ΓΡΑΜΜΙΚΗ ΑΛΓΕΒΡΑ
Διάλεξη (XI) 21&22

Λύσεις ελαχίστων τετραγώνων
Ορίζουσες μητρώων

Ε. Στεφανόπουλος & Ε. Γαλλόπουλος

Τμήμα Μηχανικών Η/Υ & Πληροφορικής, Πανεπιστήμιο Πατρών

8-11 Δεκεμβρίου 2022

Εισαγωγή

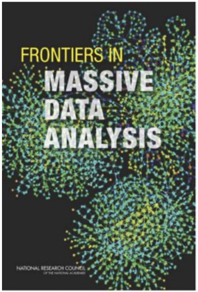
- Πολλές εφαρμογές οδηγούν σε μετρήσεις, π.χ. ζεύγη τιμών $(\xi_1, \psi_1), \ldots, (\xi_m, \psi_m)$ (τα x_j διαφορετικά) και αναζητούμε απλή συνάρτηση f που να προσαρμόζεται στις τιμές αυτές (να τις μοντελοποιεί) όσο γίνεται καλύτερα. Εννοούμε ότι θα θέλαμε οι τιμές $f(\xi_1), \ldots, f(\xi_m)$ να είναι 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.

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**.

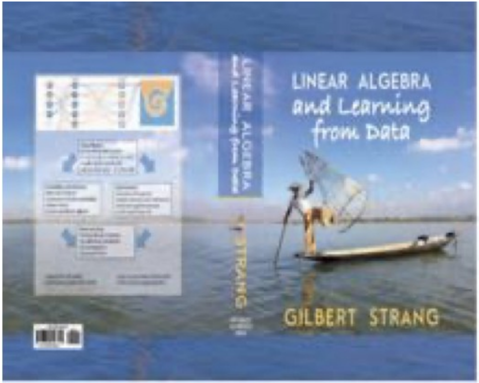
The 7 Computational Giants of Massive Data Analysis (National Research Council, 2013)



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**.



Παράδειγμα

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

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

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) $c = [\gamma_0, \ldots, \gamma_{n-1}]^\top$.

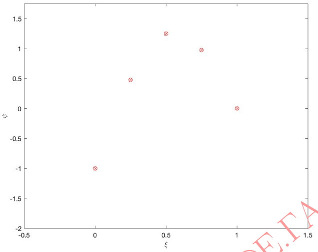
$$A = \begin{pmatrix} 1 & \xi_1 & \cdots & \xi_1^{n-1} \\ 1 & \xi_2 & & \xi_2^{n-1} \\ \vdots & & & \vdots \\ \vdots & & \vdots & \\ 1 & \xi_m & \cdots & \xi_m^{n-1} \end{pmatrix}, c = \begin{pmatrix} \gamma_0 \\ \vdots \\ \gamma_{n-1} \end{pmatrix}. \text{ και } \quad y = \begin{pmatrix} \psi_1 \\ \vdots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

- 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 $\arg\min_{c\in\mathbb{R}^n} \|y - Ac\|$

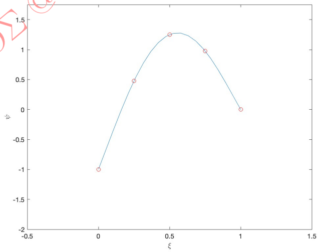
Παράδειγμα

Δίνονται

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



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

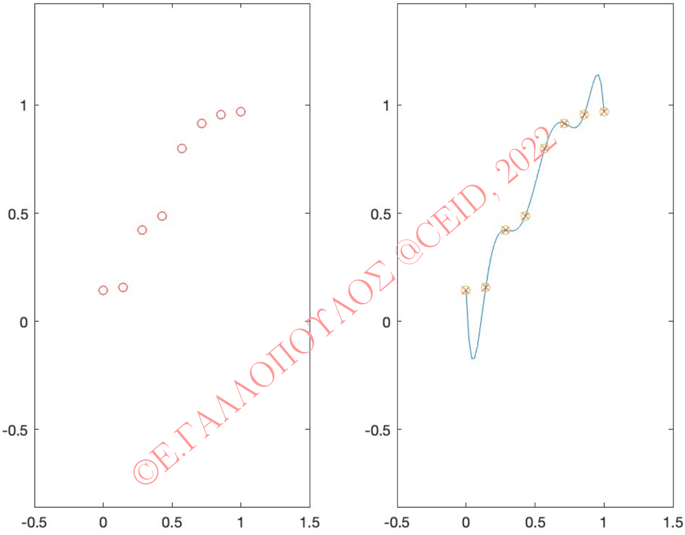


$$A^{-1}y = \begin{pmatrix} -1.0000 \\ 6.1699 \\ 2.1506 \\ -14.6409 \\ 7.3204 \end{pmatrix}$$

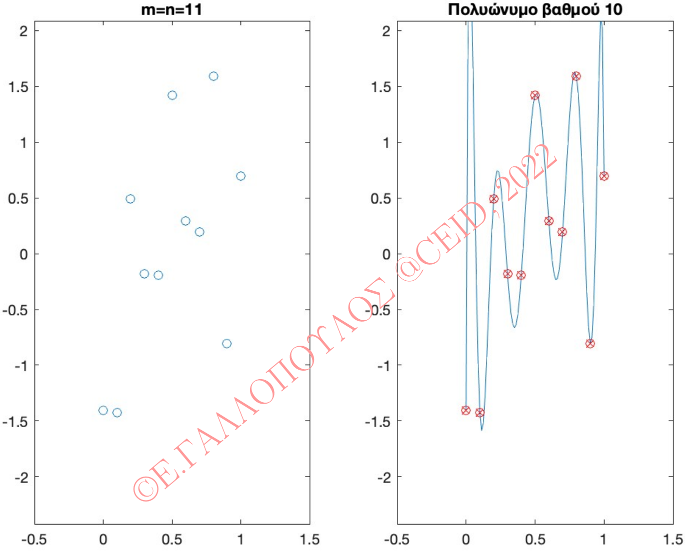
Από τα σημεία $x, y \in \mathbb{R}^5$, κατασκευάσαμε 4-βάθμιο πολυώνυμο p για το οποίο $p(\xi_i) = \psi_i, i = 1, \ldots, 5$.

