Εθνικό Μετσόβειο Πολυτεχνείο

Σχολή Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών

# Θέματα ελαχίστων τετραγώνων και μέθοδοι επίλυσης

Διπλωματική εργασία του Μάρα Ισίδωρου

Επιβλέπων: Κ. Χρυσαφίνος

# ΠΕΡΙΕΧΟΜΕΝΑ

# ΚΕΦΑΛΑΙΟ 1

## 1.1 ΠΟΛΛΑΠΛΑΣΙΑΣΜΟΣ ΠΙΝΑΚΑ-ΔΙΑΝΥΣΜΑ

Η μέθοδος πολλαπλασιασμού πίνακα επί διάνυσμα είναι γνωστή. Σκοπός αυτής της παραγράφου είναι να περιγράψουμε έναν κατάλληλο τρόπο για υπολογισμούς μεγάλων διαστάσεων. Ξεκινάμε παρατηρώντας ότι αν $b = Ax$, τότε ο $b$ είναι ένας *γραμμικός συνδυασμός των στηλών του* $A$.

### *1.1.1 ΓΝΩΣΤΟΙ ΟΡΙΣΜΟΙ*

Έστω $x$ να είναι ένα διάνυσμα μιας στήλης διάστασης $n$ και έστω $A$ να είναι ένας $m \times n$ πίνακας ( $m$ γραμμές, $n$ στήλες). Τότε το αποτέλεσμα $b = Ax$ του πολλαπλασιασμού πίνακα επί διάνυσμα είναι ένα διάνυσμα μιας στήλης $m$ διάστασης η οποία ορίζεται ως:

$$b_i = \sum_{j=1}^{n} a_{ij} x_j \ , \quad i = 1, 2, ....., m \tag{1.1}$$

Το $b_j$ δηλώνει το i-στο στοιχείο του $b$, $a_{ij}$ δηλώνει το ij-οστό στοιχείο του $A$ (i-οστή γραμμή, j-οστή στήλη) και το $x_j$ δηλώνει το j-οστό στοιχείο του $x$. Για λόγους απλότητας, υποθέτουμε ότι τέτοιες ποσότητες ανήκουν στον *C*, το σύνολο των μιγαδικών αριθμών. Το διάστημα των $m$-διανυσμάτων είναι το $C^m$ και το διάστημα των $m \times n$ πινάκων είναι το $C^{m \times n}$.
Η απεικόνιση $x \mapsto Ax$ είναι γραμμική, δηλαδή για κάθε $x, y \in C^n$ και για κάθε $a \in C$ έχουμε

$$A(x + y) = Ax + Ay$$
$$A(\alpha x) = \alpha Ax$$

Αντίστροφα, κάθε γραμμική απεικόνιση από το $C^n$ στο $C^m$ μπορεί να εκφραστεί σαν πολλαπλασιασμός με έναν πίνακα $m \times n$.

### *1.1.2 ΠΟΛΛΑΠΛΑΣΙΑΣΜΟΣ ΠΙΝΑΚΑ ΕΠΙ ΔΙΑΝΥΣΜΑ*

Έστω $a_j$ να είναι η j-οστή στήλη του $A$ ,δηλαδή $a_j$ είναι διάνυσμα διαστάσεων $m$. Τότε η (1.1 ) μπορεί να γραφτεί ως

$$b = Ax = \sum_{j=1}^{n} x_j a_j \tag{1.2}$$

Παρατηρούμε ότι στην (1.2) ο $b$ εκφράζεται ως ένας γραμμικός συνδυασμός των στηλών $a_j$ Οι διαφορές ανάμεσα στις εξισώσεις (1.1) και (1.2) είναι ότι ως μαθηματικοί έχουμε συνηθίσει να θεωρούμε ότι στην εξίσωση $Ax = b$ ο πίνακας $A$ εφαρμόζεται στο $x$ για να προκύψει το αποτέλεσμα $b$. Στην εξίσωση όμως (1.2) έχουμε ότι το διάνυσμα $x$ εφαρμόζεται στον πίνακα $A$ και προκύπτει το $b$.

**Παράδειγμα 1.1.2 (Πίνακας Vandermonde)**

'Εστω η ακολουθία αριθμών $\{x_1, x_2, ..., x_m\}$ .Αν $p$ και $q$ είναι πολυώνυμα με βαθμό μικρότερο του n και a πραγματικός αριθμός,τότε και τα $p+q$ , $ap$ είναι επίσης πολυώνυμα με βαθμό μικρότερο του n. Επίσης οι τιμές των πολυωνύμων στα σημεία $x_i$ ικανοποιούν τις παρακάτω γραμμικές ιδιότητες:

1. $(p+q)(x_i) = p(x_i) + q(x_i)$
2. $(ap)(x_i) = a(p(x_i))$

Έτσι η απεικόνιση από διανύσματα με στοιχεία τους συντελεστές πολυωνύμων p,βαθμού μικρότερο του n, σε διανύσματα $(p(x_1), p(x_2), ..., p(x_m))$ με στοιχεία τις τιμές πολυωνύμων στα σημεία $x_i$ είναι γραμμική. Κάθε γραμμική απεικόνιση μπορεί να εκφραστεί μέσω ενός γινομένου πινάκων. Στην πραγματικότητα εκφράζεται μέσω ενός $m \times n$ πίνακα Vandermonde.

$$A = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_m & x_m^2 & \cdots & x_m^{n-1} \end{bmatrix}$$

Έστω c διάνυσμα στήλη με στοιχεία τους συντελεστές πολυωνύμου p

$$c = \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ c_{n-1} \end{bmatrix}, \qquad \mathrm{p(x)} = \mathrm{c}_0 + c_1 x + c_2 x^2 + \ldots + c_{n-1} x^{n-1}$$

Το αποτέλεσμα του γινομένου Ac δίνει τις τιμές του πολυωνύμου p στα σημεία $x_i$ Για κάθε i από 1 έως m έχουμε :

$$(\mathrm{Ac})_i = \mathrm{c}_0 + c_1 x_i + c_2 x_i + \ldots + c_{n-1} x_i^{n-1} = p(x_i)$$

Συμπερασματικά ο πίνακας Α μπορεί να θεωρηθεί ως ένας πίνακας του οποίου κάθε στήλη δίνει τις τιμές των μονονύμων στα σημεία $x_i$ και το αποτέλεσμα του γινομένου *Ac* πρέπει να ερμηνεύεται ως το διάνυσμα του αθροίσματος της μορφής (1.2) το οποίο δινει έναν απευθείας γραμμικό συνδυασμό αυτών των μονονύμων.

$$A = \begin{bmatrix} | & | & | & & | \\ 1 & x & x^2 & \cdots & x^{n-1} \\ | & | & | & & | \end{bmatrix}$$

$$Ac = c_0 + c_1 x + c_2 x^2 + \ldots + c_{n-1} x^{n-1} = p(x)$$

*1.1.3 ΠΟΛΛΑΠΛΑΣΙΑΣΜΟΣ ΠΙΝΑΚΑ ΕΠΙ ΠΙΝΑΚΑ*

Για το αποτέλεσμα του πολλαπλασιασμού πίνακα επί πίνακα $B = AC$ κάθε στήλη του $B$ είναι γραμμικός συνδυασμός των στηλών του $A$. Αν ο $A$ είναι ένας $l \times m$ πίνακας και ο $C$ είναι ένας $m \times n$ πίνακας τότε ο $B$ είναι ένας $l \times n$ πίνακας του οποίου τα στοιχεία δίνονται από:

$$b_{ij} = \sum_{k=1}^{m} a_{ik} c_{kj} \qquad (1.3)$$

όπου $b_{ij}$, $a_{ik}$ και $c_{kj}$ είναι τα στοιχεία των πινάκων $A$, $B$, $C$ αντίστοιχα.

**Παράδειγμα 1.1.3.1 (Εξωτερικό γινόμενο)**
Ένα απλό παράδειγμα πολλαπλασιασμού πίνακα επί πίνακα είναι το εξωτερικό γινόμενο. Στο εξωτερικό γινόμενο πολλαπλασιάζουμε ένα διανυσμα στήλη **u** διάστασης m επί ένα διάνυσμα γραμμή **v**,διάστασης n. Το αποτέλεσμα που προκύπτει είναι ένας πίνακας διαστάσεων $m \times n$ και βαθμού 1.

$$\begin{bmatrix} \\ u \\ \\ \end{bmatrix} \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} = \begin{bmatrix} u_1 v_1 & \cdots & v_n u_1 \\ \vdots & & \vdots \\ v_1 u_m & \ldots & v_n u_m \end{bmatrix}$$

**Παράδειγμα 1.1.3.2**

Έστω το γινόμενο B=AR όπου R είναι ένας άνω τριγωνικός πίνακας $n \times n$ με στοιχεία $r_{ij} = 1$ για $\mathrm{i} \le j$ και $r_{ij} = 0$ για $\mathrm{i} \ge \mathrm{j}$.

$$\begin{bmatrix} \vert & & \vert \\ b_1 & \cdots & b_n \\ \vert & & \vert \end{bmatrix} = \begin{bmatrix} \vert & & \vert \\ a_1 & \cdots & a_n \\ \vert & & \vert \end{bmatrix} \begin{bmatrix} 1 & \cdots & 1 \\ & \ddots & \vdots \\ & & 1 \end{bmatrix}$$

Τα στοιχεία του πίνακα B δίνονται από την σχέση:

$$b_j = Ar_j = \sum_{k=1}^{j} a_k$$

Η παραπάνω υπολογιστική τεχνική είναι κατάλληλα για την δομή προβλημάτων μεγάλων διαστάσεων.

Συνεχίζουμε παρουσιάζοντας βασικές έννοιες από την γραμμική άλγεβρα.

*1.1.4 ΠΕΔΙΟ ΤΙΜΩΝ ΚΑΙ ΠΥΡΗΝΑΣ*

Το πεδίο τιμών ενός πίνακα $A$, συμβολίζεται με range($A$) είναι το σύνολο των διανυσμάτων τα οποία μπορούν να εκφραστούν ως $Ax$ για κάποια $x$. Η εξίσωση (1.2) οδηγεί στον παρακάτω χαρακτηρισμό του range($A$).

ΘΕΩΡΗΜΑ 1.1

Το πεδίο τιμών ενός πίνακα $A$, range($A$), είναι η γραμμική θήκη που παράγεται από τις στήλες του $A$.

ΑΠΟΔΕΙΞΗ

Βλέπε βιβλιογραφία [1] σελ. 7

Ο πυρήνας ενός πίνακα $A \in C^{m \times n}$, συμβολίζεται με $null(A)$, είναι το σύνολο των διανυσμάτων $x$ που ικανοποιούν την εξίσωση $Ax = 0$ όπου το μηδέν είναι το μηδενικό διάνυσμα στον $C^m$. Τα στοιχεία κάθε διανύσματος $x \in null(A)$ αποτελούν τους συντελεστές μιας επέκτασης του μηδενός ως γραμμικός συνδυασμός των στηλών του $A$, δηλαδή $0 = x_1 a_1 + x_2 a_2 + ..... + x_n a_n$.

### *1.1.5 ΤΑΞΗ*

Η τάξη των στηλών ενός πίνακα είναι η διάσταση του χώρου των στηλών του. Όμοια η τάξη των γραμμών ενός πίνακα είναι η διάσταση του χώρου που παράγεται από τις γραμμές του. Η τάξη των γραμμών είναι πάντα ίση με την τάξη των στηλών του οπότε δεν υπάρχει διαχωρισμός μεταξύ τους και αναφερόμαστε στην τάξη ενός πίνακα.
Ένας $m \times n$ πίνακας πλήρους τάξης είναι ένας πίνακας ο οποίος έχει τη μέγιστη δυνατή τάξη. Αυτό σημαίνει ότι ένας πίνακας πλήρους τάξης με $m \geq n$ πρέπει να έχει $n$ γραμμικά ανεξάρτητες στήλες.

ΘΕΩΡΗΜΑ 1.2
Ένας πίνακας $A \in C^{m \times n}$ με $m \geq n$ είναι πλήρους τάξης αν και μόνο αν απεικονίζει δυο διαφορετικά διανύσματα στο ίδιο διάνυσμα.

ΑΠΟΔΕΙΞΗ
Βλέπε βιβλιογραφία [1] σελ. 7

### *1.1.6 ΑΝΑΣΤΡΟΦΗ*

Ένας ομαλός πίνακας είναι ένας τετραγωνικός πίνακας πλήρους τάξης. Οι $m$ στήλες ενός ομαλού πίνακα $A$ αποτελούν μια βάση για όλο το χώρο $C^m$. Οπότε μπορούμε να εκφράσουμε κατά μοναδικό τρόπο οποιοδήποτε διάνυσμα σαν ένα γραμμικό συνδυασμό αυτών. Πιο συγκεκριμένα η κανονική βάση με 1 στην j-οστή θέση του διανύσματος και μηδενικά οπουδήποτε αλλού, συμβολίζεται με $e_j$, μπορεί να επεκταθεί ως

$$e_j = \sum_{i=1}^{m} z_{ij} a_i \qquad (1.4)$$

Έστω $Z$ πίνακας με στοιχεία $z_{ij}$ και έστω $z_j$ να είναι η $j-$οστή στήλη του $Z$. Τότε η σχέση (1.4) μπορεί να γραφτεί ως $e_j = Az_j$. Η εξίσωση αυτή έχει τη μορφή της (1.3), η οποία μπορεί να γραφτεί ως

$$\left[ \begin{array}{c|c|c} & & \\ e_1 & \cdots & e_m \\ & & \end{array} \right] = I = AZ$$

όπου ο πίνακας $I$ είναι ένας μοναδιαίος $m \times m$ πίνακας. Ο πίνακας $Z$ είναι ο ανάστροφος του $A$. Κάθε τετραγωνικός ομαλός πίνακας $A$ έχει μοναδικό ανάστροφο ο οποίος συμβολίζεται με $A^{-1}$ και ικανοποιεί τη σχέση $AA^{-1} = A^{-1}A = I$.

Το παρακάτω θεώρημα περιγράφει κάποιες συνθήκες οι οποίες ισχύουν όταν ένας τετραγωνικός πίνακας είναι ομαλός.

ΘΕΩΡΗΜΑ 1.3

Για $A \in C^{m \times n}$, οι παρακάτω συνθήκες είναι ισοδύναμες:

(α) ο $A$ έχει ανάστροφο τον $A^{-1}$

(β) $rank(A) = m$

(γ) $range(A) = C^m$

(δ) $null(A) = \{0\}$

(ε) το 0 είναι ιδιοτιμή του $A$

(στ)το 0 δεν είναι ομαλή τιμή του $A$

(ζ) $\det(A) \neq 0$

Επισημαίνεται ότι η χρήση της ορίζουσας σε αλγορίθμους είναι καταστροφική.

### *1.1.7 ΑΝΑΣΤΡΟΦΟΣ ΠΙΝΑΚΑΣ ΕΠΙ ΔΙΑΝΥΣΜΑ*

Όταν γράφουμε το αποτέλεσμα $x = A^{-1}b$ είναι σημαντικό να μην αφήνουμε τη δομή του ανάστροφου πίνακα να μας μπερδεύει! Αντί να βλέπουμε το $x$ σαν το αποτέλεσμα της εφαρμογής του $A^{-1}$ στον $b$, θα πρέπει να το βλέπουμε σαν το μοναδικό διάνυσμα που ικανοποιεί την ισότητα $Ax = b$. Από τη σχέση (1.2), αυτό σημαίνει ότι το $x$ είναι είναι το διάνυσμα της επέκτασης του $b$ πάνω στη βάση των στηλών του $A$. Ο πολλαπλασιασμός με $A^{-1}$ γίνεται με τη διαδικασία αλλαγής βάσης. Επειδή δεν μπορούμε να αναλύσουμε το κομμάτι αυτό περισσότερο επαναλαμβάνουμε ότι:

Το διάνυσμα $A^{-1}b$ είναι το διάνυσμα της επέκτασης του $b$ πάνω στη βάση των στηλών του $A$

## 1.2 ΟΡΘΟΓΩΝΙΑ ΔΙΑΝΥΣΜΑΤΑ ΚΑΙ ΟΡΘΟΓΩΝΙΟΙ ΠΙΝΑΚΕΣ

Από το 1960 πολλοί αλγόριθμοι της αριθμητικής γραμμικής άλγεβρας έχουν βασιστεί στην ορθογωνιότητα. Σε αυτή την παράγραφο θα ασχοληθούμε με ορθογώνια διανύσματα και ορθογώνιους πίνακες.

### *1.2.1 ΣΥΖΥΓΗΣ ΠΙΝΑΚΑΣ*

Ο μιγαδικός συζυγής ενός συζυγή $z$ συμβολίζεται με $\bar{z}$ ή με $z^*$ και προκύπτει βάζοντας αντίθετο πρόσημο στο φανταστικό μέρος του μιγαδικού. Για πραγματικό αριθμό ισχύει $\bar{z} = z$.
Ο ερμιτιανός συζυγής ή αναστροφοσυζυγής ενός πίνακα $A \quad m \times n$ συμβολίζεται με $A^*$ και είναι ένας $n \times m$ πίνακας του οποίου τα i,j στοιχεία είναι τα μιγαδικά συζυγή των j,i στοιχείων του $A$. Αν $A = A^*$ τότε ο $A$ λέγεται ερμιτιανός. Εξ ορισμού ένας ερμιτιανός πίνακας πρέπει να είναι τετραγωνικός. Για πραγματικό $A$ ο αναστροφοσυζυγής προκύπτει από απλή εναλλαγή γραμμών με στήλες. Σε αυτή την περίπτωση ο πίνακας ονομάζεται ανάστροφος και συμβολίζεται με $A^T$. Αν ο πραγματικός πίνακας είναι ερμιτιανός τότε ισχύει $A = A^T$ και λέγεται επίσης συμμετρικός.

Συμβολισμός: $z$ διάνυσμα γραμμή ή $z^T$ όταν είναι διάνυσμα πραγματικών αριθμών

### *1.2.2 ΕΣΩΤΕΡΙΚΟ ΓΙΝΟΜΕΝΟ*

Το εσωτερικό γινόμενο δύο διανυσμάτων-στηλών $x, y \in C^m$ είναι το αποτέλεσμα του συζυγή της στήλης $x$ επί τη στήλη $y$ και δίνεται από τη σχέση:

$$x^* y = \sum_{i=1}^{m} \bar{x}_i y_i \tag{1.5}$$

Το Ευκλείδειο μήκος του $x$ συμβολίζεται με $\|x\|$ και δίνεται από τη σχέση:

$$\|x\| = \sqrt{x^* x} = \left( \sum_{i=1}^{m} |x_i|^2 \right)^{1/2} \tag{1.6}$$

Η γωνία $a$ μεταξύ των $x$ και $y$ μπορεί να εκφραστεί με βάση το εσωτερικό γινόμενο και δίνεται από την παρακάτω σχέση:

$$\cos a = \frac{x^* y}{\|x\|\|y\|} \tag{1.7}$$

Το εσωτερικό γινόμενο είναι διγραμμικό, δηλαδή:

$$(x_1 + x_2)^* y = x_1^* y + x_2^* y$$
$$x^* (y_1 + y_2) = x^* y_1 + x^* y_2$$
$$(ax)^* (\beta y) = \overline{a}\beta x^* y .$$

Για οποιαδήποτε διανύσματα ή πίνακες $A, B$ ίδιας διάστασης ισχύει ότι

$$(AB)^* = B^* A^* \tag{1.8}$$

Ανάλογη είναι και η σχέση που ισχύει για αντιστρέψιμους τετραγωνικούς πίνακες:

$$(AB)^{-1} = B^{-1} A^{-1} \tag{1.9}$$

Συμβολισμός: $A^{-*} \equiv (A^*)^{-1}$ ή $(A^{-1})^*$

### *1.2.3 ΟΡΘΟΓΩΝΙΑ ΔΙΑΝΥΣΜΑΤΑ*

Δυο διανύσματα $x$ και $y$ λέγονται ορθογώνια αν $x^* y = 0$. Αν τα διανύσματα αυτά είναι πραγματικά τότε είναι κάθετα στον $R^m$.
Ένα σύνολο $S$ μη μηδενικών διανυσμάτων ονομάζεται ορθογώνιο αν τα στοιχεία του είναι ορθογώνια ανά δύο. Ένα σύνολο διανυσμάτων ονομάζεται ορθοκανονικό αν είναι ορθογώνιο και για κάθε $x \in S$ να ισχύει $\|x\| = 1$.

ΘΕΩΡΗΜΑ 1.4
Τα διανύσματα σε ένα ορθογώνιο σύνολο $S$ είναι γραμμικά ανεξάρτητα.

ΑΠΟΔΕΙΞΗ
Βλέπε βιβλιογραφία [1] σελ. 13

### *1.2.4 ΣΥΝΙΣΤΩΣΕΣ ΔΙΑΝΥΣΜΑΤΟΣ*

Το πιο σημαντικό αποτέλεσμα που μπορεί να προκύψει από τα εσωτερικά γινόμενα και την ορθογωνιότητα είναι ότι τα εσωτερικά γινόμενα μπορούν να χρησιμοποιηθούν για να αποσυνθέσουν τυχαία διανύσματα σε ορθογώνιες συνιστώσες.

Για παράδειγμα ας υποθέσουμε ότι $\{q_1, q_2, ...., q_n\}$ είναι ένα ορθογώνιο σύνολο και έστω $u$ να είναι ένα τυχαίο διάνυσμα. Η ποσότητα $q_j^* u$ είναι βαθμωτό γινόμενο. Χρησιμοποιώντας αυτά τα βαθμωτά γινόμενα σε μια επέκταση βρίσκουμε ότι το διάνυσμα

$$r = u - (q_1^* u)q_1 - (q_2^* u)q_2 - .... - (q_n^* u)q_n$$

είναι ορθογώνιο στο $\{q_1, q_2, ...., q_n\}$. Αυτό μπορεί να διαπιστωθεί υπολογίζοντας το $q_i^* r$ χρησιμοποιώντας τις ιδιότητες του εσωτερικού γινομένου:

$$q_i^* r = q_i^* u - (q_1^* u)(q_i^* q_1) - .... - (q_n^* u)(q_i^* q_n) \qquad (1.10)$$

Επειδή $q_i^* q_j = 0$ για $i \neq j$ έχουμε:

$$q_i^* r = q_i^* u - (q_1^* u)(q_i^* q_i) = 0$$

Βλέπουμε ότι το $u$ μπορεί να διαμεριστεί σε $n+1$ ορθογώνιες συνιστώσες:

$$u = r + \sum_{i=1}^{n} (q_i^* u)q_i = r + \sum_{i=1}^{n} (q_i q_i^*)u \qquad (1.11)$$

Σε αυτό το διαχωρισμό το $r$ είναι το ορθογώνιο μέρος του $u$ στο σύνολο των διανυσμάτων $\{q_1, q_2, ...., q_n\}$.

Αν $\{q_i\}$ είναι μια βάση για το $C^m$ τότε το $n$ πρέπει να είναι ίσο με το $m$ και το $r$ πρέπει να είναι το μηδενικό διάνυσμα έτσι ώστε το $u$ να διαχωριστεί πλήρως σε $m$ ορθογώνιες συνιστώσες στη διεύθυνση του $q_i$:

$$u = \sum_{i=1}^{m} (q_i^* u)q_i = \sum_{i=1}^{m} (q_i q_i^*)u \qquad (1.12)$$

Οι δυο ισότητες στις σχέσεις (1.11) και (1.12) είναι ίσες αλλά έχουν διαφορετική ερμηνεία. Στην πρώτη βλέπουμε το $u$ σαν ένα γινόμενο των συντελεστών $q_i^* u$ επί τα διανύσματα $q_i$ ενώ στη δεύτερη βλέπουμε το $u$ σαν ένα άθροισμα των ορθογώνιων απεικονίσεων του $u$ σε διάφορες κατευθύνσεις των $q_i$.

*1.2.5 ΟΡΘΟΜΟΝΑΔΙΑΙΟΙ ΠΙΝΑΚΕΣ*

Ένας τετραγωνικός πίνακας $Q \in C^{m\times m}$ είναι ορθομοναδιαίος ή ορθοκανονικός ακριβώς όταν ισχύει μία από τις επόμενες ισοδύναμες συνθήκες :

$$QQ^* = I,\ \ Q^*Q = I,\ \ Q^* = Q^{-1}$$

Από τον τελευταίο ορισμό προκύπτουν άμεσα οι ιδιότητες :

(α) Οι ορθομοναδιαίοι πίνακες είναι ομαλοί, $|\det Q| = 1$

(β)Ο πίνακας Q* είναι ορθομοναδιαίος

(γ)Το γινόμενο ορθομοναδιαίων πινάκων είναι ορθομοναδιαίος πίνακας

**ΠΑΡΑΔΕΙΓΜΑ 1.2.5.1**

$\text{Ο } \pi\acute{\iota}\nu\alpha\kappa\alpha\varsigma\ \ A = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 & 1+i \\ 1-i & -1 \end{bmatrix}$ είναι ορθομοναδιαίος,αφού είναι:

$$A^*A = AA^* = \frac{1}{3}\begin{bmatrix} 1 & 1+i \\ 1-i & -1 \end{bmatrix}\begin{bmatrix} 1 & 1+i \\ 1-i & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$$

*1.2.6 ΠΟΛΛΑΠΛΑΣΙΑΜΟΣ ΜΕ ΟΡΘΟΜΟΝΑΔΙΑΙΟ ΠΙΝΑΚΑ*

Στην παράγραφο 1.2 συζητήσαμε για την ερμηνεία των αποτελεσμάτων από τα γινόμενα $Ax$ και $A^{-1}b$. Αν $A$ είναι ένας ορθομοναδιαίος πίνακας $Q$ τότε τα παραπάνω αποτελέσματα γίνονται $Qx$ και $Q^*b$ και οι ερμηνείες παραμένουν ίδιες. Όπως και πριν, $Qx$ είναι ο γραμμικός συνδυασμός των στηλών του $Q$ με συντελεστή $x$. Αντίστροφα, $Q^*b$ είναι το διάνυσμα των συντελεστών της επέκτασης του $b$ στη βάση των στηλών του $Q$.
Αυτές οι διαδικασίες πολλαπλασιασμού με ορθομοναδιαίο πίνακα ή με τον συζυγή του διατηρούν την γεωμετρική τους δομή με την Ευκλείδεια έννοια επειδή τα εσωτερικά γινόμενα διατηρούνται. Δηλαδή για ορθομοναδιαίο πίνακα $Q$ είναι:

$$(Qx)^*(Qy) = x^*y \qquad (1.13)$$

και επαληθεύεται από την (1.8). Η μη μεταβλητότητα των εσωτερικών γινομένων σημαίνει ότι οι γωνίες μεταξύ των διανυσμάτων διατηρούνται και κατά συνέπεια και τα μήκη των διανυσμάτων

$$\|Qx\| = \|x\| \qquad (1.14)$$

Στους πραγματικούς, ο πολλαπλασιασμός με έναν ορθογώνιο πίνακα $Q$ αντιστοιχεί σε μια στάσιμη περιστροφή (αν $\det Q = 1$ ) ή απεικόνιση (αν $\det Q = -1$ ) στο χώρο των διανυσμάτων.

## 1.3 ΝΟΡΜΕΣ

Οι βασικές έννοιες του μεγέθους και της απόστασης σε ένα διανυσματικό χώρο περιλαμβάνονται στις νόρμες. Οι νόρμες είναι το εργαλείο μας με το οποίο μετράμε τις προσεγγίσεις και τη σύγκλιση στην αριθμητική γραμμική άλγεβρα.

### *1.3.1 ΝΟΡΜΑ ΔΙΑΝΥΣΜΑΤΟΣ*

Μια νόρμα είναι μια συνάρτηση $\|\cdot\| : C^m \to R$ η οποία αντιστοιχίζει σε κάθε διάνυσμα το μέτρο του. Η νόρμα ικανοποιεί τα παρακάτω:

(1) $\|x\| \geq 0$ και $\|x\| = 0$ μόνο αν $x = 0$

(2) $\|x + y\| \leq \|x\| + \|y\|$ (1.15)

(3) $\|ax\| = |a|\|x\|$

Δηλαδή τα παραπάνω προϋποθέτουν ότι (1) η νόρμα ενός μη μηδενικού διανύσματος είναι θετική, (2) ότι η νόρμα του αθροίσματος διανυσμάτων δεν είναι μεγαλύτερη από το άθροισμα των νορμών ξεχωριστά (τριγωνική ανισότητα) και (3) το ανηγμένο γινόμενο αριθμός επί διάνυσμα είναι ίσο με την ποσότητα $\|ax\|$.

Η πιο σημαντικές $k$ -νορμες είναι οι $p$ -νόρμες και οριζονται όπως παρακάτω

$$\|x\|_1 = \sum_{i=1}^{m} |x_i|, \qquad \|x\|_2 = (\sum_{i=1}^{m} |x_i|^2)^{1/2} = \sqrt{x^* x},$$

$$\|x\|_\infty = \max_{1 \leq i \leq m} |x_i|, \qquad \|x\|_p = (\sum_{i=1}^{m} |x_i|^p)^{1/p} \qquad (1 \leq p \leq \infty) \tag{1.16}$$

Μία ενδιαφέρουσα παρατήρηση εδώ είναι ότι αν

$$|x_k| = \max\{|x_1|, |x_2|, ..., |x_v|\} \text{ τότε } \lim_{p\to\infty} \|x\|_p = \lim_{p\to\infty} \left( \sum_{i}^{v} |x_i|^p \right)^{1/p} =$$

$$\lim_{p\to\infty} \left( |x_k| \left( 1 + \left|\frac{x_1}{x_k}\right|^p + ... + \left|\frac{x_v}{x_k}\right|^p \right) \right)^{1/p} = |x_k| = \|x\|_\infty,$$

$$\textit{διότι} \quad \left|\frac{x_1}{x_k}\right| < 1$$

Πέρα από τις $p$ -νόρμες οι πιο χρήσιμες νόρμες είναι οι σταθμισμένες $p$ -νόρμες, όπου κάθε συνιστώσα διανύσματος πολλαπλασιάζεται με κάποιο (δοσμένο) συντελεστή, που καλέιται βάρος . Γενικά κάθε νόρμα μπορεί να γραφεί ως σταθμισμένη νόρμα ως εξής:

$$\|x\|_w = \|Wx\| \quad , \tag{1.17}$$

όπου *W* είναι ο διαγώνιος πίνακας του οποίου το i- οστό διαγώνιο στοιχείο είναι το βάρος $w_i \neq 0$.

### *1.3.2 ΝΟΡΜΕΣ ΠΙΝΑΚΑ ΠΟΥ ΠΡΟΚΥΠΤΟΥΝ ΑΠΟ ΝΟΡΜΕΣ ΔΙΑΝΥΣΜΑΤΟΣ*

Ένας $m \times n$ πίνακας μπορεί να θεωρηθεί ως ένα διάνυσμα σε ένα χώρο $mn$ διαστάσεων: κάθε ένα από τα $mn$ στοιχεία του πίνακα είναι μια ανεξάρτητη συντεταγμένη. Κάθε νόρμα $mn$ διάστασης μπορεί να χρησιμοποιηθεί για να μετρήσουμε το «μέγεθος» ενός τέτοιου πίνακα.
Όμως, όταν έχουμε ένα σύνολο πινάκων οι νόρμες που περιγράφονται από τις σχέσεις (1.16) και (1.17) δεν είναι ιδιαίτερα χρήσιμες. Οι νόρμες που είναι χρήσιμες λέγονται νόρμες επαγόμενου πίνακα. Δοθέντων των νορμών $\|\cdot\|_{(n)}$ και $\|\cdot\|_{(m)}$ στο χώρο και στο διάστημα $A \in C^{m\times n}$, η επαγόμενη νόρμα $\|A\|_{(m,n)}$ είναι ο μικρότερος αριθμός $C$ για τον οποίο ισχύει η παρακάτω ανισότητα για όλα τα $x \in C^n$ :

$$\|Ax\|_{(m)} \leq C\|x\|_{(n)} \tag{1.18}$$

Με άλλα λόγια, $\|A\|_{(m,n)}$ είναι το supremum του $\|Ax\|_{(m)} / \|x\|_{(n)}$ σε όλα τα διανύσματα $x \in C^n$, δηλαδή ο μεγαλύτερος παράγοντας με τον οποίο το $A$ μπορεί να «επεκτείνει» ένα διάνυσμα $x$ . Λέμε ότι η νόρμα $\|\cdot\|_{(m,n)}$ είναι η νόρμα πίνακα που επάγεται από τις νόρμες $\|\cdot\|_{(n)}$ και $\|\cdot\|_{(m)}$.
Λόγω της υπόθεσης (3) στη σχέση (1.15) η επίδραση του $A$ ορίζεται από την επίδρασή του σε μεμονωμένα διανύσματα. Οπότε η νόρμα πίνακα μπορεί ισοδύναμα να οριστεί ως:

$$\|A\|_{(m,n)} = \sup_{x\in C^n, x\neq 0} \frac{\|Ax\|_{(m)}}{\|x\|_{(n)}} = \sup_{x\in C^n, \|x\|_{(n)}=1} \|Ax\|_{(m)} \tag{1.19}$$

**Παραδειγμα 1.3.2.1 (1-νόρμα πίνακα)**
Έστω $A$ ένας $m\times n$ πίνακας. Τότε $\|A\|_1$ είναι ίση με το μέγιστο άθροισμα των στοιχείων των στηλών του πίνακα $A$, δηλαδή

$$\|A\|_1 = \max_{1\leq j\leq n} \|a_j\|_1$$

**Παραδειγμα 1.3.2.2 ($\infty$ -νόρμα πίνακα)**

Έστω $A$ ένας $m\times n$ πίνακας. Τότε $\|A\|_\infty$ είναι ίση με το μέγιστο άθροισμα των στοιχείων των γραμμών του πίνακα $A$, δηλαδή

$$\|A\|_\infty = \max_{1\le i\le m} \|a_i^*\|_1$$

όπου $a_i^*$ δηλωνει το $i$-οστή γραμμή του πίνακα $A$.

Παρατήρηση: Οι νόρμες $\|\cdot\|_1$ και $\|\cdot\|_\infty$ είναι εύκολα υπολογίσιμες.

## *1.3.3 ΟΙ ΑΝΙΣΟΤΗΤΕΣ CAUCHY-SCHWARZ ΚΑΙ HÖLDER*

Ο υπολογισμός των $p$-νόρμών με $|x^*y| \le \|x\|_2 \|y\|_2$ είναι πιο δύσκολος και για να λύσουμε αυτό το πρόβλημα φράσουμε τα εσωτερικά γινόμενα χρησιμοποιώντας $p$-νόρμες. Έστω p και q να ικανοποιούν τη σχέση $1/p+1/q=1$ με $1\le p,q\le\infty$.
Η ανισότητα Hölder δίνεται από τη σχέση:

$$|x^*y| \le \|x\|_p \|y\|_q \qquad (1.20)$$

Η ανισότητα Cauchy-Schwarz είναι μια ειδική περίπτωση της ανισότητας Hölder για $p=q=2$:

$$|x^*y| \le \|x\|_2 \|y\|_2 \qquad (1.21)$$

Η σχέση $\|x\|_p = (\sum_{i=1}^{m} |x_i|^p)^{1/p}$ δεν είναι νόρμα όταν $0\le p<1$, διότι δεν ικανοποιεί η τριγωνική ανισότητα. Στην ουσία η ανισότητα Cauchy-Schwartz γενικεύεται για τις p-νόρμες με την ανισότητα Holder.

## *1.3.4 Η ΦΡΑΓΜΕΝΗ ΝΟΡΜΑ ΤΟΥ AB ΕΙΝΑΙ ΜΙΑ ΕΠΑΓΟΜΕΝΗ ΝΟΡΜΑ ΠΙΝΑΚΑ*

Η επαγόμενη νόρμα πίνακα ενός πίνακα μπορεί να φραχτεί. Έστω $\|\cdot\|_{(l)}$, $\|\cdot\|_{(m)}$ και $\|\cdot\|_{(n)}$ να είναι οι νόρμες στους $C^l$, $C^m$ και $C^n$ αντίστοιχα και έστω $A$ να είναι ένας $l\times m$ πίνακας και $B$ ένας $m\times n$ πίνακας. Για κάθε $x\in C^n$ έχουμε

$$\|ABx\|_{(l)} \le \|A\|_{(l,m)} \|Bx\|_{(m)} \le \|A\|_{(l,m)} \|B\|_{(m,n)} \|x\|_{(n)}.$$

Οπότε η επαγόμενη νόρμα του $AB$ πρέπει να ικανοποιεί

$$\|AB\|_{(l,n)} \le \|A\|_{(l,m)} \|B\|_{(m,n)} \tag{1.22}$$

Γενικά αυτή η ανισότητα δεν είναι ισότητα. Για παράδειγμα, η ανισότητα $\|A^n\| \le \|A\|^n$ ισχύει για κάθε τετραγωνικό πίνακα σε κάθε νόρμα πίνακα που επάγεται από μια νόρμα διανύσματος αλλά $\|A^n\| = \|A\|^n$ δεν ισχύει γενικά για $n \ge 2$ .

*1.3.5 ΓΕΝΙΚΕΣ ΝΟΡΜΕΣ ΠΙΝΑΚΑ*

Όπως αναφέραμε παραπάνω, οι νόρμες πινάκων δεν είναι απαραίτητο να επάγονται από νόρμες διανυσμάτων. Γενικά, μια νόρμα πίνακα πρέπει τουλάχιστον να ικανοποιεί τις τρείς υποθέσεις για τη νόρμα διανύσματος (1.15) που ισχύουν για το $mn$ διάστασης χώρο διανυσμάτων πινάκων:

(1) $\|A\| \ge 0$ και $\|A\| = 0$ μόνο αν $A = 0$

(2) $\|A + B\| \le \|A\| + \|B\|$ (1.23)

(3) $\|aA\| = |a|\|A\|$

Αν θεωρήσουμε τους πίνακες του $C^{m\times n}$ ως διανύσματα του $A^{mn}$ και στον πίνακα $A = \left[ a_{ij} \right]_{i,j=1}^{m,n}$ αντιστοιχίσουμε το διάνυσμα $\mathrm{a} = \left[ \mathrm{a}_{11}\ \mathrm{a}_{12}\ \ldots\ \mathrm{a}_{1n}\ \mathrm{a}_{21} \ldots \mathrm{a}_{2n} \ldots \mathrm{a}_{m1} \ldots \mathrm{a}_{mn} \right]^T$ η Ευκλείδεια νόρμα του **a** ορίζει την **Frobenius** νόρμα $\|A\|_F$ του πίνακα A και έχουμε

$$\|A\|_F = \left(\sum_{j=1}^{n} \|a_j\|_2^2\right)^{1/2} = \left( \sum_{i=1}^{m} \left\| \tilde{a}_i \right\|_2^2 \right)^{1/2} = \|A^*\|_F \tag{1.24}$$

Παρατηρήστε ότι η νόρμα αυτή είναι ίδια με τη νόρμα δεύτερης τάξης πίνακα σε χώρο $mn$ διάστασης. Ο τύπος της νόρμας Frobenius μπορεί να γραφτεί και ως μεμονωμένες στήλες ή γραμμές . Για παράδειγμα αν $a_j$ είναι η j-οστή στήλη του πίνακα $A$ τότε έχουμε:

$$\|A\|_F = \left(\sum_{j=1}^{n} \|a_j\|_2^2\right)^{1/2} = \left( \sum_{i=1}^{m} \left\| \tilde{a}_i \right\|_2^2 \right)^{1/2} \tag{1.25}$$

Όπου $\tilde{a}_i$ είναι οι γραμμές του A.

Αυτή η ταυτότητα καθώς και το ανάλογο αποτέλεσμα βασιζόμενο στις γραμμές αντί για τις στήλες μπορεί να εκφραστεί από την παρακάτω ισότητα:

$$\|A\|_F = \sqrt{tr(A^* A)} = \sqrt{tr(AA^*)} \tag{1.26}$$

όπου το $tr(B)$ είναι το ίχνος του πίνακα $B$ , το άθροισμα των διαγώνιων στοιχείων του.

Όπως μια επαγόμενη νόρμα πίνακα έτσι και η Frobenius νόρμα μπορεί να χρησιμοποιηθεί ως φράγμα για τα αποτελέσματα των πράξεων μεταξύ πινάκων. Έστω $C = AB$ με στοιχεία $c_{ik}$ και έστω $a_i^*$ να είναι η i-στη γραμμή του πίνακα $A$ και $b_j$ να είναι η j-οστή στήλη του πίνακα . Τότε $c_{ij} = a_i^* b_j$ τέτοιο ώστε από την ανισότητα Cauchy-Schwarz να έχουμε $\left|c_{ij}\right| \le \left\|a_i\right\|_2 \left\|b_j\right\|_2$ . Τετραγωνίζοντας και τα δυο μέλη και αθροίζοντας ως προς i και j έχουμε:

$$\begin{aligned} \left\|AB\right\|_F^2 &= \sum_{i=1}^{n} \sum_{j=1}^{m} \left|c_{ij}\right|^2 \\ &\le \sum_{i=1}^{n} \sum_{j=1}^{m} (\left\|a_i\right\|_2 \left\|b_j\right\|_2)^2 \\ &\le \sum_{i=1}^{n} (\left\|a_i\right\|_2)^2 \sum_{j=1}^{m} (\left\|b_j\right\|_2)^2 = \left\|A\right\|_F^2 \left\|B\right\|_F^2 \end{aligned}$$

**Παράδειγμα 1.3.5.1**

Για τον πίνακα

$$\mathrm{A}=\begin{bmatrix} -3 & -2 & 4 \\ 5 & -2 & -3 \\ 2 & 1 & -6 \end{bmatrix}$$

έχουμε

$\left\|a_1\right\|_1 = 10$ , $\left\|a_2\right\|_1 = 5, \left\|a_3\right\|_1 = 13 \Rightarrow \left\|A\right\|_1 = 13$

και

$$\left\|\tilde{\alpha}_1\right\|_1 = 9, \left\|\tilde{\alpha}_2\right\|_1 = 10, \left\|\tilde{\alpha}_3\right\|_1 = 9, \Rightarrow \left\|A\right\|_\infty = 10.$$

Επιπλέον για τον πίνακα

$$\mathrm{A}=\begin{bmatrix} -3 & -2 & 4 \\ 5 & -2 & -3 \\ 2 & 1 & -6 \end{bmatrix}$$

έχουμε

$\left\|A\right\|_2 = \sqrt{90.4952}$ και $\left\|A\right\|_F = \sqrt{108}$

Βλέπε σημειώσεις Ανάλυσης Πινάκων Ιωάννης Β. Μαρουλάς

## 1.4 Η ΠΑΡΑΓΟΝΤΟΠΟΙΗΣΗ ΙΔΙΑΖΟΝΤΩΝ ΤΙΜΩΝ (The Singular Value Decomposition-SVD)

Η SVD είναι μια παραγοντοποίηση πίνακα ο υπολογισμός της οποίας αποτελεί βήμα σε πολλούς αλγορίθμους. Ίσης σημαντικότητας είναι και η χρήση της SVD για εννοιολογικούς σκοπούς. Πολλά προβλήματα στο πεδίο της γραμμικής άλγεβρας μπορούν να γίνουν πιο εύκολα κατανοητά αν σκεφτούμε να εφαρμόσουμε την SVD παραγοντοποίηση.

