17-12-15

Το γραμμικό πρόβλημα ελαχίστων τετραγώνων.

Έστω $A\in\mathbb{R}^{n\times m}$, $m\le n$, $x\in\mathbb{R}^m$, $b\in\mathbb{R}^n$.
Να βρεθεί $x\in\mathbb{R}^m$ που θα πληροί κατά τον καλύτερο δυνατό τρόπο το σύστημα $Ax=b$. Αν το σύστημα δεν έχει λύση, τότε υπολογίζουμε τον x, ώστε να ελαχιστοποιείται η $\|r\|=\|b-Ax\|$. Αν ως νόρμα πάρουμε την $\|\cdot\|_2$, τότε έχουμε το πρόβλημα ελαχίστων τετραγώνων.

$$\min_{x\in\mathbb{R}^m} f(x)=\min_{x\in\mathbb{R}^m}\|r\|^2=\min_{x\in\mathbb{R}^m}\|b-Ax\|^2=\min_{x\in\mathbb{R}^m}(b-Ax,\,b-Ax)=\boxed{\min_{x\in\mathbb{R}^m}}$$

$$\boxed{\sum_{i=1}^{n}\left(b_i-(Ax)_i\right)^2}=\min_{x\in\mathbb{R}^m}\sum_{i=1}^{n}\Big(b_i-\sum_{j=1}^{m}a_{ij}x_j\Big)^2.$$ Το ελάχιστο θα βρίσκεται σε κρίσιμο σημείο. $\nabla f(x)=\left[\frac{\partial f}{\partial x_1}\;\frac{\partial f}{\partial x_2}\;\cdots\;\frac{\partial f}{\partial x_m}\right]^T=0.$

$$\frac{\partial f(x)}{\partial x_k}=\sum_{i=1}^{n}2\Big(b_i-\sum_{j=1}^{m}a_{ij}x_j\Big)\frac{\partial}{\partial x_k}\Big(b_i-\sum_{j=1}^{m}a_{ij}x_j\Big)=2\sum_{i=1}^{n}\Big(b_i-\sum_{j=1}^{m}a_{ij}x_j\Big)(-a_{ik})=0 \quad (1)$$

$\Leftrightarrow \sum_{i=1}^{n}a_{ik}\sum_{j=1}^{m}a_{ij}x_j=\sum_{i=1}^{n}a_{ik}(Ax)_i=(a_k,Ax)=a_k^TAx$, όπου a_k η k στήλη του A.

$$\sum_{i=1}^{n}b_i\cdot a_{ik}=(a_k,b)=a_k^T b \quad (2)$$ Τελικά από (1),(2): $a_k^TAx=a_k^Tb$

$$\frac{\partial f}{\partial x_k}=0\Leftrightarrow a_k^TAx=a_k^Tb$$

$$\nabla f=\begin{bmatrix}\frac{\partial f}{\partial x_1}\\ \frac{\partial f}{\partial x_2}\\ \vdots\\ \frac{\partial f}{\partial x_m}\end{bmatrix}=0\Leftrightarrow\begin{bmatrix}a_1^TAx\\ a_2^TAx\\ \vdots\\ a_m^TAx\end{bmatrix}=\begin{bmatrix}a_1^Tb\\ a_2^Tb\\ \vdots\\ a_m^Tb\end{bmatrix}=A^TAx=A^Tb$$: σύστημα των κανονικών εξισώσεων είναι $m\times m$.

Έστω x η λύση του συστήματος των κανονικών εξισώσεων και $y\in\mathbb{R}^m$, $y\neq x$.
$f(y)=\|b-Ay\|^2=(b-Ay,\,b-Ay)=(b-Ax+A(x-y),\,b-Ax+A(x-y))=$
$=(r_x+A(x-y),\,r_x+A(x-y))=(r_x,r_x)+(r_x,A(x-y))+(A(x-y),r_x)+$
$+(A(x-y),A(x-y))=f(x)+r_x^TA(x-y)+(x-y)^TA^Tr_x+(x-y)^TA^TA(x-y)=$

↳ πίσω

$$\boxed{\begin{aligned}&(x-y)^TA^Tr_x=(x-y)^TA^T(b-Ax)=(x-y)^T(A^Tb-A^TAx)=0\\ &r_x^TA(x-y)=(x-y)^TA^Tr_x=0\end{aligned}}$$

$= f(x) + (x-y, A^TA(x-y)) \geq f(x)$

Ο A^TA είναι συμμετρικός και μη αρνητικά ορισμένος.
Έστω $z \in \mathbb{R}^m \setminus \{0\}$: $(z, A^TAz) = (Az, Az) = \|Az\|^2 \geq 0 \Rightarrow$
A^TA μη αρνητικά ορισμένος

Ο πίνακας A^TA είναι θετικά ορισμένος αν οι στήλες του A είναι γραμμικά ανεξάρτητες. Έστω ότι οι στήλες του A είναι γραμ. αν., ισοδύναμα έχουμε ότι για κάθε $z \in \mathbb{R}^m \setminus \{0\}$, $Az = z_1a_1 + z_2a_2 + \dots + z_ma_m \neq 0 \Leftrightarrow$
$\|Az\|^2 > 0 \Leftrightarrow (z, A^TAz) > 0 \Leftrightarrow A^TA$ Θ.Ο.