- 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 $\xi_1, \ldots, \xi_m$, τότε θα μπορούσαμε να υπολογίσουμε ακριβώς τους συντελεστές $\gamma_0, \ldots, \gamma_{n-1}$ 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 άγνωστη 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 $\phi_j(\xi)$, δηλ.

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

- 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 $y \in \text{Range}(A)$, 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.)

Στη συνέχεια αλλάζουμε θα χρησιμοποιούμε το σύμβολο x για το διάνυσμα αγνώστων αντί του c.

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

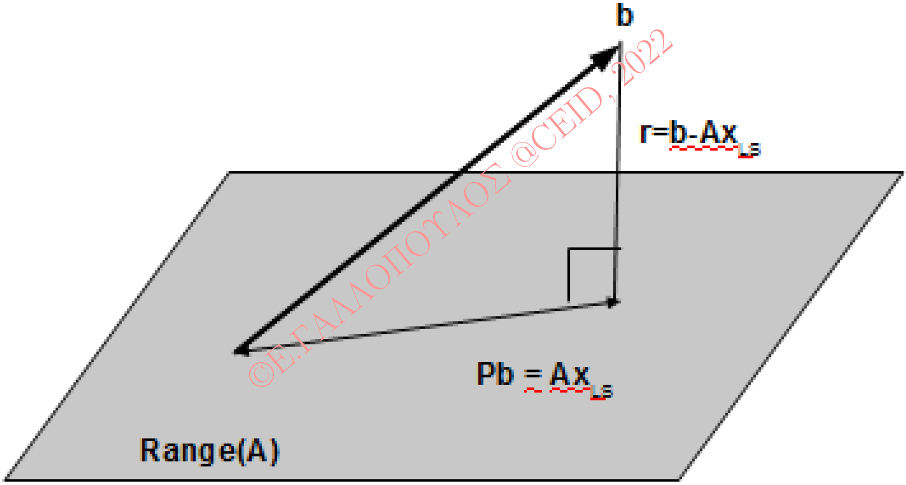
$$\mathcal{X} = \{x \in \mathbb{R}^n : \text{ ελαχιστοποιεί το } \rho(x) = \|Ax - y\|_2\}.$$

- 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 Range(A) άρα $r \in$ Null$(A^\top)$. Επομένως, για οποιοδήποτε αυθαίρετο διάνυσμα Az του 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 $A^\top Ax = A^\top b$

<u>Υπάρχει λύση? Είναι μοναδική?</u>

- <u>Επιλυσιμότητα</u> 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).
- <u>Παρατήρηση</u>: 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**.

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

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

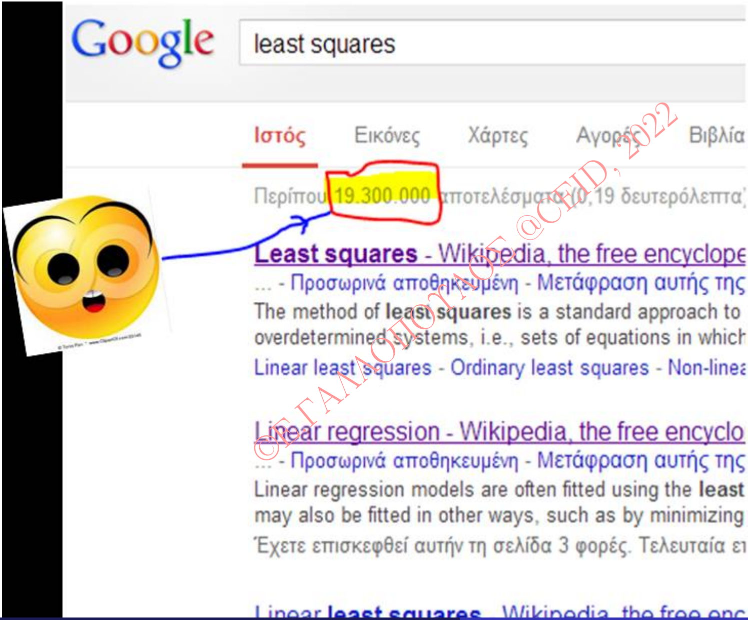
όπου $r_j(x) = \psi_j - f(t_j;x)$, $j = 1, \ldots, m$ είναι οι τιμές των υπολοίπων στα σημεία $t_1, \ldots, t_m$.

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

Ορισμός (Wikipedia)

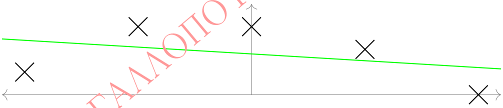
Μηχανική μάθηση είναι υποπεδίο της επιστήμης των υπολογιστών, που δίνει στους υπολογιστές την ικανότητα να μαθαίνουν, χωρίς να έχουν ρητά προγραμματιστεί.

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

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.

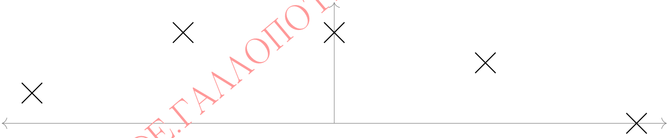
Ορολογία: Στη Στατιστική λέγεται γραμμική παλινδρόμηση (linear regression) Περί τίνος πρόκειται; Είναι η προσπαθεια να κατασκευάσουμε ένα απλό (γραμμικό) μοντέλο για τις παρατηρήσεις που αντιστοιχούν στα m ζεύγη των παραπάνω τιμών. Πώς μετράμε το σφάλμα? Το ορίζουμε ως το άθροισμα των τετραγώνων των 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

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

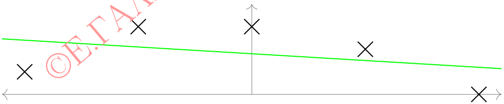
$$A^\top A = \begin{pmatrix} 5.0 & 0.0 \\ 0.0 & 2.5 \end{pmatrix}, A^\top b = \begin{pmatrix} 0.9 \\ -0.15 \end{pmatrix}$$

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



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

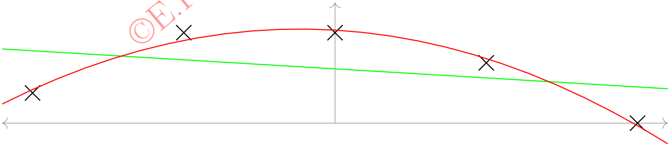
$$x_{LS} = (A^\top A)^{-1}A^\top b = \begin{pmatrix} 0.3086 \\ -0.0600 \\ -0.2571 \end{pmatrix}$$

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

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

σφάλμα $\|b - Ax_{LS}\|_2 = 0.0338$

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



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".