### *1.4.1 ΜΙΑ ΓΕΩΜΕΤΡΙΚΗ ΕΡΜΗΝΕΙΑ*

Η παραγοντοποίηση SVD ερμηνεύεται με το παρακάτω γεωμετρικό φαινόμενο:
Η εικόνα της μοναδιαίας σφαίρας σε κάθε $m \times n$ πίνακα είναι μια υπερ-έλλειψη. Η μέθοδος αυτή μπορεί να εφαρμοστεί όχι μόνο σε πίνακες με πραγματικά αλλά και σε πίνακες με μιγαδικά στοιχεία. Στη γεωμετρική όμως περιγραφή υποθέτουμε ότι ο πίνακας έχει πραγματικά στοιχεία.
Ο όρος «υπερ-έλλειψη» αντιπροσωπεύει μια γενικοποίηση $m$-διαστάσεων μιας έλλειψης. Στο χώρο $R^m$ η υπερ-έλλειψη ορίζεται ως η επιφάνεια που προκύπτει από την επέκταση της μοναδιαίας σφαίρας στον $R^m$ από κάποιους παράγοντες $\sigma_1,.....,\sigma_m$ (πιθανόν μηδενικούς) σε κάποιες ορθογώνιες κατευθύνσεις $u_1,.....,u_m \in R^m$. Για ευκολία θεωρούμε $u_i$ να είναι τα μοναδιαία διανύσματα, για παράδειγμα $\|u_i\|_2 = 1$. Τα διανύσματα $\{\sigma_i u_i\}$ είναι οι κύριοι ημιάξονες της υπερ-έλλειψης με μήκη $\sigma_1,.....,\sigma_m$. Αν ο $A$ έχει τάξη $r$, ακριβώς $r$ από τα μήκη των $\sigma_i$ θα είναι τελικά μη μηδενικά και πιο συγκεκριμένα αν $m \geq n$ το πολύ $n$ από αυτά θα είναι μη μηδενικά.
Η αρχική υπόθεση για την εικόνα της μοναδιαίας σφαίρας έχει την ακόλουθη ερμηνεία. Από τη μοναδιαία σφαίρα , εννοούμε τη συνήθη Ευκλείδεια σφαίρα σε ένα $n$-διάστημα, για παράδειγμα η μοναδιαία σφαίρα της νόρμας δεύτερης τάξης. Έστω ότι συμβολίζουμε το διάστημα αυτό με $S$. Τότε $AS$, η εικόνα του $S$ σε μια χαρτογράφηση του $A$ είναι μια υπερ-έλλειψη όπως ορίστηκε παραπάνω.
Η γεωμετρία δεν είναι εμφανής. Προς το παρόν θα θεωρήσουμε ότι είναι εμφανής και θα την αποδείξουμε αργότερα.
Έστω $S$ να είναι η μοναδιαία σφαίρα στον $R^n$ και παίρνουμε κάθε $AS \in R^{m \times n}$ με $m \geq n$. Για ευκολία υποθέτουμε προς το παρόν ότι ο $A$ έχει πλήρη βαθμό $n$. Η εικόνα $AS$ είναι μια υπερ-έλλειψη στον $R^m$. Θα ορίσουμε κάποιες ιδιότητες του $A$ με βάση το σχήμα του $AS$.

Πρώτα θα ορίσουμε τις $n$ ιδιάζουσες τιμές του $A$. Αυτές οι τιμές είναι το μήκος των $n$ κύριων ημιαξόνων του $AS$ και τις συμβολίζουμε με $\sigma_1,.....,\sigma_n$. Είναι βολικό να

υποθέσουμε ότι οι ιδιάζουσες τιμές είναι σε φθίνουσα σειρά, δηλαδή $\sigma_1 \geq \sigma_2 \geq .....\geq \sigma_n > 0$.

Το επόμενο βήμα είναι να ορίσουμε τα $n$ αριστερά ιδιάζοντα διανύσματα του $A$. Αυτά είναι τα μοναδιαία διανύσματα $\{u_1, u_2, ....u_n\}$ με προσανατολισμό στις κατευθύνσεις των κύριων ημιαξόνων του $AS$, αριθμημένα έτσι ώστε να αντιστοιχούν στις ιδιάζουσες τιμές. Γι αυτό το διάνυσμα $\sigma_i u_i$ είναι ο i-οστός μεγαλύτερος κύριος ημιάξονας του $AS$.

Τέλος, ορίζουμε τα $n$ αριστερά ιδιάζοντα διανύσματα του $A$. Αυτά είναι τα μοναδιαία διανύσματα $\{v_1, v_2, ....v_n\} \in S$ έτσι ώστε οι προ-εικόνες των ημιαξόνων του $AS$ να είναι αριθμημένοι έτσι ώστε να ισχύει $Av_j = \sigma_j u_j$.

Οι όροι «αριστερά» και «δεξιά» στους παραπάνω ορισμούς προκύπτουν από τις θέσεις των παραγόντων $U$ και $V$ στην (1.28) και (1.29) πιο κάτω. Παράξενο είναι ότι στο σχήμα 1.1 τα αριστερά ιδιάζοντα διανύσματα αντιστοιχούν στο διάστημα στο δεξί μέρος του σχήματος και τα δεξιά ιδιάζοντα διανύσματα αντιστοιχούν στο αριστερό μέρος του σχήματος. Αυτό το πρόβλημα θα μπορούσε να λυθεί εναλλάσσοντας τα δυο μισά του σχήματος με την απεικόνιση του $A$ να κατευθύνεται από δεξιά προς αριστερά αλλά αυτό θα αποτελούσε παράβαση κάποιων μαθηματικών κανόνων.

### *1.4.2 ΜΕΙΩΜΕΝΗ SVD*

Στην προηγούμενη παράγραφο αναφέραμε ότι οι εξισώσεις σχετικά με τα δεξιά ιδιάζοντα διανύσματα $\{v_j\}$ και τα αριστερά $\{u_j\}$ μπορούν να γραφτούν ως:

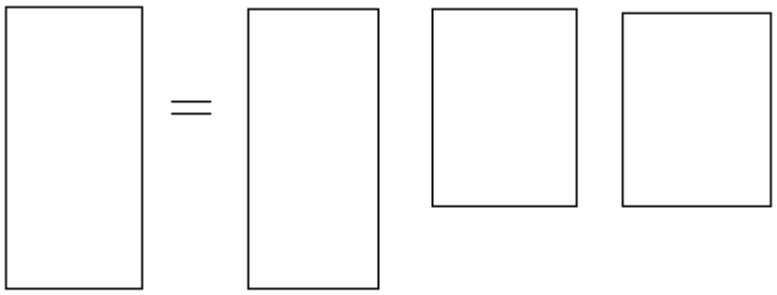
$$Av_j = \sigma_j u_j, \quad 1 \leq j \leq n \tag{1.27}.$$

Αυτό το σύστημα των εξισώσεων των διανυσμάτων μπορεί να εκφραστεί ως μια εξίσωση πίνακα ή πιο συμπαγές $AV = \hat{U}\hat{\Sigma}$. Σε αυτή την εξίσωση πίνακα $\Sigma \in R^{m\times n}$ είναι ένας $n \times n$ διαγώνιος πίνακας με θετικά πραγματικά στοιχεία (αφού υποθέσαμε ότι ο $A$ έχει πλήρη βαθμό $n$), $\hat{U}$ είναι ένας $m \times n$ πίνακας με ορθοκανονικές στήλες, και $V$ είναι ένας $n \times n$ πίνακας με ορθοκανονικές στήλες. Επειδή ο $V$ είναι ορθομοναδιαίος μπορούμε να πολλαπλασιάσουμε από δεξιά με τον αντίστροφο του $V^*$ και έχουμε σαν αποτέλεσμα:

$$A = \hat{U}\hat{\Sigma}V^* \tag{1.28}.$$

Η παραγοντοποίηση αυτή του $A$ ονομάζεται μειωμένος διαχωρισμός ιδιαζόντων τιμών ή μειωμένο SVD του $A$. Σχηματικά η παραγοντοποίηση αυτή μπορεί να αναπαρασταθεί ως εξής:

Μειωμένη SVD $(m \geq n)$



$A$ $\hat{U}$ $\hat{\Sigma}$ $V^*$

*1.4.3 ΠΛΗΡΗΣ SVD*

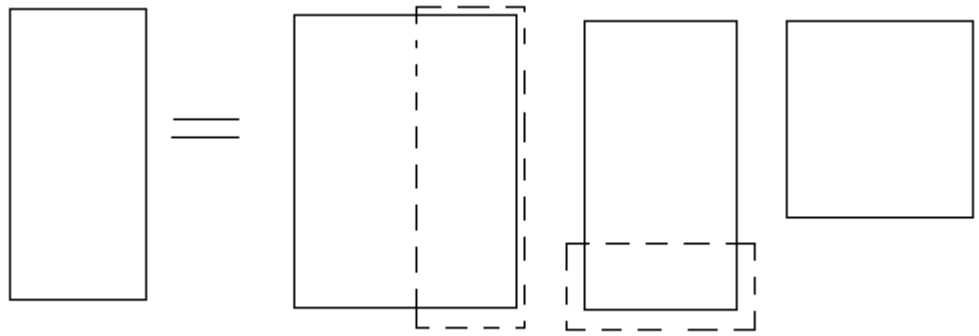
Στις περισσότερες εφαρμογές η παραγοντοποίηση SVD χρησιμοποιείται ακριβώς όπως περιγράφηκε παραπάνω. Εισάγαμε τον όρο «μειωμένο» και τα καπελάκια στους $\Sigma$ και $U$ για να διαχωρίσουμε την παραγοντοποίηση στην εξίσωση (1.28) από την πλήρη μέθοδο SVD. Η ιδέα είναι όπως περιγράφεται παρακάτω. Οι στήλες του $\hat{U}$ είναι $n$ ορθοκανονικά διανύσματα στο χώρο $m$-διάστασης $C^m$. Αν δεν ισχύει $m = n$ τότε τα διανύσματα αυτά δεν αποτελούν μια βάση για το χώρο αυτό ή ο $\hat{U}$ δεν είναι ένας ορθομοναδιαίος πίνακας. Όμως, με τις συζυγής $m - n$ ορθοκανονικές στήλες, ο $\hat{U}$ μπορεί να επεκταθεί σε ένα ορθομοναδιαίο πίνακα. Ας υποθέσουμε ότι μπορούμε να το κάνουμε αυτό κατά κάποιο τρόπο αυθαίρετα και ας ονομάσουμε το αποτέλεσμα $U$.

Αν ο $\hat{U}$ αντικατασταθεί από τον $U$ στην (1.28) τότε ο $\hat{\Sigma}$ θα πρέπει και αυτός να αλλάξει. Για να παραμείνει το αποτέλεσμα ίδιο πρέπει οι τελευταίες $m - n$ στήλες του $U$ να πολλαπλασιαστούν με το μηδέν. Οπότε έστω $\Sigma$ να είναι ένας $m \times n$ πίνακας που αποτελείται από τον $\hat{\Sigma}$ στο πάνω $n \times n$ μέρος μαζί με τις $m - n$ από κάτω. Τώρα έχουμε μια νέα παραγοντοποίηση, την πλήρη SVD του πίνακα $A$:

$$A = U\Sigma V^* \tag{1.29}$$

Όπου $U$ είναι ένας $m \times m$ ορθομοναδιαίος πίνακας, $V$ είναι ένας $n \times n$ ορθομοναδιαίος πίνακας και $\Sigma$ είναι ένας $m \times n$ διαγώνιος πίνακας αποτελούμενος από θετικά και πραγματικά στοιχεία. Σχηματικά η παραγοντοποίηση είναι:

Πλήρης SVD $(m \geq n)$



$A$ $U$ $\Sigma$ $V^*$

Οι διακεκομμένες γραμμές υποδηλώνουν τις «σιωπηλές» στήλες του $U$ και γραμμές του $\Sigma$ οι οποίες παραλείφτηκαν κατά τη μετάβαση από την εξίσωση (1.29) στην (1.28).

Έχοντας περιγράψει την πλήρη παραγοντοποίηση SVD δεν μας είναι πλέον χρήσιμη η απλοποιημένη υπόθεση ότι ο πίνακας $A$ έχει πλήρη βαθμό. Ακόμα και αν ο $A$ δεν έχει πλήρη βαθμό τότε η παραγοντοποίηση (1.29) είναι επίσης επαρκής. Αυτές οι αλλαγές αποσκοπούν στο ότι όχι το $n$ αλλά μόνο το $r$ των αριστερών ιδιαζόντων διανυσμάτων του $A$ θα καθορίζονται από τη γεωμετρία της υπερ-έλλειψης. Για να κατασκευάσουμε τον ορθομοναδιαίο πίνακα $U$ χρησιμοποιούμε $m-r$ αντί για $m-n$ διαφορετικές τυχαίες ορθοκανονικές στήλες. Ο πίνακας $V$ θα χρειαστεί επίσης $n-r$ τυχαίες ορθοκανονικές στήλες για να επεκτείνει τις $r$ στήλες που έχουν καθοριστεί από τη γεωμετρία. Ο πίνακας $\Sigma$ θα έχει τώρα $r$ θετικά διαγώνια στοιχεία και τα υπόλοιπα $n-r$ στοιχεία θα είναι μηδενικά.

Με βάση την ίδια φιλοσοφία, η μειωμένη παραγοντοποίηση SVD (1.28) θα ισχύει για πίνακες $A$ με όχι πλήρη βαθμό. Μπορούμε να θέσουμε ο πίνακας $\hat{U}$ να είναι $m \times n$ διάστασης , ο πίνακας $\hat{\Sigma}$ να έχει διάσταση $n \times n$ με κάποια μηδενικά στη διαγώνιο ή μπορούμε να πετύχουμε μια συμπιεσμένη μορφή της εξίσωσης αυτής με τον πίνακα $\hat{U}$ να είναι διάστασης $m \times r$ και ο πίνακας $\hat{\Sigma}$ να είναι διάστασης $r \times r$ με αυστηρά θετικά στοιχεία στη διαγώνιο.

### *1.4.4 ΕΠΙΣΗΜΟΣ ΟΡΙΣΜΟΣ*

Έστω ότι ορίζουμε τυχαία $m$ και $n$ και δεν απαιτούμε $m \geq n$ . Δοθέντος $A \in C^{m \times n}$ , όχι απαραιτήτων πλήρους βαθμού, μια παραγοντοποίηση SVD ενός πίνακα $A$ είναι μια παραγοντοποίηση της μορφής

$$A = U \Sigma V^* \tag{1.29}$$

όπου

$U \in C^{m\times m}$ είναι ορθομοναδιαίος

$V \in C^{n\times n}$ είναι ορθομοναδιαίος

$\Sigma \in R^{m\times n}$ είναι διαγώνιος .

Σε αντίθεση, υποθέτουμε ότι τα διαγώνια στοιχεία $\sigma_j$ του $\Sigma$ είναι μη αρνητικά σε φθίνουσα σειρά, δηλαδή $\sigma_1 \geq \sigma_2 \geq ..... \geq \sigma_p \geq 0$ όπου $p = \min(m,n)$ .

Ας σημειώσουμε ότι ο διαγώνιος πίνακας $\Sigma$ έχει το ίδιο σχήμα με τον $A$ ακόμα και αν ο $A$ δεν είναι τετραγωνικός αλλά οι πίνακες $U$ και $V$ είναι πάντα τετραγωνικοί ορθομοναδιαίοι πίνακες.

Είναι σαφές ότι η εικόνα της μοναδιαίας σφαίρας στον $R^n$ κάτω από μια απεικόνιση $A = U\Sigma V^*$ πρέπει να είναι μια υπερ-έλλειψη στον χώρο $R^m$ . Η ορθομοναδιαία απεικόνιση $V^*$ διατηρεί τη σφαίρα, ο διαγώνιος πίνακας $\Sigma$ επεκτείνει τη σφαίρα σε μια υπερ-έλλειψη σύμφωνα με την κανονική βάση και τέλος ο ορθομοναδιαίος πίνακας $U$ περιστρέφει ή αντανακλά την υπερ-έλλειψη χωρίς να αλλάζει το σχήμα της. Οπότε μπορούμε να αποδείξουμε ότι κάθε πίνακας έχει μια SVD και επίσης μπορούμε να αποδείξουμε ότι η εικόνα της μοναδιαίας σφαίρας κάτω από κάθε γραμμική απεικόνιση είναι μια υπερ-έλλειψη.

### *1.4.5 ΥΠΑΡΞΗ ΚΑΙ ΜΟΝΑΔΙΚΟΤΗΤΑ*

ΘΕΩΡΗΜΑ 1.6

Κάθε πίνακας $A \in C^{m\times n}$ έχει μια ιδιάζουσα τιμή διαχωρισμού (1.30). Επίσης οι ιδιάζουσες τιμές $\{\sigma_j\}$ είναι μοναδικά ορισμένες και αν ο $A$ είναι τετραγωνικός και οι $\sigma_j$ είναι διακεκριμένες , τα αριστερά και δεξιά ιδιάζοντα διανύσματα $\{u_j\}$ και $\{v_j\}$ είναι μοναδικά ορισμένα μιγαδικά σύμβολα (για παράδειγμα μιγαδικοί βαθμωτοί παράγοντες μιας απόλυτης τιμής 1).

ΑΠΟΔΕΙΞΗ

Βλέπε βιβλιογραφία [1] σελ. 29

## 1.5 ΠΕΡΙΣΣΟΤΕΡΑ ΣΤΗΝ ΠΑΡΑΓΟΝΤΟΠΟΙΗΣΗ SVD

Συνεχίζουμε με την παρουσίαση εισαγωγικών θεμάτων για την παραγοντοποίηση SVD δίνοντας έμφαση στο πώς σχετίζεται με την προσέγγιση πινάκων χαμηλού βαθμού στη νόρμα δεύτερης τάξης και στη νόρμα Frobenius.

### *1.5.1 ΑΛΛΑΓΗ ΒΑΣΗΣ*

Η παραγοντοποίηση SVD μας εξασφαλίζει ότι κάθε πίνακας είναι διαγώνιος μόνο αν χρησιμοποιεί τις σωστές βάσεις για τα κύρια και πεδίο τιμών διαστημάτων.
Σε αυτή την παράγραφο θα εξετάσουμε πώς λειτουργεί η αλλαγή βάσης. Κάθε $b \in C^m$ μπορεί να επεκταθεί στη βάση των αριστερών ιδιαζόντων διανυσμάτων του πίνακα $A$ (στήλες του $U$) και κάθε $x \in C^n$ μπορεί να επεκταθεί στη βάση των δεξιών ιδιαζόντων διανυσμάτων του πίνακα $A$ (στήλες του $V$). Τα διανύσματα που προκύπτουν για αυτές τις επεκτάσεις είναι:

$$b' = U^* b\,, \qquad x' = V^* x\,.$$

Από την (1.29) η σχέση $b = Ax$ μπορεί να εκφραστεί με βάση τα $b'$ και $x'$:

$$b = Ax \Leftrightarrow U^* b = U^* Ax = U^* U \Sigma V^* x \Leftrightarrow b' = \Sigma x'$$

Όπου $b = Ax$ τώρα έχουμε $b' = \Sigma x'$. Το συμπέρασμα είναι ότι ο πίνακας $A$ μειώνεται στο διαγώνιο πίνακα $\Sigma$ όταν ο βαθμός πίνακα εκφράζεται στη βάση των στηλών του $U$ και το χωρίο εκφράζεται βάση των στηλών του πίνακα $V$.

### *1.5.2 ΣΥΓΚΡΙΣΗ ΠΑΡΑΓΟΝΤΟΠΟΙΗΣΗΣ SVD ΚΑΙ ΔΙΑΧΩΡΙΣΜΟΥ ΙΔΙΟΤΙΜΩΝ*

Η διαδικασία της διαγωνιοποίησης ενός πίνακα εκφράζοντάς τον με μια νέα βάση είναι χρήσιμη στη μελέτη των ιδιοτιμών. Ένας μη ελλιπής τετράγωνος πίνακας $A$ μπορεί να εκφραστεί σαν ένας διαγώνιος πίνακας των ιδιοτιμών $\Lambda$ αν το πεδίο τιμών και η περιοχή παρουσιάζονται σε μια βάση των ιδιοτιμών.
Αν οι στήλες ενός πίνακα $X \in C^{m\times m}$ περιέχουν γραμμικά ανεξάρτητες ιδιοτιμές του $A \in C^{m\times m}$, ο διαχωρισμός των ιδιοτιμών του $A$ είναι:

$$A = X \Lambda X^{-1} \tag{1.30}$$

όπου $\Lambda$ είναι ένας $m \times m$ διαγώνιος πίνακας του οποίου τα στοιχεία είναι οι ιδιοτιμές του πίνακα $A$. Δηλαδή αν ορίσουμε για $b$, $x \in C^m$ να ικανοποιείται η $b = Ax$,

$$b' = X^{-1} b\,, \quad x' = X^{-1} x\,,$$

τότε τα καινούργια διανύσματα $b'$ και $x'$ που έχουν επεκταθεί ικανοποιούν την εξίσωση $b' = \Lambda x'$ .

Υπάρχουν θεμελιώδεις διαφορές ανάμεσα στην παραγοντοποίηση SVD και τη μέθοδο του διαχωρισμού των ιδιοτιμών. Μια διαφορά είναι ότι η παραγοντοποίηση SVD χρησιμοποιεί δυο διαφορετικές βάσεις (τα σύνολα των αριστερών και δεξιών ιδιαζόντων διανυσμάτων) ενώ η μέθοδος του διαχωρισμού των ιδιοτιμών χρησιμοποιεί μόνο μια βάση (τα ιδιοδιανύσματα). Μια άλλη διαφορά είναι ότι η παραγοντοποίηση SVD χρησιμοποιεί ορθοκανονικές βάσεις ενώ η μέθοδος του διαχωρισμού των μεταβλητών χρησιμοποιεί μια βάση η οποία γενικά δεν είναι ορθοκανονική. Μια τρίτη διαφορά είναι ότι δεν μπορεί σε όλους τους πίνακες (ακόμα και στους τετράγωνους) να γίνει διαχωρισμός ιδιοτιμών αλλά σε όλους τους πίνακες (ακόμα και στους ορθογώνιους) μπορεί να εφαρμοστεί η παραγοντοποίηση SVD όπως εξασφαλίσαμε και από το θεώρημα 1.6. Στις εφαρμογές, οι ιδιοτιμές τείνουν να είναι σχετικές με προβλήματα που περιέχουν τη συμπεριφορά των επαναλαμβανόμενων μορφών του πίνακα $A$ , όπως οι δυνάμεις πίνακα $A^k$ ή εκθετικά $e^{tA}$ , ενώ τα ιδιάζοντα διανύσματα τείνουν να είναι σχετικά με προβλήματα που περιέχουν τη συμπεριφορά του $A$ ή του ανάστροφού του.

### *1.5.3 ΙΔΙΟΤΗΤΕΣ ΠΙΝΑΚΑ ΜΕΣΩ ΤΟΥ SVD*

Η χρησιμότητα της παραγοντοποίησης SVD γίνεται φανερή όταν βλέπουμε τη σύνδεση που έχει με θεμελιώδη πεδία της γραμμικής άλγεβρας. Για τα παρακάτω θεωρήματα υποθέτουμε ότι ο πίνακας $A$ είναι διάστασης $m \times n$ . Έστω $p$ να είναι το μικρότερο από τα $m$ και $n$ , έστω $r \le p$ να είναι το πλήθος των μη μηδενικών ιδιαζόντων τιμών του $A$ και έστω $\langle x, y, ....., z \rangle$ να είναι το σύνολο που παράγεται από τα διανύσματα $x, y, ...., z$ .

ΘΕΩΡΗΜΑ 1.7

Η τάξη του πίνακα $A$ είναι $r$ , ο αριθμός των μη μηδενικών ιδιαζόντων τιμών.

ΑΠΟΔΕΙΞΗ

Ο βαθμός ενός διαγώνιου πίνακα είναι ίσος με τον αριθμό των μη μηδενικών του στοιχείων και στο διαχωρισμό $A = U\Sigma V^*$ , οι πίνακες $U$ και $V$ δεν έχουν πλήρη βαθμό. Οπότε $rank(A) = rank(\Sigma) = r$ .

ΘΕΩΡΗΜΑ 1.8

$range(A) = \langle u_1, ....., u_r \rangle$ και $null(A) = \langle v_{r+1}, ....., v_n \rangle$

ΑΠΟΔΕΙΞΗ

Είναι μια συνέπεια του ότι $range(\Sigma) = \langle e_1, e_2, ....., e_r \rangle \subseteq C^m$ και ότι $null(\Sigma) = \langle e_{r+1}, ....., e_n \rangle \subseteq C^n$.

ΘΕΩΡΗΜΑ 1.9

$\|A\|_2 = \sigma_1$ και $\|A\|_F = \sqrt{\sigma_1^2 + \sigma_2^2 + ... + \sigma_r^2}$ .

ΑΠΟΔΕΙΞΗ

Το πρώτο αποτέλεσμα προκύπτει από την απόδειξη του θεωρήματος 1.6: αφού $A = U\Sigma V^*$ με τους ορθομοναδιαίους πίνακες $U$ και $V$ , $\|A\|_2 = \|\Sigma\|_2 = \max\{|\sigma_j|\} = \sigma_1$ από το θεώρημα 1.5. Για το δεύτερο αποτέλεσμα ας σημειώσουμε ότι από το θεώρημα 1.5 και το παρακάτω σχόλιο, η νόρμα Frobenius είναι αναλλοίωτη στον πολλαπλασιασμό με ορθομοναδιαίο πίνακα, οπότε $\|A\|_F = \|\Sigma\|_F$ και από την εξίσωση (1.24) προκύπτει το ζητούμενο.

ΘΕΩΡΗΜΑ 1.10

Οι μη μηδενικές ιδιάζουσες τιμές του πίνακα $A$ είναι οι τετραγωνικές ρίζες των μη μηδενικών ιδιοτιμών του $A^*A$ ή του $AA^*$. (Οι πίνακες αυτοί έχουν τις ίδιες μη μηδενικές ιδιοτιμές.)

ΑΠΟΔΕΙΞΗ

Από τον υπολογισμό

$$A^*A = (U\Sigma V)^*(U\Sigma V^*) = V\Sigma^* U^* U\Sigma V^* = V(\Sigma^*\Sigma)V^*,$$

βλέπουμε ότι $A^*A$ είναι όμοιος με $\Sigma^*\Sigma$ και κατά συνέπεια έχει τις ίδιες $n$ ιδιοτιμές. Οι ιδιοτιμές του διαγώνιου πίνακα $\Sigma^*\Sigma$ είναι $\sigma_1^2, \sigma_2^2, ....., \sigma_p^2$ με $n - p$ επιπρόσθετες μηδενικές ιδιοτιμές αν $n > p$. Ένας παρόμοιος υπολογισμός μπορεί να γίνει για τις $m$ ιδιοτιμές του $AA^*$. □

ΘΕΩΡΗΜΑ 1.11

Αν $A = A^*$ τότε οι ιδιάζουσες τιμές του πίνακα $A$ είναι οι απόλυτες τιμές των ιδιοτιμών του πίνακα $A$ .

ΑΠΟΔΕΙΞΗ

Όπως είναι γνωστό, ένας ερμιτιανός πίνακας έχει ολοκληρωμένο σύνολο ορθογωνικών ιδιοδιανυσμάτων και όλες οι ιδιοτιμές του είναι πραγματικές. Ένας ισοδύναμος ισχυρισμός είναι ότι η εξίσωση (1.30) ισχύει για $X$ ίσο με κάποιο

ορθομοναδιαίο πίνακα $Q$ και ένα πραγματικό διαγώνιο πίνακα $A$ . Τότε μπορούμε να γράψουμε:

$$A = Q\Lambda Q^* = Q|\Lambda|\,sign(\Lambda)Q^* \tag{1.31}$$

όπου $|\Lambda|$ και $sign(\Lambda)$ είναι οι διαγώνιοι πίνακες των οποίων τα στοιχεία είναι τα νούμερα $|\lambda_j|$ και $sign(\lambda_j)$ αντίστοιχα. Αφού $sign(\Lambda)Q^*$ είναι ορθομοναδιαίος όποτε ο $Q$ είναι ορθομοναδιαίος η εξίσωση (1.30) είναι μια παραγοντοποίηση SVD για τον πίνακα $A$ με τις ιδιάζουσες τιμές να είναι ίσες με τα διαγώνια στοιχεία του $|\Lambda|$, δηλαδή τα $|\lambda_j|$. Αν χρειάζεται μπορούμε να βάλουμε τα νούμερα αυτά σε φθίνουσα σειρά εισάγοντας κατάλληλη μετάθεση στους πίνακες ως παράγοντες στον από αριστερά ορθομοναδιαίο πίνακα της εξίσωσης (1.30), $Q$ και στον από δεξιά ορθομοναδιαίο πίνακα $sign(\Lambda)Q^*$.

ΘΕΩΡΗΜΑ 1.12

Για $A \in C^{m\times m}$, $|\det(A)| = \prod_{i=1}^{m} \sigma_i$ .

ΑΠΟΔΕΙΞΗ

Η ορίζουσα ενός γινομένου τετράγωνων πινάκων είναι το γινόμενο των οριζουσών των παραγόντων. Η ορίζουσα ενός ορθομοναδιάιου πίνακα είναι πάντα 1 σε απόλυτη τιμή. Αυτό προκύπτει από τη σχέση $U^*U = I$ και την ιδιότητα $\det(U^*) = (\det(U))^*$. Οπότε,

$$|\det(A)| = |\det(U\Sigma V^*)| = |\det(U)||\det(\Sigma)||\det(V^*)| = |\det(\Sigma)| = \prod_{i=1}^{m} \sigma_i$$

### *1.5.4 ΠΡΟΣΕΓΓΙΣΕΙΣ ΧΑΜΗΛΗΣ ΤΑΞΗΣ*

Τι είναι όμως η παραγοντοποίηση SVD; Μια άλλη προσέγγιση για να δώσουμε κάποια εξήγηση είναι να σκεφτούμε πώς ένας πίνακας $A$ μπορεί να αντιπροσωπευτεί από ένα σύνολο πινάκων πρώτης τάξης 1.

ΘΕΩΡΗΜΑ 1.13

$A$ είναι το σύνολο $r$ πινάκων τάξης 1: $A = \sum_{j=1}^{r} \sigma_j u_j v_j^*$ (1.32)

ΑΠΟΔΕΙΞΗ

Αν γράψουμε το $\Sigma$ σαν ένα άθροισμα $r$ πινάκων $\Sigma_j$ , όπου $\Sigma_j = diag(0,0,.....,0,\sigma_j,0,.....,0)$ τότε η (1.32) προκύπτει από την (1.29).

Υπάρχουν πολλοί τρόποι για να εκφράσουμε ένα $m \times n$ πίνακα $A$ σαν ένα άθροισμα πινάκων τάξης 1. Για παράδειγμα, ο $A$ θα μπορούσε να γραφτεί ως το άθροισμα $m$ γραμμών του ή των $n$ στηλών του ή των $mn$ στοιχείων του. Ένα άλλο παράδειγμα είναι να εφαρμόσουμε την απαλοιφή Gauss και να μειώσουμε τον πίνακα $A$ στο άθροισμα ενός πλήρη πίνακα τάξης 1, πίνακας τάξης 1 του οποίου η πρώτη γραμμή και στήλη είναι μηδενικές, ένας πίνακας τάξης 1 του οποίου οι δυο πρώτες γραμμές και στήλες είναι μηδενικές και ούτως κάθε εξής.
Η εξίσωση (1.32) όμως αντιπροσωπεύει ένα διαχωρισμό σε πίνακες τάξης 1 με μια διαφορετική ιδιότητα: το $v$-οστό μερικό άθροισμα «παγιδεύει» όσο το δυνατόν περισσότερη ενέργεια του πίνακα $A$. Σε αυτή την ιδιότητα ο όρος «ενέργεια» ορίζεται ως είτε η νόρμα δεύτερης τάξης ή η νόρμα Frobenius.

ΘΕΩΡΗΜΑ 1.14

Για κάθε $v$ με $0 \le v \le r$ ορίζουμε,

$$A_v = \sum_{j=1}^{r} \sigma_j u_j v_j^* \quad , \tag{1.33}$$

Διαφορετικά αν $v = p = \min\{m,n\}$, ορίζουμε $\sigma_{v+1} = 0$. Τότε

$$\|A - A_v\|_2 = \inf_{\substack{B \in C^{m \times n} \\ rank(B) \le v}} \|A - B\|_2 = \sigma_{v+1}$$

ΑΠΟΔΕΙΞΗ

Έστω ότι υπάρχει κάποιος $B$ με $rank(B) \le v$ τέτοιο ώστε $\|A - B\|_2 < \|A - A_v\|_2 = \sigma_{v+1}$. Τότε υπάρχει ένα $(n-v)$ -διάστασης υποδιάστημα $W \subseteq C^n$ τέτοιο ώστε $w \in W \Rightarrow Bw = 0$. Κατά συνέπεια, για κάθε $w \in W$ έχουμε ότι $Aw = (A-B)w$ και

$$\|Aw\|_2 = \|(A-B)w\|_2 \le \|A-B\|_2 \|w\|_2 < \sigma_{v+1} \|w\|_2$$

Οπότε $W$ είναι υποδιάστημα $(n-v)$-διάστασης όπου $\|Aw\| < \sigma_{v+1} \|w\|$. Αλλά υπάρχει ένα υποδιάστημα διάστασης $(v+1)$ όπου $\|Aw\| \ge \sigma_{v+1} \|w\|$, δηλαδή το διάστημα που παράγεται από τα πρώτα $(v+1)$ δεξιά ιδιάζοντα διανύσματα του $A$. Αφού το άθροισμα των διαστάσεων αυτών των διαστημάτων υπερβαίνει το $n$ τότε πρέπει

να υπάρχει ένα μη μηδενικό διάνυσμα και στα δύο, κάτι που καταλήγει σε άτοπο.

Το θεώρημα 1.14 έχει γεωμετρική ερμηνεία. Ποια είναι η καλύτερη προσέγγιση μιας υπερέλλειψης από ένα ευθύγραμμο τμήμα; Ας υποθέσουμε ότι το ευθύγραμμο τμήμα είναι ο μεγαλύτερος άξονας. Ποια είναι η καλύτερη προσέγγιση για ένα ελλειψοειδές δυο διαστάσεων; Ας υποθέσουμε ότι είναι το ελλειψοειδές που παράγεται από τους δυο μεγαλύτερους άξονες. Αν συνεχίσουμε με αυτή τη λογική σε κάθε βήμα θα βελτιώνουμε την προσέγγισή μας προσθέτοντας στην προσέγγισή μας τον μεγαλύτερο άξονα του υπερελλειψοειδούς που δεν έχουμε συμπεριλάβει ακόμα. Ύστερα από $r$ βήματα έχουμε καλύψει όλο τον πίνακα $A$ .Η ιδέα αυτή έχει εφαρμογές στην εφαρμοσμένη ανάλυση.
Παραθέτουμε το ανάλογο αποτέλεσμα για τη νόρμα Frobenius χωρίς απόδειξη.
ΘΕΩΡΗΜΑ 1.15
Για κάθε $v$ με $0 \le v \le r$ ο πίνακας $A_v$ της (1.33) ικανοποιεί επίσης και τη σχέση:

$$\|A - A_v\|_F = \inf_{\substack{B \in C^{m \times n} \\ rank(B) \le v}} \|A - B\|_F = \sqrt{\sigma_{v+1}^2 + ..... + \sigma_r^2}$$

## *1.5.5 ΥΠΟΛΟΓΙΣΜΟΣ ΤΗΣ ΠΑΡΑΓΟΝΤΟΠΟΙΗΣΗΣ SVD*

Σε αυτή και στην προηγούμενη παράγραφο εξετάσαμε τις ιδιότητες της παραγοντοποίησης SVD αλλά δεν σκεφτήκαμε πώς αυτή μπορεί να υπολογιστεί. Ο υπολογισμός της SVD παρουσιάζει ενδιαφέρον. Οι καλύτερες μέθοδοι είναι διάφοροι αλγόριθμοι οι οποίοι χρησιμοποιούνται για τον υπολογισμό ιδιοτιμών.
Από τη στιγμή που μπορούμε να την υπολογίσουμε, η SVD μπορεί να χρησιμοποιηθεί ως εργαλείο επίλυσης για όλα τα είδη προβλημάτων. Στην πραγματικότητα τα περισσότερα από τα θεωρήματα αυτής της παραγράφου έχουν υπολογιστικές συνέπειες. Η καλύτερη μέθοδος για να βρούμε την τάξη ενός πίνακα είναι να υπολογίσουμε τον αριθμό των ιδιαζόντων τιμών που είναι μεγαλύτερες από μια τιμή ανοχής (θεώρημα 1.7) (Για τα δυο αυτά παραδείγματα η παραγοντοποίηση QR προσφέρει εναλλακτικούς αλγορίθμους οι οποίοι είναι γρηγορότεροι αλλά όχι πιο ακριβείς). Η πιο ακριβής μέθοδος για να βρούμε μια ορθοκανονική βάση ενός πεδίου τιμών ή ενός μηδενικού χώρου είναι μέσω του θεωρήματος 1.8. Το θεώρημα 1.9 αντιπροσωπεύει τη βασική μέθοδο για τον υπολογισμό της $\|A\|_2$ και τα θεωρήματα 1.14 και 1.15 αντιπροσωπεύουν τις βασικές μεθόδους για τον υπολογισμό προσεγγίσεων χαμηλής τάξης με αναφορά στις νόρμες $\|\cdot\|_2$ και $\|\cdot\|_F$. Εκτός από αυτά τα παραδείγματα η SVD είναι επίσης μέρος ανθεκτικών αλγορίθμων για ελάχιστα τετράγωνα , τομή υποδιαστημάτων, κανονικοποίηση και πολλά άλλα προβλήματα.

# ΚΕΦΑΛΑΙΟ 2
# QR ΠΑΡΑΓΟΝΤΟΠΟΙΗΣΗ ΚΑΙ ΕΛΑΧΙΣΤΑ ΤΕΤΡΑΓΩΝΑ

## 2.1 ΠΡΟΒΟΛΕΣ

Σε αυτό το κεφάλαιο το κύριο θέμα είναι η ορθογωνιότητα. Θα ξεκινήσουμε με τα θεμελιώδη εργαλεία των πινάκων προβολής ή των ορθογώνιων και μη ορθογώνιων προβολών.

### *2.1.1 ΠΡΟΒΟΛΕΙΣ*

Ένας προβολέας είναι ένας τετραγωνικός πίνακας $P$ ο οποίος ικανοποιεί την παρακάτω σχέση:

$$P^2 = P \tag{2.1}$$

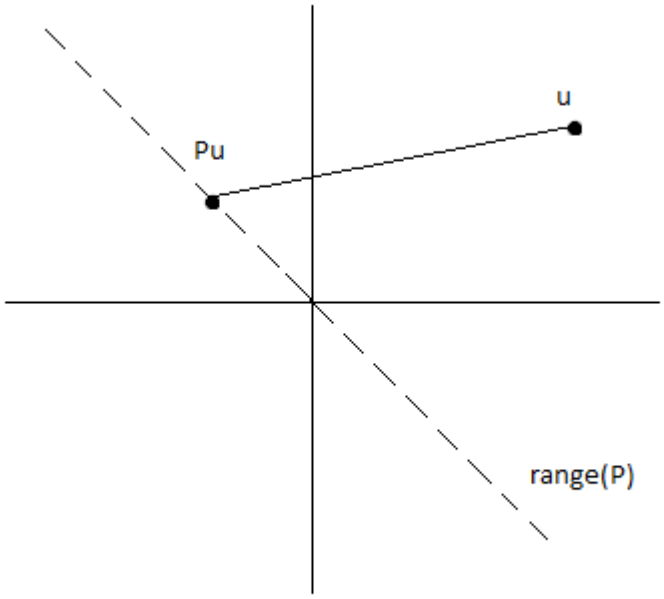
(ένας τέτοιος πίνακας λέγεται και ταυτοδύναμος.) Ο ορισμός αυτός περιλαμβάνει όχι μόνο τους ορθογώνιους αλλά και τους μη ορθογώνιους πίνακες. Για να αποφύγουμε τυχόν λάθη θα χρησιμοποιήσουμε τον όρο πλάγιος προβολέας για τους μη ορθογώνιους πίνακες.
Ο όρος προβολέας μπορεί να ερμηνευτεί αν σκεφτούμε το εξής: αν ρίχναμε φως στο υποδιάστημα $range(P)$ μόνο από τη δεξιά πλευρά τότε $P_u$ θα ήταν η σκιά που θα πρόβαλλόνταν από το διάνυσμα $u$.
Ας παρατηρήσουμε ότι αν $u \in range(P)$ τότε το $u$ ταυτίζεται με τη σκιά του και ο προβολέας είναι το ίδιο το $u$. Σε μαθηματικούς όρους έχουμε τη σχέση $u = Px$ για κάποιο $x$ και

$$Pu = P^2 x = Px = u$$

Όμως, από ποια διεύθυνση φωτίζει το φως όταν $u \neq Pu$; Γενικά η απάντηση εξαρτάται από το $u$ αλλά για κάποιο συγκεκριμένο $u$ η απάντηση προκύπτει αν τραβήξουμε την ευθεία από το $u$ στο $Pu$, $Pu - u$ (γράφημα 2.1). Εφαρμόζοντας τον προβολέα σε αυτό το διάνυσμα προκύπτει μηδενικό αποτέλεσμα:

$$P(Pu - u) = P^2 u - Pu = 0$$

Αυτό σημαίνει ότι $(Pu - u) \in null(P)$. Δηλαδή η διέθυνση του φωτός μπορεί να διαφέρει για διάφορετικό $u$ αλλά πάντα περιγράφεται από ένα διάνυσμα στο $null(P)$.




