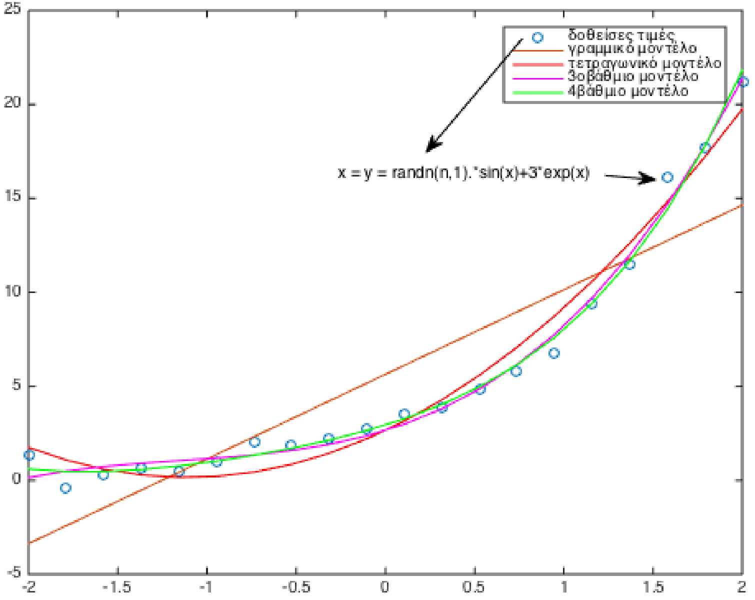
Χρησιμοποιούμε τον παρακάτω κώδικα MATLAB για διάφορες επιλογές x, y

```
% kataskeu'h twn mhtr'wwn
A1 = [ones(n,1),x]; % grammik'o mont'elo
A2 = [A1,x.^2]; % tetragwnik'o mont'elo
A3 = [A2,x.^3]; % 3b'ajmio mont'elo
A4 = [A3,x.^4]; % 4b'ajmio mont'elo
% ep'ilush kanonik'wn exis'wsewn
z1 = (A1'*A1)\(A1'*y); z2 = (A2'*A2)\(A2'*y);
z3 = (A3'*A3)\(A3'*y); z4 = (A4'*A4)\(A4'*y);
% optikopo'ihsh
plot(x,y,'o',x,A1*z1,x,A2*z2,'r',x,A3*z3,'m',x,A4*z4,'g')
% sf'almata
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)
```



```
n=20; x = linspace(-2,2,n)'; y = randn(n,1).*sin(x)+3*exp(x)
```



Προσέγγιση συναρτήσεων με ελάχιστα τετράγωνα: Εισαγωγή I

- Γνωρίζουμε ότι πολλές κατηγορίες συναρτήσεων αποτελούν δ.χ.
- Παραδείγματα: α) $\mathbb{P}_n$: τα πολυώνυμα βαθμού $\leq n$. (ο δ.χ. έχει διάσταση $n+1$). β) $C[-1,1]$: χώρος συνεχών συναρτήσεων στο κλειστό διάστημα $[-1,1]$. γ) Τα πολυώνυμα οποιουδήποτε βαθμού (οι δ.χ. είναι απειροδιάστατοι).
- Για ολοκληρώσιμες συναρτήσεις f, g στο διάστημα $[a, b]$ και θετική συνάρτηση w, το ορισμένο ολοκλήρωμα

$$\langle f, g\rangle := \int_a^b f(t)g(t)w(t)dt$$

  ορίζει εσωτερικό γινόμενο.
- Επομένως μπορούμε να ορίσουμε και καθετότητα συναρτήσεων. Για παράδειγμα, αν $w(x) = 1$, οι συναρτήσεις $\{1, x, \frac{1}{2}(3x^2 - 1)\}$ είναι κάθετες μεταξύ τους. Επίσης είναι γραμμικά ανεξάρτητες και αποτελούν βάση για τον $\mathbb{P}_2[-1,1]$ που είναι υπόχωρος του $C[-1,1]$.
- Μπορούμε να εφαρμόσουμε τη διαδικασία Gram-Schmidt σε πολυώνυμα και να κατασκευάσουμε Ορθογώνια Πολυώνυμα (σε κάποιο διάστημα και ως προς δεδομένη συνάρτηση w.)

Προσέγγιση συναρτήσεων με ελάχιστα τετράγωνα: Εισαγωγή II

Ερώτημα Δίνεται συνάρτηση f από κάποιο χώρο συναρτήσεων και θέλουμε να βρούμε βέλτιστη προσέγγισή της ως προς το κριτήριο των ελαχίστων τετραγώνων από κάποιον επιλεγμένο γραμμικό υπόχωρο, π.χ. τα πολυώνυμα, με βάση το παραπάνω εσωτερικό γινόμενο.