The BigChaos Solution to the Netflix

Andreas Töscher and Michael Jahrer
commendo research & consulting
Neuer Weg 23, A-8580 Köflach, Aust
{andreas.toescher,michael.jahrer}@comme

Robert M. Bell*
AT&T Labs - Research
Florham Park, NJ

Input: A matrix $\mathbf{P}$ with all previous probe predictions. $\mathbf{P}$ always includes a constant predictor (a column with ones).

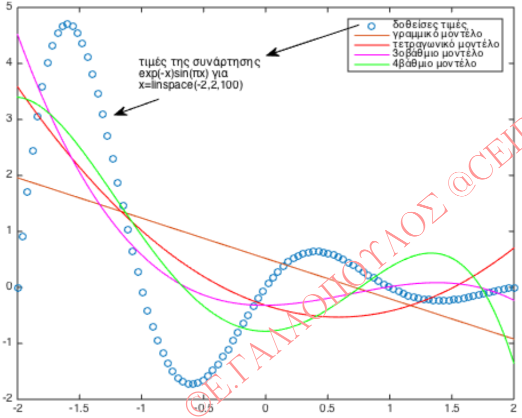
Exclude the Probe ratings $\mathbf{r}$ from the training set.
Initialize the weights.
$RMSE_{best} = \infty$
$RMSE_{epoch} = 1000$
$epochs = 0$
while $RMSE_{epoch} \leq RMSE_{best}$ **do**
  Train one epoch.
  if $RMSE_{epoch} \leq RMSE_{best}$ **then**
    $RMSE_{best} = RMSE_{epoch}$
    Save the current 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