Σχήμα 2.1 Μια πλάγια προβολή

### *2.1.2 ΣΥΜΠΛΗΡΩΜΑΤΙΚΟΙ ΠΡΟΒΟΛΕΙΣ*

Αν ο $P$ είναι ένας προβολέας τότε ο $I-P$ είναι επίσης ένας προβολέας ο οποίος είναι επίσης ταυτοδύναμος:

$$(I-P)^2 = I-2P+P^2 = I-P$$

Ο πίνακας $I-P$ ονομάζεται επίσης συμπληρωματικός προβολέας του $P$.
Σε ποιο διάστημα προβάλλει ο $I-P$; Ακριβώς στο μηδενοχώρο του $u \in C^m$! Γνωρίζουμε ότι $range(I-P) \supseteq null(P)$ επειδή αν $Pu=0$ τότε έχουμε $(I-P)u=u$. Αντίστροφα, γνωρίζουμε ότι $range(I-P) \subseteq null(P)$, επειδή για κάθε $u$ έχουμε ότι $(I-P)u = u-Pu \in null(P)$. Οπότε για κάθε προβολέα $P$

$$range(I-P) = null(P) \qquad (2.2)$$

Γράφοντας $P = I-(I-P)$ προκύπτει το συμπληρωματικό γεγονός

$$null(I-P) = range(P) \qquad (2.3)$$

Μπορούμε επίσης να δούμε ότι $null(I-P) \cap null(P) = \{0\}$: για κάθε διάνυσμα $u$ και στα δυο σύνολα ικανοποιείται η σχέση $u = u - Pu = (I-P)u = 0$. Ένας άλλος τρόπος για να διατυπώσουμε το παραπάνω είναι

$$range(P) \cap null(P) = \{0\} \tag{2.4}$$

Αυτοί οι υπολογισμοί δείχνουν ότι ένας προβολέας διαχωρίζει το χώρο $C^m$ σε δυο διαστήματα. Αντίστροφα, έστω $S_1$ και $S_2$ να είναι τα δυο υποδιαστήματα του $C^m$ τέτοια ώστε $S_1 \cap S_2 = \{0\}$ και $S_1 + S_2 = C^m$, όπου $S_1 + S_2$ είναι η θήκη των $S_1$ και $S_2$ δηλαδή το σύνολο των διανυσμάτων $s_1 + s_2$ με $s_1 \in S_1$ και $s_2 \in S_2$. (Τέτοια ζεύγη ονομάζονται συμπληρωματικά υποδιαστήματα.) Τότε υπάρχει ένας προβολέας $P$ τέτοιος ώστε $range(P) = S_1$ και $null(P) = S_2$. Λέμε ότι ο $P$ είναι ένας προβολέας πάνω στο $S_1$ κατά μήκος του $S_2$. Αυτός ο προβολέας και ο συμπληρωματικός του μπορούν να είναι η μοναδική λύση στο παρακάτω πρόβλημα:

Δοθέντος $u$ βρείτε τα διανύσματα $u_1 \in S_1$ και $u_2 \in S_2$ τέτοια ώστε $u_1 + u_2 = u$

Από την προβολή $Pu$ προκύπτει το $u_1$ και από τη συμπληρωματική προβολή του $(I-P)u$ προκύπτει το $u_2$. Αυτά τα διανύσματα είναι μοναδικά επειδή όλες οι λύσεις πρέπει να είναι της μορφής

$$(Pu + u_3) + ((I-P)u - u_3) = u,$$

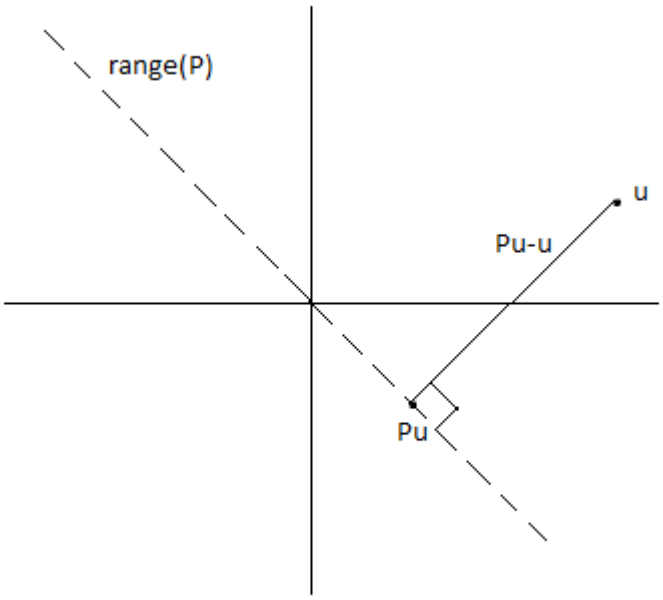
όπου είναι φανερό ότι το $u_3$ πρέπει να ανήκει και στο $S_1$ και στο $S_2$, για παράδειγμα $u_3 = 0$.

Ένα συμφραζόμενο όπου εμφανίζονται οι προβολείς και οι συμπληρωματικοί τους είναι γνωστό. Ας υποθέσουμε ότι ένας $m \times m$ πίνακας $A$ έχει ολοκληρωμένο σύνολο ιδιοδιανυσμάτων $\{u_j\}$ όπως την (1.30), με την έννοια ότι τα $\{u_j\}$ είναι μια βάση του $C^m$.

### *2.1.3 ΟΡΘΟΓΩΝΙΟΙ ΠΡΟΒΟΛΕΙΣ*

Ένας ορθογώνιος προβολέας (Σχήμα 2.2) είναι ένα διάνυσμα το οποίο προβάλλει στο υποδιάστημα $S_1$ κατά μήκος ενός διαστήματος $S_2$, όπου τα $S_1$ και $S_2$ είναι ορθογώνια. (Προσοχή: Οι ορθογώνιοι προβολείς δεν είναι ορθογώνιοι πίνακες!)

Υπάρχει επίσης και μια αλγεβρική ερμηνεία: ένας ορθογώνιος προβολέας είναι κάθε προβολέας που είναι ερμιτιανός, δηλαδή ικανοποιεί τη σχέση $P^* = P$ καθώς και τη σχέση (2.1). Φυσικά, πρέπει να εξασφαλίσουμε ότι αυτός ο ορισμός είναι ισοδύναμος με τον πρώτο.




Σχήμα 2.2 Μια ορθογώνια προβολή

ΘΕΩΡΗΜΑ 2.1

Ένας προβολέας $P$ είναι ορθογώνιος αν και μόνο αν $P = P^*$

ΑΠΟΔΕΙΞΗ

Αν $P = P^*$ τότε το εσωτερικό γινόμενο ανάμεσα σε ένα διάνυσμα $Px \in S_1$ και ένα διάνυσμα $(I - P)y \in S_2$ είναι ίσο με το μηδέν:

$$x^* P^* (I - P) y = x^* (P - P^2) y = 0$$

Οπότε ο προβολέας είναι ορθογώνιος και το ευθύ του θεωρήματος έχει αποδειχτεί.

Για να αποδείξουμε το αντίστροφο μπορούμε να χρησιμοποιήσουμε την παραγοντοποίηση SVD. Ας υποθέσουμε ότι ο $P$ προβάλλει πάνω στο $S_1$ κατά μήκος του $S_2$ όπου $S_1 \perp S_2$ και το $S_1$ έχει διάσταση $n$. Τότε μια SVD του $P$ μπορεί να κατασκευαστεί ως ακολούθως:

Έστω $\{q_1, q_2, ....., q_m\}$ να είναι μια ορθοκανονική βάση για τον $C^m$ , όπου $\{q_1, q_2, ....., q_n\}$ είναι μια βάση για το $S_1$ και $\{q_{n+1}, q_2, ....., q_m\}$ είναι μια βάση για το $S_2$. Για $j \le n$ έχουμε $P_{q_j} = q_j$ και για $j < n$ έχουμε $P_{q_j} = 0$. Τώρα έστω $Q$ να είναι ο ορθομοναδιαίος πίνακας του οποίου η j-οστή στήλη είναι το $q_j$. Τότε έχουμε:

$$PQ = \begin{bmatrix} & & & & \\ q_1 & \dots & q_n & 0 & \dots \\ & & & & \end{bmatrix}$$

έτσι ώστε

$$Q^* PQ = \begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & 0 & \\ & & & & \ddots \end{bmatrix} = \Sigma$$

ένας διαγώνιος πίνακας με άσσους τα πρώτα $n$ στοιχεία μηδενικά όλα τα υπόλοιπα. Οπότε φτιάξαμε μια παραγοντοποίηση SVD του πίνακα $P$ :

$$P = Q \Sigma Q^* \qquad (2.5)$$

(Σημειώστε ότι αυτή είναι και διαχωρισμός ιδιοτιμών (1.30).) Οπότε βλέπουμε ότι ο πίνακας $P$ είναι ερμιτιανός αφού

$$P^* = (Q \Sigma Q^*)^* = Q \Sigma^* Q^* = Q \Sigma Q^* = P$$

### *2.1.4 ΠΡΟΒΟΛΗ ΣΕ ΜΙΑ ΟΡΘΟΚΑΝΟΝΙΚΗ ΒΑΣΗ*

Αφού ένας ορθογώνιος προβολέας έχει κάποιες ιδιάζουσες τιμές ίσες με το μηδέν (εκτός από την τετριμμένη περίπτωση όπου $P = I$ ) είναι φυσιολογικό να παραλείψουμε τις σιωπηλές στήλες του Q στη σχέση (2.5) και να χρησιμοποιήσουμε τη μειωμένη παρά την πλήρη παραγοντοποίηση SVD. Προκύπτει η πολύ απλή σχέση

$$P = \hat{Q} \hat{Q}^* \qquad (2.6)$$

όπου οι στήλες του $\hat{Q}$ είναι ορθοκανονικές.

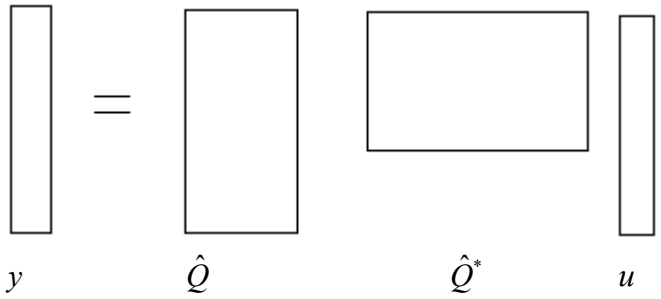
Στη σχέση (2.6) ο πίνακας $\hat{Q}$ δε χρειάζεται να προκύψει από μια SVD . Έστω $\{q_1,.....,q_n\}$ να είναι οποιοδήποτε σύνολο $n$ ορθοκανονικών διανυσμάτων στον $C^m$ και έστω $\hat{Q}$ να είναι ο αντίστοιχος $m \times n$ πίνακας. Από την (1.11) ξέρουμε ότι

$$u = r + \sum_{i=1}^{n} (q_i q_i^*) u$$

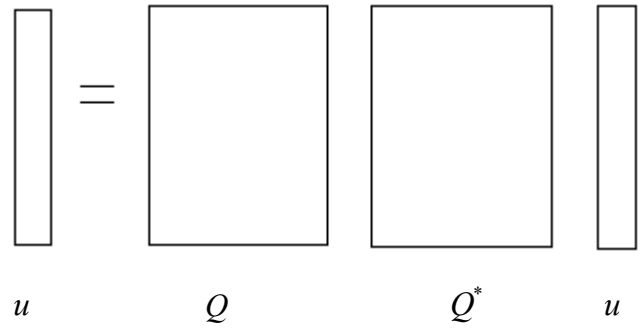
Αντιπροσωπεύει ένα διαχωρισμό ενός διανύσματος $u \in C^m$ σε μια συνιστώσα στο χώρο στηλών του $\hat{Q}$ συν μια συνιστώσα στο ορθογώνιο χώρο. Οπότε ο χάρτης

$$u \mapsto \sum_{i=1}^{n} (q_i q_i^*) u \tag{2.7}$$

είναι ένας ορθογώνιος προβολέας στο $range(\hat{Q})$ και σε μορφή πινάκων μπορεί να γραφτεί $y = \hat{Q}\hat{Q}^* u$ :




Οπότε οποιοδήποτε γινόμενο $\hat{Q}\hat{Q}^*$ είναι πάντα ένας προβολέας στο χώρο των στηλών του $\hat{Q}$, ανεξάρτητα από το πώς προέκυψε ο $\hat{Q}$, αρκεί οι στήλες του να είναι ορθοκανονικές. Ίσως, ο $\hat{Q}$ να προέκυψε παραλείποντας κάποιες στήλες και γραμμές από μια πλήρη παραγοντοποίηση $u = \hat{Q}\hat{Q}^* u$ της ταυτότητας,




αλλά μπορεί και να μην ισχύει η παραπάνω υπόθεση.

Το συμπλήρωμα ενός ορθογώνιου προβολέα είναι επίσης ένας ορθογώνιος προβολέας. Το συμπλήρωμα προβάλλει στον ορθοκανονικό χώρο $range(\hat{Q})$.

**Παράδειγμα: προβολέας τάξης 1**

Μια πολύ σημαντική ιδιαίτερη περίπτωση ορθογώνιου προβολέα είναι ο ορθογώνιος προβολέας τάξης ένα ο οποίος απομονώνει το συμπληρωματικό σε μια μονή διεύθυνση $q$, η οποία μπορεί να γραφτεί ως

$$P_q = qq^* \tag{2.8}$$

Αυτά είναι τα μέρη από τα οποία μπορούν να κατασκευαστούν προβολείς μεγαλύτερης τάξης, όπως στην (2.7). Τα συμπληρώματα είναι τάξης $m-1$ ορθογώνιοι προβολείς οι οποίοι ελαχιστοποιούν τη συνιστώσα στη διεύθυνση του $q$:

$$P_{\perp q} = I - qq^* \tag{2.9}$$

Οι εξισώσεις (2.8) και (2.9) υποθέτουν ότι το $q$ είναι μοναδιαίο διάνυσμα. Για αυθαίρετα μη μηδενικά διανύσματα $a$, οι ανάλογοι τύπο είναι:

$$P_a = \frac{aa^*}{a^*a} \tag{2.10}$$

και

$$P_{\perp q} = I - \frac{aa^*}{a^*a} \tag{2.11}$$

## *2.1.5 ΠΡΟΒΟΛΗ ΣΕ ΜΙΑ ΤΥΧΑΙΑ ΒΑΣΗ*

Ένας ορθογώνιος προβολέας πάνω σε ένα υποδιάστημα του $C^m$ μπορεί επίσης να κατασκευαστεί ξεκινώντας από μια αυθαίρετη βάση, όχι απαραίτητα ορθογώνια. Ας υποθέσουμε ότι ένα υποδιάστημα παράγεται από τα γραμμικά ανεξάρτητα διανύσματα $\{a_1, a_2, ....., a_n\}$ και έστω ο $A$ να είναι ένας $m \times n$ πίνακας του οποίου η j-οστή στήλη είναι το $a_j$.

Κατά το πέρασμα από $u$ στην ορθογώνια προβολή του $y \in range(A)$, η διαφορά $y-u$ πρέπει να είναι ορθογώνια στο $range(A)$. Αυτό είναι ισοδύναμο με το ότι το $y$ πρέπει να ικανοποιεί τη σχέση $a_j^*(y-u)=0$ για κάθε j. Αφού $y \in range(A)$

μπορούμε να θέσουμε $y = Ax$ και να γράψουμε την παραπάνω υπόθεση ως $a_j^*(Ax-u)=0$ για κάθε j ή ισοδύναμα $A^*(Ax-u)=0$ ή $A^*Ax = A^*u$. Μπορούμε εύκολα να συμπεράνουμε ότι αφού ο $A$ είναι πλήρους τάξης, ο πίνακας $A^*A$ είναι ομαλός.

Οπότε

$$x = (A^*A)^{-1}A^*u \tag{2.12}$$

Τέλος, η προβολή του $u, y = Ax$ είναι $y = A(A^*A)^{-1}A^*u$. Οπότε ο ορθογώνιος προβολέας πάνω στο $range(A)$ μπορεί να εκφραστεί από τη σχέση:

$$P = A(A^*A)^{-1}A^* \tag{2.13}$$

Ας σημειώσουμε ότι αυτό είναι μια πολυδιάστατη γενικοποίηση της σχέσης (2.10). Στην ορθοκανονική περίπτωση $A = \hat{Q}$ ο όρος στην παρένθεση καταρρέει στην ταυτότητα και παίρνουμε τη σχέση (2.6).

## 2.2 QR ΠΑΡΑΓΟΝΤΟΠΟΙΗΣΗ

Μια αλγοριθμική ιδέα στην αριθμητική γραμμική άλγεβρα είναι πιο σημαντική από όλες τις άλλες: Η παραγοντοποίηση QR.

### *2.2.1 ΜΕΙΩΜΕΝΗ QR ΠΑΡΑΓΟΝΤΟΠΟΙΗΣΗ*

Σε πολλές εφαρμογές μας ενδιαφέρει ο χώρος των στηλών ενός πίνακα $\mathrm{A}$. Ας σημειώσουμε τον πληθυντικό: αυτοί είναι οι διαδοχικοί χώροι που παράγονται από τις στήλες $\alpha_1, \alpha_2, .....$ του πίνακα $\mathrm{A}$:

$$\langle \alpha_1 \rangle \subseteq \langle \alpha_1, \alpha_2 \rangle \subseteq \langle \alpha_1, \alpha_2, \alpha_3 \rangle \subseteq .....$$

Εδώ όπως και στην προηγούμενη παράγραφο και σε όλη την έκταση της εργασίας, ο συμβολισμός $\langle ..... \rangle$ υποδηλώνει τον υπόχωρο που παράγεται από οποιαδήποτε διανύσματα που βρίσκονται μέσα στις αγκύλες. Οπότε το σύμβολο $\langle \alpha_1 \rangle$ είναι ένας χώρος μιας διάστασης που παράγεται από το διάνυσμα $\alpha_1$, $\langle \alpha_1, \alpha_2 \rangle$ είναι ένας χώρος δυο διαστάσεων που παράγεται από τα διανύσματα $\alpha_1$ και $\alpha_2$ και ούτως κάθε εξής. Η ιδέα της QR παραγοντοποίησης είναι η κατασκευή μιας ακολουθίας

ορθοκανονικών διανυσμάτων $q_1, q_2, \ldots$ τα οποία παράγουν αυτούς τους διαδοχικούς χώρους.

Για να είμαστε ακριβείς, ας υποθέσουμε ότι ο $\hat{R}$ έχει πλήρη τάξη $n$. Θέλουμε η ακολουθία $q_1, q_2, \ldots$ να έχει την ιδιότητα

$$\left\langle q_1, q_2, \ldots, q_j \right\rangle = \left\langle a_1, a_2, \ldots, a_j \right\rangle, \quad j = 1, 2, \ldots, n \tag{2.14}$$

Από τις παρατηρήσεις στην πρώτη παράγραφο, δεν είναι δύσκολο να δούμε ότι η παραπάνω σχέση είναι ισοδύναμη με την υπόθεση

$$\begin{bmatrix} & & & \\ a_1 & a_2 & \cdots & a_n \\ & & & \end{bmatrix} = \begin{bmatrix} & & & \\ q_1 & q_2 & \cdots & q_n \\ & & & \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & & \vdots \\ & & \ddots & \vdots \\ & & & r_{nn} \end{bmatrix} \tag{2.15}$$

όπου τα διαγώνια στοιχεία $r_{kk}$ είναι μη μηδενικά γιατί αν η σχέση (2.15) ισχύει τότε τα $a_1, a_2, \ldots, a_k$ μπορούν να εκφραστούν ως γραμμικοί συνδυασμοί των $q_1, q_2, \ldots, q_k$ και αντίστροφα, από την αναστρεψιμότητα του άνω αριστερού $k \times k$ υποπίνακα του τριγωνικού πίνακα προκύπτει ότι τα $r_{ij}, q_2, \ldots$ μπορούν να εκφραστούν ως γραμμικοί συνδυασμοί των $a_1, a_2, \ldots, a_k$. Από τον παραπάνω πίνακα αν γράψουμε τις ισότητες, είναι της μορφής

$$\begin{aligned} a_1 &= r_{11} q_1 \\ a_2 &= r_{12} q_1 + r_{22} q_2 \\ a_3 &= r_{13} q_1 + r_{23} q_2 + r_{33} q_3 \\ &\vdots \\ a_n &= r_{1n} q_1 + r_{2n} q_2 + \ldots + r_{nn} q_n \end{aligned} \tag{2.16}$$

Σε μορφή πίνακα έχουμε

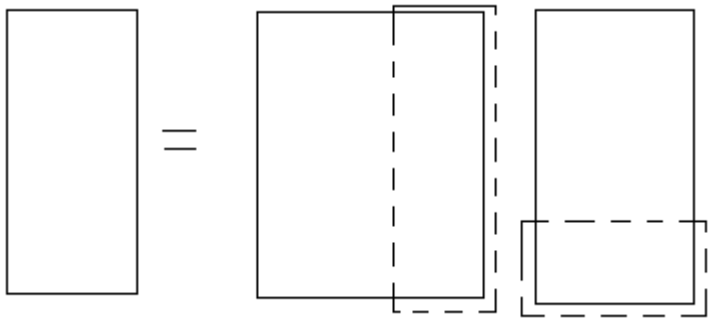
$$A = \hat{Q}\hat{R} \tag{2.17}$$

όπου ο $\hat{Q}$ είναι ένας $m \times n$ πίνακας με ορθοκανονικές στήλες και ο $\hat{R}$ είναι ένας $n \times n$ πίνακας και άνω τριγωνικός. Μια τέτοια παραγοντοποίηση ονομάζεται μειωμένη QR παραγοντοποίηση του πίνακα $A$.

### *2.2.2 ΠΛΗΡΗΣ QR ΠΑΡΑΓΟΝΤΟΠΟΙΗΣΗ*

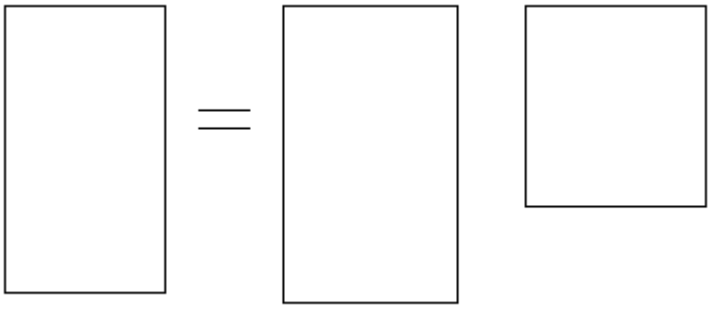
Μια πλήρης QR παραγοντοποίηση ενός πίνακα $A \in C^{m \times n} (m \geq n)$ μπορεί να επεκταθεί προσθέτοντας $m-n$ ορθοκανονικές στήλες στον πίνακα $\hat{Q}$ έτσι ώστε να γίνει ένας $m \times m$ ορθομοναδιαίος πίνακας $Q$. Η διαδικασία αυτή είναι ανάλογη με το πέρασμα από τη μειωμένη παραγοντοποίηση SVD στην πλήρη όπως αυτή περιγράφηκε στην προηγούμενη παράγραφο. Κατά τη διαδικασία, προσθέτουμε στον $\hat{R}$ γραμμές με μηδενικά έτσι ώστε να γίνει ένας $m \times n$ πίνακας $R$ και να παραμείνει άνω τριγωνικός. Η σχέση ανάμεσα στη μειωμένη και την πλήρη QR παραγοντοποίηση είναι αυτή που περιγράφεται παρακάτω.

Πλήρης QR παραγοντοποίηση $(m \geq n)$




Στην πλήρη QR παραγοντοποίηση ο $Q$ είναι ένας $m \times m$ πίνακας, ο $R$ είναι ένας $m \times n$ πίνακας και οι τελευταίες $m-n$ στήλες του $Q$ πολλαπλασιάζονται με μηδενικά στον $R$ (οι στήλες που περικλείονται στις διακεκομμένες γραμμές). Στη μειωμένη παραγοντοποίηση QR οι σιωπηλές στήλες και γραμμές αφαιρούνται. Τώρα ο $\hat{Q}$ είναι $m \times n$, ο $\hat{R}$ είναι $n \times n$ και καμία από τις γραμμές του $\hat{R}$ δεν είναι απαραίτητα μηδενική.

Μειωμένη QR παραγοντοποίηση $(m \geq n)$



$A$ $\hat{Q}$ $\hat{R}$

Ας παρατηρήσουμε ότι στην πλήρη QR παραγοντοποίηση, οι στήλες $q_j$ για $j > n$ είναι ορθογώνιες στο $range(A)$. Αν υποθέσουμε ότι ο πίνακας $A$ έχει πλήρη τάξη $n$ τότε οι στήλες αυτές αποτελούν μια ορθοκανονική βάση για το $range(A)^{\perp}$ ή ισοδύναμα για το $null(A^*)$.

### *2.2.3 GRAM-SCHMIDT ΟΡΘΟΚΑΝΟΝΙΚΟΠΟΙΗΣΗ*

Οι εξισώσεις (2.16) υποδηλώνουν μια μέθοδο για να υπολογίσουμε τη μειωμένη QR παραγοντοποίηση. Δοθέντων $a_1, a_2, .....$ μπορούμε να κατασκευάσουμε τα διανύσματα $q_1, q_2, .....$ και τα στοιχεία $r_{ij}$ με μια διαδικασία ορθοκανονικοποίησης. Η διαδικασία αυτή είναι γνωστή ως *Gram-Schmidt ορθοκανονικοποίηση*.
Η λειτουργία της διαδικασίας είναι η ακόλουθη. Στο j-οστό βήμα θέλουμε να βρούμε ένα μοναδιαίο διάνυσμα $q_j \in \langle a_1, a_2, ...., a_j \rangle$ το οποίο να είναι ορθογώνιο στα $q_1, q_2, ..., q_{j-1}$. Έχουμε ήδη δει την απαραίτητη τεχνική ορθοκανονικοποίησης στη σχέση (1.10). Από αυτή την εξίσωση βλέπουμε ότι

$$v_j = a_j - (q_1^* a_j) q_1 - (q_2^* a_j) q_2 - ..... - (q_{j-1}^* a_j) q_{j-1} \qquad (2.18)$$

είναι ένα διάνυσμα που ζητάμε αλλά δεν έχει κανονικοποιηθεί. Αν διαιρέσουμε με $\|v_j\|_2$ το αποτέλεσμα είναι το ζητούμενο διάνυσμα $q_j$.
Με αυτή τη λογική ξαναγράφουμε την εξίσωση (2.16) στην παρακάτω μορφή

$$q_1 = \frac{a_1}{r_{11}}$$

$$q_2 = \frac{a_2 - r_{12} q_1}{r_{22}}$$

$$q_3 = \frac{a_3 - r_{13} q_1 - r_{23} q_2}{r_{33}}$$

. (2.19)

.

.

$$q_n = \frac{a_n - \sum_{i=1}^{n-1} r_{in} q_i}{r_{nn}}$$

Από τη σχέση (2.18) είναι φανερό ότι ένας κατάλληλος ορισμός για τους συντελεστές $r_{ij}$ στους αριθμητές της (2.19) είναι:

$$r_{ij} = q_i^* a_j \qquad (i \neq j) \tag{2.20}$$

Οι συντελεστές $r_{jj}$ στους παρονομαστές έχουν επιλεγεί για ορθοκανονικοποίηση:

$$\left| r_{jj} \right| = \left\| a_j - \sum_{i=1}^{j-1} r_{ij} q_i \right\|_2 \tag{2.21}$$

Ας σημειώσουμε ότι το σύμβολο $r_{jj}$ δεν έχει οριστεί. Τυχαία, μπορούμε να διαλέξουμε $r_{jj} > 0$ και σε αυτή την περίπτωση μπορούμε να ολοκληρώσουμε με μια παραγοντοποίηση $A = \hat{Q}\hat{R}$ στην οποία ο $\hat{R}$ έχει θετικά στοιχεία σε όλη τη διαγώνιο. Ο αλγόριθμος που περιγράφηκε στις σχέσεις (2.19)-(2.21) είναι η επανάληψη Gram-Schmidt. Μαθηματικά με μια απλή διαδρομή στον αλγόριθμο μπορούμε να καταλάβουμε και να αποδείξουμε διάφορες ιδιότητες της παραγοντοποίησης QR. Αριθμητικά, η μέθοδος αυτή αποδεικνύεται ασταθής λόγω των σφαλμάτων του υπολογιστή. Για να δώσουν έμφαση στην αστάθεια, οι ερευνητές στην περιοχή της αριθμητικής ανάλυσης αναφέρονται σε αυτή ως την *κλασσική μέθοδο Gram-Schmidt* για να την ξεχωρίσουν από την *τροποποιημένη μέθοδο Gram-Schmidt*, την οποία θα δούμε στην επόμενη παράγραφο.

**ΑΣΤΑΘΗΣ ΜΕΘΟΔΟΣ Gram-Schmidt**

ΑΛΓΟΡΙΘΜΟΣ 2.1 Κλασσική μέθοδος (ασταθής)

Για $j = 1$ μέχρι $n$

$v_j = a_j$

για $i = 1$ μέχρι $j - 1$

$r_{ij} = q_i^* a_j$

$v_j = v_j - r_{ij} q_i$

$r_{jj} = \left\| v_j \right\|_2$

$q_j = v_j / r_{jj}$

***ΠΑΡΑΔΕΙΓΜΑ 2.2.1***

Έστω $S=\left\{\mathbf{v}_1=\begin{pmatrix}3\\1\end{pmatrix},\mathbf{v}_2=\begin{pmatrix}2\\2\end{pmatrix}\right\}$

Εκτελώντας την ασταθή μέθοδο Gram-Schmidt

$$\mathbf{u}_1=\mathbf{v}_1=\begin{pmatrix}3\\1\end{pmatrix},\ \mathbf{u}_2=\mathbf{v}_2-\text{proj}_{\mathbf{u}_1}$$

$$\mathbf{u}_2=\mathbf{v}_2-\text{proj}_{\mathbf{u}_1}(\mathbf{v}_2)=\begin{pmatrix}2\\2\end{pmatrix}-\text{proj}_{\begin{pmatrix}3\\1\end{pmatrix}}\begin{pmatrix}2\\2\end{pmatrix}=\begin{pmatrix}-2/5\\6/5\end{pmatrix}$$

τα οποία είναι πράγματι ορθογώνια αφού:

$$\langle\mathbf{u}_1,\mathbf{u}_2\rangle=\left\langle\begin{pmatrix}3\\1\end{pmatrix},\begin{pmatrix}-2/5\\6/5\end{pmatrix}\right\rangle=-\frac{6}{5}+\frac{6}{5}=0$$

Τελικά προκύπτουν τα

$$\mathbf{e}_1=\frac{1}{\sqrt{10}}\begin{pmatrix}3\\1\end{pmatrix}\ \text{και}\ \mathbf{e}_2=\frac{1}{\sqrt{\frac{40}{25}}}\begin{pmatrix}-2/5\\6/5\end{pmatrix}=\frac{1}{\sqrt{10}}\begin{pmatrix}-1\\3\end{pmatrix}$$

## *2.2.4 ΥΠΑΡΞΗ ΚΑΙ ΜΟΝΑΔΙΚΟΤΗΤΑ*

Σε όλους τους πίνακες μπορούμε να εφαρμόσουμε την QR παραγοντοποίηση και σε ορισμένες περιπτώσεις αυτή η παραγοντοποίηση είναι μοναδική. Πρώτα θα δούμε το θεώρημα που αφορά την ύπαρξη.

ΘΕΩΡΗΜΑ 2.2

Σε κάθε πίνακα $A\in C^{m\times n}\quad (m\geq n)$ μπορούμε να εφαρμόσουμε την QR παραγοντοποίηση καθώς και τη μειωμένη QR παραγοντοποίηση.

ΑΠΟΔΕΙΞΗ

Αρχικά υποθέτουμε ότι ο πίνακας $A$ έχει πλήρη τάξη και ότι θέλουμε ως αποτέλεσμα την μειωμένη QR παραγοντοποίηση. Σε αυτή την περίπτωση, η απόδειξη της ύπαρξης προκύπτει από τον αλγόριθμο Gram-Schmidt. Από κατασκευή, από αυτή τη διαδικάσία προκύπτουν ορθοκανονικές στήλες του $\hat{Q}$ και στοιχεία του $\hat{R}$ έτσι ώστε να ισχύει η σχέση (2.17). Αποτυχία μπορεί να προκύψει μόνο αν σε κάποια βήματα το $v_j$ είναι μηδενικό και κατά συνέπεια να μη μπορεί να κανονικοποιηθεί για να προκύψει το $q_j$. Όμως από αυτό θα παίρναμε ότι $a_j\in\langle q_1,q_2,....,q_{j-1}\rangle=\langle a_1,a_2,....,a_{j-1}\rangle$, κάτι που είναι αντίθετο με την αρχική μας υπόθεση ότι ο πίνακας $r_{jj}$ έχει πλήρη τάξη.

Τώρα υποθέτουμε ότι ο πίνακας $A$ δεν έχει πλήρη τάξη. Τότε σε ένα ή περισσότερα βήματα j, θα βρίσκαμε ότι η (2.18) δίνει $v_j = 0$. Σε αυτό το σημείο διαλέγουμε αυθαίρετα ένα $q_j$ να είναι κάθε κανονικοποιημένο διάνυσμα ορθογώνιο στο $\langle q_1, q_2, ...., q_{j-1} \rangle$ και μετά συνεχίζουμε την μέθοδο Gram-Schmidt.

Τέλος, η πλήρης παρά η μειωμένη QR παραγοντοποίηση ενός $m \times n$ πίνακα με $m > n$ μπορεί να κατασκευαστεί εισάγοντας τυχαία ορθοκανονικά διανύσματα με τον ίδιο τρόπο. Ακολουθούμε τη διαδικασία Gram-Schmidt μέχρι το βήμα $n$ και μετά συνεχίζουμε με τα επιπρόσθετα $m-n$ βήματα εισάγοντας διανύσματα $q_j$ σε κάθε βήμα.

Τώρα θα ασχοληθούμε με τη μοναδικότητα. Ας υποθέσουμε ότι $A = \hat{Q}\hat{R}$ είναι μια μειωμένη QR παραγοντοποίηση. Αν η i-οστή στήλη του $\hat{Q}$ πολλαπλασιαστεί με $z$ και η i-οστή γραμμή του $\hat{R}$ πολλαπλασιαστεί με $z^{-1}$ για κάποιο βαθμωτό $z$ με $|z| = 1$ παιρνουμε άλλη μια παραγοντοποίηση QR του πίνακα $A$. Το επόμενο θεώρημα μας εξασφαλίζει ότι αν ο πίνακας $A$ έχει πλήρη τάξη τότε αυτός είναι ο μόνος τρόπος για να προκύψει διακεκριμένες QR παραγοντοποιήσεις.

ΘΕΩΡΗΜΑ 2.3

Σε κάθε πίνακα $A \in C^{m \times n} (m \geq n)$ ο οποίος έχει πλήρη τάξη μπορούμε να εφαρμόσουμε μοναδική μειωμένη QR παραγοντοποίηση $A = \hat{Q}\hat{R}$ με $r_{jj} > 0$.

ΑΠΟΔΕΙΞΗ

Και σε αυτό το θεώρημα η απόδειξη προκύπτει από την επανάληψη Gram-Schmidt. Από την (2.17), την ορθοκανονικότητα των στηλών του $\hat{Q}$ και την ανω τριγωνικότητα του $\hat{R}$ προκύπτει ότι κάθε μειωμένη QR παραγοντοποίηση του $A$ πρέπει να ικανοποιεί τις (2.19)-(2.21). Από την υπόθεση της πλήρους τάξης, οι παρονομαστές (2.21) από την (2.19) είναι μη μηδενικοί οπότε σε κάθε διαδοχικό βήμα j από αυτούς τους τύπους προκύπτουν τα $r_{ij}$ και $q_j$ εκτός από την περίπτωση όπου το σύμβολο $r_{jj}$ δεν διευκρινίζεται στην (2.21). Αυτό διορθώνεται από την υπόθεση ότι $r_{jj} > 0$, όπως και στον αλγόριθμο 2.1, και η παραγοντοποίηση ολοκληρώνεται.

### *2.2.5 ΟΤΑΝ ΤΑ ΔΙΑΝΥΣΜΑΤΑ ΓΙΝΟΝΤΑΙ ΣΥΝΕΧΕΙΣ ΣΥΝΑΡΤΗΣΕΙΣ*

Η QR παραγοντοποίηση είναι ανάλογη με τις ορθοκανονικές επεκτάσεις συναρτήσεων παρά με των διανυσμάτων.

Ας υποθέσουμε ότι αντικαθιστούμε τον $C^m$ με τον $L^2[-1,1]$, ένα διανυσματικό χώρο με μιγαδικές συναρτήσεις στο [-1,1]. Δεν θα εισάγουμε τις ιδιότητες αυτού

του χώρου επίσημα, θα πούμε ότι του εσωτερικό γινόμενο των $f$ και $g$ παίρνει τη μορφή

$$(f,g)=\int_{-1}^{1}\overline{f(x)}g(x)dx \tag{2.22}$$

Ας δούμε για παράδειγμα τον παρακάτω «πίνακα» του οποίου οι «στήλες» είναι τα μονώνυμα $x^j$:

$$A=\begin{bmatrix} & & & & \\ 1 & x & x^2 & \cdots & x^{n-1} \\ & & & & \end{bmatrix} \tag{2.23}$$

Κάθε στήλη είναι μια συνάρτηση στον $L^2[-1,1]$ και κατά συνέπεια ο $A$ είναι διακριτός στην οριζόντια διεύθυνση και συνεχής στην κατακόρυφη διεύθυνση.
Η «συνεχής QR παραγοντοποίηση» του πίνακα $A$ παίρνει τη μορφή:

$$A=QR=\begin{bmatrix} & & & & \\ q_0(x) & q_1(x) & q_3(x) & \cdots & q_{n-1}(x) \\ & & & & \end{bmatrix}\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & & \vdots \\ & & \ddots & r_{nn} \end{bmatrix},$$

όπου οι στήλες του $Q$ είναι συναρτήσεις του $x$, ορθοκανονικές σε σχέση με το εσωτερικό γινόμενο (2.22):

$$\int_{-1}^{1}\overline{q_i(x)}q_j(x)dx=\delta_{ij}=\begin{cases}1, i=j\\ 0, i\neq j\end{cases}.$$

Από την κατασκευή Gram-Schmidt μπορούμε να δούμε ότι το $q_j$ είναι ένα πολυώνυμο βαθμού j. Αυτά τα πολυώνυμα βαθμωτά πολλαπλάσια των πολυωνύμων Legendre, $P_j$, τα οποία κατά συνθήκη κανονικοποιούνται έτσι ώστε $P_j(1)=1$. Τα πρώτα $P_j$ είναι

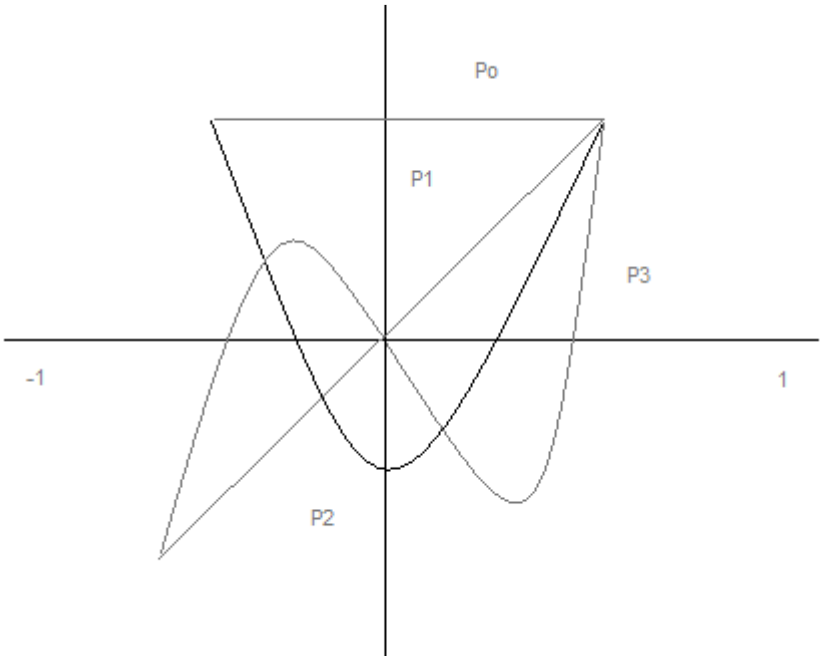
$$P_0(x)=1,\qquad P_1(x)=x,\qquad P_2(x)=\frac{3}{2}x^2-\frac{1}{2},\qquad P_3(x)=\frac{5}{2}x^3-\frac{3}{2}x \tag{2.24}$$

Δείτε το σχήμα 2.3. Όπως τα πολυώνυμα $1, x, x^2, \ldots\ldots$ αυτή η ακολουθία πολυωνύμων παράγει χώρους πολυωνύμων με διαδοχικά μεγαλύτερο βαθμό. Όμως τα $P_0(x), P_1(x), P_2(x), \ldots\ldots$ έχουν το πλεονέκτημα ότι είναι ορθογώνια και κατά συνέπεια πολύ πιο κατάλληλα για κάποιους υπολογισμούς. Στην πραγματικότητα, υπολογισμοί με τέτοια πολυώνυμα είναι η βάση των φασματικών μεθόδων, οι οποίες είναι από τις πιο σημαντικές τεχνικές για την αριθμητική λύση μερικών διαφορικών εξισώσεων.