Επίλυση όπως πριν, θα αξιοποιήσουμε το ότι η διαφορά της υπό προσέγγιση συνάρτησης από τη βέλτιστη προσέγγιση θα είναι κάθετη σε όλες τις συναρτήσεις του υποχώρου απ' όπου «χτίζουμε» την προσέγγιση.

Ιδέα

Από το θεώρημα προβολής, αν αφαιρέσουμε από τη συνάρτηση το γραμμικό συνδυασμό που παράγει τη συνάρτηση βέλτιστης προσέγγισης, το κατάλοιπο θα είναι κάθετο σε όλες τις συναρτήσεις ϕ_j, επομένως $\langle f - \sum_{j=1}^{n} \gamma_j \phi_j, \phi_i \rangle = 0$ για $i = 1, \dots, n$. Επομένως

$$\begin{pmatrix} \langle f, \phi_1 \rangle \\ \vdots \\ \langle f, \phi_n \rangle \end{pmatrix} = \underbrace{\begin{pmatrix} \langle \phi_1, \phi_1 \rangle & \cdots & \langle \phi_1, \phi_n \rangle \\ \vdots & & \vdots \\ \langle \phi_n, \phi_1 \rangle & \cdots & \langle \phi_n, \phi_n \rangle \end{pmatrix}}_{\text{Μητρώο Gram}} \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_n \end{pmatrix}.$$

Παράδειγμα

Να υπολογιστεί η βέλτιστη προσέγγιση της συνάρτησης x^2 από τον υπόχωρο $\mathrm{span}\langle 1, x\rangle$ με εσωτερικό γινόμενο $\langle f, g\rangle := \int_{-1}^{1} f(x)g(x)dx$.
Έστω $\phi(x) = \gamma_1 + \gamma_2 x$ η ζητούμενη προσέγγιση, τότε αν $r(x) = x^2 - \phi(x)$ θα πρέπει από το θεώρημα ορθογωνιότητας $\langle r, 1\rangle = \langle r, x\rangle = 0$. Επομένως

$$\begin{aligned}
\langle x^2, 1\rangle &= \int_{-1}^{1} x^2 dx = \frac{2}{3} \\
\langle x^2, x\rangle &= \int_{-1}^{1} x^3 dx = 0 \\
\langle 1, 1\rangle &= 2, \langle 1, x\rangle = 0 \\
\langle x, 1\rangle &= 0, \langle x, x\rangle = \frac{2}{3}
\end{aligned}$$

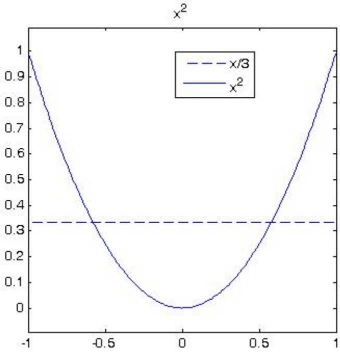
$$\begin{pmatrix} \frac{2}{3} \\ 0 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & \frac{2}{3} \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix}. \Rightarrow \gamma_1 = \frac{1}{3}, \gamma_2 = 0.$$

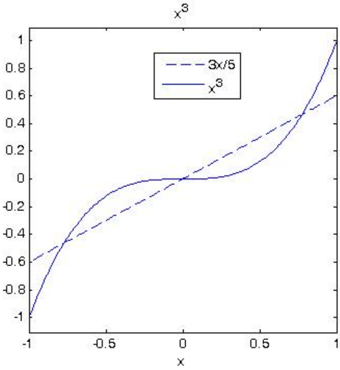
Άρα η βέλτιστη προσέγγιση είναι $\phi(x) = \frac{1}{3}$.
Προσέξτε επίσης ότι $\langle 1, x \rangle = 0$ δηλ, οι συναρτήσεις $\phi_1(x) = 1$ και $\phi_2(x) = x$ είναι κάθετες ως προς το παραπάνω εσωτερικό γινόμενο.
Προσοχή: Με τον ίδιο τρόπο αλλά λύνοντας το σύστημα

$$\begin{pmatrix} 0 \\ \frac{2}{5} \\ 0 \end{pmatrix} = \begin{pmatrix} 2 & 0 & \frac{2}{3} \\ 0 & \frac{2}{3} & 0 \\ \frac{2}{3} & 0 & \frac{2}{5} \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{pmatrix}. \Rightarrow \gamma_1 = 0, \gamma_1 = \frac{3}{5}, \gamma_3 = 0.$$

προκύπτει ότι η βέλτιστη προσέγγιση του x^3 από το χώρο πολυωνύμων $\langle 1, x, x^2 \rangle$ ως προς παραπάνω εσωτερικό γινόμενο είναι $\phi(x) = \frac{3}{5}x$.