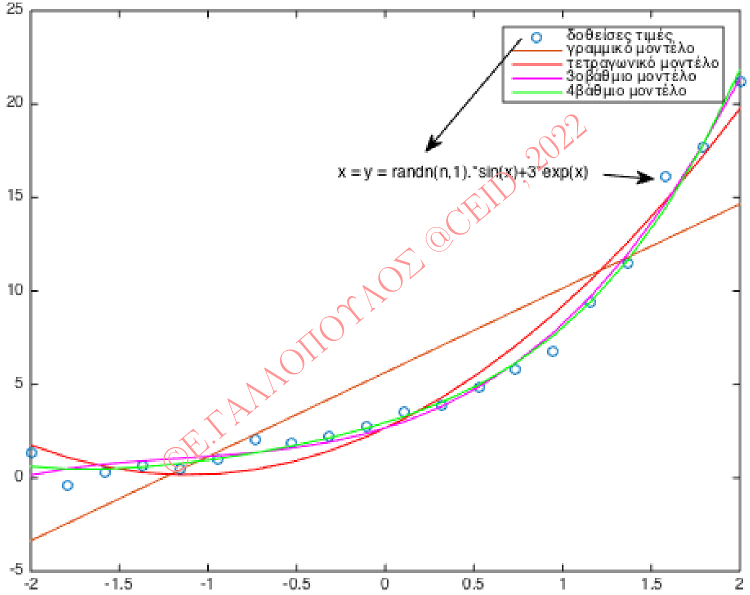
Χρησιμοποιούμε τον παρακάτω κώδικα MATLAB για διάφορες επιλογές $x, \ y$

```
% 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)



Ειδικές τιμές μητρώων

Παρόλο που ένα μητρώο $A \in \mathbb{R}^{m\times n}$ περιέχει $m \times n$ αριθμούς, σε αυτό αντιστοιχούν επίσης ορισμένες ειδικές τιμές συνδεδεμένες με αυτό. Αυτές είναι βαθμωτές συναρτήσεις μητρώου ($f : \mathbb{R}^{n\times n}\mathbb{R}$), και αποκαλύπτουν σημαντικές πληροφορίες για το μητρώο.

είδος	*συμβολισμός*	*παρατηρήσεις*
διαστάσεις	m, n	αριθμ. γραμμών και αριθμ. στηλών
αραιότητα	$\text{nnz}(A)$	πλήθος μη μηδενικών
τάξη	$\text{rank}(A)$	μέγιστος αριθμ. γ.α. γραμμών (ή στηλών)
νόρμα	$\|A\|$	μετρική ((βάρους)) , π.χ. $\sup_{x\neq 0} \frac{\|Ax\|}{\|x\|}$
ίχνος	$\text{trace}(A)$	$(m = n)$???
ορίζουσα	$\det(A)$	$(m = n)$???
ιδιοτιμές	???	$(m = n)$???
ιδιάζουσες τιμές	???	???

Ίχνος

τετραγωνικού μητρώου

- Χαρακτηριστική τιμή για κάθε τετραγωνικό μητρώο ίση με το άθροισμα των διαγώνιων στοιχείων του.
- Βαθμωτή συνάρτηση

$$\text{trace} : \mathbb{R}^{n\times n} \longrightarrow \mathbb{R}$$

τέτοια ώστε

$$\text{trace}(A) := \alpha_{1,1} + \alpha_{2,2} + \cdots + \alpha_{n,n}$$

<u>Ιδιότητες</u>

1. $\text{trace}(\gamma A + \delta B) = \gamma \text{trace}(A) + \delta \text{trace}(B)$
2. $\text{trace}(AB) = \text{trace}(BA)$

<u>Σημαντικό επακόλουθο:</u> $\text{trace}(^{-1}) = \text{trace}(A)$

- Μοναδικό χαρακτηριστικό μέγεθος (βαθμωτός) που υπάρχει για κάθε τετραγωνικό μητρώο.

$$\det : \mathbb{R}^{n \times n} \longrightarrow \mathbb{R}$$

- ((Συμπυκνώνει)) πληροφορίες σχετικά με το μητρώο.
- Η πιο γνωστή: Ένα μητρώο A είναι αντιστρέψιμο αν και μόνον αν $\det(A) \neq 0$.
- Υπολογιστικά θέματα: Ο πιο πρακτικός τρόπος υπολογισμού του για μέτριο ή μεγάλο n είναι ως παραπροϊόν της παραγοντοποίησης LU με το αντίστοιχο κόστος.
- Έχει χρήσιμη γεωμετρική ερμηνεία ως ((όγκος)) στο n-διάστατο χώρο.

Ιστορικά

- Η μελέτη των οριζουσών προηγήθηκε της Γραμμικής Άλγεβρας: Ο Gauss (1801) χρησιμοποίησε τον όρο εννοώντας την διακρίνουσα πολυωνύμων 4ου βαθμού. Ο όρος με τη σημερινή έννοια εισήχθη από τον Cauchy (1812).
- Ο Sylvester ((βάπτισε)) τις "μήτρες" για να αναδείξει ότι ((γεννούν)) ορίζουσες.

150 *On a new Class of Theorems.* [25

its form of greatest generality. For this purpose we must commence, not with a square, but with an oblong arrangement of terms consisting, suppose, of m lines and n columns. This will not in itself represent a determinant, but is, as it were, a Matrix out of which we may form various systems of determinants by fixing upon a number p, and selecting at will p lines and p

37] *Linearly Equivalent Quadratic Functions.* 247

I have in previous papers defined a "Matrix" as a rectangular array of terms, out of which different systems of determinants may be engendered, as from the womb of a common parent; these cognate determinants being by no means isolated in their relations to one another, but subject to certain simple laws of mutual dependence and simultaneous deperition. The con-

Τύπος ορίζουσας όταν $n = 2$

$n = 2$

$$A = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} \Rightarrow \det(A) = \begin{vmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{vmatrix} := \alpha_{11}\alpha_{22} - \alpha_{21}\alpha_{12}.$$

Στη συνέχεια θα χρησιμοποιήσουμε έναν αξιωματικό ορισμό της ορίζουσας από τον οποίο θα προκύψει αυτός και άλλο τύποι υπολογισμού.

Παρατήρηση: Να επαληθεύσετε την παρακάτω παραγοντοποίηση.

$$\begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \alpha_{21}\alpha_{11}^{-1} & 1 \end{pmatrix} \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ 0 & \alpha_{22} - \alpha_{21}\alpha_{11}^{-1}\alpha_{12} \end{pmatrix}$$

Βοηθητικοί ορισμοί

- Έστω δ.χ. $U_1, \ldots, U_n$ και W επί του ιδίου σώματος $\mathcal{F}$ και έστω η απεικόνιση

$$f : U_1 \times U_2 \times \cdots \times U_n \to W$$

Τότε η απεικόνιση ονομάζεται πλειογραμμική (ή n-γραμμική) αν είναι γραμμική ως πρός κάθε όρισμα αν τα υπόλοιπα ορίσματα μείνουν αμετάβλητα, π.χ. αν για $i = 1, \ldots, n$, και $u_j \in U_j$, η απεικόνιση

$$x \to f\big(u_1, \ldots, u_{i-1}, x, u_{i+1}, \ldots, u_n\big)$$

είναι γραμμική.

- Έστω μια μετάθεση $\pi = (\pi_1, \ldots, \pi_n)$ των αριθμών 1 ως n. Ονομάζουμε φυσική διάταξη την n-άδα $(1, 2, \ldots, n)$. Τότε πρόσημο της μετάθεσης ονομάζεται ο αριθμός