Ποιος είναι ο «πίνακας προβολής» $\hat{Q}\hat{Q}^*$ (2.6) που συνδέεται με τον $\hat{Q}$; Είναι ένας «$[-1,1]\times[-1,1]$ πίνακας», δηλαδή ένας τελεστής ολοκληρώματος

$$f(\cdot) \mapsto \sum_{j=0}^{n-1} q_j(\cdot) \int_{-1}^{1} \overline{q_j(x)} f(x) dx \qquad (2.25)$$

απεικονίζει συναρτήσεις στον $L^2[-1,1]$ σε συναρτήσεις στον $L^2[-1,1]$.




Σχήμα 2.3 Τα πρώτα 4 πολυώνυμα Legendre (2.24). Εκτός από τους συντελεστές κλίμακας, αυτά μπορούν να ερμηνευτούν σαν τις στήλες του $\hat{Q}$ σε μια μειωμένη QR παραγοντοποίηση στον «$[-1,1]\times 4 matrix$ » $[1, x, x^2, x^3]$.

### *2.2.6 ΛΥΣΗ ΤΟΥ Ax=b ΜΕ QR ΠΑΡΑΓΟΝΤΟΠΟΙΗΣΗ*

Στο κλείσιμο αυτής της παραγράφου θα αναφερθούμε στους διακριτούς πεπερασμένους πίνακες. Ας υποθέσουμε ότι θέλουμε να λύσουμε την $Ax = b$ όπου ο $A \in C^{m \times m}$ είναι ομαλός. Αν $A = QR$ είναι η QR παραγοντοποίηση τότε μπορούμε να γράψουμε ότι $QRx = b$ ή

$$Rx = Q^* b \tag{2.26}$$

Το δεξί μέρος αυτής της εξίσωσης υπολογίζεται εύκολα αν γνωρίζουμε το QR και το πεπλεγμένο σύστημα των γραμμικών εξισώσεων στο αριστερό μέρος είναι εύκολα υπολογίσιμο επειδή είναι τριγωνικό. Τα παραπάνω μας οδηγούν σε μια μέθοδο για τον υπολογισμό του $Ax = b$:

1. Υπολόγισε μια QR παραγοντοποίηση $A = QR$.
2. Υπολόγισε $y = Q^* b$.
3. Λύσε $Rx = y$ ως προς $x$.

Ο συνδυασμός των βημάτων 1-3 είναι μια πολύ καλή μέθοδος για την εύρεση της λύσης γραμμικών συστημάτων . Παρόλα αυτά δεν είναι η βασική μέθοδος για τη λύση τέτοιων προβλημάτων. Ο αλγόριθμος που χρησιμοποιείται συνήθως είναι η απαλοιφή του Gauss επειδή η μέθοδος αυτή απαιτεί τις μισές πράξεις για την εύρεση της λύσης.

## 2.3 GRAM-SCHMIDT ΟΡΘΟΚΑΝΟΝΙΚΟΠΟΙΗΣΗ

Η μέθοδος Gram-Schmidt είναι βάση για έναν από τους δυο πιο κύριους αριθμητικούς αλγορίθμους για τον υπολογισμό QR παραγοντοποιήσεων. Είναι μια διαδικασία «τριγωνικής ορθοκανονικοποίησης» κάνοντας τις στήλες ενός πίνακα ορθοκανονικές μέσω μιας ακολουθίας πράξεων πίνακα οι οποίες μπορούν να ερμηνευτούν σαν έναν πολλαπλασιασμό από τα δεξιά άνω τριγωνικών πινάκων.

### *2.3.1 GRAM-SCHMIDT ΠΡΟΒΟΛΕΣ*

Στα προηγούμενα παρουσιάσαμε τη μέθοδο Gram-Schmidt και την κλασσική αλγοριθμική της μορφή. Εδώ θα περιγράψουμε τον ίδιο αλγόριθμο με ένα διαφορετικό τρόπο χρησιμοποιώντας ορθογώνιους προβολείς.

Έστω $A \in C^{m \times n}$, $m \geq n$ να είναι ένας πίνακας πλήρους τάξης με στήλες $\{a_j\}$. Πριν παρουσιάσαμε τη μέθοδο Gram-Schmidt με τους τύπους (2.19)-(2.21).

Έστω η ακολουθία των τύπων

$$q_1 = \frac{P_1 a_1}{\|P_1 a_1\|},\ q_2 = \frac{P_2 a_2}{\|P_2 a_2\|}, \ldots\ldots,\ q_n = \frac{P_n a_n}{\|P_n a_n\|} \tag{2.27}$$

Σε αυτούς τους τύπους κάθε $P_j$ είναι ένας ορθογώνιος προβολέας. Ειδικά ο $P_j$ είναι ένας $m \times n$ πίνακας τάξης $m-(j-1)$ ο οποίος προβάλλει το χώρο $C^m$ ορθογώνια πάνω σε ένα ορθογώνιο χώρο του $\langle q_1, \ldots\ldots, q_{j-1} \rangle$. (Στην περίπτωση που $j=1$ η περιγραφή αυτή είναι ισοδύναμη με $P_1 = I$.) Ας παρατηρήσουμε ότι το $q_j$ έτσι όπως ορίζεται στην (2.27) είναι ορθογώνιο στα $q_1, \ldots\ldots, q_{j-1}$, βρίσκεται στο χώρο $\langle a_1, \ldots\ldots, a_j \rangle$ και η νόρμα του είναι ίση με 1. Δηλαδή βλέπουμε ότι η (2.27) είναι ισοδύναμη με τις (2.19)-(2.21) και στον αλγόριθμο 2.1.

Ο προβολέας $P_j$ μπορεί να παρουσιαστεί σε λελυμένη μορφή. Έστω $\hat{Q}_{j-1}$ να είναι ένας $m \times (j-1)$ πίνακας ο οποίος περιέχει τις πρώτες $j-1$ στήλες του $\hat{Q}$,

$$\hat{Q}_{j-1} = \begin{bmatrix} & & & \\ q_1 & q_2 & \cdots & q_{j-1} \\ & & & \end{bmatrix} \tag{2.28}$$

Τότε ο $P_j$ δίνεται από τη σχέση

$$P_j = I - \hat{Q}_{j-1} \hat{Q}_{j-1}^* \tag{2.29}$$

Μπορούμε αμέσως να αναγνωρίσουμε ότι η σχέση (2.29) είναι ο συντελεστής του $a_j$ (2.18).

### *2.3.2 ΤΡΟΠΟΠΟΙΗΜΕΝΟΣ ΑΛΓΟΡΙΘΜΟΣ GRAM-SCHMIDT*

Στην πραγματικότητα οι τύποι για τη μέθοδο Gram-Schmidt δεν εφαρμόζονται όπως τους γράψαμε στον αλγόριθμο 2.1 και στη (2.27) και γι αυτό η μέθοδος καταλήγει να είναι αριθμητικά ασταθής. Ευτυχώς υπάρχει μια απλή τροποποίηση η οποία βελτιώνει την αστάθεια. Ένας σταθερός αλγόριθμος είναι ένας αλγόριθμος ο οποίος δεν επηρεάζεται εύκολα από τα σφάλματα στρογγυλοποίησης των πράξεων μηχανής.

Για κάθε τιμή του j ο αλγόριθμος 2.1 υπολογίζει μια μονή ορθογώνια προβολή τάξης $m-(j-1)$,

$$v_j = P_j a_j \tag{2.30}$$

Σε αντίθεση ο τροποποιημένος αλγόριθμος Gram-Schmidt υπολογίζει το ίδιο αποτέλεσμα με μια ακολουθία $j-1$ προβολών τάξης $m-1$. Από την (2.9) έχουμε ότι $P_{\perp q}$ συμβολίζει την τάξη $m-1$ ορθογώνιων προβολών στο χώρο που είναι ορθογώνιος σε ένα μη μηδενικό διάνυσμα $q \in C^m$. Από τον ορισμό του $P_j$ δεν είναι δύσκολο να δούμε ότι

$$P_j = P_{\perp q_{j-1}} \cdots P_{\perp q_2} P_{\perp q_1} \tag{2.31}$$

με $P_1 = I$. Οπότε μια ισοδύναμη σχέση με την (2.30) είναι

$$v_j = P_{\perp q_{j-1}} \cdots P_{\perp q_2} P_{\perp q_1} a_j \tag{2.32}$$

Ο τροποποιημένος αλγόριθμος Gram-Schmidt βασίζεται στη χρήση της (2.32) παρά της (2.30).
Από μαθηματικής άποψης, οι σχέσεις (2.32) και (2.30) είναι ισοδύναμες. Όμως, οι ακολουθίες των αριθμητικών πράξεων που προκύπτουν από αυτούς τους τύπους είναι διαφορετικές. Ο τροποποιημένος αλγόριθμος υπολογίζει τα $v_j$ εκτιμώντας τους παρακάτω τύπους στη σειρά:

$$\begin{aligned}
v_j^{(1)} &= a_j \\
v_j^{(2)} &= P_{\perp q_1} v_j^{(1)} = v_j^{(1)} - q_1 q_1^* v_j^{(1)} \\
v_j^{(3)} &= P_{\perp q_1} v_j^{(2)} = v_j^{(2)} - q_2 q_2^* v_j^{(2)} \\
&\;\;\vdots \qquad\qquad \vdots \\
v_j = v_j^{(j)} &= P_{\perp q_{j-1}} v_j^{(j-1)} = v_j^{(j-1)} - q_{j-1} q_{j-1}^* v_j^{(j-1)}
\end{aligned} \tag{2.33}$$

Σε αριθμητική υπολογιστή πεπερασμένης ακρίβειας παρατηρούμε ότι από τη (2.33) έχουμε μικρότερο σφάλμα απ' ότι από τη (2.30).
Όταν εκτελούμε τον αλγόριθμο τότε ο προβολέας $P_{\perp q_i}$ μπορεί να εφαρμοστεί στο $v_j$ για κάθε $j > i$ αμέσως μόλις υπολογιστεί το $q_i$.

ΕΥΣΤΑΘΗΣ ΜΕΘΟΔΟΣ GRAM-SCHMIDT

ΑΛΓΟΡΙΘΜΟΣ 2.2 Τροποποιημένος αλγόριθμος Gram-Schmidt

Για $i = 1$ μέχρι $n$

$v_i = a_i$

Για $i = 1$ μέχρι $n$

$r_{ii} = \|v_i\|$

$q_i = v_i / r_{ii}$

Για $j = i + 1$ μέχρι $n$

$r_{ij} = q_i^* v_j$

$v_j = v_j - r_{ij} q_i$

Στην πράξη, είναι σύνηθες να αφήνουμε το $v_i$ να υπερκαλύπτει το $a_i$ και το $q_i$ ναι υπερικαλύπτει το $v_i$ για να κερδίζουμε χώρο μνήμης, δηλαδ.η το $v_j$ καταλαμβάνει το χώρο μνήμης των $a_j$ ενώ το $q_i$ καταλαμβάνει το χώρο μνήμης των $v_i$.

## *2.3.3 ΠΟΛΥΠΛΟΚΟΤΗΤΑ*

Ο αλγόριθμος Gram-Schmidt είναι ο πρώτος αλγόριθμος που παρουσιάσαμε σε αυτή την εργασία και είναι σημαντικό να εξετάσουμε το κόστος του. Για να το κάνουμε αυτό ακολουθούμε τα βήματα και μετράμε το πλήθος των πράξεων που γίνονται κατά τη διάρκεια εκτέλεσης του αλγορίθμου. Κάθε άθροισμα, πολλαπλασιαμός, αφαίρεση, διαίρεση ή τετραγωνική ρίζα υπολογίζεται σαν ένα βήμα.

Δεν κάνουμε καμία διάκριση ανάμεσα σε πράξεις με πραγματικούς αριθμούς και πράξεις με μιγαδικούς αριθμούς αν και στην πράξη υπάρχει κάποια διαφορά.

ΘΕΩΡΗΜΑ 2.4

Οι αλγόριθμοι 2.1 και 2.2 απαιτούν $\sim 2mn^2$ βήματα για να υπολογίσουν μια QR παραγοντοποίηση ενός $m \times n$ πίνακα.

Ας σημειώσουμε ότι το θεώρημα εκφράζει μόνο τον ηγετικό όρο της αρίθμησης του βήματος.

Το σύμβολο « $\sim$ » έχει τη συνήθη ασυμπτωτική σημασία:

$$\lim_{m,n\to\infty} \frac{number.of.flops}{2mn^2} = 1$$

Το θεώρημα 2.4 μπορεί να διαπιστωθεί ως εξής:

Για παράδειγμα τον τροποποιημένο αλγόριθμο Gram-Schmidt, δηλαδή τον αλγόριθμο 2.2 όταν τα $m$ και $n$ είναι μεγάλα οι περισσότερες πράξεις γίνονται στην εσωτερική επανάληψη:

$$r_{ij} = q_i^* v_j$$
$$v_j = v_j - r_{ij} q_i$$

Στην πρώτη γραμμή υπολογίζουμε το εσωτερικό γινόμενο $q_i^* v_j$ απαιτώντας $m$ πολλαπλασιασμούς και $m-1$ προσθέσεις και στη δεύτερη γραμμή υπολογίζουμε το $v_j - r_{ij} q_i$ απαιτώντας $m$ πολλαπλασιασμούς και $m$ διαιρέσεις. Οπότε η συνολική πολυπλοκότητα μιας μονής εσωτερικής επανάληψης είναι $\sim 4m$ βήματα ή 4 βήματα ανά στήλη διανύσματος στοιχείου. Όλα μαζί, ο αριθμός των βημάτων που απαιτούνται από τον αλγόριθμο είναι ασυμπτωτικά

$$\sum_{i=1}^{n} \sum_{j=i+1}^{n} 4m \sim \sum_{i=1}^{n} (i) 4m \sim 2mn^2 \qquad (2.34)$$

## *2.3.5 Η ΜΕΘΟΔΟΣ GRAM-SCHMIDT ΩΣ ΤΡΙΓΩΝΙΚΗ ΟΡΘΟΚΑΝΟΝΙΚΟΠΟΙΗΣΗ*

Κάθε εξωτερικό βήμα του τροποποιημένου αλγορίθμου Gram-Schmidt μπορεί να ερμηνευτεί ως ένας πολλαπλασιαμός από δεξιά με ένα τετράγωνικό άνω τριγωνικό πίνακα. Για παράδειγμα, ξεκινώντας με τον πίνακα $A$, στην πρώτη επανάληψη πολλαπλασιάζουμε την πρώτη στήλη $a_1$ με $1/r_{11}$ και μετά αφαιρούμε $r_{1j}$ φορές το αποτέλεσμα από κάθε μια από τις υπόλοιπες στήλες $a_j$. Αυτό είναι ισοδύναμο με τον πολλαπλασιαμό από δεξιά ενός πίνακα $R_1$:

$$\begin{bmatrix} & & & \\ u_1 & u_2 & \cdots & u_n \\ & & & \end{bmatrix} \begin{bmatrix} \frac{1}{r_{11}} & \frac{-r_{12}}{r_{11}} & \frac{-r_{13}}{r_{11}} & \cdots \\ & 1 & & \\ & & 1 & \\ & & & \ddots \end{bmatrix} = \begin{bmatrix} & & & \\ q_1 & u_2^{(2)} & \cdots & u_n^{(2)} \\ & & & \end{bmatrix}$$

Γενικά, στο i-οστό βήμα του αλγορίθμου 2.2 αφαιρούμε $r_{ij}/r_{ii}$ φορές τη στήλη $i$ από του τρέχοντα πίνακα $A$ από τις στήλες $j > i$ και αντικαθιστούμε τη στήλη $i$ με

$1/r_{ii}$ φορές τον εαυτό της. Αυτό αντιστοιχεί στον πολλαπλασιασμό με έναν άνω τριγωνικό πίνακα $R_i$:

$$R_2 = \begin{bmatrix} 1 & & & \\ & \frac{1}{r_{22}} & \frac{-r_{23}}{r_{22}} & \cdots \\ & & 1 & \\ & & & \ddots \end{bmatrix}, \qquad R_3 = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \frac{1}{r_{33}} & \cdots \\ & & & \ddots \end{bmatrix}, \ldots..$$

Και στο τέλος της επανάληψης έχουμε:

$$A \underbrace{R_1 R_2 \ldots\ldots R_n}_{\hat{R}^{-1}} = \hat{Q} \tag{2.36}$$

Η παραπάνω μορφοποίηση δηλώνει ότι ο αλγόριθμος Gram-Schmidt είναι μια μέθοδος τριγωνικής ορθοκανονικοποίησης. Εφραμόζει τριγωνικές πράξεις στο δεξί μέρος του πίνακα για να το μειώσει σε ένα πίνακα σε ορθοκανονικές στήλες. Βέβαια, στην πράξη, δεν φτιάχνουμε τους πίνακες $R_i$ και δεν τους πολλαπλασιάζουμε αναλυτικά. Ο σκοπός όσων αναφέραμε είναι να εξετάσουμε σε μεγαλύτερο βάθος τη δομή του αλγορίθμου Gram-Schmidt.

## 2.4 HOUSEHOLDER ΤΡΙΓΩΝΟΠΟΙΗΣΗ

Μια άλλη μέθοδος για να υπολογίσουμε QR παραγοντοποίηση είναι η Householder τριγωνοποίηση, η οποία αριθμητικά είναι πιο ευσταθής από την ορθοκανονικοποίηση Gram-Schmidt αλλά το μειονέκτημά της είναι ότι δεν είναι τόσο εύκολα εφαρμόσιμη όσο ο τροποποιημένος αλγόριθμος Gram-Schmidt. Ο αγλόριθμος της Householder τριγωνοποίησης είναι μια διαδικασία «ορθογώνιας τριγωνοποίησης», κάνοντας ένα πίνακα τριγωνικό μέσω μιας ακολουθίας πράξεων με ορθομοναδιαίους πίνακες.

### *2.4.1 HOUSEHOLDER KAI GRAM-SCHMIDT*

Όπως είδαμε στη μέθοδο Gram-Schmidt εφαρμόζουμε διαδοχικούς στοιχειώδεις τριγωνικούς πίνακες $R_k$ στον πίνακα $A$ από τα δεξιά., έτσι ώστε ο πίνακας που θα προκύψει, δηλαδή ο

$$A \underbrace{R_1 R_2 \cdots R_n}_{\hat{R}^{-1}} = \hat{Q}$$

να έχει ορθοκανονικές στήλες. Το γινόμενο $\hat{R} = R_n^{-1} \cdots R_2^{-1} R_1^{-1}$ είναι επίσης άνω τριγωνικό και οπότε ο $A = \hat{Q}\hat{R}$ είναι μια QR παραγοντοποίηση του πίνακα $A$.
Σε αντίθεση, στη μέθοδο Householder εφαρμόζουμε διαδοχικούς στοιχειώδεις πίνακες $Q_k$ στον πίνακα $A$ από τα αριστερά έτσι ώστε ο πίνακας που θα προκύψει, δηλαδή ο

$$\underbrace{Q_n \cdots Q_2 Q_1}_{Q^*} A = R$$

να είναι άνω τριγωνικός. Το γινόμενο $\hat{Q}$ είναι επίσης ορθομοναδιαίο και οπότε ο $A = QR$ είναι μια πλήρης QR παραγοντοποίηση του πίνακα $A$.
Οι δύο μέθοδοι μπορούν να παρουσιαστούν ως εξής:

Αλγόριθμος Gram-Schmidt: τριγωνική ορθοκανονικοποίηση
Αλγόριθμος Householder: ορθογώνια τριγωνοποίηση.

## *2.4.2 ΤΡΙΓΩΝΟΠΟΙΗΣΗ ΕΙΣΑΓΟΝΤΑΣ ΜΗΔΕΝΙΚΑ*

Η βασική ιδέα της μεθόδου Householder προτάθηκε από τον Alston Householder το 1958. Η ιδέα αυτή είναι ένας έξυπνος τρόπος σχεδιασμού των ορθομοναδιαίων πινάκων $Q_k$ έτσι ώστε το γινόμενο $Q_n \cdots Q_2 Q_1 A$ να είναι άνω τριγωνικός πίνακας.
Ο πίνακας $Q_k$ επιλέγεται έτσι ώστε να έχει μηδενικά κάτω από τη διαγώνιο στην k-οστή στήλη και παράλληλα να διατηρεί όλα τα μηδενικά που εισάγαμε προηγουμένως. Για παράδειγμα, στην περίπτωση ενός $5 \times 3$ πίνακα γίνονται τρείς πράξεις όπως περιγράφουμε παρακάτω. Στους πίνακες παρακάτω το σύμβολο $\times$ αντιπροσωπεύει κάθε είσοδο η οποία δεν είναι απαραίτητα μηδενική και το σύμβολο $*$ αντιπροσωπεύει τα στοιχεία που μόλις άλλαξαν. Τα κενά είναι όλα μηδενικά.

$$\underset{A}{\begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \end{bmatrix}} \xrightarrow{Q_1} \underset{Q_1 A}{\begin{bmatrix} * & * & * \\ 0 & * & * \\ 0 & * & * \\ 0 & * & * \\ 0 & * & * \end{bmatrix}} \xrightarrow{Q_2} \underset{Q_2 Q_1 A}{\begin{bmatrix} \times & \times & \times \\ & * & * \\ & 0 & * \\ & 0 & * \\ & 0 & * \end{bmatrix}} \xrightarrow{Q_3} \underset{Q_3 Q_2 Q_3 A}{\begin{bmatrix} \times & \times & \times \\ & \times & \times \\ & & * \\ & & 0 \\ & & 0 \end{bmatrix}} \qquad (2.37)$$

Πρώτα ο $Q_1$ εφαρμόζεται στις γραμμές 1 έως 5 εισάγοντας μηδενικά στις θέσεις (2,1), (3,1), (4,1) και (5,1). Στη συνέχεια ο $Q_2$ εφαρμόζεται στις γραμμές 2 έως 5 εισάγοντας μηδενικά στις θέσεις (3,2), (4,2) και (5,2) χωρίς να αλλοιώνει τα μηδενικά που είχαν εισαχθεί πριν. Τέλος, ο $Q_3$ εφαρμόζεται στις γραμμές 3 έως 5

εισάγοντας μηδενικά στις θέσεις (4,3) και (5,3) χωρίς να αλλοιώνει κανένα από τα μηδενικά που είχαν εισαχθεί πριν.

Γενικά, ο $Q_k$ εφαρμόζεται στις γραμμές $k,.....,m$. Στην αρχή του βήματος $k$ υπάρχει ένα τμήμα με μηδενικά στις πρώτες $k-1$ στήλες αυτών των γραμμών. Η εφαρμογή του $Q_k$ δημιουργεί γραμμικούς συνδυασμούς αυτών των γραμμών και οι γραμμικοί συνδυασμοί των μηδενικών παραμένουν μηδενικά. Ύστερα από $n$ βήματα όλα τα στοιχεία κάτω από τη διαγώνιο θα έχουν απαλειφτεί και ο $Q_n \cdots Q_2 Q_1 A = R$ θα είναι άνω τριγωνικός.

### *2.4.3 HOUSEHOLDER ΑΝΑΚΛΑΣΤΕΣ*

Πώς μπορούμε να κατασκευάσουμε ορθομοναδιαίους πίνακες $Q_k$ και να εισάγουμε μηδενικά με τον τρόπο που μας υποδεικνύει η (2.37); Η βασική μέθοδος είναι η παρακάτω. Κάθε $Q_k$ επιλέγεται έτσι ώστε να είναι ένας ορθομοναδιαίος πίνακας της μορφής

$$Q_k = \begin{bmatrix} I & 0 \\ 0 & F \end{bmatrix} \tag{2.38}$$

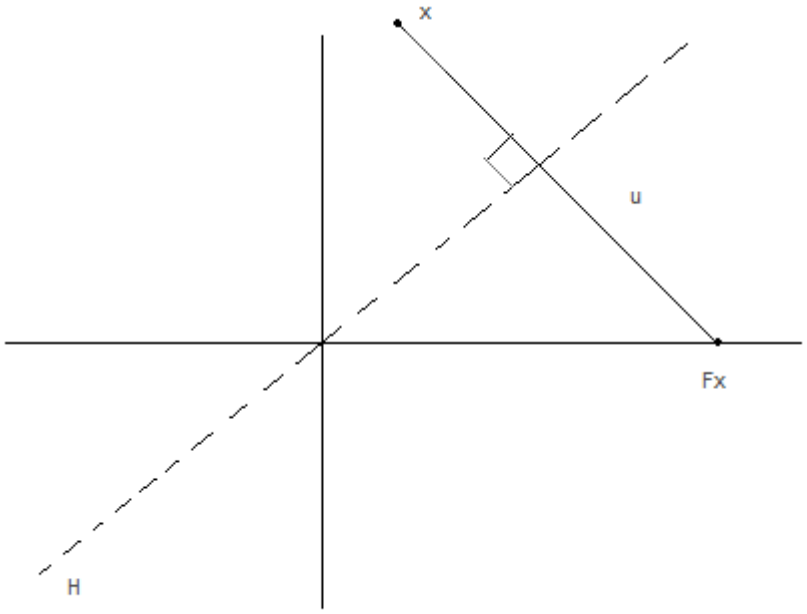
όπου I είναι ο ταυτοτικός πίνακας διάστασης $(k-1)\times(k-1)$ και $F$ είναι ένας $(m-k+1)\times(m-k+1)$ ορθομοναδιαίος πίνακας. Ο πολλαπλασιαμός με τον $F$ πρέπει να έχει σαν αποτέλεσμα την εισαγωγή μηδενικων στην $k$-οστή στήλη. Ο αλγόριθμος Householder επιλέγει τον $F$ έτσι ώστε να είναι ένας συγκεκριμένος πίνακας ο οποίος ονομάζεται Householder ανακλαστής.

Ας υποθέσουμε ότι στην αρχή του $k$ βήματος, τα στοιχεία $k,.....,m$ της $k$-οστής στήλης δίνονται από το διάνυσμα $x \in C^{m-k+1}$. Για να εισάγουμε τα σωστά μηδενικά στην $k$-οστή στήλη ο Householder ανακλαστής $F$ θα πρέπει να έχει επίδραση στην ακόλουθη απεικόνιση:

$$x = \begin{bmatrix} \times \\ \times \\ \times \\ \vdots \\ \times \end{bmatrix} \quad \overset{F}{\rightarrow} \quad Fx = \begin{bmatrix} \|x\| \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \|x\| e_1 \tag{2.39}$$

Η ιδέα για να το πετύχουμε αυτό παρουσιάζεται στο γράφημα 2.4. Ο ανακλαστής $P_y = (I - \frac{uu^*}{u^*u})y = y - u(\frac{u^*y}{u^*u})$ ανακλά το χώρο $C^{m-k+1}$ κατά μήκος του υπερεπιπέδου $H$ που είναι ορθογώνιο στο $u = \|x\| e_1 - x$. Ένα υπερπεπίπεδο είναι μια γενίκευση μεγαλύτερης διάστασης ενός χώρου δυο διαστάσεων στον τρισδιάστατο χώρο-ένας τρισδιάστατος υπόχωρος ενός χώρου τεσσάρων διαστάσεων και ούτως κάθε εξής.

Γενικά, ένα υπερεπίπεδο μπορεί να χαρακτηριστεί σαν ένα σύνολο σημείων ορθογώνιων σε ένα σταθερό μημηδενικό διάνυσμα. Στο σχήμα 2.4 αυτό το διάνυσμα είναι το $u = \|x\| e_1 - x$ και μπορούμε να θεωρήσουμε τη διακεκομμένη γραμμή σαν μια απεικόνιση του $H$ υπό γωνία.




Σχήμα 2.4 Μια ανάκλαση Householder

Όταν εφαρμοζεται ο ανακλαστής κάθε σημείο στη μια πλευρά του υπερεπιπεδου $H$ απεικονίζεται στην εικόνα του κατόπτρου του στην άλλη πλευρά του υπερεπιπέδου. Πιο συγκεκριμένα, το $x$ απεικονίζεται στο $\|x\| e_1$. Ο τύπος για αυτή την ανάκλαση μπορεί να προκύψει όπως ακολουθεί. Στη σχέση (2.11) είδαμε ότι για κάθε $y \in C^m$, το διάνυσμα

$$P_y = (I - \frac{uu^*}{u^* u})y = y - u(\frac{u^* y}{u^* u})$$

είναι η ορθογώνια προβολή του $y$ στο χώρο $H$. Για να γίνει η ανάκλαση του $y$ κατά μήκος του $H$ δεν πρέπει να σταματήσουμε σε αυτό το σημείο. Πρέπει να καλύψουμε τη διπλάσια απόσταση στην ίδια διεύθυνση. Οπότε η ανάκλαση του $Fy$ θα πρέπει να είναι

$$Fy = (I - 2\frac{uu^*}{u^* u})y = y - 2u(\frac{u^* y}{u^* u})$$

Οπότε ο πίνακας $F$ είναι ο

$$F = I - 2\frac{uu^*}{u^*u} \tag{2.40}$$

Ας σημειώσουμε ότι ο προβολέας $P$ (τάξης $m-1$) και ο ανακλαστής $F$ (πλήρης τάξη, ορθομοναδιαίος) για την ώρα διαφέρουν μόνο στον παράγοντα 2.

### *2.4.4 Ο ΚΑΛΥΤΕΡΟΣ ΑΠΟ ΤΟΥΣ ΔΥΟ ΑΝΑΚΛΑΣΤΕΣ*

Στη σχέση (2.39) και στο σχήμα 2.4 είναι απλά τα πράγματα. Για παράδειγμα, υπάρχουν πολλοί Householder ανακλαστές οι οποίοι εισάγουν τα απαραίτητα μηδενικά. Το διάνυσμα $x$ μπορεί να ανακλαστεί στο $z\|x\|e_1$ όπου το $z$ είναι κάθε βαθμωτό μέγεθος με $|z|=1$. Σε πιο περίπλοκες περιπτώσεις, υπάρχει ένας κύκλος από πιθανές ανακλάσεις και ακόμα και στους πραγματικούς, υπάρχουν δυο εναλλακτικές, παρουσιάζονται από ανακλάσεις κατά μήκους δυο διαφορετικών υπερεπίπεδα, $H^+$ και $H^-$.
Από μαθηματικής άποψης, οποιαδήποτε επιλογή προσήμου είναι ικανοποιητική. Όμως, σε αυτή την περίπτωση όπου ο στόχος είναι αριθμητική ευστάθεια-μη ευαισθησία στα σφάλματα μηχανής-μας ορίζει ότι πρέπει να κάνουμε μια επιλογή ανάμεσα στα δυο. Για αριθμητική ευστάθεια είναι επιθυμητό να γίνει ανάκλαση του $x$ στο διάνυσμα $z\|x\|e_1$ το οποίο δεν είναι πολύ κοντά στο $x$. Για να το επιτύχουμε αυτό μπορούμε να διαλέξουμε $z=-sign(x_1)$ όπου το $x_1$ υποδηλώνει την πρώτη συνιστώσα του $x$, έτσι ώστε το διάνυσμα που ανακλάται να γίνει $u=-sign(x_1)\|x\|e_1-x$ ή εξαλείφοντας τους παράγοντες -1 να έχουμε

$$u = sign(x_1)\|x\|e_1 + x \tag{2.41}$$

Για να κάνουμε μια ολοκληρωμένη περιγραφή μπορούμε αυθαίρετα να υποθέσουμε ότι $sign(x_1)=1$ αν $x_1=0$.
Δεν είναι δύσκολο να δούμε γιατί η επιλογή προσήμου κάνει τη διαφορά στην ευστάθεια. Ας υποθέσουμε ότι στο σχήμα 2.5 η γωνία ανάμεσα στο $H^+$ και τον άξονα $e_1$ είναι πολύ μικρή. Τότε το διάνυσμα $u=\|x\|e_1-x$ είναι πολύ μικρότερο από το $x$ ή το $\|x\|e_1$. Οπότε ο υπολογισμός του $u$ αντιπροσωπεύει μια αφαίρεση ανάμεσα σε κοντινές ποσότητες και θα αλλοιωθεί από σφάλματα στρογγυλοποίησης. Αν διαλέξουμε το πρόσημο όπως στην (2.39) αποφεύγουμε

τέτοιες παρεμβάσεις εξασφαλίζοντας ότι το $\|u\|$ δεν θα είναι ποτέ μικρότερο από το $\|x\|$.

### *2.4.5 Ο ΑΛΓΟΡΙΘΜΟΣ*

Τώρα θα παρουσιάσουμε τον αλγόριθμο του Householder. Έστω $A$ ένας πίνακας, ορίζουμε $A_{i:i',j:j'}$ να είναι ένας $(i'-i+1)\times(j'-j+1)$ υποπίνακας του $A$ με άνω δεξί στοιχείο το $a_{ij}$ και κάτω αριστερό στοιχείο το $a_{i'j'}$. Στην ειδική περίπτωση όπου ο υποπίνακας μειώνεται σε ένα υποδιάνυσμα γραμμής ή στήλης γράφουμε $A_{i,j:j'}$ ή $A_{i:i',j}$ αντίστοιχα.
Ο ακόλουθος αλγόριθμος υπολογίζει τον παράγοντα $R$ μίας QR παραγοντοποίησης ενός $m\times n$ πίνακα $A$ με $m\geq n$. Κατά την εκτέλεση του αλγορίθμου $n$ διανύσματα ανάκλασης $u_1,.....,u_n$ αποθηκεύονται για μετέπειτα χρήση.

ΑΛΓΟΡΙΘΜΟΣ 2.3 Householsder τριγωνοποίηση μέσω QR παραγοντοποίησης
Για $k=1$ μέχρι $n$

$$x = A_{k:m,k}$$
$$u_k = sign(x_1)\|x\|_2 e_1 + x$$
$$u_k = u_k / \|u_k\|_2$$
$$A_{k:m,k:n} = A_{k:m,k:n} - 2u_k(u_k^* A_{k:m,k:n})$$

### *2.4.6 ΕΦΑΡΜΟΖΟΝΤΑΣ Η' ΔΗΜΙΟΥΡΓΩΝΤΑΣ ΤΟΝ Q*

Για την πλήρωση του αλγορίθμου 2.3 ο πίνακας $A$ έχει μειωθεί σε άνω τριγωνική μορφή. Αυτός είναι ο πίνακας $R$ της QR παραγοντοποίησης $A=QR$. Ο ορθομοναδιαίος πίνακας $Q$ όμως ούτε έχει δημιουργηθεί ούτε η $n$-οστή στήλη του υποπίνακά του $\hat{Q}$ αντιστοιχεί σε κάποιο μειωμένη QR παραγοντοποίηση. Υπάρχει κάποιος λόγος που συμβαίνει αυτό. Η κατασκευή του $Q$ ή του $\hat{Q}$ χρειάζεται επιπλέον πράξεις και σε πολλές εφαρμογές μπορούμε να αποφύγουμε αυτές τις πράξεις εφαρμόζοντας απευθείας τον τύπο

$$Q^* = Q_n \cdots Q_2 Q_1 \tag{2.42}$$

ή τον συζυγή της

$$Q = Q_1 Q_2 \cdots Q_n \tag{2.43}$$

Για παράδειγμα, στην προηγούμενη παράγραφο είδαμε ότι ένα τετραγωνικό σύστημα εξισώσεων $Ax = b$ μπορεί να λυθεί με QR παραγοντοποίηση του πίνακα $A$. Η μόνη περίπτωση στην οποία χρειαστήκαμε τον $Q$ ήταν για τον υπολογισμό του γινομένου $Q^*b$. Από τη σχέση (2.42) μπορούμε να υπολογίσουμε το $Q^*b$ σαν μια ακολουθία $n$ πράξεων που εφαρμόζονται στον $b$, οι ίδιες πράξεις που γίνονται στον πίνακα $A$ για να τον κάνουμε τριγωνικό. Ο αλγόριθμος είναι αυτός που ακολουθεί

ΑΛΓΟΡΙΘΜΟΣ 2.4 Έμεσος υπολογισμός του γινομένου $Q^*b$.

Για $k = 1$ μέχρι $n$

$$b_{k:m} = b_{k:m} - 2u_k(u_k^* b_{k:m}) .$$

Όμοια, ο υπολογισμός του γινομένου $Qx$ μπορεί να γίνει ακολουθώντας την ίδια διαδικασία σε αντιστροφη σειρά.

ΑΛΓΟΡΙΘΜΟΣ 2.5 Έμμεσος υπολογισμός του γινομένου $Qx$

Για $k = 1$ μέχρι και το 1

$$x_{k:m} = x_{k:m} - 2u_k(u_k^* x_{k:m}) .$$

Η πολυπλοκότητα και των δυο αλγορίθμων είναι της τάξης του $O(mn)$ και όχι $O(mn^2)$ όπως του αλγορίθμου 2.3.

Μερικές φορές, μπορεί να θέλουμε να κατασκευάσουμε τον πίνακα $Q$ αναλυτικά. Αυτό μπορούμε να το επιτύχουμε με διάφορους τρόπους. Μπορούμε να κατασκευάσουμε τον $QI$ μέσω του αλγορίθμου 2.5 υπολογίζοντας τις στήλες του $Qe_1, Qe_2, ....., Qe_m$. Εναλλακτικά, μπορούμενα κατασκευάσουμε τον $Q^*I$ μέσω του αλγορίθμου 2.4 και μετά να βρούμε το συζυγή του αποτελέσματος. Μια παραλλαγή αυτής της ιδέας είναι να βρίσκουμε το συζυγή σε κάθε βήμα αντί στο τελικό αποτέλεσμα., δηλαδή να κατασκευάσουμε τον $IQ$ υπολογίζοντας τις στήλες $e_1^*Q, e_2^*Q, ...., e_m^*Q$ όπως προτάθηκε από την (2.43). Από αυτές τις διάφορες ιδέες η καλύτερη είναι η πρώτη η οποία βασίζεται στον αλγόριθμο 2.5. Ο λόγος είναι ότι ξεκινάει με πράξεις που περιέχουν τους $Q_n, Q_{n-1}, ....$ οι οποίες τροποποιούν ένα μικρό μόνο κομμάτι του διανύσματος στο οποίο εφαρμόζονται. Αν καταφέρουμε και εκμεταλευτούμε αυτή την ιδιότητα τότε αυξάνεται η ταχύτητα του αλγορίθμου.

Αν χρειάζεται μόνο ο $\hat{Q}$ και όχι ο $Q$ είναι αρκετό να υπολογίσουμε τις στήλες $Qe_1, Qe_2, ..., Qe_n$.

### *2.4.7 ΠΟΛΥΠΛΟΚΟΤΗΤΑ*

Η πολυπλοκότητα του αλγορίθμου 2.3 κυριαρχείται από την εσωτερική επανάληψη,

$$A_{k:m,j} - 2u_k(u_k^* A_{k:m,j}) \qquad (2.44)$$

Αν το μήκος του διανύσματος είναι $l = m - k + 1$ ο υπολογισμός αυτός απαιτεί $4l - 1 \approx 4l$ βαθμωτές πράξεις: $l$ για την αφαίρεση, $l$ για το βαθμωτό πολλαπλασιασμό και $2l - 1$ για το εσωτερικό γινόμενο. Αυτά είναι $\approx 4$ βήματα για κάθε στοιχείο που εισάγεται.

Ίσως να να αθροίσουμε αυτά τα τέσσερα βήματα κάθε στοιχείου με γεωμετρικό τρόπο όπως στην προηγούμενη παράγραφο. Κάθε διαδοχικό βήμα της εξωτερικής επανάληψης εφαρμόζεται σε λιγότερες γραμμές επειδή κατά την εκτέλεση του βήματος $k$ οι γραμμές $1,....,k-1$ δεν αλλάζουν. Επίσης, κάθε βήμα εκτελείται σε λιγότερες στήλες επειδή οι στήλες $1,....,k-1$ των γραμμών που εφαρμόζεται ο αλγόριθμος είναι μηδενικές και κατά συνέπεια παραλείπονται.

Πολυπλοκότητα της Householder τριγωνοποίησης $\sim 2mn^2 - \frac{2}{3}n^3$ βήματα (2.45)

## 2.5 ΠΡΟΒΛΗΜΑΤΑ ΕΛΑΧΙΣΤΩΝ ΤΕΤΡΑΓΩΝΩΝ

Η μέθοδος των ελαχίστων τετραγώνων είναι ένα πολύ σημαντικό εργαλείο επίλυσης προβλημάτων. Εφευρέθηκε από τους Gauss και Lagrange γύρω στο 1800. Στη γλωσσα της γραμμικής άλγεβρας, το πρόβλημα του οποίου ζητάμε τη λύση είναι ένα υπερκαθορισμένο σύστημα εξισώσεων $Ax = b$, ορθογώνιο με περισσότερες γραμμές απ' ότι στήλες. Η ιδέα των ελαχίστων τετραγώνων είναι να «λύσουμε» ένα σύστημα έτσι ώστε να ελαχιστοποιήσουμε τη νόρμα δεύτερης τάξης του υπολοίπου $b - Ax$.