Όταν οι στήλες του A είναι γρ. αν. λέμε ότι το σύστημα των κανονικών εξισώσεων είναι πλήρους βαθμού, αλλιώς είναι ελλειπούς βαθμού.

Στην περίπτωση ελλειπούς βαθμού, αποδεικνύεται ότι στο σύστημα των κανονικών εξισώσεων έχει απειρία λύσεων.

Άσκηση

$$A = \begin{bmatrix} 1 & 3 & 4 \\ 1 & 3 & 2 \\ 1 & 1 & 0 \\ -1 & -1 & 2 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad A^T A = \begin{bmatrix} 4 & 8 & 4 \\ 8 & 20 & 16 \\ 4 & 16 & 24 \end{bmatrix}, \quad A^T b = \begin{bmatrix} 2 \\ 6 \\ 8 \end{bmatrix}$$

$$A^T A = L \cdot L^T = \begin{bmatrix} 2 & & \\ 4 & 2 & \\ 2 & 4 & 2 \end{bmatrix} \begin{bmatrix} 2 & 4 & 2 \\ & 2 & 4 \\ & & 2 \end{bmatrix}, \quad A^T A x = A^T b \Leftrightarrow L L^T x = A^T b \Leftrightarrow$$

$$\begin{cases} Ly = A^T b \\ L^T x = y \end{cases}$$

$$\begin{bmatrix} 2 & & \\ 4 & 2 & \\ 2 & 4 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 6 \\ 8 \end{bmatrix} \qquad \begin{bmatrix} 2 & 4 & 2 \\ & 2 & 4 \\ & & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \qquad b - Ax = \begin{bmatrix} 1 - 3/2 \\ 1 - 1/2 \\ 1 - 1/2 \\ 1 - 1/2 \end{bmatrix} = \begin{bmatrix} -1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix}$$

$y^T = [1 \ 1 \ 1]$ $\qquad x = [1 \ -\frac{1}{2} \ \frac{1}{2}]$ $\qquad \|r_x\| = 1$

Συνεπώς για μεγάλα m, ο δείκτης κατάστασης του $A^T A$ είναι πολύ μεγάλος, οπότε λόγω σφαλμάτων στρογγύλευσης, η μέθοδος δεν είναι αξιόπιστη.

Παραγοντοποίηση QR

Θεώρημα: Έστω $Q \in \mathbb{C}^{n,m}$, $m \leq n$, ορθοκανονικός πίνακας, δηλαδή $(q_i, q_j) = \delta_{ij}$, q_i η i στήλη του Q.

Τότε $\|Qx\| = \|x\| \ \forall x \in \mathbb{C}^m$

$$\|Qx\|^2 = (Qx, Qx) = (x, Q^T Q x) = (x, I_m x) = (x, x) = \|x\|^2$$

Έστω ότι $A = QR$, $Q \in \mathbb{R}^{n,m}$ με $(q_i, q_j) = \delta_{ij}$ και $R \in \mathbb{R}^{m,m}$ άνω τριγωνικός με θετικά διαγώνια στοιχεία.

Πρόβλημα: εύρεση x, ώστε να ελαχιστοποιεί την $\|r_x\| = \|b - Ax\|$

Έστω $\tilde{Q} \in \mathbb{R}^{n,n-m}$ τέτοιος ώστε να συμπληρώνει τον Q, δηλαδή ο $[Q | \tilde{Q}]$ να 'ναι ορθογώνιος. Αποδεικνύεται ότι υπάρχει τέτοιος $\tilde{Q}$, τότε και ο ανάστροφός του: $\begin{bmatrix} Q^T \\ \tilde{Q}^T \end{bmatrix}$ είναι ορθογώνιος.

$$\|r_x\| = \|b - Ax\| = \|b - QRx\| = \left\| \begin{bmatrix} Q^T \\ \tilde{Q}^T \end{bmatrix} (b - QRx) \right\| = \left\| \begin{matrix} Q^T b - Rx \\ \tilde{Q}^T b \end{matrix} \right\|$$

$$\|r_x\|^2 = \left\| \begin{matrix} Q^T b - Rx \\ \tilde{Q}^T b \end{matrix} \right\|^2 = \|Q^T b - Rx\|^2 + \|\tilde{Q}^T b\|^2$$