$$\sigma(\pi) = \begin{cases} +1 & \text{αν η } \pi \text{ επανέρχεται στη φυσική διάταξη} \\ & \text{με άρτιο πλήθος εναλλαγών} \\ -1 & \text{αν η } \pi \text{ επανέρχεται στη φυσική διάταξη} \\ & \text{με περιττό πλήθος εναλλαγών.} \end{cases}$$

Ορίζοντας την ορίζουσα (αξιωματικά)

Η ορίζουσα $\det(A)$ ενός τετραγωνικού μητρώου $A \in \mathbb{R}^{n\times n}$ είναι μία συνάρτηση

$$\det : \mathbb{R}^{n\times n} \longrightarrow \mathbb{R}$$

που έχει τις εξής 3 ιδιότητες για ένα μητρώο[α΄] $= [a_1, \cdots, a_n] \in \mathbb{R}^{n\times n}$:

1. Αν $a_i = a_j$ για $i \neq j$ τότε $\det([a_1, \ldots, a_n]) = 0$ (εκφυλισμός).
2. $\det([e_1, \ldots, e_n]) = 1$ (κανονικότητα).
3. Η συνάρτηση είναι γραμμική ως προς τις n στήλες:

$$\begin{aligned}\det([a_1, \ldots, a_{j-1}, \gamma a + \delta b, a_{j+1}, \ldots, a_n]) = &\ \gamma \det([a_1, \ldots, a_{j-1}, a, a_{j+1}, \ldots, a_n]) \\ &+ \delta \det([a_1, \ldots, a_{j-1}, b, a_{j+1}, \ldots, a_n])\end{aligned}$$

Συναρτήσεις με αυτή την ιδιότητα λέγονται πλειογραμμικές.

[α΄] Αναφερόμαστε στις στήλες, αντίστοιχες διατυπώσεις υπάρχουν και ως προς τις γραμμές).

Σημαντικά επακόλουθα

Έστω το μητρώο (σε μορφή στηλών) $= [a_1, \cdots, a_n] \in \mathbb{R}^{n \times n}$.

- Αν εναλλάξουμε 2 στήλες ενός μητρώου, η ορίζουσα του νέου μητρώου έχει ίδιο μέτρο αλλά αντίθετο πρόσημο. Γι' αυτό η ορίζουσα χαρακτηρίζεται ως εναλλασσόμενη απεικόνιση (alternating map).

Σημαντικά επακόλουθα

Έστω το μητρώο (σε μορφή στηλών) $= [a_1, \cdots, a_n] \in \mathbb{R}^{n \times n}$.

- Αν εναλλάξουμε 2 στήλες ενός μητρώου, η ορίζουσα του νέου μητρώου έχει ίδιο μέτρο αλλά αντίθετο πρόσημο. Γι' αυτό η ορίζουσα χαρακτηρίζεται ως εναλλασσόμενη απεικόνιση (alternating map).
- δείγμα απόδειξης

$$
\begin{aligned}
0 &= \det([a_1 + a_2, a_1 + a_2]) = \det([a_1, a_1]) + \det([a_1, a_2]) + \det([a_2, a_1]) + \det([a_2, a_2]) \\
0 &= \det([a_1, a_2]) + \det([a_2, a_1])
\end{aligned}
$$

Σημαντικά επακόλουθα

Έστω το μητρώο (σε μορφή στηλών) $= [a_1, \cdots, a_n] \in \mathbb{R}^{n\times n}$.

- Αν εναλλάξουμε 2 στήλες ενός μητρώου, η ορίζουσα του νέου μητρώου έχει ίδιο μέτρο αλλά αντίθετο πρόσημο. Γι' αυτό η ορίζουσα χαρακτηρίζεται ως εναλλασσόμενη απεικόνιση (alternating map).
- δείγμα απόδειξης

$$\begin{aligned} 0 &= \det([a_1 + a_2, a_1 + a_2]) = \det([a_1, a_1]) + \det([a_1, a_2]) + \det([a_2, a_1]) + \det([a_2, a_2]) \\ 0 &= \det([a_1, a_2]) + \det([a_2, a_1]) \end{aligned}$$

- Αν οι στήλες μητρώου είναι γραμμικά εξαρτημένες, η ορίζουσα είναι 0.

Σημαντικά επακόλουθα

Έστω το μητρώο (σε μορφή στηλών) $= [a_1, \cdots, a_n] \in \mathbb{R}^{n\times n}$.

- Αν εναλλάξουμε 2 στήλες ενός μητρώου, η ορίζουσα του νέου μητρώου έχει ίδιο μέτρο αλλά αντίθετο πρόσημο. Γι΄ αυτό η ορίζουσα χαρακτηρίζεται ως εναλλασσόμενη απεικόνιση (alternating map).
- δείγμα απόδειξης

$$\begin{aligned} 0 &= \det([a_1 + a_2, a_1 + a_2]) = \det([a_1, a_1]) + \det([a_1, a_2]) + \det([a_2, a_1]) + \det([a_2, a_2]) \\ 0 &= \det([a_1, a_2]) + \det([a_2, a_1]) \end{aligned}$$

- Αν οι στήλες μητρώου είναι γραμμικά εξαρτημένες, η ορίζουσα είναι 0.
- δείγμα απόδειξης

$$\det([a_1, a_2, \gamma a_1 + \delta a_2]) \quad = \quad \gamma\det([a_1, a_2, a_1]) + \delta\det([a_1, a_2, a_2]) = 0$$

Σημαντικά επακόλουθα

Έστω το μητρώο (σε μορφή στηλών) $= [a_1, \cdots, a_n] \in \mathbb{R}^{n\times n}$.

- Αν εναλλάξουμε 2 στήλες ενός μητρώου, η ορίζουσα του νέου μητρώου έχει ίδιο μέτρο αλλά αντίθετο πρόσημο. Γι' αυτό η ορίζουσα χαρακτηρίζεται ως εναλλασσόμενη απεικόνιση (alternating map).
- δείγμα απόδειξης

$$
\begin{aligned}
0 &= \mathrm{det}([a_1 + a_2, a_1 + a_2]) = \mathrm{det}([a_1, a_1]) + \mathrm{det}([a_1, a_2]) + \mathrm{det}([a_2, a_1]) + \mathrm{det}([a_2, a_2]) \\
0 &= \mathrm{det}([a_1, a_2]) + \mathrm{det}([a_2, a_1])
\end{aligned}
$$

- Αν οι στήλες μητρώου είναι γραμμικά εξαρτημένες, η ορίζουσα είναι 0.
- δείγμα απόδειξης

$$
\mathrm{det}([a_1, a_2, \gamma a_1 + \delta a_2]) = \gamma \mathrm{det}([a_1, a_2, a_1]) + \delta \mathrm{det}([a_1, a_2, a_2]) = 0
$$

- Αν προσθέσουμε δύο στήλες η ορίζουσα παραμένει αμετάβλητη.

Σημαντικά επακόλουθα

Έστω το μητρώο (σε μορφή στηλών) $= [a_1, \cdots, a_n] \in \mathbb{R}^{n \times n}$.

- Αν εναλλάξουμε 2 στήλες ενός μητρώου, η ορίζουσα του νέου μητρώου έχει ίδιο μέτρο αλλά αντίθετο πρόσημο. Γι᾽ αυτό η ορίζουσα χαρακτηρίζεται ως εναλλασσόμενη απεικόνιση (alternating map).