### *2.5.1 ΤΟ ΠΡΟΒΛΗΜΑ*

Ας υποθέσουμε ότι έχουμε ένα γραμμικό σύστημα εξισώσεων με $n$ αγνώστους αλλά $m > n$ εξισώσεις. Δηλαδή θέλουμε να βρούμε ένα διάνυσμα $x \in C^n$ που να ικανοποιεί την $Ax = b$, όπου $A \in C^{m \times n}$ και $b \in C^n$. Γενικά, ένα τέτοιο πρόβλημα δεν έχει λύση. Ένα κατάλληλο διάνυσμα $x$ υπάρχει μόνο αν το $b$ ανήκει στο πεδίο τιμών του $A$ και αν το $b$ είναι ένα διάνυσμα $m$ διάστασης όπου το $range(A)$ έχει μέγιστη διάσταση $n$, το πρόβλημα έχει λύση για συγκεκριμένα $b$. Τότε λέμε ότι ένα ορθογώνιο σύστημα εξισώσεων με $m > n$ είναι υπερκαθορισμένο. Ένα διάνυσμα ονομάζεται υπόλοιπο αν ισχύει

$$r = b - Ax \in C^m \qquad (2.46)$$

και μπορεί να είναι αρκετά μικρό για μια κατάλληλη επιλογή του $x$ αλλά γενικά δεν μπορεί να είναι ίσο με το μηδέν.
Τι μπορεί να σημαίνει το να λύσουμε ένα πρόβλημα το οποίο δεν έχει λύση; Στην περίπτωση ενός υπερκαθορισμένου σύστηματος εξισώσεων υπάρχει λογική απάντηση. Αφού το υπόλοιπο $r$ δεν μπορεί να είναι ίσο με το μηδέν, ας υποθέσουμε ότι είναι όσο μικρό γίνεται. Μετρώντας το μέγεθος του $r$ πρέπει να διαλέξουμε και μια νόρμα. Αν διαλέξουμε τη νόρμα δεύτερης τάξης, το πρόβλημα παίρνει την παρακάτω μορφη:

$$\text{Δοθέντος } A \in C^{m\times n},\ m \geq n,\ b \in C^m$$
$$\text{Βρείτε } x \in C^n \text{ τέτοιο ώστε } \|b - Ax\|_2 \text{ να είναι ελάχιστη} \qquad (2.47)$$

Αυτή η διατύπωση είναι του γενικού (γραμμικού) προβλήματος. Η επιλογή της νόρμας δεύτερης τάξης μπορεί να δικαιολογηθεί από διάφορα γεωμετρικά και στατιστικά επιχειρήματα και οδηγεί σε απλούς αλγόριθμους, επειδή η παράγωγος μιας τετραγωνικής συνάρτησης, η οποία πρέπει να τεθεί ίση με το μηδέν για ελαχιστοποίηση, είναι γραμμική.
Η νόρμα δεύτερης τάξης αντιπροσωπεύει την Ευκλείδεια απόσταση οπότε υπάρχει μια απλή γεωμετρική ερμηνεία της (2.47). Ψάχνουμε ένα διάνυσμα $x \in C^n$ τέτοιο ώστε το διάνυσμα $Ax \in C^m$ να είναι το πιο κοντινό σημείο του $b$ στο $range(A)$.

Παράδειγμα: Πολυωνυμική προσαρμογή δεδομένων

Σαν παράδειγμα, ας συγκρίνουμε την πολυωνυμική παρεμβολή η οποία μας οδηγεί σε ένα τετραγωνικό σύστημα εξισώσεων και την πολυωνυμική προσαρμογή δεδομένων ελαχίστων τετραγώνων όπου το σύστημα είναι ορθογώνιο.

Παράδειγμα:Πολυωνυμική παρεμβολή

Έστω ότι έχουμε $\|r\|_2^2$ $m$ διακριτά σημεία $x_1,.....,x_m \in C$ και δεδομένα $y_1,.....,y_m \in C$ σε αυτά τα σημεία. Τότε υπάρχει μοναδική πολυωνυμιακή παρεμβολή των δεδομένων αυτών σε αυτά τα σημεία η οποία είναι με βαθμό πολυωνύμου το πολύ $m-1$,

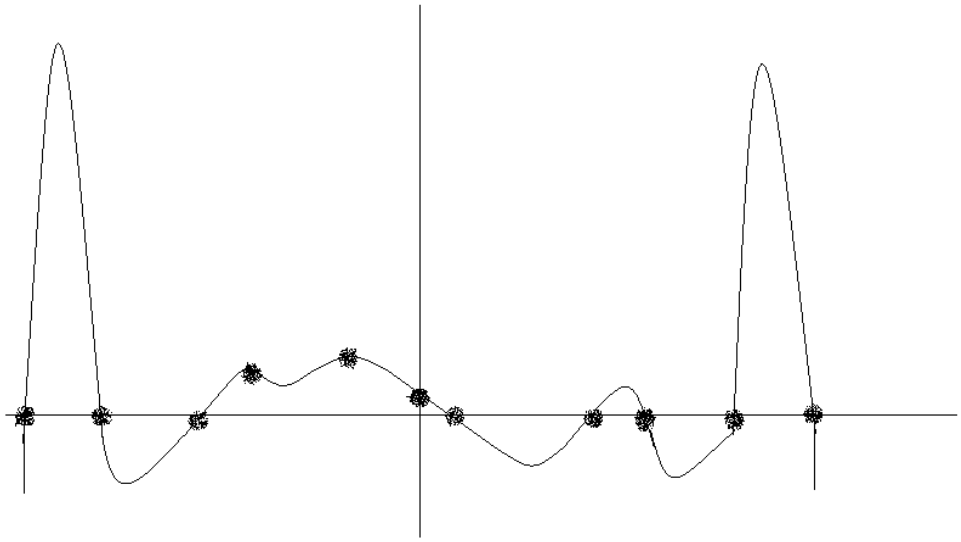
$$p(x) = c_0 + c_1 x + ..... + c_{m-1} x^{m-1} \qquad (1)$$

με την ιδιότητα ότι σε κάθε $x_i, p(x_i) = y_i$. Η σχέση των δεδομένων $\{x_i\},\{y_i\}$ με τους συντελεστές $\{c_i\}$ μπορεί να εκφραστεί από το τετραγωνικό σύστημα Vandermonde που έχουμε ήδη δει στο παράδειγμα

$$\begin{bmatrix} 1 & x_1 & x_1^2 & & x_1^{m-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{m-1} \\ 1 & x_3 & x_3^3 & & x_3^{m-1} \\ & \vdots & & & \vdots \\ 1 & x_m & x_m^2 & \cdots & x_m^{m-1} \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ c_{m-1} \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_m \end{bmatrix} \quad (2)$$

Για να προσδιορίσουμε τους συντελεστές $\{c_i\}$ για ένα δεδομένο σύνολο δεδομένων μπορούμε να λύσουμε αυτό το σύστημαεξισώσεων το οποίο δεν είναι σίγουρα ομαλό όσο τα σημεία $\{x_i\}$ είναι διακριτά.

Στο παρακάτω γράφημα βλέπουμε ένα παράδειγμα της διαδικασίας της πολυωνυμικής παρεμβολής. Έχουμε έντεκα σημεία δεδομένων στη μορφή ενός διακριτού τετράγωνου κύματος που παρουσιάζεται με σταυρούς και η καμπύλη $p(x)$ τα διαπερνά. Όμως, η προσαρμογή δεν είναι και τόσο καλή. Κοντά στα άκρα της προσαρμογής το πολυώνυμο $p(x)$ ταλαντώνεται έντονα, φαινόμενο που έρχεται σε αντίθεση με τη διαδικασία της παρεβολής και δεν αποτελεί λογική αντανάκλαση των δεδομένων.



Γράφημα 1 Βαθμού 10 πολυωνυμική παρεμβολή σε έντεκα σημεία δεδομένων

Αυτή η μη ικανοποιητική συμπεριφορά είναι σύνηθης στην πολυωνυμική παρεμβολή. Τα σχήματα που προκύπτουν συνήθως δεν είναι ικανοποιητικά και τείνουν να μην βελτιώνονται παρά να χειροτερεύουν καθώς τα δεδοδένα αυξάνονται. Ακόμα και αν το σχήμα είναι ικανοποιητικό, η διαδικασία παρεμβολής μπορεί να είναι κακής κατάστασης, πχ ευαισθησία στις μεταβολές των δεδομένων. Για να αποφύγουμε τέτοια προβλήματα, μπορούμε να χρησιμοποιήσουμε μη ομοιόμορφο σύνολο σημείων παρεμβολής όπως τα σημεία Chebysev στο διάστημα $[-1,1]$. Στις εφαρμογές όμως δεν είναι πάντα εύκολο να διαλέξουμε τα σημεία παρεμβολής.

Παράδειγμα: Πολυωνυμιακή προσαρμογή ελαχίστων τετραγώνων

Χωρίς να αλλάζουμε τα σημεία δεδομένων, μπορομε να βελτιώσουμε την παρεμβολή μειώνοντας το βαθμό του πολυωνύμου. Δεδομένων $x_1,.....,x_m$ και $y_1,.....,y_m$ έστω ότι έχουμε ένα πολυώνυμο βαθμού $n-1$

$$p(x)=c_0+c_1x+.....+c_{n-1}x^{n-1}$$

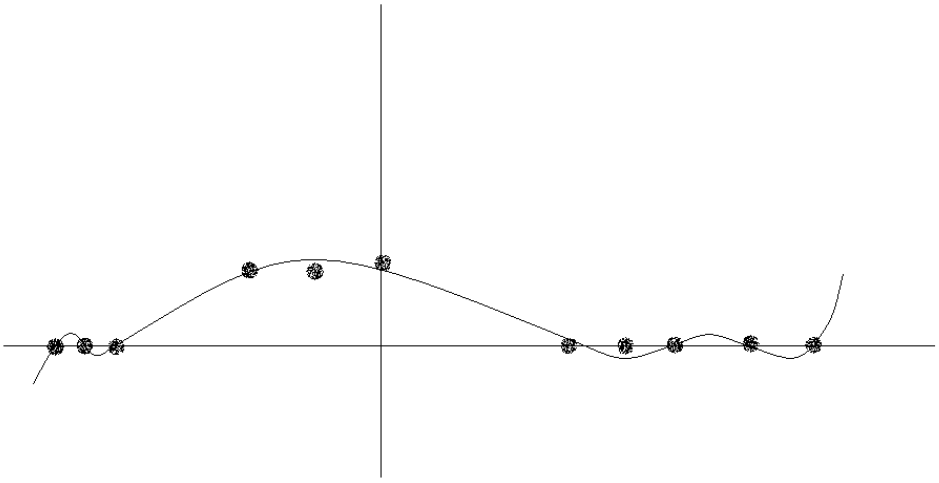
για κάποια $n<m$. Ένα τέτοιο πολυώνυμο είναι η εφαρμογή των ελαχίστων τετραγώνων στα δεδομένα αν ελαχιστοποιείται το άθροισμα των τετραγώνων στην απόκλιση των δεδομένων

$$\sum_{i=1}^{m}\left|p(x_i)-y_i\right|^2$$

Το άθροισμα των τετραγώνων είναι ίσο με το τετράγωνο της νόρμας του υπολοίπου $\|r\|_2^2$ για το ορθογώνιο σύστημα Vandermonde

$$\begin{bmatrix} 1 & x_1 & x_1^2 & & x_1^{m-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{m-1} \\ 1 & x_3 & x_3^3 & & x_3^{m-1} \\ & \vdots & & & \vdots \\ 1 & x_m & x_m^2 & \cdots & x_m^{m-1} \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{n-1} \end{bmatrix} \approx \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_m \end{bmatrix}$$

Στο γράφημα 2 βλέπουμε τι συμβαίνει αν πάρουμε τα ίδια έντεκα σημεία από το προηγούμενο παράδειγμα με βαθμό πολυωνύμου ίσο με 7. Το νέο πολυώνυμο δεν παρεμβάλεται στα σημεία αλλά απεικονίζει τη συνολική συμπεριφορά τους πολύ καλύτερα από το πολυώνυμο του προηγούμενου παραδείγματος. Το γράφημα είναι το παρακάτω.



Γράφημα 2 Βαθμού 7 πολυώνυμο ελαχίστων τετραγώνων εφαρμοσμένο στα έντεκα ίδια σημεία δεδομένων.

## *2.5.2 ΟΡΘΟΓΩΝΙΑ ΠΡΟΒΟΛΗ ΚΑΙ ΚΑΝΟΝΙΚΕΣ ΕΞΙΣΩΣΕΙΣ*

Το ερώτημα είναι πώς λύνονται τα προβλήματα ελαχίστων τετραγώνων γενικά; Το κλειδί στο να βρούμε αλγορίθμους για τη λύση είναι η ορθογώνια προβολή.
Η ιδέα απεικονίζεται στο σχήμα 2.6. Στόχος μας είναι να βρούμε το πιο κοντινό σημείο $Ax$ στο $b$ στο $range(A)$ έτσι ώστε η νόρμα του υπολοίπου $r = b - Ax$ να είναι ελάχιστη. Από γεωμετρική άποψη, είναι φανερό ότι αυτό θα προκύψει αν $Ax = Pb$, όπου $P \in C^{m\times m}$ είναι ο ορθογώνιος προβολέας ο οποίος απεικονίζει το $C^m$ πάνω στο $range(A)$. Με άλλα λόγια, το υπόλοιπο $r = b - Ax$ πρέπει να είναι ορθογώνιο στο $range(A)$. Διατυπώνουμε αυτή την υπόθεση στο παρακάτω θεώρημα.

ΘΕΩΡΗΜΑ 2.5
Έστω $A \in C^{m\times n}$ $(m \geq n)$ και $b \in C^m$. Ένα διάνυσμα $x \in C^n$ ελαχιστοποιεί τη νόρμα του υπολοίπου $\|r\|_2 = \|b - Ax\|_2$ και κατά συνέπεια λύνεται το πρόβλημα ελαχίστων τετραγώνων (2.47) αν και μόνο αν $r \perp range(A)$, δηλαδή αν

$$A^* r = 0 \tag{2.48}$$

ή ισοδύναμα

$$A^* Ax = A^* b \tag{2.49}$$

ή ισοδύναμα

$$Pb = Ax \tag{2.50}$$

όπου $P \in C^{m\times m}$ είναι ο ορθογώνιος προβολέας πάνω στο $range(A)$. Το $n \times n$ σύστημα εξισώσεων (2.49), γνωστές και ως κανονικές εξισώσεις, είναι ομαλό αν και μόνο αν ο πίνακας $A$ έχει πλήρη τάξη. Συνεπώς, η λύση $x$ είναι μοναδική αν και μόνο αν ο πίνακας $A$ έχει πλήρη τάξη.

ΑΠΟΔΕΙΞΗ
Η ισοδυναμία ανάμεσα στις (2.48) και (2.50) προκύπτει από τις ιδιότητες των ορθογώνιων προβολέων και η ισοδυναμία των (2.48) και (2.49) προκύπτει από τον ορισμό του $r$. Για να δείξουμε ότι $y = Pb$ είναι το μοναδικό σημείο στο $range(A)$ που ελαχιστοποιεί τη $\|b - y\|_2$, υποθέτουμε ότι $z \neq y$ είναι ένα άλλο σημείο στο $range(A)$. Επειδή το $z - y$ είναι ορθογώνιο στο $b - y$ από το Πυθαγόρειο Θεώρημα έχουμε ότι $\|b - z\|_2^2 = \|b - y\|_2^2 + \|y - z\|_2^2 > \|b - y\|_2^2$, όπως απαιτείται. Τέλος, ας σημειώσουμε ότι αν ο $A^* A$ είναι μη ομαλός τότε $A^* Ax = 0$ για κάποια μη μηδενικά $x$, συνεπάγεται ότι $x^* A^* Ax = 0$. Οπότε $Ax = 0$ το οποίο συνεπάγεται ότι ο $A$ είναι

μειωμένης τάξης. Αντίστροφα, αν ο $A$ είναι μειωμένης τάξης, τότε $Ax = 0$ για κάποιο μη μηδενικό $x$, συνεπάγεται ότι επίσης $A^* Ax = 0$ οπότε ο $A^* A$ είναι μη ομαλός. Από την (2.49) αυτός ο χαρακτηρισμός των ομαλών πινάκων $A^* A$ συνεπάγεται ότι την υπόθεση για τη μοναδικότητα του $x$.

### *2.5.3 ΨΕΥΔΟΑΝΤΙΣΤΡΟΦΗ*

Είδαμε ότι αν ο πίνακας $A$ έχει πλήρη τάξη τότε η λύση $x$ του προβλήματος ελαχίστων τετραγώνων (2.47) είναι μοναδική και δίνεται από τον τύπο $x = (A^* A)^{-1} A^* b$. Ο πίνακας $(A^* A)^{-1} A^*$ είναι γνωστός ως και ψευδοαντίστροφος του $A$ και συμβολίζεται με $A^+$,

$$A^+ = (A^* A)^{-1} A^* \in C^{n,m} \tag{2.51}$$

Αυτός ο πίνακας απεικονίζει διανύσματα $b \in C^m$ σε διανύσματα $x \in C^n$, κάτι που εξηγεί γιατί έχει διαστάσεις $n \times m$-περισσότερες γραμμές από στήλες.
Μπορούμε να διατυπώσουμε το πρόβλημα ελαχίστων τετραγώνων πλήρους τάξης (2.47) όπως ακολούθως. Το πρόβλημα είναι να υπολογίσουμε ένα ή και τα δυο διανύσματα

$$x = A^+ b, \qquad y = Pb \tag{2.52}$$

όπου ο $A^+$ είναι ο ψευδοαντίστροφος του $A$ και $P$ είναι ο ορθογώνιος προβολέας πάνω στο $range(A)$.
Στη συνέχεια θα περιγράψουμε τρείς αλγορίθμους με τους οποίους μπορούμε να το κάνουμε αυτό.

### *2.5.4 ΚΑΝΟΝΙΚΕΣ ΕΞΙΣΩΣΕΙΣ*

Ένας κλασσικός τρόπος επίλυσης προβλημάτων ελαχίστων τετραγώνων είναι να λύσουμε τις κανονικές εξισώσεις (2.49). Αν ο πίνακας $A$ είναι πλήρους τάξης τότε έχουμε ένα τετραγωνικό, ερμιτιανό, θετικά ορισμένο σύστημα εξισώσεων διάστασης $n$. Η βασική μέθοδος για την επίλυση τέτοιων προβλημάτων είναι η παραγοντοποίηση Cholesky. Η μέθοδος αυτή κατασκευάζει μια παραγοντοποίηση $A^* A = R^* R$, όπου ο $R$ είναι άνω τριγωνικός και μειώνει το (2.49) στις εξισώσεις

$$R^* Rx = A^* b \tag{2.53}$$

Ο αλγόριθμος είναι ο παρακάτω.

ΑΛΓΟΡΙΘΜΟΣ 2.6 Ελάχιστα τετράγωνα μέσω κανονικών εξισώσεων

1. Κατασκευάστε τον πίνακα $A^*A$ και τα διανύσματα $A^*b$.
2. Υπολογίστε την παραγοντοποίηση Cholesky $A^*A = R^*R$.
3. Λύστε το κάτω τριγωνικό σύστημα $R^*w = A^*b$ ως προς $w$.
4. Λύστε το άνω τριγωνικό σύστημα $R^*x = w$ ως προς $x$.

Τα βήματα που συνεισφέρουν περισσότερο στην πολυπλοκότητα του αλγορίθμου είναι τα δύο πρώτα. Λόγω συμμετρίας, ο υπολογισμός του $A^*A$ απαιτεί μόνο $mn^2$ υπολογισμούς, τους μισούς απ' όσους θα απαιτούνταν αν οι $A$ και $A^*$ ήταν τυχαίοι πίνακες ίδιας διάστασης. Η παραγοντοποίηση Cholesky , η οποία εκμεταλεύεται τη συμμετρία επίσης, απαιτεί $n^3/3$ υπολογισμούς. Συνολικά, η επίλυση προβλημάτων ελαχίστων τετραγώνων μέσω κανονικών εξισώσεων απαιτεί συνολικό πλήθος πράξεων

$$\text{Πολυπλοκότητα αλγορίθμου 2.6: } \sim mn^2 + \frac{1}{3}n^3 \text{ βήματα.} \qquad (2.54)$$

## *2.5.5 QR ΠΑΡΑΓΟΝΤΟΠΟΙΗΣΗ*

Η «μοντέρνα κλασσική» μέθοδος για την επίλυση προβλημάτων ελαχίστων τετραγώνων, δημοφιλής στη δεκαετία του 1960, βασίζεται στη μειωμένη QR παραγοντοποίηση. Από την Gram-Schmidt ορθογωνοποίηση ή πιο συχνά από την Householder τριγωνοποίηση μπορούμε να κατασκευάσουμε μια παραγοντοποίηση $A = \hat{Q}\hat{R}$. Ο ορθογώνιος προβολέας $P$ μπορεί να γραφτεί ως $P = \hat{Q}\hat{Q}^*$ (2.6), οπότε έχουμε

$$y = Pb = \hat{Q}\hat{Q}^*b \quad . \qquad (2.55)$$

Επειδή $y \in range(A)$, το σύστημα $Ax = y$ έχει μια ακριβή λύση. Συνδυάζοντας την QR παραγοντοποίηση και την (2.55) έχουμε

$$\hat{Q}\hat{R}x = \hat{Q}\hat{Q}^*b \qquad (2.56)$$

και με έναν πολλαπλασιασμό από δεξιά με τον $\hat{Q}^*$ έχουμε τελικά

$$\hat{R}x = \hat{Q}^*b \qquad (2.57)$$

(Πολλαπλασιάζοντας με $\hat{R}^{-1}$ προκύπτει ο τύπος $A^{+} = \hat{R}^{-1}\hat{Q}^{*}$ για τον ψευδοαντίστροφο.) Η ισότητα (2.57) είναι ένα άνω τριγωνικό σύστημα, ομαλό αν ο $A$ είναι πλήρους τάξης και μπορεί να λυθεί με προς τα πίσω αντικατάσταση.

ΑΛΓΟΡΙΘΜΟΣ 2.7 Ελάχιστα τετράγωνα μέσω QR παραγοντοποίησης

1. Υπολογίστε τη μειωμένη QR παραγοντοποίηση $A = \hat{Q}\hat{R}$.
2. Υπολογίστε το διάνυσμα $\hat{Q}^{*}b$.
3. Λύστε το άνω τριγωνικό σύστημα $\hat{R}x = \hat{Q}^{*}b$ ως προς $x$.

Μπορούμε να παρατηρήσουμε ότι η (2.57) μπορεί να προκύψει και από τις κανονικές εξισώσεις. Αν $A^{*}Ax = A^{*}b$ τότε $\hat{R}^{*}\hat{Q}^{*}\hat{Q}\hat{R}x = \hat{R}^{*}\hat{Q}^{*}b$ όπου συνεπάγεται ότι $\hat{R}x = \hat{Q}^{*}b$.
Η παράμετρος που καθορίζει την πολυπλοκότητα του αλγορίθμου 2.7 είναι το κόστος της QR παραγοντοποίησης. Αν είχαμε χρησιμοποιήσει Householder ανακλαστές σε αυτό το βήμα θα είχαμε από την (2.45)

Πολυπλοκότητα αλγορίθμου 2.7: $\sim 2mn^{2} - \frac{2}{3}n^{3}$ βήματα . (2.58)

### *2.5.6 SVD*

Στο σημείο αυτό θα δούμε άλλο ένα αλγόριθμο για την επίλυση προβλημάτων ελαχίστων τετραγώνων. Έστω $P$ να είναι $P = \hat{U}\hat{U}^{*}$ έχουμε

$$y = Pb = \hat{U}\hat{U}^{*}b \tag{2.59}$$

και τα ανάλογα των (2.56) και (2.57) είναι

$$\hat{U}\hat{\Sigma}V^{*}x = \hat{U}\hat{U}^{*}b \tag{2.60}$$

και

$$\hat{\Sigma}V^{*}x = \hat{U}^{*}b \tag{2.61}$$

(Πολλαπλασιάζοντας με $V\hat{\Sigma}^{-1}$ προκύπτει $A^{+} = V\hat{\Sigma}^{-1}\hat{U}^{*}$.)
Ο αλγόριθμος είναι ο παρακάτω.

ΑΛΓΟΡΙΘΜΟΣ 2.8 Ελάχιστα τετράγωνα μέσω SVD

1. Υπολογίστε τη μειωμένη SVD $A = \hat{U}\hat{\Sigma}V^*$.
2. Υπολογίστε το διάνυσμα $\hat{U}^* b$.
3. Λύστε το διαγώνιο σύστημα $\hat{\Sigma}w = \hat{U}^* b$ ως προς $w$.
4. Θέστε $x = Vw$.

Μπορούμε να παρατηρήσουμε ότι ενώ η QR παραγοντοποίηση μειώνει το πρόβλημα ελαχίστων τετραγώνων σε ένα τριγωνικό σύστημα εξισώσεων, η SVD το μειώνει σε ένα διαγώνιο σύστημα εξισώσεων το οποίο είναι επιλύσιμο. Αν ο πίνακας $A$ έχει πλήρη τάξη, τότε το διαγώνιο σύστημα είναι ομαλό.
Όπως πριν, η (2.61) μπορεί να προκύψει από τις κανονικές εξισώσεις. Αν $A^*Ax = A^*b$, τότε $V\hat{\Sigma}^*\hat{U}^*\hat{U}\hat{\Sigma}V^*x = V\hat{\Sigma}^*\hat{U}^*b$, και συνεπάγεται ότι $\hat{\Sigma}V^*x = \hat{U}^*b$.
Η κύρια παράμετρος για τον υπολογισμό της πολυπλοκότητας του αλγορίθμου 2.8 είναι ο υπολογισμός της SVD. Ένας τυπικός υπολογισμός είναι

$$\text{Πολυπλοκότητα αλγορίθμου 2.8: } \sim 2mn^2 + 11n^3 \text{ βήματα} \qquad (2.62)$$

## *2.5.7 ΣΥΓΚΡΙΣΗ ΤΩΝ ΑΛΓΟΡΙΘΜΩΝ*

Κάθε μια από τις μεθόδους που περιγράψαμε παραπάνω ξεχωρίζει σε συγκεκριμένες περιπτώσεις. Όταν αυτό που μας ενδιαφέρει είναι η ταχύτητα, ο αλγόριθμος ελαχίστων τετραγώνων μέσω κανονικών εξισώσεων (2.6) είναι ο καλύτερος. Όμως, κατά την επίλυση κανονικών εξισώσεων ο αλγόριθμος αυτός δεν είναι πάντα ευσταθής λόγω σφαλμάτων μηχανής και γι' αυτό από πολλά χρόνια οι επιστήμονες προτιμούσαν τον αλγόριθμο ελαχίστων τετραγώνων μέσω QR παραγοντοποίησης (2.7) αντί για τη βασική μέθοδο των ελαχίστων τετραγώνων. Ο αλγόριθμος αυτός είναι πράγματι ένας καλός αλγόριθμος και προτείνεται για συχνή χρήση. Αν ο πίνακας $A$ είναι χαμηλής τάξης, τότε ο αλγόριθμος 2.7 δεν είναι ιδιαίτερα σταθερός και σε αυτές τις περιπτώσεις είναι καλύτερο να χρησιμοποιούμε τον αλγόριθμο ελαχίστων τετραγώνων μέσω SVD (2.8), ο οποίος βασίζεται στην SVD.

# ΚΕΦΑΛΑΙΟ 3
# ΚΑΤΑΣΤΑΣΗ ΚΑΙ ΕΥΣΤΑΘΕΙΑ

## 3.1 ΚΑΤΑΣΤΑΣΗ ΚΑΙ ΔΕΙΚΤΕΣ ΚΑΤΑΣΤΑΣΗΣ

Στο τρίτο κεφάλαιο της εργασίας θα εξετάσουμε πιο αναλυτικά κάποια πεδία της αριθμητικής ανάλυσης των προβλημάτων που παρουσιάστηκαν στο προηγούμενο κεφάλαιο. Συγκεκριμένα θα μελετήσουμε τις έννοιες της κατάστασης και ευστάθειας που αφορούν τις διαταραχές στη συμπεριφορά ενός μαθηματικού προβλήματος και αλγορίθμου αντίστοιχα.

### *3.1.1 ΚΑΤΑΣΤΑΣΗ ΕΝΟΣ ΠΡΟΒΛΗΜΑΤΟΣ*

Γενικά μπορούμε να δούμε ένα πρόβλημα σαν μια συνάρτηση $f : X \to Y$ από ένα χώρο δεδομένο με νόρμα $X$ σε ένα χώρο διανυσμάτων με νόρμα $Y$ ο οποίος είναι ο χώρος λύσεων. Η συνάρτηση αυτή είναι συνήθως μη γραμμική (ακόμα και στη γραμμική άλγεβρα) αλλά τις περισσότερες φορές είναι τουλάχιστον συνεχής.
Τυπικά, μας απασχολεί η συμπεριφορά ενός προβλήματος $f$ σε συγκεκριμένα σημεία $x \in X$ (η συμπεριφορά αυτή μπορεί να διαφέρει πολύ από σημείο σε σημείο). Ο συνδυασμός ενός προβλήματος $f$ με προδιαγραμμένα σημεία $x$ μπορεί να ονομαστεί καλά ορισμένο πρόβλημα αλλά τις πιο πολλές φορές χρησιμοποιούμε τον όρο πρόβλημα και σε αυτή την περίπτωση.
Ένα πρόβλημα καλής κατάστασης είναι ένα πρόβλημα του οποίου όλες οι διαταραχές των $x$ έχουν σαν συνέπεια μικρές αλλαγές στην $f(x)$. Ένα μη καλά ορισμένο πρόβλημα είναι ένα πρόβλημα στο οποίο μικρές διαταραχές στα σημεία $x$ έχουν σαν συνέπεια μεγάλες αλλαγές στην $f(x)$.
Οι όροι «μικρές» και «μεγάλες» στους παραπάνω ορισμούς εξαρτώνται από την εφαρμογή. Πιο συγκεκριμένα, κάποιες φορές είναι πιο σωστό να μετράμε διαταραχές σε απόλυτη κλίμακα και κάποιες άλλες είναι πιο σωστό να τις μετράμε σε σχέση με τη νόρμα του αντικείμενου το οποίο διαταράσσεται.

### *3.1.2 ΑΠΟΛΥΤΟΣ ΔΕΙΚΤΗΣ ΚΑΤΑΣΤΑΣΗΣ*

Έστω $\delta x$ να είναι μια μικρή διαταραχή του $x$ και γράφουμε $\delta f = f(x+\delta x) - f(x)$.
Ο απόλυτος δείκτης κατάστασης του προβλήματος $f$ στο $x$ είναι $\hat{\kappa} = \hat{\kappa}(x)$ και ορίζεται ως

$$\hat{\kappa} = \lim_{\delta \to 0} \sup_{\|\delta x\| \le \delta} \frac{\|\delta f\|}{\|\delta x\|} \qquad (3.1)$$

Για τα περισσότερα προβλήματα, το όριο του άνω φράγματος στον παραπάνω τύπο μπορεί να ερμηνευτεί ως ένα άνω φράγμα όλων των απειροελάχιστων διαταραχών του $\delta x$ και γράφουμε τον παραπάνω τύπο ως

$$\hat{\kappa} = \sup_{\delta x} \frac{\|\delta f\|}{\|\delta x\|} \tag{3.2}$$

με την προϋπόθεση ότι τα $\delta x$ και $\delta f$ είναι απειροελάχιστα.
Αν η $f$ είναι παραγωγίσιμη, μπορούμε να εκτιμήσουμε τον δείκτη κατάστασης σε σχέση με την παράγωγο της $f$. Έστω $J(x)$ να είναι ένας πίνακας του οποίου το $i, j$ στοιχείο να είναι η μερική παράγωγος $\partial f_i / \partial x_j$ υπολογισμένη στο $x$, γνωστή και ως Ιακωβιανή της $f$ στο $x$. Ο ορισμός της παραγώγου μας δίνει $\delta f \approx J(x)\delta x$ με ισότητα όταν το όριο $\|\delta x\| \to 0$. Ο απόλυτος δείκτης κατάστασης γίνεται:

$$\hat{\kappa} = \|J(x)\| \tag{3.3}$$

όπου η ποσότητα $\|J(x)\|$ αντιπροσωπεύει τη νόρμα του $J(x)$ που παράγεται από τις νόρμες των χώρων $X$ και $Y$.

*3.1.3 ΣΧΕΤΙΚΟΣ ΔΕΙΚΤΗΣ ΚΑΤΑΣΤΑΣΗΣ*

Όταν μας ενδιαφέρουν σχετικές αλλαγές τότε εξετάζουμε το σχετικό δείκτη κατάστασης. Ο σχετικός δείκτης κατάστασης $\kappa = \kappa(x)$ ορίζεται ως

$$\kappa = \lim_{\delta \to 0} \sup_{\|\delta x\| \le \delta} \left( \frac{\|\delta f\|}{\|f(x)\|} \Big/ \frac{\|\delta x\|}{\|x\|} \right) \tag{3.4}$$

ή αν υποθέσουμε ότι τα $\delta x$ και $\delta f$ είναι απειροελάχιστα έχουμε

$$\kappa = \sup_{\delta x} \left( \frac{\|\delta f\|}{\|f(x)\|} \Big/ \frac{\|\delta x\|}{\|x\|} \right). \tag{3.5}$$

Αν η $f$ είναι παραγωγίσιμη τότε μπορούμε να εκφράσουμε την ποσότητα αυτή με βάση την Ιακωβιανή

$$\kappa = \frac{\|J(x)\|}{\|f(x)\| / \|x\|} \tag{3.6}$$

Ο σχετικός και απόλυτος δείκτης κατάστασης χρησιμοποιούνται σε διάφορες περιπτώσεις αλλά ο σχετικός δείκτης κατάστασης είναι πιο σημαντικός στην αριθμητική ανάλυση. Αυτό είναι θεμελιώδες επειδή στην αριθμητική κινητής υποδιαστολής που χρησιμοποιείται προκύπτουν σχετικά σφάλματα και όχι απόλυτα. Ένα πρόβλημα είναι καλά ορισμένο αν ο $\kappa$ είναι μικρός (π.χ. $10^6, 10^{16}$).

*3.1.4 ΣΥΝΘΗΚΗ ΠΟΛΛΑΠΛΑΣΙΑΣΜΟΥ ΠΙΝΑΚΑ-ΔΙΑΝΥΣΜΑ*

Έστω $A \in C^{m\times n}$ και έστω ότι έχουμε το πρόβλημα υπολογισμού της ποσότητας $Ax$ με δεδομένο το $x$, δηλαδή θέλουμε να βρούμε ένα δείκτη κατάστασης με βάση τις διαταραχές του $x$ αλλά όχι του $A$. Δουλεύοντας με τον ορισμό του $\kappa$, με τη $\|\cdot\|$ να δηλώνει μια τυχαία νόρμα διανύσματος και να αντιστοιχεί στην παραγόμενη νόρμα πίνακα, βρίσκουμε ότι

$$\kappa = \sup_{\delta x}\left(\frac{\|A(x+\delta x)-Ax\|}{\|Ax\|} \Big/ \frac{\|\delta x\|}{\|x\|}\right) = \sup_{\delta x} \frac{\|A\delta x\|}{\|\delta x\|} \Big/ \frac{\|Ax\|}{\|x\|},$$

δηλαδή

$$\kappa = \|A\| \frac{\|x\|}{\|Ax\|} \tag{3.7}$$

Αυτός είναι ένα ακριβής τύπος για το δείκτη κατάστασης $\kappa$, ο οποίος εξαρτάται και από τον $A$ και από το $x$.
Έστω ότι στον παραπάνω υπολογισμό ο πίνακας $A$ είναι τετραγωνικός και ομαλός. Τότε μπορούμε να χρησιμοποιήσουμε το γεγονός ότι $\|x\|/\|Ax\| \le \|A^{-1}\|$ βρίσκουμε για την (3.7) ένα όριο για το $\kappa$ ανεξάρτητο από το $x$:

$$\kappa \le \|A\|\|A^{-1}\| \quad , \tag{3.8}$$

ή αλλιώς μπορούμε να γράψουμε

$$\kappa = \alpha \|A\|\|A^{-1}\| \tag{3.9}$$

με

$$\alpha = \frac{\|x\|}{\|Ax\|} \Big/ \|A^{-1}\| \tag{3.10}$$

Για συγκεκριμένες τιμές του $x$ είναι $a=1$ και κατά συνέπεια $\kappa=\|A\|\|A^{-1}\|$. Αν $\|\cdot\|=\|\cdot\|_2$ τότε η παραπάνω σχέση θα ισχύει όποτε το $x$ είναι πολλαπλάσιο ενός ελάχιστου ιδιάζοντος διανύσματος του πίνακα $A$.

Ο πίνακας $A$ δε χρειάζεται να είναι τετραγωνικός. Αν ο $A\in C^{m\times n}$ με $m\geq n$ είναι πλήρους τάξης, οι ισότητες (3.8)-(3.10) ισχύουν αντικαθιστώντας τον $A^{-1}$ με τον ψευδοαντίστροφο του $A^{+}$, όπως αυτός ορίστηκε στην (11.11)

Στο αντίστροφο πρόβλημα , δηλαδή δοθέντος του πίνακα $A$ και του $b$ μπορούμε να υπολογίσουμε το $A^{-1}b$; Από μαθηματικής άποψης, το πρόβλημα αυτό είναι ίδιο με το πρόβλημα που συζητήσαμε στην παράγραφο αυτή αλλά ο πίνακας $A$ έχει αντικατασταθεί από τον $A^{-1}$. Κατά συνέπεια έχουμε αποδείξει το παρακάτω θεώρημα.

ΘΕΩΡΗΜΑ 3.1

Έστω $A\in C^{m\times m}$ να είναι ομαλός και έστω ότι έχουμε την εξίσωση $Ax=b$. Το πρόβλημα υπολογισμού του $b$ δοθέντος του $x$ έχει δείκτη κατάστασης

$$\kappa=\|A\|\frac{\|x\|}{\|b\|}\leq\|A\|\|A^{-1}\| \tag{3.11}$$

ως προς τις διαταραχές του $x$. Το πρόβλημα υπολογισμού του $x$ δοθέντος του $b$ έχει δείκτη κατάστασης

$$\kappa=\|A^{-1}\|\frac{\|b\|}{\|x\|}\leq\|A\|\|A^{-1}\| \tag{3.12}$$

ως προς τις διαταραχές του $b$. Αν $\|\cdot\|=\|\cdot\|_2$ τότε ισχύει ισότητα στην (3.11) αν το $x$ είναι πολλαπλάσιο ενός δεξιού ιδιάζοντος διανύσματος του πίνακα $A$ που αντιστοιχεί σε μια ελάχιστη ιδιάζουσα τιμή $\sigma_m$ και ισχύει ισότητα στην (3.12) αν ο $b$ είναι πολλαπλάσιο ενός αριστερού ιδιάζοντος διανύσματος του πίνακα $A$ που αντιστοιχεί στη μέγιστη ιδιάζουσα τιμή $\sigma_1$.

### *3.1.5 ΑΡΙΘΜΟΣ ΚΑΤΑΣΤΑΣΗΣ ΕΝΟΣ ΠΙΝΑΚΑ*

Το γινόμενο $\|A\|\|A^{-1}\|$ ονομάζεται δείκτης κατάστασης του $A$ και συμβολίζεται με $\kappa(A)$:

$$\kappa(A)=\|A\|\|A^{-1}\| \tag{3.13}$$

Οπότε σε αυτή την περίπτωση ο όρος «δείκτης κατάστασης» σχετίζεται με πίνακα και όχι με πρόβλημα. Αν ο $\kappa(A)$ είναι μικρός τότε ο πίνακας $A$ είναι καλά ορισμένος. Αν ο $\kappa(A)$ είναι μεγάλος τότε ο πίνακας $A$ είναι ασθενώς ορισμένος. Αν ο πίνακας $A$ είναι μη ομαλός τότε $\kappa(A)=\infty$.

Ας σημειώσουμε ότι αν $\|\cdot\|=\|\cdot\|_2$ τότε $\|A\|=\sigma_1$ και $\|A^{-1}\|=1/\sigma_m$. Οπότε

$$\kappa(A)=\frac{\sigma_1}{\sigma_m} \tag{3.14}$$

Ισχύει για τη νόρμα δεύτερης τάξης. Ο λόγος $\frac{\sigma_1}{\sigma_m}$ μπορεί να ερμηνευτεί ως η εκκεντρότητα την υπερ-έλλειψης που είναι η εικόνα την μοναδιαίας σφαίρας του $C^m$ κάτω από τον $A$.

Για ένα ορθογώνιο πίνακα $A\in C^{m\times n}$ πλήρους τάξης, $m\geq n$ ο δείκτης κατάστασης ορίζεται σε σχέση με τον ψευδοαντίστροφο πίνακα: $\kappa(A)=\|A\|\|A^{+}\|$. Στην περίπτωση $\|\cdot\|=\|\cdot\|_2$ έχουμε:

$$\kappa(A)=\frac{\sigma_1}{\sigma_n} \tag{3.15}$$

### *3.1.6 ΚΑΤΑΣΤΑΣΗ ΕΝΟΣ ΣΥΣΤΗΜΑΤΟΣ ΕΞΙΣΩΣΕΩΝ*

Στο θεώρημα 3.1 κρατάγαμε τον $A$ σταθερό και διαταράσσαμε το $x$ ή το $b$. Τι γίνεται αν διαταράξουμε τον $A$; Ειδικότερα ας κρατήσουμε το $b$ σταθερό και ας δούμε τη συμπεριφορά του προβλήματος $A\mapsto x=A^{-1}b$ όταν ο $A$ διαταράσσεται απειροελάχιστα κατά $\delta A$. Τότε το $x$ διαταράσσεται κατά $\delta x$, όπου

$$(A+\delta A)(x+\delta x)=b$$

Χρησιμοποιώντας την ισότητα $Ax=b$ και παραλείποντας τον όρο $(\delta A)(\delta x)$ παίρνουμε $(\delta A)x+A(\delta x)=0$, δηλαδή $\delta x=-A^{-1}(\delta A)x$. Από την εξίσωση αυτή προκύπτει ότι $\|\delta x\|\leq\|A^{-1}\|\|\delta A\|\|x\|$ ή ισοδύναμα

$$\frac{\|\delta x\|}{\|x\|}\Big/\frac{\|\delta A\|}{\|A\|}\leq\|A^{-1}\|\|A\|=\kappa(A)$$

Η ισότητα σε αυτή τη σχέση ισχύει αν η $\delta A$ είναι τέτοια ώστε

$$\left\| A^{-1}(\delta A)x \right\| = \left\| A^{-1} \right\| \left\| \delta A \right\| \left\| x \right\|$$

και μπορεί να δειχτεί ότι για κάθε $A$ και $b$ και νόρμα $\|\cdot\|$ τέτοια διαταραχή $\delta A$ ισχύει. Αυτό μας οδηγεί στο παρακάτω θεώρημα.