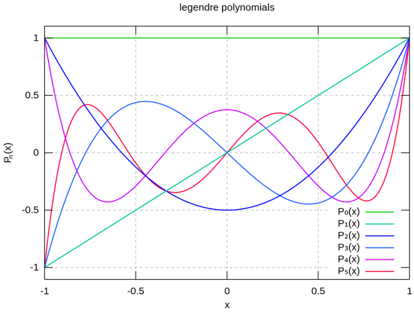
Οπτικοποίηση των βέλτιστων προσεγγίσεων





Ορθογώνια πολυώνυμα Legendre

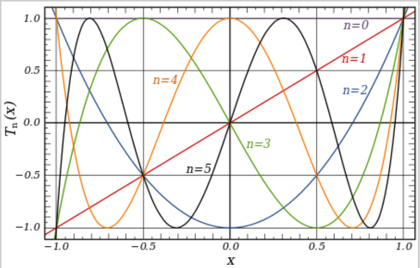
n	$P_n(x)$
0	1
1	x
2	$\frac{1}{2}(3x^2-1)$
3	$\frac{1}{2}(5x^3-3x)$
4	$\frac{1}{8}(35x^4-30x^2+3)$
5	$\frac{1}{8}(63x^5-70x^3+15x)$
6	$\frac{1}{16}(231x^6-315x^4+105x^2-5)$
7	$\frac{1}{16}(429x^7-693x^5+315x^3-35x)$
8	$\frac{1}{128}(6435x^8-12012x^6+6930x^4-1260x^2+35)$
9	$\frac{1}{128}(12155x^9-25740x^7+18018x^5-4620x^3+315x)$
10	$\frac{1}{256}(46189x^{10}-109395x^8+90090x^6-30030x^4+3465x^2-63)$



- Παράγονται με διαδικασία Gram-Schmidt επί των $\{1, x, x^2, \ldots\}$ με το εσωτερικό γινόμενο $\int_{-1}^{1} f(x)g(x)\mathrm{d}x$
- Ισχύει ότι $\int_{-1}^{1} P_n(x)P_m(x)dx = \frac{2}{2n+1}\delta_{nm}$ (δέλτα Kronecker)
- Υπάρχουν πολλές οικογένειες ορθογωνίων πολυωνύμων (π.χ. Chebyshev, Gegenbauer, Legendre, Laguerre, Hermite...).
- Η βασική διαφορά τους αφορά στη συνάρτηση βάρους.

Ορθογώνια πολυώνυμα Chebyshev

$T_0(x) = 1$
$T_1(x) = x$
$T_2(x) = 2x^2 - 1$
$T_3(x) = 4x^3 - 3x$
$T_4(x) = 8x^4 - 8x^2 + 1$
$T_5(x) = 16x^5 - 20x^3 + 5x$
$T_6(x) = 32x^6 - 48x^4 + 18x^2 - 1$
$T_7(x) = 64x^7 - 112x^5 + 56x^3 - 7x$
$T_8(x) = 128x^8 - 256x^6 + 160x^4 - 32x^2 + 1$
$T_9(x) = 256x^9 - 576x^7 + 432x^5 - 120x^3 + 9x.$
$T_{10}(x) = 512x^{10} - 1280x^8 + 1120x^6 - 400x^4 + 50x^2 - 1.$
$T_{11}(x) = 1024x^{11} - 2816x^9 + 2816x^7 - 1232x^5 + 220x^3 - 11x.$



The first few Chebyshev polynomials of the first kind in the domain $-1 < x < 1$: The flat T_0, T_1, T_2, T_3, T_4 and T_5.

- Παράγονται με διαδικασία Gram-Schmidt επί των $\{1, x, x^2, \dots\}$ με το εσωτερικό γινόμενο $\int_{-1}^{1} f(x)g(x)\frac{1}{\sqrt{1-x^2}}\mathrm{d}x$
- Ισχύει ότι $\int_{-1}^{1} T_n(x)T_m(x)dx = 0$ αν $m \neq n$.
- Ένα σημαντικό στοιχείο των ορθογωνίων πολυωνύμων που αξίζει να αναφερθεί (εκτός ύλης) είναι ότι μπορούν να παραχθούν με "κοντή" αναδρομική σχέση (τριών όρων), π.χ. τα Chebyshev: $T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x)$ όπου $T_0(x) = 1, T_1(x) = x$.

Προσέγγιση ελαχίστων τετραγώνων χρησιμοποιώντας διαφορικό λογισμό

Για το γραμμικό μοντέλο

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

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

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

$$A^\top Ac = A^\top y$$

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