- δείγμα απόδειξης

$$\begin{aligned} 0 &= \det([a_1 + a_2, a_1 + a_2]) = \det([a_1, a_1]) + \det([a_1, a_2]) + \det([a_2, a_1]) + \det([a_2, a_2]) \\ 0 &= \det([a_1, a_2]) + \det([a_2, a_1]) \end{aligned}$$

- Αν οι στήλες μητρώου είναι γραμμικά εξαρτημένες, η ορίζουσα είναι 0.
- δείγμα απόδειξης

$$\det([a_1, a_2, \gamma a_1 + \delta a_2]) = \gamma \det([a_1, a_2, a_1]) + \delta \det([a_1, a_2, a_2]) = 0$$

- Αν προσθέσουμε δύο στήλες η ορίζουσα παραμένει αμετάβλητη.
- δείγμα απόδειξης

$$\det([a_1, a_1 + a_2]) = \det([a_1, a_1]) + \det([a_1, a_2]) = \det([a_1, a_2])$$

Μπορούμε να την υπολογίσουμε βάσει των παραπάνω ιδιοτήτων:

$$\begin{aligned}\det([a_1, a_2]) &= \det([\alpha_{11} e_1 + \alpha_{21} e_2, \alpha_{12} e_1 + \alpha_{22} e_2]) \\ &= \alpha_{11} \det([e_1, \alpha_{12} e_1 + \alpha_{22} e_2]) + \alpha_{21} \det([e_2, \alpha_{12} e_1 + \alpha_{22} e_2])\end{aligned}$$

$$= \alpha_{11}\alpha_{12} \det([e_1, e_1]) + \alpha_{11}\alpha_{22} \det([e_1, e_2]) + \alpha_{21}\alpha_{12} \det([e_2, e_1]) + \alpha_{21}\alpha_{22} \det([e_2, e_2])$$

$$\det([a_1, a_2]) = \alpha_{11}\alpha_{22} + \alpha_{21}\alpha_{12} \det([e_2, e_1]) = \boldsymbol{\alpha_{11}\alpha_{22} - \alpha_{21}\alpha_{12}}$$

Προσοχή: πολλοί όροι μηδενίζονται (λόγω ((εκφυλισμού))).

- Όταν $n = 2$, αντί για $4 = 2^2$ όρους, έχουμε μόνον 2.
- με $n = 3$, αντί για $27 = 3^3$ όρους έχουμε 6 ...
- ... γενικά, αντί για για n^n όρους, έχουμε $n!$.

Υπολογισμός ορίζουσας από αλγεβρικά συμπληρώματα

Ανάπτυγμα Laplace

$$\det(A) = \alpha_{11}\det(A_{11}) + (-1)^{1+2}\alpha_{12}\det(A_{12}) + \cdots + (-1)^{1+n}\alpha_{1n}\det(A_{1n})$$

- Προσωρινός συμβολισμός Με $A_{ij} \in \mathbb{R}^{(n-1)\times(n-1)}$ συμβολίζεται το μητρώο που προκύπτει αν διαγράψουμε από το A τη γραμμή i και τη στήλη j. Αποκαλείται ελάσσον (minor) του στοιχείου α_{ij}.
- Το $\psi_{i,j} := (-1)^{i+j}\det(A_{ij})$ λέγεται αλγεβρικό συμπλήρωμα του στοιχείου στη θέση (i, j).

Γενικά ισχύει ότι μπορούμε να εκφράσουμε την ορίζουσα ως προς οποιαδήποτε γραμμή (ή στήλη):

$$\det(A) = \alpha_{i,1}\psi_{i,1} + \alpha_{i,2}\psi_{i,2} + \cdots + \alpha_{i,n}\psi_{i,n}$$

Προσοχή: Στον παραπάνω τύπο εκφράσαμε την ορίζουσα βάσει οριζουσών μεγέθους $(n-1)\times(n-1)$.
Σύσταση: Στην εφαρμογή του τύπου, αξίζει να επιλέγουμε τη γραμμή ή στήλη με τα περισσότερα μηδενικά!

Τύπος για $n = 3$

$$\begin{vmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{vmatrix} := \alpha_{11} \begin{vmatrix} \alpha_{22} & \alpha_{23} \\ \alpha_{32} & \alpha_{33} \end{vmatrix} - \alpha_{12} \begin{vmatrix} \alpha_{21} & \alpha_{23} \\ \alpha_{31} & \alpha_{33} \end{vmatrix} + \alpha_{13} \begin{vmatrix} \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \end{vmatrix}$$

Άρα

Τύπος για $n = 3$

$$\begin{vmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{vmatrix} := \alpha_{11} \begin{vmatrix} \alpha_{22} & \alpha_{23} \\ \alpha_{32} & \alpha_{33} \end{vmatrix} - \alpha_{12} \begin{vmatrix} \alpha_{21} & \alpha_{23} \\ \alpha_{31} & \alpha_{33} \end{vmatrix} + \alpha_{13} \begin{vmatrix} \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \end{vmatrix}$$

Άρα

$$\det(A) = \alpha_{11}\alpha_{22}\alpha_{33} - \alpha_{11}\alpha_{23}\alpha_{32} - \alpha_{12}\alpha_{21}\alpha_{33} + \alpha_{12}\alpha_{23}\alpha_{31} + \alpha_{13}\alpha_{21}\alpha_{32} - \alpha_{13}\alpha_{22}\alpha_{31}$$

Τύπος για $n = 3$

$$\begin{vmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{vmatrix} := \alpha_{11} \begin{vmatrix} \alpha_{22} & \alpha_{23} \\ \alpha_{32} & \alpha_{33} \end{vmatrix} - \alpha_{12} \begin{vmatrix} \alpha_{21} & \alpha_{23} \\ \alpha_{31} & \alpha_{33} \end{vmatrix} + \alpha_{13} \begin{vmatrix} \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \end{vmatrix}$$

Άρα

$$\det(A) = \alpha_{11}\alpha_{22}\alpha_{33} - \alpha_{11}\alpha_{23}\alpha_{32} - \alpha_{12}\alpha_{21}\alpha_{33} + \alpha_{12}\alpha_{23}\alpha_{31} + \alpha_{13}\alpha_{21}\alpha_{32} - \alpha_{13}\alpha_{22}\alpha_{31}$$

Προσέξτε:

$$\det(A) = \alpha_{1\color{red}{1}}\alpha_{2\color{red}{2}}\alpha_{3\color{red}{3}} - \alpha_{1\color{red}{1}}\alpha_{2\color{red}{3}}\alpha_{3\color{red}{2}} - \alpha_{1\color{red}{2}}\alpha_{2\color{red}{1}}\alpha_{3\color{red}{3}} + \alpha_{1\color{red}{2}}\alpha_{2\color{red}{3}}\alpha_{3\color{red}{1}} + \alpha_{1\color{red}{3}}\alpha_{2\color{red}{1}}\alpha_{3\color{red}{2}} - \alpha_{1\color{red}{3}}\alpha_{2\color{red}{2}}\alpha_{3\color{red}{1}}$$