ΘΕΩΡΗΜΑ 3.2
Έστω $b$ να είναι σταθερό και έστω το πρόβλημα υπολογισμού $x = A^{-1}b$, όπου ο $A$ είναι τετραγωνικός και ομαλός. Ο δείκτης κατάστασης αυτού του προβλήματος ως προς τις διαταραχές του $A$ είναι

$$\kappa = \left\| A \right\| \left\| A^{-1} \right\| = \kappa(A) \tag{3.16}$$

Τα θεωρήματα 3.1 και 3.2 είναι θεμελιώδους σημαντικότητας για την αριθμητική γραμμική άλγεβρα επειδή καθορίζουν την ακρίβεια λύσης ενός συστήματος εξισώσεων. Αν ένα πρόβλημα $Ax = b$ περιέχει έναν ασθενώς ορισμένο πίνακα $A$ τότε πρέπει πάντα να περιμένουμε ότι θα έχουμε απώλεια $\log_{10} \kappa(A)$ στοιχείων κατά τον υπολογισμό της λύσης, με εξαίρεση κάποιων πολύ ειδικών περιπτώσεων.

## 3.2 ΑΡΙΘΜΗΤΙΚΗ ΚΙΝΗΤΗΣ ΥΠΟΔΙΑΣΤΟΛΗΣ

Με την εφεύρεση των υπολογιστών προέκυψε και το θέμα ότι έπρεπε να παρουσιαστούν οι πραγματικοί αριθμοί σε ψηφιακή μορφή. Το μυστικό για αυτό είναι η αριθμητική της κινητής υποδιαστολής, το εργαλείο για τον ψηφιακό συμβολισμό.

### *3.2.1 ΠΕΡΙΟΡΙΣΜΟΙ ΣΤΗΝ ΨΗΦΙΑΚΗ ΑΝΑΠΑΡΑΣΤΑΣΗ*

Από τη στιγμή που οι ψηφιακοί υπολογιστές χρησιμοποιούν πεπερασμένο πλήθος bits για να αναπαραστήσουν ένα πραγματικό αριθμό, μπορούν να αναπαραστήσουν μόνο ένα πεπερασμένο υποσύνολο των πραγματικών αριθμών (ή και των μιγαδικών). Αυτός ο περιορισμός παρουσιάζει δυο δυσκολίες. Πρώτον, οι αριθμοί που αναπαριστάνονται δεν μπορούν να είναι αυθαίρετα μεγάλοι ή μικροί. Δεύτερον, πρέπει να υπάρχουν κενά μεταξύ τους.
Οι υπολογιστές τελευταίας τεχνολογίας αναπαριστάνουν αριθμούς αρκούντως μεγάλους ή μικρούς έτσι ώστε να μην προκύπτουν δυσκολίες από τον πρώτο

περιορισμό. Για παράδειγμα, αν η ευρέως χρησιμοποιούμενη IEEE αριθμητική διπλής ακρίβειας επιτρέπει αριθμούς τόσο μεγάλους όσο $1.79\times10^{308}$ και τόσο μικρούς όσο $2.23\times10^{-308}$. Με άλλα λόγια, η υπερχείλιση και η υποχείλιση δεν αποτελούν συνήθως κίνδυνο.
Σε αντίθεση, το πρόβλημα των κενών ανάμεσα στους αναπαριστάμενους αριθμούς αποτελεί σημείο άξιο προσοχής κατά τον επιστημονικό υπολογισμό. Για παράδειγμα, στην IEEE αριθμητική διπλής ακρίβειας, το διάστημα $[1,2]$ μπορεί να αναπαρασταθεί από το διακριτό υποδιάστημα

$$1,\quad 1+2^{-52},\quad 1+2\times2^{-52},\quad 1+3\times2^{-52},\quad \ldots\ldots,\quad 2 \qquad (3.17)$$

Το διάστημα [2,4] μπορεί να αναπαρασταθεί από τους ίδιους αριθμούς πολλαπλασιαζόμενους επί 2

$$2,\quad 2+2^{-51},\quad 2+2\times2^{-51},\quad 2+3\times2^{-51},\quad \ldots..,\quad 4$$

και γενικά τα διαστήματα $[2^j,2^{j+1}]$ αναπαριστάνονται από την (3.17) πολλαπλασιασμένη επί $2^j$. Οπότε στην IEEE αριθμητική διπλής ακρίβειας τα κενά μεταξύ των γειτονικών αριθμών δεν είναι ποτέ μεγαλύτερα από $2^{-52}\approx2.22\times10^{-16}$.

### *3.2.2 ΑΡΙΘΜΟΙ ΚΙΝΗΤΗΣ ΥΠΟΔΙΑΣΤΟΛΗΣ*

Η IEEE αριθμητική είναι ένα παράδειγμα ενός αριθμητικού συστήματος το οποίο βασίζεται στην παρουσίαση με κινητή υποδιαστολή των πραγματικών αριθμών. Αυτή η μέθοδος χρησιμοποιείται γενικά από τους υπολογιστές για διάφορους σκοπούς. Σε ένα σύστημα αριθμών κινητής υποδιαστολής η θέση του δεκαδικού στοιχείου αποθηκεύεται ξεχωριστά από τα ψηφία και τα κενά μεταξύ των γειτονικών αριθμών ποικίλουν ανάλογα με το μέγεθος των αριθμών. Αυτό είναι διαφορετικό από μια αναπαράσταση σταθερού στοιχείου όπου τα κενά έχουν όλα το ίδιο μέγεθος.
Ειδικότερα, ας μελετήσουμε ένα ιδανικό σύστημα αριθμών κινητής υποδιαστολής όπως ακολουθεί. Το σύστημα αποτελείται από ένα διακριτό υποσύνολο $F$ των πραγματικών αριθμών $R$ καθορισμένο από έναν ακέραιο $\beta\geq2$ γνωστό και ως βάση ή βάση λογαρίθμου και ένα ακέραιο $t\geq1$ γνωστό ως ακρίβεια. Τα στοιχεία του $F$ είναι το νούμερο μηδέν μαζί με όλα τα νούμερα της μορφής

$$x=\pm(m/\beta^t)\beta^e \qquad (3.18)$$

όπου το $m$ είναι ένας ακέραιος στο διάστημα $1 \le m \le \beta^t$ και το $e$ είναι ένας τυχαίος ακέραιος. Ισοδύναμα, μπορούμε να περιορίσουμε το διάστημα στο $\beta^{t-1} \le m \le \beta^t - 1$ και οπότε να κάνουμε μοναδική την επιλογή του $m$. Η ποσότητα $\pm(m/\beta^t)$ είναι τότε γνωστή ως το κλάσμα ή κλασματικό μέρος του $x$ και το $e$ είναι ο εκθέτης.

Το παραπάνω σύστημα αριθμών κινητής υποδιαστολής είναι ιδανικό στο ότι αγνοεί την υπερχείλιση και την υποχείλιση. Το αποτέλεσμα είναι το $F$ να είναι ένα αριθμήσιμο πεπερασμένο σύνολο και να είναι και αυτοόμοιο: $F = \beta F$.

### *3.2.3 ΕΨΙΛΟΝ ΜΗΧΑΝΗΣ*

Η ευκρίνεια του $F$ περιγράφεται από ένα νούμερο γνωστό ως έψιλον μηχανής. Ας ορίσουμε το νούμερο αυτό ως

$$\varepsilon_{\mu\eta\chi\alpha\nu\acute{\eta}\varsigma} = \frac{1}{2}\beta^{1-t} \tag{3.19}$$

Το νούμερο αυτό είναι η μισή απόσταση ανάμεσα στο 1 και τον επόμενο μεγάλο αριθμό κινητής υποδιαστολής. Δηλαδή το $\varepsilon_{machine}$ έχει την παρακάτω ιδιότητα

$$\text{Για όλα τα } x \in R \text{ υπάρχει } x' \in F \text{ τέτοιο ώστε } |x - x'| \le \varepsilon_{\mu\eta\chi\alpha\nu\acute{\eta}\varsigma} |x| \tag{3.20}$$

Για τις τιμές του $\beta$ και του $t$ σε διάφορους υπολογιστές συνήθως το $\varepsilon_{\mu\eta\chi\alpha\nu\acute{\eta}\varsigma}$ είναι ανάμεσα στα $10^{-6}$ και $10^{-35}$. Στην IEEE αριθμητική μονής και διπλής ακρίβειας το $\varepsilon_{\mu\eta\chi\alpha\nu\acute{\eta}\varsigma}$ είναι $2^{-24} \approx 5.96 \times 10^{-8}$ και $2^{-53} \approx 1.11 \times 10^{-16}$ αντίστοιχα.

Έστω $Fl : R \to F$ να είναι μια συνάρτηση η οποία δίνει την κοντινότερη προσέγγιση ενός αριθμού κινητής υποδιαστολής σε ένα πραγματικό αριθμό, τη στρογγυλοποίηση του αριθμού αυτού. Η ανισότητα (3.20) μπορεί να γραφεί με βάση τη συνάρτηση $FL$:

Για όλα τα $x \in R$ υπάρχει $\varepsilon$ με $|\varepsilon| \le \varepsilon_{\mu\eta\chi\alpha\nu\acute{\eta}\varsigma}$ τέτοιο ώστε $Fl(x) = x(1+\varepsilon)$ (3.21).

Δηλαδή, η διαφορά ανάμεσα σε έναν πραγματικό αριθμό και την κοντινότερή του προσέγγιση κινητής υποδιαστολής είναι πάντα μικρότερη από το $\varepsilon_{\mu\eta\chi\alpha\nu\acute{\eta}\varsigma}$.

### *3.2.4 ΑΡΙΘΜΗΤΙΚΗ ΚΙΝΗΤΗΣ ΥΠΟΔΙΑΣΤΟΛΗΣ*

Δεν είναι αρκετό μόνο να αναπαριστούμε πραγματικούς αριθμούς-πρέπει να τους υπολογίζουμε κιόλας. Σε έναν υπολογιστή όλοι οι μαθηματικοί υπολογισμοί μειώνονται στις βασικές μαθηματικές πράξεις, όπου το κλασσικό σύνολο είναι το

$+,-,\times,\div$ . Αυτές είναι οι πράξεις στον $R$ . Στον $F$ οι πράξεις αυτές συμβολίζονται με $\oplus,\otimes$

Ένας υπολογιστής είναι πιθανό να λειτουργεί με βάση την παρακάτω αρχή. Έστω $x$ και $y$ να είναι τυχαίοι αριθμοί κινητής υποδιαστολής, δηλαδή $x, y \in F$ . Έστω $\cdot$ να είναι μια από τις πράξεις-πρόσθεση, αφαίρεση, πολλαπλασιασμός ή διαίρεση-και έστω $\cdot$ να είναι το ανάλογό του κινητής υποδιαστολής. Τότε το $x \cdot y$ πρέπει να είναι ακριβώς ίσο με

$$x \cdot y = Fl(x \cdot y) \tag{3.22}$$

Αν ισχύει αυτή η ιδιότητα τότε από τις σχέσεις (3.21) και (3.22) συμπεραίνουμε ότι ο υπολογιστής έχει μια απλή και ισχυρή ιδιότητα.

**Θεμελιώδες Αξίωμα της Αριθμητικής Κινητής Υποδιαστολής**

Για όλα τα $x, y \in F$ υπάρχει $\varepsilon$ με $|\varepsilon| \le \varepsilon_{μηχανής}$ τέτοιο ώστε

$$x \cdot y = (x \cdot y)(1+\varepsilon) \tag{3.23}$$

Δηλαδή κάθε συνάρτηση κινητής υποδιαστολής παρουσιάζει σχετικό σφάλμα το πολύ ίσο με $\varepsilon_{μηχανής}$ .

### *3.2.5 ΕΨΙΛΟΝ ΜΗΧΑΝΗΣ,ΠΑΛΙ*

Στην εργασία αυτή, η ανάλυση των σφαλμάτων που προκύπτουν από στρογγυλοποίηση βασίζεται στις (3.21) και (3.27) και όχι σε άλλες λεπτομέρειες της αριθμητικής κινητής υποδιαστολής. Αυτό σημαίνει ότι δεν μας πειράζει αν οι υπολογισμοί κινητής υποδιαστολής δεν γίνονται με τόση ακρίβεια όση προσδιορίζεται από τη σχέση (3.22). Σε αυτή την περίπτωση, οι (3.21) και (3.23) μπορούν να ικανοποιούνται ακόμα και αν το $\varepsilon_{μηχανής}$ αντικατασταθεί από μια μεγαλύτερη τιμή. Για παράδειγμα, αν σε έναν υπολογιστή οι μέσες ποσότητες αποκόβονται και δεν στρογγυλοποιούνται η σχέση (3.23) ισχύει αν το $\varepsilon_{μηχανής}$ αντικατασταθεί από το $2\,\varepsilon_{μηχανής}$ .

Ο πιο απλός τρόπος για να επιτραπούν τέτοιες παραλλαγές είναι να διατηρήσουμε τις (3.21) και (3.23) αλλά να τροποποιήσουμε τον ορισμό του $\varepsilon_{μηχανής}$ . Από εδώ και πέρα ας υποθέσουμε ότι το $\varepsilon_{μηχανής}$ δεν ορίζεται από την (3.19) αλλά από ένα μικρότερο νούμερο για το οποίο ισχύουν οι (3.21) και (3.23). Στους περισσότερους υπολογιστές όλες αυτές οι αλλαγές στην IEEE αριθμητική δεν έχουν κάποια σημαντική αλλαγή στην τιμή του $\varepsilon_{μηχανής}$ .

Κάποιες φορές μια αναπάντεχα μεγάλη τιμή του $\varepsilon_{μηχανής}$ μπορεί να είναι απαραίτητη για να ισχύει η (3.23). Για την ακρίβεια, υπάρχουν μηχανές στις οποίες η (3.23) ισχύει για $\varepsilon_{μηχανής} = 1$.

Ευτυχώς, τα πλεονεκτήματα του αξιώματος (3.23) και η υιοθέτηση ομαλών βάσεων στην αριθμητική των υπολογιστών έχουν εξαλείψει τους υπολογιστές που αποτυγχάνουν να ικανοποιήσουν την (3.23) για μικρή τιμή του $\varepsilon_{μηχανής}$.

### *3.2.6 ΜΙΓΑΔΙΚΟΙ ΑΡΙΘΜΟΙ ΚΙΝΗΤΗΣ ΥΠΟΔΙΑΣΤΟΛΗΣ*

Οι μιγαδικοί αριθμοί κινητής υποδιαστολής παρουσιάζονται σαν ζευγάρια πραγματικών αριθμών κινητής υποδιαστολής και οι πράξεις γίνονται χωριστά στο πραγματικό και το φανταστικό κομμάτι. Το αποτέλεσμα είναι ότι το αξίωμα (3.23) ισχύει τόσο για μιγαδικούς όσο και για πραγματικούς αριθμούς κινητής υποδιαστολής εκτός από τις πράξεις της διαίρεσης και του πολλαπλασιασμού όπου το $\varepsilon_{μηχανής}$ στην (3.19) πρέπει να αυξηθεί κατά τους παράγοντες $2^{5/2}$ και $2^{3/2}$ αντίστοιχα. Αν το $\varepsilon_{μηχανής}$ προσαρμοστεί σε αυτά τα δεδομένα τότε η ανάλυση για τα σφάλματα που προκύπτουν από στρογγυλοποίηση είναι ίδια που κάναμε και στους πραγματικούς αριθμούς.

## 3.3 ΕΥΣΤΑΘΕΙΑ

Θα ήταν θετικό αν οι αριθμητικοί αλγόριθμοι μπορούσαν να παράγουν ακριβείς λύσεις στα αριθμητικά προβλήματα. Αλλά τα προβλήματα είναι συνεχή ενώ οι αλγόριθμοι είναι διακριτοί οπότε το παραπάνω δεν είναι εφικτό. Η χρήση του όρου της ευστάθειας είναι ο βασικός τρόπος για να χαρακτηρίζουμε τι είναι δυνατό, δηλαδή τι σημαίνει να παίρνουμε τη «σωστή απάντηση» ακόμα και αν αυτή δεν είναι ακριβής.

### *3.3.1 ΑΛΓΟΡΙΘΜΟΙ*

Σε προηγούμενη παράγραφο ορίσαμε ένα μαθηματικό πρόβλημα ως μια συνάρτηση $f : X \to Y$ από ένα χώρο διανυσμάτων $X$ σε ένα χώρο διανυσμάτων λύσεων $Y$.

Ένας αλγόριθμος μπορεί να θεωρηθεί σαν μια απεικόνιση $\overline{f} : X \to Y$ ανάμεσα σε δυο χώρους όπως τους περιγράψαμε παραπάνω. Έστω ότι έχουμε ένα πρόβλημα $f$ και έναν υπολογιστή του οποίου το σύστημα κινητής υποδιαστολής ικανοποιεί την (3.23) (αλλά όχι απαραίτητα και την (3.22)), έναν αλγόριθμο για την $f$ και μια εφαρμογή του αλγορίθμου αυτού. Έστω $x \in X$ να είναι τα στρογγυλοποιημένα

δεδομένα είναι στρογγυλοποιημένα που ικανοποιούν την (3.21) και έστω ότι αποτελούν τα δεδομένα εισαγωγής ενός αλγόριθμου. Τρέχουμε τον αλγόριθμο. Το αποτέλεσμα είναι μια συλλογή αριθμών κινητής υποδιαστολής οι οποίοι ανήκουν στο χώρο διανυσμάτων $Y$ (αφού ο αλγόριθμος σχεδιάστηκε για να λύνει την $f$). Έστω ότι αυτό το υπολογίσιμο αποτέλεσμα το ονομάζουμε $\overline{f}(x)$.

Το $\overline{f}(x)$ θα επηρεαστεί τουλάχιστον από τα σφάλματα στρογγυλοποίησης. Ανάλογα με τις συνθήκες, μπορεί να επηρεαστεί από όλων των ειδών επιπλοκές, όπως ανοχές σύγκλισης ή και από άλλα προγράμματα που τρέχουν στον υπολογιστή. Οπότε η «συνάρτηση» $\overline{f}(x)$ μπορεί να παίρνει διάφορες τιμές διαφορετικές από αυτές που έχουμε προβλέψει. Παρόλο που εμφανίζονται αυτές οι δυσκολίες μπορούμε να κάνουμε σταθερές υποθέσεις για την $\overline{f}(x)$ και για την ακρίβεια των αλγορίθμων της αριθμητικής γραμμικής άλγεβρας, οι οποίες βασίζονται στα αξιώματα (3.21) και (3.23).

Η $\overline{f}$ συμβολίζει το υπολογισμένο ανάλογο της $f$ και όλες οι υπολογισμένες ποσότητες θα συμβολίζονται με αυτόν τον τρόπο. Για παράδειγμα, η υπολογισμένη λύση του συστήματος εξισώσεων $Ax = b$ θα συμβολίζεται με $\overline{x}$.

### *3.3.2 ΑΚΡΙΒΕΙΑ*

Εκτός από τετριμμένες περιπτώσεις, η $\overline{f}$ δεν μπορεί να είναι συνεχής. Παρόλα αυτά ένας καλός αλγόριθμος προσεγγίζει το συσχετισμένο πρόβλημα $f$. Για να κάνουμε αυτή την ιδέα ποσοτική μπορούμε να σκεφτούμε το απόλυτο σφάλμα ενός υπολογισμού, $\left\| \overline{f}(x) - f(x) \right\|$ ή το σχετικό σφάλμα

$$\frac{\left\| \overline{f}(x) - f(x) \right\|}{\left\| f(x) \right\|} \tag{3.24}$$

Αν ο αλγόριθμος $\overline{f}$ είναι καλός, τότε περιμένουμε ότι το σχετικό σφάλμα να είναι μικρό, της τάξης του $\varepsilon_{μηχανής}$. Μπορούμε να πούμε ότι ένας αλγόριθμος $\tilde{f}$ για ένα πρόβλημα $f$ είναι ακριβής για κάθε $x \in X$,

$$\frac{\left\| \overline{f}(x) - f(x) \right\|}{\left\| f(x) \right\|} = O_{\varepsilon_{μηχανής}} \tag{3.25}$$

*3.3.3 ΕΥΣΤΑΘΕΙΑ*

Αν το πρόβλημα $f$ είναι ασθενώς ορισμένο δεν μπορούμε να επιτύχουμε την ακρίβεια όπως αυτή περιγράφεται στην (3.25). Η στρογγυλοποίηση των δεδομένων του αλγορίθμου είναι αναπόφευκτη σε ένα ψηφιακό υπολογιστή και ακόμα και αν οι υπολογισμοί γίνουν τέλεια, η στρογγυλοποίηση θα οδηγήσει σε αξιοπρόσεκτη διαφορά στο αποτέλεσμα. Αντί να στοχεύουμε στην ακρίβεια σε όλες τις περιπτώσεις, είναι πιο βολικό να έχουμε σαν στόχο την ευστάθεια. Λέμε ότι ένας αλγόριθμος $\overline{f}$ για ένα πρόβλημα $f$ είναι ευσταθής αν για κάθε $x \in X$ ,

$$\frac{\left\| \overline{f}(x) - f(\overline{x}) \right\|}{\left\| f(\overline{x}) \right\|} = O(\varepsilon_{\mu\eta\chi\alpha\nu\acute{\eta}\varsigma}) \tag{3.26}$$

για κάποια $\overline{x}$ με

$$\frac{\left\| \overline{x} - x \right\|}{\left\| x \right\|} = O(\varepsilon_{\mu\eta\chi\alpha\nu\acute{\eta}\varsigma}) \tag{3.27}$$

Με λόγια:

Ένας ευσταθής αλγόριθμος δίνει λύση κοντά στη σωστή
σε ένα πρόβλημα κοντά στο σωστό.

*3.3.4 ΠΡΟΣ ΤΑ ΠΙΣΩ ΕΥΣΤΑΘΕΙΑ*

Πολλοί αλγόριθμοι της αριθμητικής γραμμικής άλγεβρας ικανοποιούν μια υπόθεση η οποία είναι πιο απλή και ισχυρή από την ευστάθεια. Λέμε ότι ένας αλγόριθμος $\overline{f}$ για ένα πρόβλημα $f$ είναι προς τα πίσω ευσταθής αν για κάθε $x \in X$ ,

$$\overline{f}(x) = f(\overline{x}) \text{ για κάποια } \overline{x} \text{ με } \frac{\left\| \overline{x} - x \right\|}{\left\| x \right\|} = O(\varepsilon_{\mu\eta\chi\alpha\nu\acute{\eta}\varsigma}) \tag{3.28}$$

Αυτός είναι ένα πιο αυστηρός ορισμός της ευστάθειας στον οποίο ο όρος $O(\varepsilon_{\mu\eta\chi\alpha\nu\acute{\eta}\varsigma})$ της (3.26) έχει αντικατασταθεί με μηδέν. Με λόγια,

Ένας προς τα πίσω ευσταθής αλγόριθμος δίνει ακριβώς τη σωστή λύση
σε ένα πρόβλημα είναι σχεδόν στο σωστό.

*3.3.5 Η ΣΗΜΑΣΙΑ ΤΟΥ Ο(ε)*

Εδώ θα εξηγήσουμε την ακριβή σημασία του «$O(\varepsilon_{\mu\eta\chi\alpha\nu\acute{\eta}\varsigma})$» στις (3.25)-(3.28).
Ο συμβολισμός

$$\varphi(t) = O(\psi(t)) \tag{3.29}$$

είναι βασικός με ακριβή ορισμό. Από την εξίσωση αυτή προκύπτει ότι υπάρχει θετική σταθερά $C$ τέτοια ώστε για όλα τα $t$ κοντά σε ένα όριο (πχ. $t \to 0$ ή $t \to \infty$)

$$\left|\varphi(t)\right| \le C\psi(t) \tag{3.30}$$

Για παράδειγμα από την πρόταση $\sin^2 t = O(t^2)$ καθώς $t \to 0$ προκύπτει ότι υπάρχει μια σταθερά $C$ τέτοια ώστε για κάθε αρκετά μικρό $t$, $\left|\sin^2 t\right| \le Ct^2$.
Επίσης, βασικές είναι οι προτάσεις της μορφής

$$\phi(s,t) = O(\psi(t)) \text{ ομαλά στο } s, \tag{3.31}$$

όπου η $\phi$ είναι μια συνάρτηση η οποία εξαρτάται όχι μόνο από το $t$ αλλά και από μια άλλη μεταβλητή $s$. Η λέξη «ομαλά» φανερώνει ότι υπάρχει μια μονότιμη σταθερά $C$ όπως στην (3.30) η οποία ισχύει για κάθε $s$. Για παράδειγμα,

$$(\sin^2 t)(\sin^2 s) = O(t^2)$$

ισχύει ομαλά καθώς $t \to 0$ αλλά η ομαλότητα χάνεται αν αντικαταστήσουμε το $\sin^2 s$ με $s^2$.

Στην εργασία αυτή εμφανίζεται συνήθως η σχέση για την υπολογιζόμενη απόσταση:

$$\left\|\text{computed quantity}\right\| = O(\varepsilon_{\mu\eta\chi\alpha\nu\acute{\eta}\varsigma}) \tag{3.32}$$

Παραθέτουμε την ερμηνεία της σχέσης αυτής. Πρώτον, ποσότητα $\left\|\text{computed quantity}\right\|$ αντιπροσωπεύει τη νόρμα ενός αριθμού ή μιας παράστασης έτσι όπως αυτή έχει καθοριστεί από έναν αλγόριθμο $\bar{f}$ ενός προβλήματος $f$ εξαρτώμενος από τα δεδομένα $x \in X$ για $f$ και από το $\varepsilon_{\mu\eta\chi\alpha\nu\acute{\eta}\varsigma}$. Ένα παράδειγμα είναι το σχετικό σφάλμα (3.24). Δεύτερον, η έμμεση διαδικασία ορίου είναι $\varepsilon_{\mu\eta\chi\alpha\nu\acute{\eta}\varsigma} \to 0$ (πχ το $\varepsilon_{\mu\eta\chi\alpha\nu\acute{\eta}\varsigma}$ είναι η μεταβλητή για το $t$ στη σχέση (3.31)). Τρίτον, η

πολυπλοκότητα «Ο» εφαρμόζεται σε όλα τα δεδομένα $x \in X$ (πχ το $x$ είναι η μεταβλητή για το $s$ ).
Σε κάθε αριθμητική μηχανής, ο αριθμός $\varepsilon_{μηχανής}$ είναι μια σταθερή ποσότητα. Η εξίσωση (3.31) σημαίνει ότι αν έπρεπε να υλοποιήσουμε έναν αλγόριθμο ο οποίος να ικανοποιεί τις σχέσεις (3.21) και (3.23) για μια ακολουθία τιμών του $\varepsilon_{μηχανής}$ η οποία να συγκλίνει στο μηδέν τότε η ποσότητα $\|\text{computed quantity}\|$ θα μειώνονταν με το ρυθμό του $\varepsilon_{μηχανής}$ ή και γρηγορότερα. Αυτοί οι ιδανικοί υπολογισμοί απαιτούνται για να ικανοποιούνται μόνο οι σχέσεις (3.21) και (3.23).

### *3.3.6 ΕΞΑΡΤΗΣΗ ΑΠΟ ΤΑ m, n ΚΑΙ ΟΧΙ ΑΠΟ ΤΑ A ΚΑΙ b*

Θα συζητήσουμε τη σημασία του $O(\varepsilon_{μηχανής})$ στις (3.25)-(3.28) λίγο περισσότερο. Η ομαλότητα της σταθεράς «Ο» μπορεί να περιγραφεί από το παρακάτω παράδειγμα. Έστω ότι έχουμε έναν αλγόριθμο για τη λύση ως προς $x$ του ομαλού $m \times m$ συστήματος των εξισώσεων $Ax = b$ και υποθέτουμε ότι το αποτέλεσμα $\bar{x}$ που υπολογίστηκε από τον αλγόριθμο αυτό ικανοποιεί τη σχέση

$$\frac{\|\bar{x} - x\|}{\|x\|} = O(\kappa(A)\varepsilon_{μηχανής}) \tag{3.33}$$

Η υπόθεση αυτή σημαίνει ότι το όριο

$$\frac{\|\bar{x} - x\|}{\|x\|} \le C\kappa(A)\varepsilon_{μηχανής} \tag{3.34}$$

ισχύει για μια σταθερά $C$ ανεξάρτητη από τον πίνακα $A$ ή το αριστερό μέρος $b$, για κάθε $\varepsilon_{μηχανής}$ αρκετά μικρό.
Αν ο παρονομαστής σε μια σχέση όπως η (3.34) είναι μηδενικός, το νόημά του ορίζεται από την παρακάτω παραδοχή. Όταν γράφουμε την (3.34) αυτό που εννοούμε είναι

$$\|\bar{x} - x\| \le C\kappa(A)\varepsilon_{μηχανής} \tag{3.35}$$

Δεν υπάρχει διαφορά αν $\|x\| \neq 0$ αλλά αν $\|x\| = 0$ από την (3.35) είναι φανερό ότι η ακριβής σημασία της (3.33) είναι ότι $\|\bar{x} - x\| = 0$ για κάθε αρκετά μικρό $\varepsilon_{μηχανής}$.
Παρόλο που η σταθερά $C$ της (3.34) ή της (3.35) δεν εξαρτάται ούτε από τον $A$ ή τον $b$ εξαρτάται γενικά από τη διάσταση $m$. Αυτό είναι μια συνέπεια του ορισμού

του προβλήματος που δώσαμε σε προηγούμενη παράγραφο. Αν οι διαστάσεις όπως η $m$ ή η $n$ που ορίζουν ένα πρόβλημα $f$ αλλάξουν τότε οι διανυσματικοί χώροι $X$ και $Y$ πρέπει να αλλάξουν και αυτοί και κατά συνέπεια θα έχουμε ένα νέο πρόβλημα $f'$. Και στην πράξη, οι επιδράσεις από τα σφάλματα στρογγυλοποίησης των αλγορίθμων της αριθμητικής γραμμικής άλγεβρας γενικά αλλάζουν όταν αλλάζουν τα $m$ και $n$. Όμως, αυτή η αλλαγή είναι πολύ μικρή για να τη λάβουμε υπόψη. Η εξάρτηση από τα $m$ και $n$ είναι τυπικά γραμμική, τετραγωνική ή κυβική στη χειρότερη περίπτωση (ο εκθέτης εξαρτάται από την επιλογή της νόρμας καθώς και από την επιλογή του αλγορίθμου) και τα σφάλματα για τα περισσότερα δεδομένα είναι πολύ μικρότερα απ' ότι στη χειρότερη περίπτωση χάρη σε στατιστική απαλοιφή.

*3.3.7 ΑΝΕΞΑΡΤΗΣΙΑ ΝΟΡΜΑΣ*

*(θ*εμελιώσης ιδιότητα χώρων πεπερασμένης διάστασης)

Οι ορισμοί που περιέχουν την πολυπλοκότητα $O(\varepsilon_{μηχανής})$ έχουν την ιδιότητα ότι δοθέντων $\mathrm{X}$ και $\mathrm{Y}$ πεπερασμένης διάστασης τότε είναι ανεξάρτητοι νόρμας.

ΘΕΩΡΗΜΑ 3.3

Για προβλήματα $f$ και αλγόριθμους $\bar{f}$ που ορίζονται στους χώρους πεπερασμένης διάστασης $X$ και $Y$, οι ιδιότητες της ακρίβειας, της ευστάθειας και της προς τα πίσω ευστάθειας ισχύουν ή όχι ανάλογα με την επιλογή της νόρμας στους $X$ και $Y$.

ΑΠΟΔΕΙΞΗ

Είναι γνωστό ότι σε ένα διανυσματικός χώρος όλες οι νόρμες είναι ισοδύναμες με την έννοια ότι αν $\|\cdot\|$ και $\|\cdot\|'$ είναι δυο νόρμες στον ίδιο χώρο τότε υπάρχουν θετικές σταθερές $C_1$ και $C_2$ τέτοιες ώστε $C_1\|x\| \le \|x\|' \le C_2\|x\|$ για όλα τα $x$ στο χώρο αυτό. Συνεπώς, μια αλλαγή στη νόρμα μπορεί να έχει επίδραση στο μέγεθος της σταθεράς $C$ που περιέχεται σε μια υπόθεση της $O(\varepsilon_{μηχανής})$ αλλά όχι στην ύπαρξη μιας τέτοιας σταθεράς.

## 3.4 ΠΕΡΙΣΣΟΤΕΡΑ ΣΤΗΝ ΕΥΣΤΑΘΕΙΑ

Στην παράγραφο αυτή θα εξετάσουμε τους ευσταθείς και ασταθείς αλγόριθμους. Στη συνέχεια, θα συζητήσουμε μια θεμελιώδη ιδέα η οποία σχετίζει την ευστάθεια και την κατάσταση ενός αλγόριθμου. Η ιδέα αυτή έχει αποδειχτεί σε αμέτρητες εφαρμογές από το 1950: είναι η ανάλυση των προς τα πίσω σφαλμάτων.

### *3.4.1 ΕΥΣΤΑΘΕΙΑ ΤΗΣ ΑΡΙΘΜΗΤΙΚΗΣ ΚΙΝΗΤΗΣ ΥΠΟΔΙΑΣΤΟΛΗΣ*

Τα τέσσερα πιο απλά προβλήματα υπολογισμού είναι η πρόσθεση, η αφαίρεση, ο πολλαπλασιασμός και η διαίρεση. Από τα αξιώματα (3.21) και (3.23) προκύπτει ότι οι τέσσερεις αλγόριθμοι των πράξεων αυτών είναι προς τα πίσω ευσταθείς.
Ας αποδείξουμε τον παραπάνω ισχυρισμό για την αφαίρεση αφού αυτή η στοιχειώδης πράξη αναμένεται ότι θα παρουσιάζει το μεγαλύτερο ρίσκο αστάθειας. Ο χώρος δεδομένων $X$ αποτελείται από το σύνολο 2 διανυσμάτων, $C^2$, και ο χώρος λύσεων $Y$ είναι το σύνολο των μονόμετρων μεγεθών, $C$. Από το θεώρημα 3.3 δε χρειάζεται να καθορίσουμε τις νόρμες σε αυτούς τους χώρους.
Για τα δεδομένα $x = (x_1, x_2)^* \in X$ το πρόβλημα της αφαίρεσης αντιστοιχεί στη συνάρτηση $f(x_1, x_2) = x_1 - x_2$ και ο αλγόριθμος για τον οποίο συζητάμε μπορεί να γραφτεί ως

$$\overline{f}(x_1, x_2) = Fl(x_1) - Fl(x_2)$$

Η ισότητα αυτή σημαίνει ότι πρώτα στρογγυλοποιούμε τα $x_1$ και $x_2$ σε τιμές κινητής υποδιαστολής και μετά εφαρμόζουμε την πράξη «». Τώρα από την (3.21) έχουμε

$$Fl(x_1) = x_1(1+\varepsilon_1), \qquad Fl(x_2) = x_2(1+\varepsilon_2)$$

για κάποια $|\varepsilon_1|, |\varepsilon_2| \le \varepsilon_{μηχανής}$. Από την (3.23) έχουμε

$$|\varepsilon_3| \le \varepsilon_{μηχανής}$$

για κάποιο $|\varepsilon_3| \le \varepsilon_{μηχανής}$. Συνδυάζοντας αυτές τις εξισώσεις έχουμε ότι

$$\begin{aligned} Fl(x_1) - Fl(x_2) &= [x_1(1+\varepsilon_1) - x_2(1+\varepsilon_2)](1+\varepsilon_3) \\ &= x_1(1+\varepsilon_1)(1+\varepsilon_3) - x_2(1+\varepsilon_2)(1+\varepsilon_3) \\ &= x_1(1+\varepsilon_4) - x_2(1+\varepsilon_5) \end{aligned}$$

για κάποια $|\varepsilon_4|,|\varepsilon_5| \le 2\varepsilon_{μηχανής} + O(\varepsilon^2_{μηχανής})$. Με άλλα λόγια, το υπολογισμένο αποτέλεσμα $\overline{f}(x) = Fl(x_1) - Fl(x_2)$ είναι ακριβώς ίσο με τη διαφορά $\overline{x}_1 - \overline{x}_2$ όπου τα $\overline{x}_1$ και $\overline{x}_2$ ικανοποιούν τις σχέσεις

$$\frac{|\overline{x}_1 - x_1|}{|x_1|} = O(\varepsilon_{μηχανής}), \qquad \frac{|\overline{x}_2 - x_2|}{|x_2|} = O(\varepsilon_{μηχανής})$$

και κάθε $C > 2$ θα αρκεί για τις σταθερές στα σύμβολα «Ο». Για οποιαδήποτε επιλογή νόρμας $\|\cdot\|$ στο χώρο $C^2$ θα προκύπτει η (3.28).

Παράδειγμα Εσωτερικό Γινόμενο

Έστω δοθέντα διανύσματα $x, y \in C^m$ και έστω ότι θέλουμε να υπολογίσουμε το εσωτερικό γινόμενο $\alpha = x^* y$. Ο προφανής αλγόριθμος με βάση τον οποίο μπορούμε να κάνουμε τον υπολογισμό αυτό είναι να υπολογίσουμε ξεχωριστά κάθε γινόμενο $\overline{x_i y_i}$ μέσω της πράξης $\otimes$ και να τα προσθέσουμε μέσω της πράξης $\oplus$ και να καταλήξουμε στο αποτέλεσμα $\overline{\alpha}$. Ο αλγόριθμος που χρησιμοποιούμε είναι προς τα πίσω ευσταθής.

Παράδειγμα Εξωτερικό Γινόμενο

Έστω ότι θέλουμε να υπολογίσουμε το πρώτης τάξης εξωτερικό γινόμενο $A = xy^*$ για δοθέντα διανύσματα $x \in C^m, y \in C^n$. Ο προφανής αλγόριθμος που μπορούμε να χρησιμοποιήσουμε για τον υπολογισμό είναι να υπολογίσουμε τα $mn$ γινόμενα $x_i \overline{y_i}$ μέσω της πράξης $\otimes$ και να τα ορίσουμε σαν στοιχεία ενός πίνακα $\overline{A}$. Ο αλγόριθμος αυτός είναι ευσταθής αλλά δεν είναι προς τα πίσω ευσταθής. Η εξήγηση δίνεται από το γεγονός ότι ο πίνακας $\overline{A}$ δεν έχει τάξη ακριβώς ίση με 1 και κατά συνέπεια δεν μπορεί να γραφτεί στη μορφή $(x+\delta x)(y+\delta y)^*$. Κατά κανόνα, στα προβλήματα όπου η διάσταση του χώρου λύσεων $Y$ είναι μεγαλύτερη από τη διάσταση του χώρου $X$ του προβλήματος, η προς τα πίσω ευστάθεια είναι σπάνια.

### *3.4.2 ΕΝΑΣ ΑΣΤΑΘΗΣ ΑΛΓΟΡΙΘΜΟΣ*

Έχουμε το πρόβλημα: η χρήση ενός χαρακτηριστικού πολυωνύμου για την εύρεση των ιδιοτιμών ενός πίνακα.

Αφού το $z$ είναι μια ιδιοτιμή του πίνακα $A$ αν και μόνο αν $p(z) = 0$, όπου $p(z)$ είναι το χαρακτηριστικό πολυώνυμο $\det(zI - A)$, οι ρίζες του $p$ είναι οι ιδιοτιμές του $A$. Από τα παραπάνω προκύπτει μια μέθοδος για τον υπολογισμό των ιδιοτιμών:

1. Βρείτε τους συντελεστές του χαρακτηριστικού πολυωνύμου
2. Βρείτε τις ρίζες του

Ο αλγόριθμος αυτός δεν είναι μόνο προς τα πίσω ασταθής αλλά είναι γενικά ασταθής και δεν πρέπει να χρησιμοποιείται. Ακόμα και στις περιπτώσεις όπου ο υπολογισμός των ιδιοτιμών είναι ένα καλά ορισμένο πρόβλημα, μπορεί τα αποτελέσματα που θα προκύψουν να έχουν σχετικά σφάλματα πολύ πιο μεγάλα από το $\varepsilon_{μηχανής}$.

Η αστάθεια προκύπτει στο δεύτερο βήμα, όπου ζητάμε την εύρεση των ριζών. Το πρόβλημα της εύρεσης των ριζών ενός πολυωνύμου, δοθέντων των συντελεστών, είναι γενικά ασθενώς ορισμένο. Κατά συνέπεια, μικρά σφάλματα στους συντελεστές του χαρακτηριστικού πολυωνύμου θα αυξηθούν όταν θα βρεθούν οι ρίζες ακόμα και αν η διαδικασία εύρεσης των ριζών έχει εκτελεστεί με απόλυτη ακρίβεια.