Χρησιμοποιήσαμε πυθαγόρειο: $\left\| \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \right\|^2 = \left\| \begin{bmatrix} z_1 \\ 0 \end{bmatrix} \right\|^2 + \left\| \begin{bmatrix} 0 \\ z_2 \end{bmatrix} \right\|^2 = \|z_1\|^2 + \|z_2\|^2$

$$\left\| \begin{matrix} z_1 \\ z_2 \end{matrix} \right\| = \begin{bmatrix} z_1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ z_2 \end{bmatrix}$$

$$\min_{x \in \mathbb{R}^m} \|r_x\|^2 = \min_{x \in \mathbb{R}^m} \|Q^T b - Rx\|^2 + \|\tilde{Q}^T b\|^2.$$

$\min_{x \in \mathbb{R}^m} \|Q^T b - Rx\| = 0 \Leftrightarrow Q^T b - Rx = 0 \Leftrightarrow Rx = Q^T b$: x η μοναδική λύση του συστήματος: $x = R^{-1} Q^T b$.

Αν οι στήλες του A είναι γραμμ. ανεξ. τότε υπάρχει $Q \in \mathbb{R}^{n,m}$ με $(q^{(i)}, q^{(j)}) = \delta_{ij}$ και άνω τριγωνικός πίνακας R με $r_{ii} > 0$ ώστε $A = QR$. Η παραγοντοποίηση είναι μοναδική.

Απόδειξη

$$A = QR \iff [a_1, a_2 \dots a_m] = [q^{(1)}\, q^{(2)} \dots q^{(m)}] \cdot \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ & r_{22} & \cdots & r_{2m} \\ & & \ddots & \vdots \\ & & & r_{mm} \end{bmatrix}$$

$a_1 = r_{11} q^{(1)} \Rightarrow \|a_1\| = \|r_{11} q^{(1)}\| = r_{11} \|q^{(1)}\| = r_{11}, \; q^{(1)} = \frac{a_1}{r_{11}}$

$a_2 = r_{12} q^{(1)} + r_{22} q^{(2)} \Rightarrow (a_2, q^{(1)}) = r_{12}(q^{(1)}, q^{(1)}) + r_{22}(q^{(2)}, q^{(1)}) \iff$

$r_{12} = (a_2, q^{(1)}), \; r_{22} q^{(2)} = a_2 - r_{12} q^{(1)}$

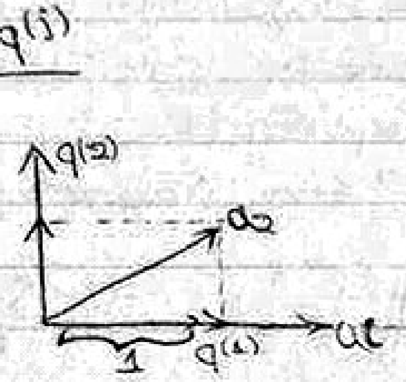
$r_{22} = \|a_2 - r_{12} q^{(1)}\| \;, \; q^{(2)} = \frac{a_2 - r_{12} q^{(1)}}{r_{22}}$

Επαγωγικά: $a_i = r_{1i} q^{(1)} + r_{2i} q^{(2)} + \dots + r_{i-1,i} q^{(i-1)} + r_{ii} q^{(i)}$.

$r_{ji} = (a_i, q^{(j)}) \;, \; j = 1, 2, \dots, i-1 \;, \; r_{ii} = \|a - \sum_{j=1}^{i-1} r_{ji} q^{(j)}\|$,

$q^{(i)} = \frac{a_i - \sum_{j=1}^{i-1} r_{ji} q^{(j)}}{r_{ii}}$

$\boxed{\text{Gram-Schmidt}}$



Αλγόριθμος Gram-Schmidt ορθογωνοποίησης

Δεδομένα: $A \in \mathbb{R}^{n \times m}$

Για $i = 1(1)n$
    $q^{(i)} = a_i$
    Για $j = 1(1)i-1$
        $r_{ji} = (a_i, q^{(j)})$
        $q^{(i)} = q^{(i)} - r_{ji} q^{(j)}$
    Τέλος 'Για'
    $r_{ii} = \|q^{(i)}\|$
    $q^{(i)} = q^{(i)} / r_{ii}$
Τέλος 'για'

$(q^{(i)}, q^{(j)}) = (a_i - \sum_{k=1}^{j-1} r_{ki} q^{(k)}, q^{(j)})$
$= (a_i, q^{(j)}) + \sum_{k=1}^{j-1} (q^{(k)}, q^{(j)}) \to 0$

Σημείωση.