Μεγάλος τύπος ορίζουσας

... και πρόσημο μετάθεσης

Προσοχή: Για μία μετάθεση π των στοιχείων $(1, \ldots, n)$, δηλ. $(\pi(1), \ldots, \pi(n))$, ορίζουμε ως πρόσημο της μετάθεσης, $\sigma(\pi)$, μία τιμή ± 1. Η τιμή θα είναι +1 αν η μετάθεση π προέρχεται από άρτιο αριθμό εναλλαγών ή -1 αν προέρχεται από περιττό αριθμό εναλλαγών. Το αντίστοιχο μητρώο μετάθεσης P_π θα είναι εκείνο για το οποίο $P_\pi \begin{pmatrix} 1 \\ 2 \\ \vdots \\ n \end{pmatrix} = \begin{pmatrix} \pi(1) \\ \pi(2) \\ \cdots \\ \pi(n) \end{pmatrix}$.

Παρατηρήστε ότι το P_π είναι μετάθεση του ταυτοτικού, επομένως $\det(P_\pi) = \pm 1$. Η τιμή του συμπίπτει με το πρόσημο της μετάθεσης.

Μεγάλος τύπος (γιατί έχει $n!$ όρους)

$\det(A) = \sum_{\pi \in \mathcal{S}} \sigma(\pi) \alpha_{1\pi(1)} \cdots \alpha_{n\pi(n)}$ όπου $\mathcal{S}$ είναι το σύνολο των $n!$ μεταθέσεων των στοιχείων $(1, \ldots, n)$ και $\sigma(\pi)$ είναι το πρόσημο της μετάθεσης π.

Ορίζουσες μητρώων με ειδική δομή

Τριγωνικά: Η ορίζουσα κάθε τριγωνικού μητρώου είναι ίση με το γινόμενο όλων των στοιχείων της διαγωνίου του.

Κατά πλοκάδες τριγωνικά Αν A, D τετραγωνικά μητρώα, τότε

$$\det\begin{pmatrix} A & B \\ 0 & D \end{pmatrix} = \det(A)\det(D)$$

Ειδική περίπτωση αν $\hat{A} \in \mathbb{R}^{(n-1)\times(n-1)}$,

$$\det\left(\begin{array}{c|ccc} 1 & \star & \cdots & \star \\ \hline 0_{n-1,1} & \hat{A} & & \end{array}\right) = \det(\hat{A})$$

Προσοχή: Γενικά

$$\begin{aligned} \det\begin{pmatrix} A & B \\ C & D \end{pmatrix} &\neq \det(A)\det(D) \text{ καθώς επίσης} \\ &\neq \det(A)\det(D) - \det(B)\det(C) \end{aligned}$$

΄Αλλες χρήσιμες ιδιότητες

- Το είναι αντιστρέψιμο αν και μόνον αν $\det(A) \neq 0$.
- $\det(A^{-1}) = 1/\det(A^\top)$ αν $\det(A) \neq 0$.
- $\det(A) = \det(A^\top)$
- $\det(AB) = \det(A)\det(B)$

<u>ΠΡΟΣΟΧΗ</u> Η τελευταία ιδιότητα αποκαλύπτει έναν τρόπο υπολογισμού του $\det(A)$ που έχει περίπου όσες πράξεις χρειάζεται η παραγοντοποίηση LU.

<u>Πώς?</u> Αν $A = PLU$ τότε

$$\begin{aligned} \det(A) &= \det(P)\,\underbrace{\det(L)}_{=1}\,\det(U) \\ &= (-1)^\kappa \det(U) = (-1)^\kappa \upsilon_{11}\upsilon_{22}\cdots\upsilon_{nn}. \end{aligned}$$

όπου κ είναι ο αριθμός των εναλλαγών που μετατρέπουν το I σε P. Το $(-1)^\kappa$ θα είναι το πρόσημο της αντίστοιχης μετάθεσης.

- Ο υπολογισμός της ορίζουσας με τους παραπάνω τύπους είναι ακριβός
- ... εκτός αν την υπολογίσουμε μέσω LU
- ... οπότε δεν υπάρχει λόγος να χρησιμοποιήσουμε τον κανόνα
- ... γιατί από την LU βρίσκουμε τη λύση με μπρος και πίσω αντικατάσταση, πιο φθηνά!
- Εκτός αν επικρατούν ειδικές συνθήκες! (π.χ. κάτω τριγωνικό μητρώο)

Ορίζουσες και τάξη μητρώου

Ενδιαφέρον: Ένα μητρώο $A \in \mathbb{R}^{m \times n}$ έχει τάξη ακριβώς r ανν και μόνον αν το μεγαλύτερο τετραγωνικό (υπο)μητρώο με μη μηδενική ορίζουσα είναι $r \times r$.

Αν $Ax = b$ τότε κάθε στοιχείο του x μπορεί να υπολογιστεί με τον κανόνα Cramer που εκφράζεται ως εξής:

Έστω οι στήλες $a_1, \ldots, a_n$ τ.ώ. $\det([a_1, \ldots, a_n]) \neq 0$. Έστω επίσης βαθμωτοί $\xi_1, \ldots, \xi_n$ τ.ώ.

$$b = \xi_1 e_1 + \cdots + \xi_n e_n.$$

Τότε για κάθε $j = 1, \ldots, n$ ισχύει ότι

$$\xi_j = \frac{\det([a_1, \ldots, a_{j-1}, b, a_{j+1}, \ldots, a_n])}{\det([a_1, \ldots, a_n])}$$

Τύπος αντιστρόφου

Θεώρημα

Αν το A είναι αντιστρέψιμο, τότε

$$
\begin{aligned}
(A^{-1})_{i,j} &= (-1)^{i+j}\frac{\det(A_{j,i})}{\det(A)} \\
&= \frac{\psi_{j,i}}{\det(A)}
\end{aligned}
$$

Παραδείγματα οριζουσών

εμφανίζονται συχνά στις εφαρμογές

Μητρώο τάξης 1 Έστω ότι από τα διανύσματα (στήλες) u, v κατασκευάζουμε το μητρώο $A = uv^\top \in \mathbb{R}^{n\times n}$. Ποιά θα είναι η ορίζουσα?

Παραδείγματα οριζουσών

εμφανίζονται συχνά στις εφαρμογές

Μητρώο τάξης 1 Έστω ότι από τα διανύσματα (στήλες) u, v κατασκευάζουμε το μητρώο $A = uv^\top \in \mathbb{R}^{n\times n}$. Ποιά θα είναι η ορίζουσα?

Απάντηση: 0 (γιατί?)

Παραδείγματα οριζουσών

εμφανίζονται συχνά στις εφαρμογές

Μητρώο τάξης 1 Έστω ότι από τα διανύσματα (στήλες) u, v κατασκευάζουμε το μητρώο $A = uv^\top \in \mathbb{R}^{n\times n}$. Ποιά θα είναι η ορίζουσα?

Απάντηση: 0 (γιατί?)

Μητρώο προβολής σε (γνήσιο) υπόχωρο?

Απάντηση: 0 (γιατί?)

Παραδείγματα οριζουσών

εμφανίζονται συχνά στις εφαρμογές

Μητρώο τάξης 1 Έστω ότι από τα διανύσματα (στήλες) u, v κατασκευάζουμε το μητρώο $A = uv^\top \in \mathbb{R}^{n\times n}$. Ποιά θα είναι η ορίζουσα?