Για παράδειγμα, ας υποθέσουμε ότι $A = I$ είναι ο $2 \times 2$ ταυτοτικός πίνακας. Οι ιδιοτιμές του $A$ δεν επηρεάζονται από τις διαταραχές των στοιχείων και ένας ευσταθής αλγόριθμος θα μπορούσε να τις υπολογίσει με σφάλματα $O(\varepsilon_{μηχανής})$. Όμως, ένας αλγόριθμος, όπως περιγράφηκε παραπάνω παράγει σφάλματα της τάξης του $\sqrt{\varepsilon_{μηχανής}}$. Για να το εξηγήσουμε αυτό ας πάρουμε το χαρακτηριστικό πολυώνυμο $x^2 - 2x + 1$. Όταν υπολογίζονται οι συντελεστές του πολυωνύμου αυτού αναμένουμε σφάλματα της τάξης του $\varepsilon_{μηχανής}$ και τα σφάλματα αυτά μπορούν να προκαλέσουν αλλαγή στις ρίζες της τάξης του $\sqrt{\varepsilon_{μηχανής}}$. Για παράδειγμα αν $\varepsilon_{μηχανής} = 10^{-16}$ τότε οι ρίζες του χαρακτηριστικού πολυωνύμου μπορούν να διαταραχθούν από τις πραγματικές ιδιοτιμές με ακρίβεια υπολογίζονται $10^{-8}$, δηλαδή έχουμε απώλεια ακρίβειας κατά οχτώ ψηφία.

### *3.4.3 ΑΚΡΙΒΕΙΑ ΕΝΟΣ ΠΡΟΣ ΤΑ ΠΙΣΩ ΣΤΑΘΕΡΟΥ ΑΛΓΟΡΙΘΜΟΥ*

Ας υποθέσουμε ότι έχουμε έναν προς τα πίσω ευσταθή αλγόριθμο $\bar{f}$ για ένα πρόβλημα $f : X \to Y$. Θα είναι τα αποτελέσματα ακριβή; Η απάντηση εξαρτάται από το δείκτη κατάστασης $\kappa = \kappa(x)$ του $f$. Αν ο $\kappa(x)$ είναι μικρός τότε τα αποτελέσματα θα είναι σχετικά ακριβή αλλά αν είναι μεγάλος τότε θα έχουμε απώλεια ακρίβειας.

ΘΕΩΡΗΜΑ 3.4

Έστω ότι εφαρμόζουμε σε έναν υπολογιστή έναν προς τα πίσω ευσταθή αλγόριθμο για να λύσουμε ένα πρόβλημα $f : X \to Y$ με δείκτη κατάστασης $\kappa$ και έστω ότι

ικανοποιούνται τα αξιώματα (3.21) και (3.23). Τότε το σχετικό σφάλμα ικανοποιεί τη σχέση

$$\frac{\left\|\overline{f}(x)-f(x)\right\|}{\left\|f(x)\right\|}=O(\kappa(x)\varepsilon_{\mu\eta\chi\alpha\nu\acute{\eta}\varsigma}) \tag{3.36}$$

ΑΠΟΔΕΙΞΗ

Από τον ορισμό (3.28) της προς τα πίσω ευστάθειας, έχουμε $\overline{f}(x)-f(\overline{x})$ για κάποιο $\overline{x}\in X$ που ικανοποιεί τη σχέση

$$\frac{\left\|\overline{x}-x\right\|}{\left\|x\right\|}=O(\varepsilon_{\mu\eta\chi\alpha\nu\acute{\eta}\varsigma})$$

Από τον ορισμό (3.5) του $\kappa(x)$ προκύπτει ότι

$$\frac{\left\|\overline{f}(x)-f(x)\right\|}{\left\|f(x)\right\|}\le(\kappa(x)+o(1))\frac{\left\|\overline{x}-x\right\|}{\left\|x\right\|} \tag{3.37}$$

όπου το $o(1)$ είναι η ποσότητα που συγκλίνει στο μηδέν καθώς $\varepsilon_{\mu\eta\chi\alpha\nu\acute{\eta}\varsigma}\to 0$. Από το συνδυασμό αυτών των ορίων έχουμε τη σχέση (3.36).

### *3.4.4 ΑΝΑΛΥΣΗ ΤΩΝ ΠΡΟΣ ΤΑ ΠΙΣΩ ΣΦΑΛΜΑΤΩΝ*

Η διαδικασία που ακολουθήσαμε για να αποδείξουμε το θεώρημα 3.4 είναι γνωστή και ως ανάλυση των προς τα πίσω σφαλμάτων. Εξασφαλίσαμε μια εκτίμηση ακρίβειας με δυο βήματα. Το ένα βήμα ήταν να εξετάσουμε την κατάσταση του προβλήματος. Το άλλο ήταν να εξετάσουμε την ευστάθεια του αλγορίθμου. Το συμπέρασμα ήταν ότι αν ο αλγόριθμος είναι ευσταθής τότε η τελική ακρίβεια αντιστοιχεί στο δείκτη κατάστασης.

Από μαθηματικής απόψεως η ιδέα αυτή προκύπτει άμεσα αλλά δεν είναι το πρώτο πράγμα που θα μπορούσαμε να σκεφτούμε. Η πρώτη ιδέα που θα σκεφτόμασταν θα ήταν ανάλυση των προς τα εμπρός σφαλμάτων.

Η εμπειρία έχει δείξει ότι στους περισσότερους αλγόριθμους της αριθμητικής γραμμικής άλγεβρας η ανάλυση των προς τα εμπρός σφαλμάτων είναι πιο δύσκολη απ’ ότι η ανάλυση των προς τα πίσω. Δεν είναι δύσκολο να δείξουμε γιατί συμβαίνει αυτό. Ας υποθέσουμε ότι χρησιμοποιούμε ένα δοκιμασμένο και αληθή αλγόριθμο για να λύσουμε το πρόβλημα $Ax=b$ σε έναν υπολογιστή. Είναι αποδεδειγμένο ότι τα αποτελέσματα που θα λάβουμε είναι λιγότερο ακριβή όταν ο $A$ είναι ασθενώς ορισμένος. Πώς θα μπορούσε μια ανάλυση των προς τα εμπρός

σφαλμάτων να περιγράψει αυτό το φαινόμενο; Ο δείκτης κατάστασης του $A$ είναι μια ιδιότητα η οποία μπορεί να είναι περισσότερο ή λιγότερο εμφανής στο επίπεδο των ατομικών πράξεων κινητής υποδιαστολής που πραγματοποιούνται κατά την επίλυση του $Ax = b$. Οπότε με τον έναν ή τον άλλο τρόπο με την προς τα εμπρός ανάλυση θα πρέπει να εντοπίσουμε αυτό το δείκτη κατάστασης αν θέλουμε να έχουμε ένα σωστό αποτέλεσμα.
Οπότε είναι γεγονός ότι οι καλύτεροι αλγόριθμοι για την επίλυση των περισσοτέρων προβλημάτων δεν κάνουν κάτι περισσότερο από το να υπολογίζουν ακριβείς λύσεις για δεδομένα που έχουν διαταραχθεί ελάχιστα.

## 3.5 ΕΥΣΤΑΘΕΙΑ ΤΗΣ HOUSEHOLDER ΤΡΙΓΩΝΟΠΟΙΗΣΗΣ

Σε αυτή την παράγραφο θα δούμε πώς λειτουργεί η ανάλυση των προς τα πίσω σφαλμάτων. Θα δούμε πώς το βήμα της τριγωνοποίησης μπορεί να συνδυαστεί με άλλα προς τα πίσω ευσταθή μέρη για να προκύψει ένας ευσταθής αλγόριθμος που να λύνει το $Ax = b$.

*ΘΕΩΡΗΜΑ*
Η τριγωνοποίηση Householder είναι προς τα πίσω ευσταθής για όλους τους πίνακες $A$ και για υπολογιστές που ικανοποιούν τις (3.21) και (3.23) . ΄Εχουμε το παρακάτω θεώρημα. Το αποτέλεσμα που θα πάρουμε θα είναι της μορφής

$$\bar{Q}\bar{R} = A + \delta A \tag{3.38}$$

όπου το $\delta A$ είναι μικρό. Με λόγια, το γινόμενο του υπολογισμένου $Q$ επί τον υπολογισμένο $R$ είναι ίσο με μια μικρή διαταραχή του δοθέντος πίνακα $A$. Με το σύμβολο $\bar{R}$ συμβολίζουμε έναν άνω τριγωνικό πίνακα ο οποίος έχει κατασκευαστεί από μια Householder τριγωνοποίηση με την αριθμητική κινητής υποδιαστολής. Με $\bar{Q}$ συμβολίζουμε ένα συγκεκριμένο πίνακα ο οποίος είναι ακριβώς ορθομοναδιαίος. Ας θυμηθούμε ότι ο $Q$ είναι ίσος με το γινόμενο $Q = Q_1 Q_2 \cdots Q_n$ όπου $Q_k$ είναι ο Householder ανακλαστής που ορίζεται από το διάνυσμα $u_k$ όπως αυτό ορίστηκε στο $k$-οστό βήμα του αλγόριθμου 2.3. Κατά τον υπολογισμό κινητής υποδιαστολής παίρνουμε μια ακολουθία διανυσμάτων $\bar{u}_k$. Έστω $\bar{Q}_k$ να είναι ο ακριβώς ορθομοναδιαίος ανακλαστής όπως ορίστηκε από το διάνυσμα κινητής υποδιαστολής $\bar{u}_k$. Τώρα ορίζουμε

$$\bar{Q} = \bar{Q}_1 \bar{Q}_2 \cdots \bar{Q}_n \tag{3.39}$$

Αυτός ο ακριβώς ορθομοναδιαίος πίνακας $\bar{Q}$ θα είναι ο «υπολογισμένος $Q$».

ΘΕΩΡΗΜΑ 3.5

Έστω η $QR$ παραγοντοποίηση $A = QR$ ενός πίνακα $A \in C^{m\times n}$ να είναι υπολογισμένη από μια Householder τριγωνοποίηση (Αλγόριθμος 2.3) σε έναν υπολογιστή και να ικανοποιεί τα αξιώματα (3.21) και (3.23) και έστω οι υπολογισμένοι παράγοντες $\bar{Q}$ και $\bar{R}$ να ορίζονται όπως παραπάνω. Τότε έχουμε

$$\bar{Q}\bar{R} = A + \delta A, \qquad \frac{\|\delta A\|}{\|A\|} = O(\varepsilon_{\mu\eta\chi\alpha\nu\acute{\eta}\varsigma}) \tag{3.40}$$

για κάποιο $\delta A \in C^{m\times n}$.

*3.5.1 ΑΝΑΛΥΟΝΤΑΣ ΤΟΝ ΑΛΓΟΡΙΘΜΟ ΓΙΑ ΝΑ ΛΥΣΟΥΜΕ ΤΟ Ax=b*

Είδαμε ότι η Householder τριγωνοποίηση είναι προς τα πίσω ευσταθής αλλά δεν είναι πάντα ακριβής όταν έχει φορά προς τα εμπρός. Αυτό ισχύει για τις περισσότερες παραγοντοποιήσεις πινάκων στην αριθμητική γραμμική άλγεβρα. Η QR παραγοντοποίηση χρησιμοποιείται για την επίλυση συστημάτων εξισώσεων, προβλημάτων ελαχίστων τετραγώνων ή προβλημάτων ιδιοτιμών. Είναι η παραγοντοποίηση αυτή αρκετά προς τα πίσω ευσταθής για να αποτελέσει ένα ικανοποιητικό κομμάτι ενός μεγαλύτερου αλγόριθμου; Δηλαδή, είναι η ακρίβεια του QR γινομένου αρκετή για εφαρμογές ή χρειαζόμαστε ξεχωριστά την ακρίβεια του Q και του R; Η απάντηση είναι η ακρίβεια του QR είναι αρκετή για τους περισσότερους σκοπούς.

Το παράδειγμα που θα μελετήσουμε είναι η χρήση της Householder τριγωνοποίησης στην επίλυση ενός ομαλού $m\times m$ γραμμικού συστήματος $Ax = b$. Παραθέτουμε το κομμάτι του αλγόριθμου που μας ενδιαφέρει.

ΑΛΓΟΡΙΘΜΟΣ 3.1 Επίλυση του $Ax = b$ με QR παραγοντοποίηση

| | |
|---|---|
| $QR = A$ | Παραγωγίζουμε τον $A$ μέσω του αλγόριθμου 2.3, όπου το $Q$ είναι το γινόμενο των ανακλαστών. |
| $y = Q^* b$ | Κατασκευάζουμε τον $Q^* b$ μέσω του αλγορίθμου 2.4 |
| $x = R^{-1} y$ | Λύνουμε το τριγωνικό σύστημα $Rx = y$ με τη χρήση προς τα πίσω αντικατάστασης (Αλγόριθμος 3.2) |

Ο αλγόριθμος είναι προς τα πίσω ευσταθής και η απόδειξη αυτού του ισχυρισμού είναι άμεση αν σκεφτούμε ότι καθένα από τα τρία βήματα είναι προς τα πίσω ευσταθές.

Το πρώτο βήμα του παραπάνω αλγόριθμου είναι μια QR παραγοντοποίηση του πίνακα $A$ και οδηγεί στους υπολογισμένους πίνακες $\bar{Q}$ και $\bar{R}$. Η προς τα πίσω ευστάθεια αυτής της διαδικασίας έχει εκφραστεί με τη σχέση (3.40).

Το δεύτερο βήμα είναι ο υπολογισμός του $\bar{Q}^* b$ μέσω του αλγορίθμου 2.4 (Ας σημειώσουμε ότι δε γράφουμε $Q^* b$ γιατί σε αυτό το σημείο του υπολογισμού το πρώτο βήμα έχει ολοκληρωθεί και ο πίνακας έχει παραχθεί όχι από το $Q$αλλά από το $\bar{Q}$.) Όταν έχουμε υπολογίσει το $\bar{Q}^* b$ από τον αλγόριθμο 2.4 προκύπτουν σφάλματα από στρογγυλοποίηση οπότε το αποτέλεσμα δεν θα είναι ακριβώς $\bar{Q}^* b$. Αλλά θα είναι ένα διάνυσμα $\bar{y}$. Μπορούμε να δείξουμε ότι το διάνυσμα αυτό ικανοποιεί την παρακάτω εκτίμηση για την προς τα πίσω ευστάθεια:

$$(\bar{Q} + \delta Q)\bar{y} = b, \qquad \|\delta Q\| = O(\varepsilon_{\mu\eta\chi\alpha\nu\acute{\eta}\varsigma}) \tag{3.41}$$

Όπως και η (3.40), η εξίσωση αυτή είναι ακριβής. Με λόγια, το αποτέλεσμα της εφαρμογής των Householder ανακλαστών στην αριθμητική κινητής υποδιαστολής είναι ακριβώς ίση με τον πολλαπλασιασμό του $b$ επί έναν ελάχιστα διαταραγμένο πίνακα $(\bar{Q} + \delta Q)^{-1}$.

Το τελευταίο βήμα του αλγόριθμου 3.1 είναι προς τα πίσω αντικατάσταση για τον υπολογισμό του $\bar{R}^{-1}\bar{y}$. Στο βήμα αυτό θα έχουμε νέα σφάλματα στρογγυλοποίησης αλλά και πάλι ο υπολογισμός θα είναι προς τα πίσω ευσταθής. Αυτή τη φορά η εκτίμηση έχει τη μορφή:

$$(\bar{R} + \delta R)\bar{x} = \bar{y}, \qquad \frac{\|\delta R\|}{\|\bar{R}\|} = O(\varepsilon_{\mu\eta\chi\alpha\nu\acute{\eta}\varsigma}) \tag{3.42}$$

Η αριστερή εξίσωση είναι ακριβής. Προκύπτει ότι το αποτέλεσμα κινητής υποδιαστολής $\bar{x}$ είναι η ακριβής λύση μια μικρής διαταραχής του συστήματος $\bar{R}x = \bar{y}$.

Τώρα, παίρνοντας τις (3.40)-(3.42) έχουμε το παρακάτω θεώρημα. Είναι ένα τυπικό θεώρημα προς τα πίσω ευστάθειας και μπορεί να αποδειχθεί για πολλούς αλγόριθμους της αριθμητικής γραμμικής άλγεβρας.

ΘΕΩΡΗΜΑ 3.6

Ο αλγόριθμος 3.1 είναι προς τα πίσω ευσταθής και ικανοποιεί τις παρακάτω σχέσεις:

$$(A + \Delta A)\bar{x} = b, \qquad \frac{\|\Delta A\|}{\|A\|} = O(\varepsilon_{\mu\eta\chi\alpha\nu\acute{\eta}\varsigma}) \tag{3.43}$$

για κάποιο $\Delta A \in C^{m\times m}$.

ΑΠΟΔΕΙΞΗ
Από τις (3.41) και (3.42) έχουμε ότι

$$b = (\bar{Q} + \delta Q)(\bar{R} + \delta R)\bar{x} = [\bar{Q}\bar{R} + (\delta Q)\bar{R} + \bar{Q}(\delta R) + (\delta Q)(\delta R)]\bar{x}$$

Οπότε από την (3.40) έχουμε

$$b = [A + \delta A + (\delta Q)\bar{R} + \bar{Q}(\delta R) + (\delta Q)(\delta R)]\bar{x}$$

Η εξίσωση αυτή είναι της μορφής

$$b = (A + \Delta A)\bar{x} \ ,$$

όπου το $\Delta A$ είναι ένα άθροισμα τεσσάτων όρων. Για να προκύψει η (3.43) πρέπει να δείξουμε ότι κάθε ένας από αυτούς τους όρους είναι μικρός σχετικά με τον $A$. Αφού $\bar{Q}\bar{R} = A + \delta A$ και ο $\bar{Q}$ είναι ορθομοναδιαίος, έχουμε

$$\frac{\|\bar{R}\|}{\|A\|} \le \|Q^*\| \frac{\|A + \delta A\|}{\|A\|} = O(1)$$

καθώς $\varepsilon_{μηχανής} \to 0$, από την (3.40). (Είναι $1 + O(\varepsilon_{μηχανής})$ αν $\|\cdot\| = \|\cdot\|_2$ αλλά δεν έχουμε κάνει καμία υπόθεση για τη νόρμα $\|\cdot\|$. ) Οπότε έχουμε

$$\frac{\|(\delta Q)\bar{R}\|}{\|A\|} \le \|\delta Q\| \frac{\|\bar{R}\|}{\|A\|} = O(\varepsilon_{μηχανής})$$

από την (3.42). Ομοίως,

$$\frac{\|\bar{Q}(\delta R)\|}{\|A\|} \le \|\bar{Q}\| \frac{\|\delta R\|}{\|\bar{R}\|} \frac{\|\bar{R}\|}{\|A\|} = O(\varepsilon_{μηχανής})$$

από την (3.43).Τέλος,

$$\frac{\|(\delta Q)(\delta R)\|}{\|A\|} \le \|\delta Q\| \frac{\|\delta R\|}{\|A\|} = O(\varepsilon^2{}_{μηχανής})$$

Η τελική διαταραχή $\Delta A$ ικανοποιεί την παρακάτω σχέση:

$$\frac{\|\Delta A\|}{\|A\|} \le \frac{\|\delta A\|}{\|A\|} + \frac{\|(\delta Q)\bar{R}\|}{\|A\|} + \frac{\|\bar{Q}(\delta R)\|}{\|A\|} + \frac{\|(\delta Q)(\delta R)\|}{\|A\|} = O(\varepsilon_{\mu\eta\chi\alpha\nu\acute{\eta}\varsigma})$$

όπως υποθέσαμε.

Συνδυάζοντας τα θεωρήματα 3.2, 3.3 και 3.6 προκύπτει το ακόλουθο βασικό συμπέρασμα σχετικά με την ακρίβεια των λύσεων του $Ax = b$.

ΘΕΩΡΗΜΑ 3.7
Η λύση $\bar{x}$ που υπολογίστηκε από τον αλγόριθμο 3.1 ικανοποιεί τη σχέση

$$\frac{\|\bar{x} - x\|}{\|x\|} = O(\kappa(A)\varepsilon_{\mu\eta\chi\alpha\nu\acute{\eta}\varsigma}) \qquad (3.44)$$

## 3.6 ΕΥΣΤΑΘΕΙΑ ΤΗΣ ΠΡΟΣ ΤΑ ΠΙΣΩ ΑΝΤΙΚΑΤΑΣΤΑΣΗΣ

Ένα από τα πιο εύκολα προβλήματα της αριθμητικής γραμμικής άλγεβρας είναι η εύρεση λύσης ενός τριγωνικού συστήματος εξισώσεων. Ο κλασσικός αλγόριθμος βασίζεται στις διαδοχικές αντικαταστάσεις και ονομάζεται προς τα πίσω αντικατάσταση όταν το σύστημα είναι άνω τριγωνικό. Στην παράγραφο αυτή θα δείξουμε ότι αυτός ο αλγόριθμος είναι προς τα πίσω ευσταθής και περιέχει τετραγωνικά όρια για τις επιδράσεις των σφαλμάτων στρογγυλοποίησης χωρίς « $O(\varepsilon_{\mu\eta\chi\alpha\nu\acute{\eta}\varsigma})$ ».

### *3.6.1 ΤΡΙΓΩΝΙΚΑ ΣΥΣΤΗΜΑΤΑ*

Είδαμε ότι ένα γενικό σύστημα εξισώσεων $Ax = b$ μπορεί να μειωθεί σε ένα άνω τριγωνικό σύστημα $Rx = y$ με QR παραγοντοποίηση. Κάτω και άνω τριγωνικά συστήματα μπορούν να προκύψουν και σε άλλες μεθόδους της αριθμητικής γραμμικής άλγεβρας όπως η απαλοιφή του Gauss και η παραγοντοποίηση Cholesky. Αυτά τα συστήματα μπορούν να επιλυθούν εύκολα με μια διαδικασία διαδοχικών αντικαταστάσεων η οποία ονομάζεται προς τα εμπρός αντικατάσταση αν το σύστημα είναι κάτω τριγωνικό και προς τα πίσω αντικατάσταση αν το σύστημα είναι άνω τριγωνικό. Παρόλο που και οι δύο περιπτώσεις είναι ίδιες από μαθηματικής απόψεως, σε αυτή την παράγραφο θα δούμε την προς τα πίσω αντικατάσταση.
Ας υποθέσουμε ότι θέλουμε να λύσουμε το σύστημα $Rx = b$, δηλαδή,

$$\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ & r_{22} & & \\ & & \ddots & \vdots \\ & & & r_{mm} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} \qquad (3.45),$$

όπου $b \in C^m$ και $R \in C^{m \times m}$, ομαλός και άνω τριγωνικός είναι γνωστά και $x \in C^m$ είναι άγνωστος. Μπορούμε να λύσουμε το σύστημα αυτό ως προς τις συνιστώσες του $x$, ξεκινώντας από την $x_m$ και καταλήγοντας στην $x_1$. Έχουμε τον παρακάτω αλγόριθμο.

ΑΛΓΟΡΙΘΜΟΣ 3.2 Προς τα πίσω αντικατάσταση

$$\begin{aligned} x_m &= b_m / r_{mm} \\ x_{m-1} &= (b_{m-1} - x_m r_{m-1,m}) / r_{m-1,m-1} \\ x_{m-2} &= (b_{m-2} - x_{m-1} r_{m-2,m-1} - x_m r_{m-2,m}) / r_{m-2,m-2} \\ &\vdots \\ x_j &= \left(b_j - \sum_{k=j+1}^{m} x_k r_{jk}\right) \Big/ r_{jj} \end{aligned}$$

Η διαδικασία είναι τριγωνική, με μια αφαίρεση και έναν πολλαπλασιασμό σε κάθε βήμα. Είναι

Πολυπλοκότητα για προς τα πίσω αντικατάσταση: $\sim m^2$ βήματα. (3.44)

## *3.6 2 ΘΕΩΡΗΜΑ ΠΡΟΣ ΤΑ ΠΙΣΩ ΕΥΣΤΑΘΕΙΑΣ*

Σε αυτή την παράγραφο, η προς τα πίσω αντικατάστση είναι ένα από τα τρία βήματα κατά την επίλυση του $Ax = b$ με QR παραγοντοποίηση. Στις σχέσεις (3.40)-(3.42) ισχυριστήκαμε ότι το καθένα από τα βήματα αυτά είναι προς τα πίσω ευσταθές αλλά δεν αποδείξαμε αυτόν τον ισχυρισμό. Στην παράγραφο αυτή δώσουμε αυτή την απόδειξη βρίσκοντας ένα όριο από το οποίο προκύπτει η (3.42).

Πριν αποδείξουμε ότι ο αλγόριθμος 3.2 είναι προς τα πίσω ευσταθής πρέπει να τονίσουμε μια λεπτομέρια του αλγορίθομου η οποία δεν έχει διευκρινιστεί. Έστω ότι στις παραπάνω παρανθέσεις οι αφαιρέσεις γίνονται από αριστερά προς τα δεξιά. (Άλλες φορές δεν είναι ευσταθείς. Μόνο οι λεπτομέριες στις εκτιμήσεις είναι διαφορετικές.)

Έχουμε το παρακάτω θεώρημα.

ΘΕΩΡΗΜΑ 3.8

Έστω ότι ο αλγόριθμος 3.2 εφαρμόζεται σε ένα πρόβλημα (3.45) αποτελούμενο από αριθμούς κινητής υποδιαστολής σε έναν υπολογιστή και ικανοποιεί την (3.23). Ο αλγόριθμος αυτός είναι προς τα πίσω ευσταθής και η υπολογισμένη λύση $\overline{x} \in C^m$ ικανοποιεί τη σχέση

$$(R + \delta R)\overline{x} = b \qquad (3.47)$$

για κάποιο άνω τριγωνικό $\delta R \in C^{m\times m}$ με

$$\frac{\|\delta R\|}{\|R\|} = O(\varepsilon_{\mu\eta\chi\alpha\nu\acute{\eta}\varsigma}) \tag{3.48}$$

Ειδικότερα, για κάθε $i, j$

$$\frac{|\delta r_{ij}|}{|r_{ij}|} \le m\varepsilon_{\mu\eta\chi\alpha\nu\acute{\eta}\varsigma} + O(\varepsilon^2{}_{\mu\eta\chi\alpha\nu\acute{\eta}\varsigma}) \tag{3.49}$$

<u>Σχόλιο</u>

Στην (3.49) και σε όλη την έκταση της παραγράφου αυτής διατηρούμε την παραδοχή της (3.35) ότι αν ο παρονομαστής είναι μηδενικός τότε υποθέτουμε ότι ο αριθμητής είναι επίσης μηδενικός (για κάθε αρκετά μικρό $\varepsilon_{\mu\eta\chi\alpha\nu\acute{\eta}\varsigma}$ ).

Για να διατηρήσουμε τις ιδέες ξεκάθαρες και ενδιαφέρουσες, η απόδειξη του θεωρήματος θα είναι αναλυτική.

ΑΠΟΔΕΙΞΗ

<u>*m=1*</u>

Σύμφωνα με τη σχέση (3.47), στόχος μας είναι να εκφράσουμε κάθε σφάλμα αριθμού κινητής υποδιαστολής σαν μια διαταραχή του δεδομένου του αλγόριθμου. Ξεκινάμε με την πιο απλή περίπτωση όπου ο $R$ έχει διάσταση $1\times 1$. Η προς τα πίσω αντικατάσταση σε αυτή την περίπτωση αποτελείται από ένα βήμα

$$\overline{x}_1 = b_1 \div r_{11}$$

Το αξίωμα (3.23) για την πράξη της διαίρεσης μας εξασφαλίζει ότι η υπολογισμένη λύση είναι κοντά στη σωστή λύση

$$\overline{x}_1 = \frac{b_1}{r_{11}}(1+\varepsilon_1), \qquad |\varepsilon_1| \le \varepsilon_{\mu\eta\chi\alpha\nu\acute{\eta}\varsigma}$$

Όμως, θα θέλαμε να εκφράσουμε το σφάλμα σαν να προέκυπτε από μια διαταραχή στον $R$. Γι' αυτό θέτουμε $\varepsilon_1' = -\varepsilon_1/(1+\varepsilon_1)$ και ο τύπος γίνεται

$$\overline{x}_1 = \frac{b_1}{r_{11}(1+\varepsilon_1')}, \qquad |\varepsilon_1'| \le \varepsilon_{\mu\eta\chi\alpha\nu\acute{\eta}\varsigma} + O(\varepsilon^2{}_{\mu\eta\chi\alpha\nu\acute{\eta}\varsigma}) \tag{3.50}$$

Ας σημειώσουμε ότι το $\varepsilon_1'$ είναι ισοδύναμο με το $-\varepsilon_1$ συν ένος όρου της τάξη του $\varepsilon_1^2$. Μπορούμε ελεύθερα να απομακρύνουμε μικρές διαταραχές από αριθμητές σε παρονομαστές ή και αντίστροφα και το αποτέλεσμα να αλλάξει σε τάξη του $\varepsilon^2{}_{μηχανής}$.

Στην (3.50) η ανισότητα είναι σαξής. Η διαίρεση είναι μαθηματική αλλά όχι κινητής υποδιαστολής. Ο τύπος ορίζει ότι $1\times1$ προς τα πίσω αντικατάσταση είναι προς τα πίσω ευσταθής, $\overline{x}_1$ είναι ακριβώς η σωστή λύση σε ένα διαταραγμένο πρόβλημα, δηλαδή,

$$(r_{11}+\delta r_{11})\overline{x}_1 = b_1,$$

με $\delta r_{11} = \varepsilon_1' r_{11}$. Οπότε

$$\frac{|\delta r_{11}|}{|r_{11}|} \le \varepsilon_{μηχανής} + O(\varepsilon^2{}_{μηχανής})$$

*m=2*

Η περίπτωση $2\times2$ είναι λιγότερο τετριμμένη. Ας υποθέσουμε ότι έχουμε έναν άνω τριγωνικό πίνακα $R\in C^{2\times2}$ και ένα διάνυσμα $b\in C^2$. Ο υπολογισμός του $\overline{x}\in C^2$ γίνεται σε δυο βήματα. Το πρώτο βήμα είναι ίδιο με την προηγούμενη περίπτωση:

$$\overline{x}_2 = b_2 \div r_{22} = \frac{b_2}{r_{22}(1+\varepsilon_1)}, \qquad |\varepsilon_1| \le \varepsilon_{μηχανής} + O(\varepsilon^2{}_{μηχανής}) \tag{3.51}$$

Το δεύτερο βήμα δίνεται από τον τύπο:

$$\overline{x}_1 = (b_1 - (\overline{x}_2 \otimes r_{12})) \div r_{11}$$

Για να εξασφαλίσουμε προς τα πίσω ευστάθεια, πρέπει να εκφράσουμε τα σφάλματα ως προς αυτές τις τρεις πράξεις κινητής υποδιαστολής σαν διαταραχές των στοιχείων $r_{ij}$.

Ο πολλαπλασιασμός είναι εύκολος. Χρησιμοποιούμε το αξίωμα (3.23) ερμηνεύοντας τον πολλαπλασιαμό κινητής υποδιαστολής σαν μια διαταραχή του $r_{12}$:

$$\overline{x}_1 = (b_1 - \overline{x}_2 r_{12}(1+\varepsilon_2)) \div r_{11}, \qquad |\varepsilon_2| \le \varepsilon_{μηχανής}$$

Για την αφαίρεση και τον πολλαπλασιασμό πρώτα γράφουμε τον τύπο σύμφωνα με το αξίωμα (3.23)

$$\overline{x}_1 = (b_1 - \overline{x}_2 r_{12}(1+\varepsilon_2))(1+\varepsilon_3) \div r_{11} \tag{3.52}$$

$$= \frac{(b_1 - \overline{x}_2 r_{12}(1+\varepsilon_2))(1+\varepsilon_3)}{r_{11}}(1+\varepsilon_4) \tag{3.53}$$

Το σξίωμα (3.23) μας εξασφαλίζει ότι $|\varepsilon_3|, |\varepsilon_4| \le \varepsilon_{μηχανής}$. Τώρα μεταφέρουμε τους όρους $\varepsilon_3$ και $\varepsilon_4$ από τον αριθμητή στον παρονομαστή όπως πριν. Οπότε έχουμε

$$\overline{x}_1 = \frac{b_1 - \overline{x}_2 r_{12}(1+\varepsilon_2)}{r_{11}(1+\varepsilon_3')(1+\varepsilon_4')}$$

με $|\varepsilon_3'|, |\varepsilon_4'| \le \varepsilon_{μηχανής} + O(\varepsilon^2{}_{μηχανής})$ ή ισοδύναμα

$$\overline{x}_1 = \frac{b_1 - \overline{x}_2 r_{12}(1+\varepsilon_2)}{r_{11}(1+2\varepsilon_5)}$$

με $|\varepsilon_5| \le \varepsilon_{μηχανής} + O(\varepsilon^2{}_{μηχανής})$. Ο τύπος αυτός μας δείχνει ότι το $\overline{x}_1$ θα ήταν ακριβώς σωστό αν τα στοιχεία $r_{22}$, $r_{12}$ και $r_{11}$ διαταρασσόντουσαν από τους παράγοντες $(1+\varepsilon_1)$, $(1+\varepsilon_2)$ και $(1+2\varepsilon_5)$ αντίστοιχα. Αυτές οι διαταραχές μπορούν να περιληφθούν στην εξίσωση

$$(R+\delta R)\overline{x} = b,$$

όπου τα στοιχεία $\delta r_{ij}$ του πίνακα $\delta R$ ικανοποιούν τη σχέση

$$\begin{bmatrix} |\delta r_{11}|/|r_{11}| & |\delta r_{12}|/|r_{12}| \\ & |\delta r_{22}|/|r_{22}| \end{bmatrix} = \begin{bmatrix} 2|\varepsilon_5| & |\varepsilon_2| \\ & |\varepsilon_1| \end{bmatrix} \le \begin{bmatrix} 2 & 1 \\ & 1 \end{bmatrix} \varepsilon_{μηχανής} + O(\varepsilon^2{}_{μηχανής})$$

Η σχέση αυτή μας εξασφαλίζει ότι $\|\delta R\|/\|R\| = O(\varepsilon_{μηχανής})$ σε κάθε νόρμα πίνακα και γι' αυτό η $2\times 2$ προς τα πίσω αντικατάσταση είναι προς τα πίσω ευσταθής.

*m=3*

Η ανάλυση για ένα πίνακα $3\times 3$ περιλαμβάνει τη λογική που θα ακολουθήσουμε και στη γενική περίπτωση. Τα πρώτα δυο βήματα είναι ίδια με πριν:

$$\overline{x}_3 = b_3 \div r_{33} = \frac{b_3}{r_{33}(1+\varepsilon_1)} \tag{3.55}$$

$$\overline{x}_2 = (b_2 - (\overline{x}_3 \otimes r_{23})) \div r_{22} = \frac{b_2 - \overline{x}_3 r_{23}(1+\varepsilon_2)}{r_{22}(1+2\varepsilon_3)} \tag{3.56}$$

όπου

$$\begin{bmatrix} 2|\varepsilon_3| & |\varepsilon_2| \\ & |\varepsilon_1| \end{bmatrix} \le \begin{bmatrix} 2 & 1 \\ & 1 \end{bmatrix} \varepsilon_{\mu\eta\chi\alpha\nu\acute{\eta}\varsigma} + O(\varepsilon^2{}_{\mu\eta\chi\alpha\nu\acute{\eta}\varsigma})$$

Το τρίτο βήμα περιλαμβάνει τον υπολογισμό

$$\overline{x}_1 = [(b_1 - (\overline{x}_2 \otimes r_{12})) - (\overline{x}_3 \otimes r_{13})] \div r_{11} \qquad (3.57)$$

Μετατρέπουμε δυο πράξεις $\otimes$ στην (3.57) σε μαθηματικό πολλαπλασιαμό εισάγωντας τις διαταραχές $\varepsilon_4$ και $\varepsilon_5$ :

$$\overline{x}_1 = [(b_1 - \overline{x}_2 r_{12}(1+\varepsilon_4)) - \overline{x}_3 r_{13}(1+\varepsilon_5)] \div r_{11}$$

Μετατρέπουμε την πράξη του – σε μαθηματικές αφαιρέσεις εισάγωντας τις διαταραχές $\varepsilon_6$ και $\varepsilon_7$ :

$$\overline{x}_1 = [(b_1 - \overline{x}_2 r_{12}(1+\varepsilon_4))(1+\varepsilon_6) - \overline{x}_3 r_{13}(1+\varepsilon_5)](1+\varepsilon_7) \div r_{11}$$

Τέλος απαλοίφουμε τη διαίρεση χρησιμοποιώντας τη διαταραχή $\varepsilon_8$ την οποία αντικαθιστούμε αμέσως με την $\varepsilon_8^{'}$ με $|\varepsilon_8| \le \varepsilon_{\mu\eta\chi\alpha\nu\acute{\eta}\varsigma} + O(\varepsilon^2{}_{\mu\eta\chi\alpha\nu\acute{\eta}\varsigma})$ και βάζουμε το αποτέλεσμα στον παρονομαστή:

$$\overline{x}_1 = \frac{[(b_1 - \overline{x}_2 r_{12}(1+\varepsilon_4))(1+\varepsilon_6) - \overline{x}_3 r_{13}(1+\varepsilon_5)](1+\varepsilon_7)}{r_{11}(1+\varepsilon_8^{'})}$$

Η παραπάνω έκφραση έχει ό,τι χρειαζόμαστε εκτός από τους όρους που περιέχουν τις $\varepsilon_6$ και $\varepsilon_7$ . Αν αυτοί οι όροι διαταραχτούν τότε θα επηρεαστεί ο αριθμός $b_1$ ενώ ο στόχος μας είναι να διαταράξουμε μόνο τα στοιχεία $r_{ij}$ . Στον όρο που περιλαμβάνει το $\varepsilon_7$ μπορούμε να αντικαταστήσουμε το $\varepsilon_7$ με το $\varepsilon_7'$ και να το μετακινήσουμε στον παρονομαστή. Ο όρος που περιλαμβάνει το $\varepsilon_6$ χρειάζεται διαφορετική μεταχείρηση. Το μετακινούμε στον παρονομαστή αλλά κρατάμε την ισχύουσα ανισότητα και εισάγουμε τον παράγοντα $(1+\varepsilon_6')$ στον όρο $r_{13}$ .

Οπότε έχουμε

$$\overline{x}_1 = \frac{b_1 - \overline{x}_2 r_{12}(1+\varepsilon_4) - \overline{x}_3 r_{13}(1+\varepsilon_5)(1+\varepsilon_6')}{r_{11}(1+\varepsilon_6')(1+\varepsilon_7')(1+\varepsilon_8^{'})}$$

Τώρα το $r_{13}$ έχει δυο διαταραχές μεγέθους το πολύ $\varepsilon_{μηχανής}$ και το $r_{11}$ έχει τρεις διαταραχές. Στη σχέση αυτή όλα τα υπολογιστικά σφάλματα έχουν εκφραστεί σαν διαταραχές των στοιχείων του $R$.

Το αποτέλεσμα μπορεί να περιληφθεί ως

$$(R+\delta R)\overline{x}=b,$$

όπου τα στοιχεία του $\delta r_{ij}$ ικανοποιούν

$$\begin{bmatrix} |\delta r_{11}|/|r_{11}| & |\delta r_{12}|/|r_{12}| & |\delta r_{13}|/|r_{13}| \\ & |\delta r_{22}|/|r_{22}| & |\delta r_{23}|/|r_{23}| \\ & & |\delta r_{33}|/|r_{33}| \end{bmatrix} \le \begin{bmatrix} 3 & 1 & 2 \\ & 2 & 1 \\ & & 1 \end{bmatrix} \varepsilon_{μηχανής} + O(\varepsilon^2{}_{μηχανής})$$

*<u>ΓΕΝΙΚΟ m</u>*

Η ανάλυση σε περιπτώσεις μεγαλύτερης τάξης είναι παρόμοια. Για παράδειγμα, σε μια περίπτωση διάστασης $5\times 5$ παίρνουμε το όριο

$$\frac{|\delta R|}{|R|} \le \begin{bmatrix} 5 & 1 & 2 & 3 & 4 \\ & 4 & 1 & 2 & 3 \\ & & 3 & 1 & 2 \\ & & & 2 & 1 \\ & & & & 1 \end{bmatrix} \varepsilon_{μηχανής} + O(\varepsilon^2{}_{μηχανής}) \qquad (3.58)$$

Τα στοιχεία του πίνακα προκύπτουν από τρεις συνιστώσες. Οι πολλαπλασιασμοί $\overline{x}_k r_{jk}$ εισάγουν τις διαταραχές του $\varepsilon_{μηχανής}$ στη μορφή

$$\otimes: \begin{bmatrix} 0 & 1 & 1 & 1 & 1 \\ & 0 & 1 & 1 & 1 \\ & & 0 & 1 & 1 \\ & & & 0 & 1 \\ & & & & 0 \end{bmatrix} \qquad (3.59)$$

Οι διαιρέσεις με $r_{kk}$ εισάγουν διαταραχές στη μορφή

$$\div:\begin{bmatrix}1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & 1\end{bmatrix} \tag{3.60}$$

Τέλος, οι αφαιρέσεις μπορούν επίσης να εισάγουν διαταραχές της μορφής της (3.59) και λόγω του ότι οι υπολογισμοί γίνονται από αριστερά προς τα δεξιά κάθε αφαίρεση εισάγει μια διαταραχή στη διαγώνιο και σε κάθε θέση προς τα δεξιά. Οπότε προκύπτει η μορφή

$$-:\begin{bmatrix}4 & 0 & 1 & 2 & 3\\ & 3 & 0 & 1 & 2\\ & & 2 & 0 & 1\\ & & & 0 & 1\\ & & & & 0\end{bmatrix} \tag{3.61}$$

Προσθέτοντας τις (3.59),(3.60) και (3.61) προκύπτει η (3.58). Εδώ ολοκληρώνεται η απόδειξη του θεωρήματος 3.8.