Απάντηση: 0 (γιατί?)

Μητρώο προβολής σε (γνήσιο) υπόχωρο?

Απάντηση: 0 (γιατί?)

Ορθογώνιο μητρώο?

Απάντηση: $\det(A) = \pm 1$ (γιατί?)

Παραδείγματα οριζουσών

εμφανίζονται συχνά στις εφαρμογές

Μητρώο τάξης 1 Έστω ότι από τα διανύσματα (στήλες) u, v κατασκευάζουμε το μητρώο $A = uv^\top \in \mathbb{R}^{n\times n}$. Ποιά θα είναι η ορίζουσα?

Απάντηση: 0 (γιατί?)

Μητρώο προβολής σε (γνήσιο) υπόχωρο?

Απάντηση: 0 (γιατί?)

Ορθογώνιο μητρώο?

Απάντηση: $\det(A) = \pm 1$ (γιατί?)

Ταυτοτικό συν μητρώο "τάξης-1" $A = I + uv^\top$. Ποιά θα είναι η ορίζουσα?

Παραδείγματα οριζουσών

εμφανίζονται συχνά στις εφαρμογές

Μητρώο τάξης 1 Έστω ότι από τα διανύσματα (στήλες) u, v κατασκευάζουμε το μητρώο $A = uv^\top \in \mathbb{R}^{n\times n}$. Ποιά θα είναι η ορίζουσα?

Απάντηση: 0 (γιατί?)

Μητρώο προβολής σε (γνήσιο) υπόχωρο?

Απάντηση: 0 (γιατί?)

Ορθογώνιο μητρώο?

Απάντηση: $\det(A) = \pm 1$ (γιατί?)

Ταυτοτικό συν μητρώο "τάξης-1" $A = I + uv^\top$. Ποιά θα είναι η ορίζουσα? Απάντηση: $\det(A) = 1 + v^\top u$. (γιατί?)

Παραδείγματα οριζουσών

εμφανίζονται συχνά στις εφαρμογές

Μητρώο τάξης 1 Έστω ότι από τα διανύσματα (στήλες) u, v κατασκευάζουμε το μητρώο $A = uv^\top \in \mathbb{R}^{n\times n}$. Ποιά θα είναι η ορίζουσα?

Απάντηση: 0 (γιατί?)

Μητρώο προβολής σε (γνήσιο) υπόχωρο?

Απάντηση: 0 (γιατί?)

Ορθογώνιο μητρώο?

Απάντηση: $\det(A) = \pm 1$ (γιατί?)

Ταυτοτικό συν μητρώο "τάξης-1" $A = I + uv^\top$. Ποιά θα είναι η ορίζουσα? Απάντηση: $\det(A) = 1 + v^\top u$. (γιατί?)

$$B = \begin{pmatrix} 1 & -v^\top \\ u & I \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ u & I \end{pmatrix} \begin{pmatrix} 1 & -v^\top \\ 0 & I + uv^\top \end{pmatrix}, \text{ και } C = \begin{pmatrix} I & u \\ -v^\top & 1 \end{pmatrix} = \begin{pmatrix} I & 0 \\ -v^\top & 1 \end{pmatrix} \begin{pmatrix} I & u \\ 0 & 1 + v^\top u \end{pmatrix}$$

και αν $P = \begin{pmatrix} 0_{1,n} & 1 \\ I_n & 0_{n,1} \end{pmatrix}$ τότε $\det(P) = (-1)^{n+2}$ και $P^\top BP = C$ άρα $\det(P^\top BP) = \det(B) = \det(C)$. Από τις παραγοντοποιήσεις, $\det(I + uv^\top) = 1 + v^\top u$.

Γενίκευση και παράδειγμα

- **Ταυτότητα Sylvester** Γενικότερα ισχύει ότι αν $C \in \mathbb{R}^{m\times n}$, $D \in \mathbb{R}^{n\times m}$

$$\det\big(I_m + CD\big) = \det\big(I_n + DC\big)$$

- Συνιστάται να την αποδείξετε!

Παράδειγμα: Να υπολογίσετε την ορίζουσα του

$$= \begin{pmatrix} 4 & 2 & 2 & 2 \\ 2 & 4 & 2 & 2 \\ 2 & 2 & 4 & 2 \\ 2 & 2 & 2 & 4 \end{pmatrix}$$

Επίλυση: Παρατηρούμε ότι το μητρώο μπορεί να γραφτεί ως $A = 2I + 2ee^\top$ όπου $e = (1,1,1,1)^\top$. Επομένως

$$\det(A) = \det(2(I + ee^\top)) = 2^4\det(I + ee^\top) = 2^4(1 + e^\top e) = 16 \cdot 5 = 80.$$

☺ Θα ήταν αρκετά πιο χρονοβόρο αν δουλεύατε απευθείας με το A.

How to find a good submatrix

S. A. Goreinov, I. V. Oseledets, D. V. Savostyanov,
E. E. Tyrtyshnikov, N. L. Zamarashkin

October 17, 2008

Abstract

Pseudoskeleton approximation and some other problems require the knowledge of sufficiently well-conditioned submatrix in a large-scale matrix. The quality of a submatrix can be measured by modulus of its determinant, also known as volume. In this paper we discuss a search algorithm for the maximum-volume submatrix which already proved to be useful in several matrix and tensor approximation algorithms. We investigate the behavior of this algorithm on random matrices and present some its applications, including maximization of a bivariate functional.

$$\operatorname{perm}(A) := \sum_{\pi \in \mathcal{S}} \alpha_{1\pi(1)} \cdots \alpha_{n\pi(n)}$$

όπου $\mathcal{S}$ είναι το σύνολο των $n!$ μεταθέσεων των στοιχείων $(1, \ldots, n)$.

1. Ενώ η ορίζουσα μπορεί να υπολογιστεί με $O(n^3)$ πράξεις, μέσω της παραγοντοποίησης *LU*, ο υπολογισμός της $\mathrm{perm}(A)$ απαιτεί πολύ μεγάλο πλήθος πράξεων (μεγαλύτερο από πολυωνυμικό) καθώς αφορά την άθροιση εκθετικού πλήθους ακεραίων. Θυμηθείτε ότι $n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$ (προσέγγιση Stirling).
2. Ο πιο γνωστός αλγόριθμος (Ryser) για τον ακριβή υπολογισμό απαιτεί $O(n2^n)$ πράξεις.
3. Έχει αποδειχθεί ότι ο υπολογισμός είναι #P-hard (sharp-P hard). Για μητρώα στο GF(2) (δηλ με στοιχεία $\{0, 1\}$) ο υπολογισμός είναι #P-complete.

Ερευνητικό πρόβλημα: Επινόηση γρήγορων αλγορίθμου για την αποτελεσματική προσέγγιση της permanent σε πολυωνυμικό χρόνο.