*3.6.3 ΣΧΟΛΙΑ*

Η ανάλυση που μας οδήγησε στην (3.58) είναι κλασσική ανάλυση προς τα πίσω σφαλμάτων για όλων των ειδών υπολογισμών κινητής υποδιαστολής. Το βασικό στοιχείο είναι το αξίωμα (3.23) με το οποίο εξασφαλίζουμε ότι από κάθε πράξη προκύπτει ένα μικρό σχετικό σφάλμα. Οι διαταραχές της τάξης του $\varepsilon_{μηχανής}$ δημιουργούνται προσθετικά και μετακινούνται ελεύθερα ανάμεσα στους αριθμητές και τους παρονομαστές, αφού η διαφορά είναι της τάξης του $\varepsilon^2_{μηχανής}$.

Περισσότερα από ένα φράγματα σφάλματων μπορούν να προκύψουν από ένα δεδομένο αλγόριθμο. Στη συγκεκριμένη περίπτωση θα μπορούσαμε να έχουμε διαταράξει το $b_j$ καθώς και το $r_{ij}$, αποφεύγοντας τη μορφή της (3.61). Από την άλλη, ένα τελικό αποτέλεσμα στο οποίο ο $R$έχει διαταράχθεί είναι ξεκάθαρο.

Η εξίσωση (3.49) είναι ως προς τις συνιστώσες ένα φράγμα προς τα πίσω σφάλματος, δηλαδή το στοιχείο $r_{ij}$ διαταράσεται από μια ποσότητα η οποία είναι μικρή σε σχέση με τον εαυτό της, όχι μόνο σχέση με τη νόρμα του $R$. Για παράδειγμα, αν $r_{ij}=0$, το στοιχείο αυτό δεν διαταράσσεται καθόλου: η διαταραχή $\delta R$ έχει την ίδια μορφή με το $R$. Κάποιοι αλγόριθμοι της αριθμητικής γραμμικής άλγεβρας ικανοποιούν εκτιμήσεις προς τα πίσω σφαλμάτων ως προς τις συνιστώσες

αλλά κάποιοι άλλοι αλγόριθμοι δεν τις ικανοποιούν. Στους αλγόριθμους που δεν τις ικανοποιούν πρέπει να συμβιβαστούμε με μια σχέση της μορφής (3.48). Στις πρώτες δεκαετίες ανάπτυξης της αριθμητικής άλγεβρας μετά τον Δεύτερο Παγκόσμιο πόλεμο, οι περισσότερες εκτιμήσεις σφαλμάτων γινόντουσαν με τη μορφή νόρμας αλλά τα τελευταία χρόνια γίνεται μια ανάλυση ως προς τις συνιστώσες αφού τα αποτελέσματα είναι πιο ακριβή και οι αλγόριθμοι ικανοποιούν φράγματα ως προς τις συνιστώσες τα οποία είναι λιγότερα ευαίσθητα στην αναγωγή των μεταβλητών.
Κλείνουμε την παράγραφο αυτή με ένα σχόλιο σχετικά με τη σχέση ανάμεσα στα ποιοτικά φράγματα όπως το (3.49) ή το (3.58) και αυτά όπως το (3.48).
Γιατί όμως δεν μας αρκούν σχέσεις όπως η (3.48); Ο λόγος είναι ότι τα ποσοτικά φράγματα πρέπει να περιέχουν όρους όπως $\sqrt{m}$ ή $m$ , οι οποίοι είναι ανεξάρτητοι από νόρμες, ασήμαντοι και όχι πρακτικοί λόγω στατιστικής απαλοιφής. Προτιμούμε να αποφεύγουμε τέτοιες περιπλοκές εκφράζοντας τα περισσότερα αποτελέσματα σε όρους του $O(\varepsilon_{\mu\eta\chi\alpha\nu\acute{\eta}\varsigma})$ ο οποίος είναι πιο εύκολος στην απομνημόνευση. $m < n$

## 3.7 ΚΑΤΑΣΤΑΣΗ ΤΩΝ ΠΡΟΒΛΗΜΑΤΩΝ ΕΛΑΧΙΣΤΩΝ ΤΕΤΡΑΓΩΝΩΝ

Η κατάσταση των προβλημάτων ελαχίστων τετραγώνων συνδυάζεται με τη γεωμετρία των ορθογώνιων προβολών. Είναι ένα σημαντικό πεδίο των μαθηματικών διότι έχει μη τετριμμένες εφαρμογές στην ευστάθεια των αλγόρίθμων των προβλημάτων ελαχίστων τετραγώνων.

### *3.7.1 ΤΕΣΣΕΡΑ ΠΡΟΒΛΗΜΑΤΑ ΣΥΝΘΗΚΩΝ*

Στην παράγραφο αυτή θα χρησιμοποιήσουμε το πρόβλημα ελαχίστων τετραγώνων (2.47), το οποίο φαίνεται στο γράφημα 3.1. Υποθέτουμε ότι ο πίνακας που ορίζει το πρόβλημα είναι πλήρους τάξης και στην παράγραφο αυτή θα ισχύει $\|\cdot\| = \|\cdot\|_2$ :

Δοθέντος $A \in C^{m\times n}$ πλήρους τάξης, $m \geq n, b \in C^m$, (3.62)

βρείτε $x \in C^n$ τέτοιο ώστε η ποσότητα $\|b - Ax\|$ να ελαχιστοποιείται. .

Η λύση $x$ και το αντίστοιχο σημείο $y = Ax$ το οποίο είναι το πιο κοντινό στο $b$ στο πεδίο τιμών του $A$, $range(A)$, δίνονται από τις σχέσεις

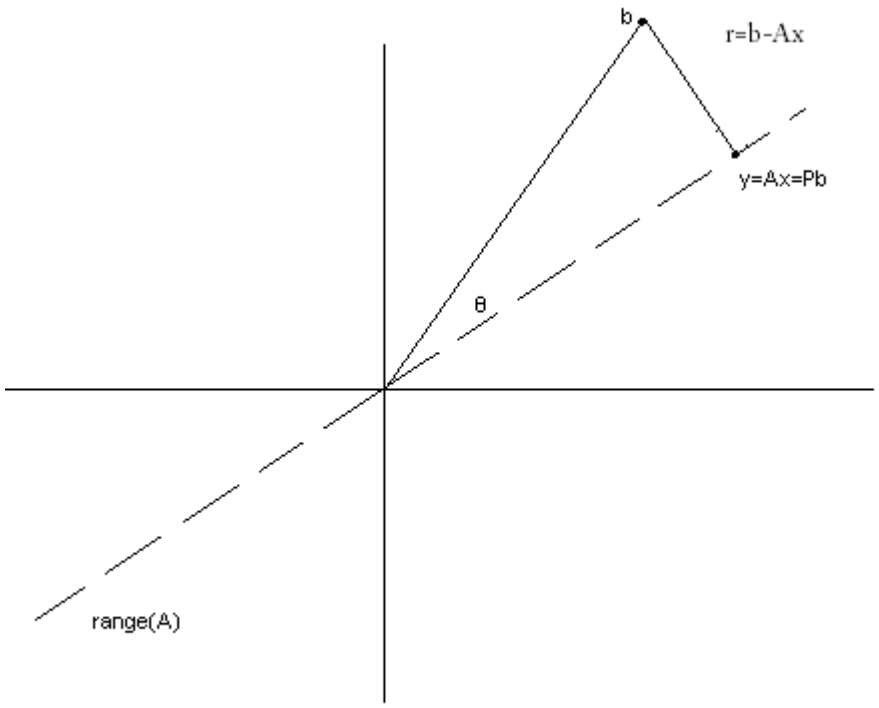
$$x = A^+ b\,, \qquad y = Pb \tag{3.63}$$

όπου $A^+ \in C^{n\times m}$ είναι ο ψευδοανάστροφος (2.56)++ του πίνακα $A$ και $P = AA^+ \in C^{m\times m}$ είναι ο ορθογώνιος προβολέας στο $range(A)$. Εξετάζουμε την κατάσταση του προβλήματος (3.62) ως προς τις διαταραχές. Επιλέγουμε το

πρόβλημα ελαχίστων τετραγώνων επειδή οι λεπτομέριες παρουσιάζουν ενδιαφέρον και επειδή έχουν σημαντικές πρακτικές συνέπειες τις οποίες θα δούμε στην επόμενη παράγραφο: η αστάθεια των κανονικών εξισώσεων σαν ένας γενικός σκοπός του αλγορίθμου των ελαχίστων τετραγώνων.
Η κατάσταση αναφέρεται στην ευαισθησία των λύσεων σε διαταραχές των δεδομένων. Για το (3.62) θα εξετάσουμε δυο επιλογές κάθε λύσης.Τα δεδομένα για το πρόβλημα είναι ο $m \times n$ πίνακας $A$ και το $m-$ διάνυσμα $b$ . Η λύση είναι είτε ο συντελεστής του διανύσματος $x$ είτε το αντίστοιχο σημείο $y = Ax$ . Οπότε:

Δεδομένα: $A, b$ Λύση: $x, y$

Μαζί, τα δυο ζεύγη επιλογών μας ορίζουν τέσσερα προβλήματα κατάστασης τα οποία θα εξετάσουμε και όλα έχουν εφαρμογές σε συγκεκριμένα πεδία.




Γράφημα 3.1 Το πρόβλημα των ελαχίστων τετραγώνων

*3.7.2 ΘΕΩΡΗΜΑ*

Το κύριο σημείο της παραγράφου αυτής είναι το θεώρημα 3.9. Τα αποτελέσματα που εκφράζονται με όρους τριών αδιάστατων παραμέτρων εμφανίζονται επανηλημένα στην ανάλυση των προβλημάτων των ελαχίστων τετραγώνων. Η πρώτη παράμετρος είναι ο δείκτης κατάστασης του πίνακα $A$. Για ένα τετραγωνικό πίνακα ο δείκτης κατάστασης είναι $\kappa(A) = \|A\|\|A^{-1}\|$ και στην τριγωνική περίπτωση, ο ορισμός γενικοποιείται στον (3.17):

$$\kappa(A) = \|A\|\|A^{+}\| \tag{3.64}$$

Η δεύτερη παράμετρος είναι η γωνία $\theta$ όπως φαίνεται στο γράφημα 3.1, η οποία είναι ένα μέτρο της εγγύτητας της προσαρμογής:

$$\theta = \cos^{-1}\frac{\|y\|}{\|b\|} \tag{3.65}$$

Η τρίτη παράμετρος είναι το κατά πόσο το $\|y\|$ αποκλίνει από τη μέγιστη δυνατή τιμή που μπορεί να πάρει δοθέντων των $\|A\|$ και $\|x\|$:

$$\eta = \frac{\|A\|\|x\|}{\|y\|} = \frac{\|A\|\|x\|}{\|Ax\|} \tag{3.67}$$

ΘΕΩΡΗΜΑ 3.9

Έστω $b \in C^m$ και $A \in C^{m\times n}$ πλήρους τάξης να είναι σταθερά. Το πρόβλημα ελαχίστων τετραγώνων (3.62) έχει τους ακόλουθους σχετικούς δείκτες κατάστασης νόρμας δεύτερης τάξης (3.5) οι οποίοι περιγράφουν τις ευαισθησίες των $y$ και $x$ σε σχέση με τις διαταραχές των $b$ και $A$:

| $y$ | $x$ | |
|---|---|---|
| $\dfrac{1}{\cos\theta}$ | $\dfrac{\kappa(A)}{\eta\cos\theta}$ | $b$ |
| $\dfrac{\kappa(A)}{\cos\theta}$ | $\kappa(A)+\dfrac{\kappa(A)^2\tan\theta}{\eta}$ | $A$ |

*Τα αποτελέσματα των πρώτων δυο γραμμών είναι ακριβή, έχουν πραγματοποιηθεί για συγκεκριμένες διαταραχές $\delta b$ και τα αποτελέσματα της δεύτερης σειράς είναι άνω φράγματα.*

Στην ειδική περίπτωση όπου $m = n$ η (3.62) μειώνετα σε ένα τετραγωνικό, ομαλό σύστημα εξισώσεων με $\theta = 0$. Στην περίπτωση αυτή, οι αριθμοί στη δεύτερη στήλη του θεωρήματος μειώνονται σε $\kappa(A)/\eta$ και $\kappa(A)$, οι οποίοι είναι τα ποτελέσματα που είχαν προκύψει στις (3.14) και (3.18) και ο αριθμός στην κάτω-αριστερή θέση μπορεί να αντικατασταθεί με το μηδέν.

### *3.7.3 ΜΕΤΑΤΡΟΠΗ ΣΕ ΕΝΑ ΔΙΑΓΩΝΙΟ ΠΙΝΑΚΑ*

Σαν ένα πρώτο βήμα για την απόδειξη του θεωρήματος 3.9 παρατηρούμε ότι το πρόβλημα ελαχίστων τετραγώνων μπορεί να αναλυθεί πιο εύκολα αν το τροποποιήσουμε σε μια βολική επιλογή βάσεων. Έστω $A$ να έχει μια παραγοντοποίηση SVD της μορφής $A = U\Sigma V^*$, όπου $\Sigma$ είναι ένας $m \times n$ διαγώνιος πίνακας με θετικά στοιχεία στη διαγώνιο. Επειδή οι διαταραχές υπολογίζονται σε νόρμα δεύτερης τάξης, τα μεγέθη τους δεν επηρεάζονται από μια ορθομοναδιαία αλλαγή βάσης, οπότε η διαταραγμένη συμπεριφορά του $A$ είναι ίδια με αυτή του $\Sigma$. Οπότε, χωρίς βλάβη της γενικότητας, μπορούμε να δουλέψουμε με τον $\Sigma$. Για να κλείσουμε τη συζήτης αυτή υποθέτουμε ότι $A = \Sigma$ και γράφουμε

$$A = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{bmatrix} = \begin{bmatrix} A_1 \\ 0 \end{bmatrix} \tag{3.68}$$

Ο $A_1$ είναι $n \times n$ και διαγώνιος και τα υπόλοιπα στοιχεία του $A$ είναι μηδενικά.
Η ορθογώνια προβολή του $b$ πάνω στο $range(A)$ είναι τετριμμένη. Γράφουμε

$$b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \quad ,$$

όπου ο $b_1$ περιέχει τα πρώτα $n$ στοιχεία του $b$. Τότε η προβολή του $y = Pb$ είναι

$$y = \begin{bmatrix} b_1 \\ 0 \end{bmatrix}$$

Για να βρούμε το αντίστοιχο για το $x$ μπορούμε να γράψουμε $Ax = y$ σαν

$$\begin{bmatrix} A_1 \\ 0 \end{bmatrix} x = \begin{bmatrix} b_1 \\ 0 \end{bmatrix} \quad ,$$

όπου προκύπτει

$$x = A_1^{-1} b_1 \tag{3.69}$$

Από τις σχέσεις αυτές είναι προφανές ότι ο ορθογώνιος προβολέας και ψευδοανάστροφος είναι πίνακες διάστασης $2\times 2$ και $1\times 2$

$$P = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}, \qquad A^+ = \begin{bmatrix} A_1^{-1} & 0 \end{bmatrix} \tag{3.70}$$

*3.7.4 ΕΥΑΙΣΘΗΣΙΑ ΤΟΥ y ΣΕ ΔΙΑΤΑΡΑΧΕΣ ΤΟΥ b*

Ξεκινάμε με το πιο απλό από τα τέσσερα προβλήματα κατάστασης. Από την (3.63), η σχέση ανάμεσα στα $b$ και $y$ είναι η γραμμική εξίσωση $y = Pb$. Ο Ιακωβιανός πίνακας της απεικόνισης $P$ είναι ο ίδιος ο $P$, με $\|P\| = 1$ από την (3.70). Από τις (3.6) και (3.65), ο δείκτης κατάστασης του $y$ ως προς τις διαταραχές του $b$ είναι

$$\kappa_{b\mapsto y} = \frac{\|P\|}{\|y\|/\|b\|} = \frac{1}{\cos\theta}$$

Αυτή η σχέση μας εξασφαλίζει το άνω-αριστερό αποτέλεσμα του θεωρήματος 3.9. Ο δείκτης κατάστασης πραγματοποιείται (δηλαδή το supremum στην (3.5) επιτυγχάνεται) για διαταραχές $\delta b$ με $\|P(\delta b)\| = \|\delta b\|$, το οποίο προκύπτει όταν η $\delta b$ είναι μηδενική εκτός από τα πρώτα $n$ στοιχεία.

*3.7.5 ΕΥΑΙΣΘΗΣΙΑ ΤΟΥ x ΣΕ ΔΙΑΤΑΡΑΧΕΣ ΤΟΥ b*

Η σχέση μεταξύ των $b$ και $x$ είναι επίσης γραμμική, $x = A^+ b$, με Ιακωβιανό τον $A^+$. Από τις (3.6), (3.65) και (3.66), ο δείκτης κατάστασης του $x$ σε σχέση με τις διαταραχές του $b$ είναι

$$\kappa_{b\mapsto x} = \frac{\|A^+\|}{\|x\|/\|b\|} = \|A^+\| \frac{\|b\|}{\|y\|} \frac{\|y\|}{\|x\|} = \|A^+\| \frac{1}{\cos\theta} \frac{\|A\|}{\eta} = \frac{\kappa(A)}{\eta\cos\theta}$$

Η σχέση αυτή μας εξασφαλίζει το άνω-δεξί αποτέλεσμα του θεωρήματος 3.9. Εδώ, ο δείκτης κατάστασης πραγματοποιείται από τις διαταραχές $\delta b$ ικανοποιώντας τη

σχέση $\left\|A^{+}(\delta b)\right\| = \left\|A^{+}\right\|\left\|\delta b\right\| = \left\|\delta b\right\|/\sigma_n$, η οποία προκύπτει όταν η $\delta b$ είναι μηδενική εκτός από το $n$-οστό στοιχείο (ή πιθανόν και άλλα στοιχεία αν ο πίνακας $A$ έχει παραπάνω από μια ιδιάζουσες τιμές ίσες με $\sigma_n$).

### *3.7.6 ΜΕΤΑΒΑΛΛΟΝΤΑΣ ΤΟ ΠΕΔΙΟ ΤΙΜΩΝ ΤΟΥ Α*

Η ανάλυση των διαταραχών του $A$ είναι ένα μη γραμμικό πρόβλημα. Θα μπορούσαμε να το λύσουμε υπολογίζοντας τις Ιακωβιανές αλγεβρικά αλλά θα το εξετάσουμε γεωμετρικά. Το σημείο εκκίνησης είναι η παρατήρηση ότι οι διαταραταχές του $A$ επηρεάζουν το πρόβλημα ελαχίστων τετραγώνων με δυο τρόπους: επηρεάζουν την απεικόνιση $C^n$ του πάνω στο $range(A)$ και αλλάζουν το ίδιο το $range(A)$. Ας μελετήσουμε τη δεύτερη περίπτωση.

Μπορούμε να φανταστούμε τις μικρές αλλαγές στο $range(A)$ σαν μικρές μεταβολές αυτού του διαστήματος. Το ερώτημα είναι ποιά είναι η μέγιστη τιμή της γωνίας που μπορεί να μεταβληθεί από μια μικρή διαταραχή του $\delta A$; Η απάντηση είναι η ακόλουθη. Η εικόνα στο $A$ της μοναδιαίας $n$-σφαίρας είναι μια υπερ-έλλειψη η οποία βρίσκεται σε οριζόντια θέση στο πεδίο τιμών του $A$. Για να αλλάξουμε το πεδίο τιμών του $A$ όσο το δυνατόν πιο αποτελεσματικά παίρνοθμε ένα σημείο $p = Au$ στην υπερ-έλλειψη (οπότε $\left\|u\right\| = 1$) και το τοποθετούμε σε μια διεύθυνση $\delta p$ ορθογώνιο στο $range(A)$. Μια διαταραχή πίνακα που μπορεί να το επιτύχει αυτό όσο πιο αποτελεσματικά γίνεται είναι η $\delta A \neq (\delta p)u^{*}$ η οποία μας δίνει $(\delta A)u = \delta p$ με $\left\|\delta A\right\| = \left\|\delta p\right\|$. Τώρα είναι προφανές ότι προκύπτει η μέγιστη μεταβολή για δοθέν $\left\|\delta p\right\|$ παίρνουμε το $p$ να είναι όσο το δυνατόν πιο κοντά στην αρχή. Δηλαδή, θέλουμε $p = \sigma_n u_n$, όπου το $\sigma_n$ είναι η μικρότερη ιδιάζουσα τιμή του πίνακα $A$ και το $u_n$ είναι το αντίστοιχο αριστερό ιδιάζον διάνυσμα. Όταν ο πίνακας $A$ είναι στη διαγώνια μορφή (3.68) τότε το $p$ είναι ίσο με την τελευταία στήλη του $A$, $u^{*}$ είναι το $n$-διάνυσμα (0,0,0,….,0,1) και $\delta A$ είναι μια διαταραχή των στοιχείων του πίνακα $A$ κάτω από τη διαγώνιο σε αυτή τη στήλη. Μια τέτοια διαταραχή μεταβάλλει $range(A)$ κατά μια γωνία $\dfrac{\left\|\delta y\right\|}{\left\|y\right\|} \Big/ \dfrac{\left\|\delta A_1\right\|}{\left\|A\right\|} \leq \dfrac{\kappa(A)}{\cos\theta}$ η οποία δίνεται από $(\delta a) = \left\|\delta p\right\|/\sigma_n$. Αφού $\left\|\delta A\right\| = \left\|\delta p\right\|$ και $\delta a \leq \tan(\delta a)$, έχουμε

$$\delta a \leq \frac{\left\|\delta A\right\|}{\sigma_n} = \frac{\left\|\delta A\right\|}{\left\|A\right\|}\kappa(A) \tag{3.71}$$

και ισότητα έχουμε όταν επιλέγουμε τέτοιες διαταραχές $\delta A$ όπως περιγράψαμε παραπάνω με την προϋπόθεση ότι είναι απειροελάχιστες (έτσι ώστε $\delta a = \tan(\delta a)$).

*3.7.7 ΕΥΑΙΣΘΗΣΙΑ ΤΟΥ y ΣΕ ΔΙΑΤΑΡΑΧΕΣ ΤΟΥ A*

Τώρα θα εξετάσουμε πώς τη δεύτερη γραμμή του πίνακα στο θεώρημα 3.9. Θα ξεκινήσουμε με το αριστερό στοιχείο. Αφού το $y$ είναι η ορθογώνια προβολή του $b$ στο $range(A)$ προκύπτει μόνο από το $b$ και το $range(A)$. Οπότε για να αναλύσουμε την ευαισθησία του $y$ στις διαταραχές του $A$ μπορούμε απλά να εξετάσουμε την επίδραση στο $y$ της αλλαγής του $range(A)$ για κάποια γωνία $\delta a$.
Μια γεωμετρική ιδιότητα γίνεται εμφανής όταν φανταζόμαστε ότι κατασκευάζουμε το $b$ και παρακολουθούμε το $y$ να διαφέρει καθώς μεταβάλλεται το $range(A)$ (σχήμα 3.2). Όπως και αν μεταβληθεί το $range(A)$ το διάνυσμα $y \in range(A)$ πρέπει πάντα να είναι ορθογώνιο στο $y-b$. Δηλαδή, η γραμμή $b-y$ πρέπει να είναι στη δεξιά γωνία της γραμμής $0-y$. Με άλλα λόγια, καθώς το $range(A)$ προσαρμόζεται, το $y$ κινείται κατά μήκος της σφαίρας ακτίνας $\|b\|/2$ με κέντρο το σημείο $b/2$.
Μεταβάλλοντας το $range(A)$ στο επίπεδο $0-b-y$ κατά μια γωνία $\delta a$ μεταβάλλεται η γωνία $2\theta$ στο κεντρικό σημείο $b/2$ κατά $2\delta a$. Οπότε η αντίστοιχη διαταραχή $\delta y$ είναι η βάση ενός ισοσκελούς τριγώνου με κεντρική γωνία $2\delta a$ και μήκος πλευράς $\|b\|/2$. Από αυτό προκύπτει ότι $\|\delta y\| = \|b\|\sin(\delta a)$. Μεταβάλλοντας το $range(A)$ σε οποιαδήποτε άλλη κατεύθυνση έχουμε σαν αποτέλεσμα μια παρόμοια γεωμετρία σε ένα διαφορετικό επίπεδο και διαταραχές μικρότερες από ένα παράγοντα τόσο μικρό όσο το $\sin\theta$. Οπότε για τυχαίες διαταραχές κατά μια γωνία $\delta a$ έχουμε

$$\|\delta y\| \le \|b\|\sin(\delta a) \le \|b\|\,\delta a \tag{3.72}$$

Από τις (3.65) και (3.71) έχουμε $\|\delta y\| \le \|\delta A\|\kappa(A)\|y\|/\|A\|\cos\theta$, δηλαδή

$$\frac{\|\delta y\|}{\|y\|} \Big/ \frac{\|\delta A_1\|}{\|A\|} \le \frac{\kappa(A)}{\cos\theta} \tag{3.73}$$

Αυτό μας εξασφαλίζει το κάτω-αριστερά αποτέλεσμα του θεωρήματος 3.9.

Σχήμα 3.2 Δυο κύκλοι στη σφαίρα όπου το $y$ κινείται καθώς το $range(A)$ μεταβάλλεται. Ο μεγάλος κύκλος ακτίνας $\|b\|/2$ αντιστοιχεί στη μεταβολή του $range(A)$ στο επίπεδο $0-b-y$, και ο μικρός κύκλος ακτίνας $(\|b\|/2)\sin\theta$, αντιστοιχεί σε μια μεταβολή κατά μια ορθογώνια διεύθυνση. Όμως το $range(A)$ μεταβάλλεται, το $y$ παραμένει στη σφαίρα ακτίνας $\|b\|/2$ με κέντρο το $b/2$.

*3.7.8 ΕΥΑΙΣΘΗΣΙΑ ΤΟΥ x ΣΕ ΔΙΑΤΑΡΑΧΕΣ ΤΟΥ A*

Τώρα θα αναλύσουμε την πιο ενδιαφέρουσα σχέση του θεωρήματος 3.9: την ευαισθησία του x στις διαταραχές του πίνακα $A$.
Μια διαταραχή $\delta a$ χωρίζεται με φυσικό τρόπο σε δυο μέρη: το ένα μέρος είναι το $\delta A_1$ στις πρώτες $n$ γραμμές του $A$ και το άλλο μέρος είναι το $\delta A_2$ στις υπόλοιπες $m-n$ γραμμές:

$$\delta A = \begin{bmatrix} \delta A_1 \\ \delta A_2 \end{bmatrix} = \begin{bmatrix} \delta A_1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \delta A_2 \end{bmatrix}$$

Πρώτα, ας εξετάσουμε την επιρροή των διαταραχών $\delta A_1$. Μια τέτοια διαταραχή μεταβάλλει την απεικόνιση του πίνακα $A$ στο πεδίο τιμών της αλλά όχι στο πεδίο τιμών του $A$ ή του $y$. Διαταράσσει το $A_1$ κατά $\delta A_1$ στο τετράγωνο σύστημα (3.69) χωρίς να αλλάζει το $b_1$. Ο δείκτης κατάστασης για τέτοιες διαταραχές δίνεται από τη σχέση (3.18) όπου εδώ έχει τη μορφή:

$$\frac{\|\delta x\|}{\|x\|} \Big/ \frac{\|\delta A_1\|}{\|A\|} \le \kappa(A_1) = \kappa(A) \qquad (3.74)$$

Στη συνέχεια εξετάζουμε την επιρροή των (απειροελάχιστων) διαταραχών $\delta A_2$. Μια τέτοια διαταραχή μεταβάλλει το $range(A)$ χωρίς να αλλάζει την απεικόνιση του πίνακα $A$ σε αυτό το διάστημα. Το σημείο $y$ και οπότε το διάνυσμα $b_1$ διαταρράσονται αλλά όχι το $A_1$. Αυτό αντιστοιχεί στη διαταραχή του $b_1$ στην (3.69) χωρίς να αλλάζει το $A_1$. Ο δείκτης κατάστασης για αυτή τη διαταραχή δίνεται από τη σχέση (3.14) και έχει τη μορφή:

$$\frac{\|\delta x\|}{\|x\|} \Big/ \frac{\|\delta b_1\|}{\|b_1\|} \le \frac{\kappa(A_1)}{\eta(A_1; x)} = \frac{\kappa(A)}{\eta} \qquad (3.75)$$

Για να ολοκληρώσουμε χρειάζεται να σχετίσουμε τα $\delta b_1$ και $\delta A_2$. Τώρα το διάνυσμα $y$ εκφράζεται από τις συνιστώσες του $range(A)$. Οπότε, οι μόνες μεταβολές του $y$ που πραγματοποιούνται σαν αλλαγές στο $b_1$ είναι αυτές που είναι παράλληλες στο $range(A)$. Οι ορθογώνιες αλλαγές δεν επηρεάζουν. Πιο συγκεκριμένα, αν το $range(A)$ μεταβάλλεται κατά μια γωνία $\delta a$ στο επίπεδο $0-b-y$ η διαταραχή $\delta y$ που θα προκύψει δεν θα είναι παράλληλη στο $range(A)$

αλλά σε μια γωνία $\pi/2-\theta$. Κατά συνέπεια, η μεταβολή στο $b_1$ ικανοποιεί τη σχέση $\|\delta b_1\| = \sin\theta \|\delta y\|$. Από τη (3.72) έχουμε

$$\|\delta b_1\| = (\|b\|\delta a)\sin\theta \tag{3.76}$$

Αν το $range(A)$ μεταβληθεί κατά τη διεύθυνση που είναι ορθογώνια στο επίπεδο $0-b-y$, προκύπτει το ίδιο φράγμα αλλά για διαφορετικό λόγο. Τώρα η $\delta y$ είναι παράλληλη στο $range(A)$ αλλά είναι κατά ένα παράγοντα $\sin\theta$ μικρότερη. Οπότε έχουμε $\|\delta y\| = (\|b\|\delta a)\sin\theta$ και αφού $\|\delta b_1\| \le \|\delta y\|$ προκύπτει πάλι η (3.76).

Αφού $\|b_1\| = \|b\|\cos\theta$ μπορούμε να ξαναγράψουμε την (3.76) ως

$$\frac{\|\delta b_1\|}{\|b_1\|} \le (\delta a)\tan\theta \tag{3.77}$$

Σχετίζοντας τη $\delta a$ με το $\|\delta A_2\|$ στην (3.71) και συνδυάζοντας τις (3.75) και (3.76) παίρνουμε σαν αποτέλεσμα

$$\frac{\|\delta x\|}{\|x\|} \Big/ \frac{\|\delta A_2\|}{\|A\|} \le \frac{\kappa(A)^2\tan\theta}{\eta}$$

Προσθέτοντας αυτό το αποτέλεσμα στην (3.74) εξασφαλίζουμε το κάτω-δεξί αποτέλεσμα του θεωρήματος 3.9

## 3.8 ΕΥΣΤΑΘΕΙΑ ΤΩΝ ΑΛΓΟΡΙΘΜΩΝ ΤΩΝ ΠΡΟΒΛΗΜΑΤΩΝ ΤΩΝ ΕΛΑΧΙΣΤΩΝ ΤΕΤΡΑΓΩΝΩΝ

Τα προβλήματα ελαχίστων τετραγώνων μπορούν να επιλυθούν με διάφορες μεθόδους όπως είδαμε στην παράγραφο 2.5. Κάποιες από τις μεθόδους αυτές είναι οι κανονικές εξισώσεις, η Householder τριγωνοποίηση, η Gram-Schmidt ορθοκανονικοποίηση και η παραγοντοποίηση SVD. Στην παράγραφο αυτή θα συγκρίνουμε τις μεθόδους αυτές και θα δείξουμε ότι η χρήση της μεθόδου των κανονικών εξισώσεων είναι γενικά ασταθής.

### *3.8.1 HOUSEHOLDER ΤΡΙΓΩΝΟΠΟΙΗΣΗ*

Όπως είδαμε και στην παράγραφο 2.5, ο κλασσικός αλγόριθμος για την επίλυση προβλημάτων ελαχίστων τετραγώνων είναι η QR παραγοντοποίηση μέσω της Householder τριγωνοποίησης (Αλγόριθμος 2.7).

ΘΕΩΡΗΜΑ 3.10
Έστω ότι λύνουμε το πρόβλημα ελαχίστων τετραγώνων πλήρους τάξης (2.47) χρησιμοποιώντας την Householder τριγωνοποίηση (Αλγόριθμος 2.7) σε έναν υπολογιστή ικανοποιώντας τα αξιώματα (3.21) και (3.23). Ο αλγόριθμος είναι προς τα πίσω ευσταθής με την έννοια ότι η υπολογισμένη λύση $\overline{x}$ έχει την ιδιότητα

$$\left\|(A+\delta A)\overline{x}-b\right\| = \min\,, \qquad \frac{\left\|\delta A\right\|}{\left\|A\right\|} = O(\varepsilon_{\mu\eta\chi\alpha\nu\acute{\eta}\varsigma}) \tag{3.78}$$

για κάποια $\delta A \in C^{m\times n}$. Αυτό ισχύει είτε το $\hat{Q}^* b$ έχει υπολογιστεί μέσω αναλυτικής μορφοποίησης του $\hat{Q}$ ή μέσω της εφαρμογής του αλγορίθμου 2.4. Επίσης ισχύει για την Householder τριγωνοποίηση με επιλογή τυχαίας οδηγού-στήλης.

### 3.8.2 *GRAM-SCHMIDT ΟΡΘΟΚΑΝΟΝΙΚΟΠΟΙΗΣΗ*

Μια άλλη μέθοδος για την επίλυση προβλημάτων ελαχίστων τετραγώνων είναι η τροποποιημένη ορθοκανονικοποίηση Gram-Schmidt (Αλγόριθμος 2.2). Για $m \approx n$, η μέθοδος αυτή απαιτεί περισσότερες πράξεις απ' ότι η Householder προσέγγιση αλλά για $m >> n$ η πολυπλοκότητα και των δυο αλγορίθμων είναι ασυμπτωτικά ίση με $2mn^2$.

ΘΕΩΡΗΜΑ 3.11
Η λύση του προβλήματος ελαχίστων τετραγώνων πλήρους τάξης (2.47) χρησιμοποιώντας την Gram-Schmidt ορθοκανονικοποίηση είναι επίσης προς τα πίσω ευσταθής και ικανοποιεί την (3.78) με την προϋπόθεση ότι το γινόμενο $\hat{Q}^* b$ έχει μορφοποιηθεί αναλυτικά.

### *3.8.3 ΚΑΝΟΝΙΚΕΣ ΕΞΙΣΩΣΕΙΣ*

Μια θεμελιωδώς διαφορτική προσέγγιση για την επίλυση προβλημάτων ελαχίστων τετραγώνων είναι η λύση που προκύπτει από την εφαρμογή της μεθόδου των

κανονικών εξισώσεων (Αλγόριθμος 2.6), τυπικά από την παραγοντοποίηση Cholesky. Για $m >> n$ η μέθοδος αυτή είναι δυο φορές πιο γρήγορη από τις μεθόδους που απαιτούν αναλυτική ορθοκανονικοποίηση και απαιτούν ασυμπτωτική πολυπλοκότητα ίση με $mn^2$ (2.54).

Θα εξηγήσουμε γιατί η μέθοδος των κανονικών εξισώσεων είναι μια ασταθής μέθοδος για την επίλυση προβλημάτων ελαχίσρων τετραγώνων.
Ας υποθέσουμε ότι έχουμε ένα προς τα πίσω ευσταθή αλγόριθμο για το πρόβλημα ελαχίστων τετραγώνων (2.47) πλήρους τάξης και έστω ότι ο αλγόριθμος αυτός μας δίνει μια λύση $\bar{x}$ η οποία ικανοποιεί τη σχέση $\|(A+\delta A)\bar{x}-b\| = \min$ για κάποιο $\delta A$ με $\frac{\|\delta A\|}{\|A\|} = O(\varepsilon_{μηχανής})$ . Από τα θεωρήματα (3.3) και (3.9) έχουμε

$$\frac{\|\bar{x}-x\|}{\|x\|} = O((\kappa + \frac{\kappa^2 \tan\theta}{\eta})\varepsilon_{μηχανής}) \qquad (3.79),$$

όπου $\kappa = \kappa(A)$ . Τώρα ας υποθέσουμε ότι ο $A$ είναι ασθενώς ορισμένος (π.χ. $\kappa >> 1$) και $\theta$ να είναι φραγμένη με $\pi/2$ . Εξαρτώμενοι από τις τιμές των διαφόρων παραμέτρων μπορούν να προκύψουν δυο πολύ διαφορετικές περιπτώσεις. Αν $\tan\theta$ είναι της τάξης του 1 και $\eta << \kappa$ τότε το αριστερό μέλος της (3.79) θα είναι $O(\kappa^2 \varepsilon_{μηχανής})$ . Από την άλλη, αν $\tan\theta$ είναι κοντά στο μηδέν ή το $\eta$ είναι πολύ κοντά στο $\kappa$ , το φράγμα γίνεται $O(\kappa\varepsilon_{μηχανής})$ . *Ο δείκτης κατάστασης του πρόβληματος ελαχίστων τετραγώνων βρίσκεται στο διάστημα ανάμεσα στο $\kappa$ και το $\kappa^2$.*
Τώρα ας σκεφτούμε τι συμβαίνει όταν λύνουμε το πρόβλημα (2.47) με τη χρήση κανονικών εξισώσεων, $(A^*A)x = A^*b$ . Η παραγοντοποίηση Cholesky είναι ένας σταθερός αλγόριθμος για αυτό το σύστημα εξισώσεων, με την έννοια ότι παράγει μια λύση $\bar{x}$ η οποία ικανοποιεί $(A^*A+\delta H)\bar{x} = A^*b$ για κάποιο $\delta H$ με $\frac{\|\delta H\|}{\|A^*A\|} = O(\varepsilon_{μηχανής})$ . Όμως, ο πίνακας $A^*A$ έχει δείκτη κατάστασης $\kappa^2$ και όχι $\kappa$.
Οπότε το καλύτερο που μπορούμε να περιμένουμε για τις κανονικές εξισώσεις είναι

$$\frac{\|\bar{x}-x\|}{\|x\|} = O(\kappa^2 \varepsilon_{μηχανής}) \qquad (3.80)$$

*Η συμπεριφορά των κανονικών εξισώσεων κυριαρχείται από το $\kappa^2$ και όχι από το $\kappa$.*

Τώρα το συμπερασμα είναι ξεκάθαρο. Αν $\tan\theta$ είναι της τάξης του 1 και $\eta << \kappa$ ή αν ο $\kappa$ είναι της τάξης του 1 τότε οι (3.79) και (3.80) είναι της ίδιας τάξης και η

μέθοδος των κανονικών εξισώσεων είναι ευσταθής. Αν το $\kappa$ είναι μεγάλο και είτε η $\tan\theta$ είναι κοντά στο μηδέν ή το $\eta$ είναι κοντά στο $\kappa$ τότε η (3.80) είναι πολύ πιο μεγάλη από την (3.79) και η μέθοδος των κανονικών εξισώσεων είναι ασταθής. *Η μέθοδος των κανονικών εξισώσεων είναι ασταθής για ασθενώς ορισμένα προβλήματα.*

Συμφωνα με τους ορισμούς που δώσαμε, ένας αλγόριθμος είναι ευσταθής αν έχει ικανοποιητικά ομαλή συμπεριφορά σε όλα τα προβλήματα. Το επόμενο θεώρημα είναι μια μορφοποίηση των παρατηρήσεων που κάναμε πριν.

ΘΕΩΡΗΜΑ 3.12

Η επίλυση του προβλήματος ελαχίστων τετραγώνων πλήρους τάξης (2.47) μέσω των κανονικών εξισώσεων (Αλγόριθμος 2.6) είναι ασταθής. Η ευστάθεια μπορεί να επιτευχθεί με τον περιορισμό σε μια κατηγορία προβλημάτων στα οποία ο $\kappa(A)$ είναι ομαλά άνω φραγμένος ή $\tan\theta/\eta$ είναι ομαλά κάτω φραγμένο.

### *3.8.4 SVD*

Ένας άλλος αλγόριθμος που μπορεί να χρησιμοποιηθεί για την επίλυση προβλημάτων ελαχίστων τετραγώνων είναι αυτός την παραγοντοποίησης SVD.

ΘΕΩΡΗΜΑ 3.13

Η προβλημάτων ελαχίστων τετραγώνων πλήρους τάξης (2.47) μέσω της παραγοντοποίησης SVD (Αλγόριθμος 2.8) είναι προς τα πίσω ευσταθής και ικανοποιεί την εκτίμηση (3.78).

### *3.8.5 ΠΡΟΒΛΗΜΑΤΑ ΕΛΑΧΙΣΤΩΝ ΤΕΤΡΑΓΩΝΩΝ ΧΑΜΗΛΗΣ ΤΑΞΗΣ*

Στην παράγραφο αυτή είδαμε τέσσερεις προς τα πίσω ευσταθείς αλγόριθμους που εφαρμόζονται για την επίλυση γραμμικών προβλημάτων ελαχίστων τετραγώνων. Η τριγωνοποίηση Householder, η τριγωνοποίηση Householder με τη χρήση οδηγού-στήλης, η τροποποιημένη μέθοδος Gram-Schmidt με αναλυτικό υπολογισμό του $\hat{Q}^* b$ και η παραγοντοποίηση SVD. Από τη σκοπιά της κλασσικής ανάλυσης του προβλήματος πλήρους τάξης (2.47) με βάση τη νόρμα και την ευστάθεια, οι διαφορές ανάμεσα στους αλγόριθμους είναι αμελητέες και μπορούμε να εφαρμόσουμε τον αλγόριθμο που είναι πιο γρήγορος και έχει τη μικρότερη πολυπλοκότητα, ο οποίος είναι η τριγωνοποίηση Householder χωρίς οδήγηση.

Όμως, υπάρχουν και άλλα είδη προβλημάτων ελαχίστων τετραγώνων όπου η οδηγός-στήλη και η παραγοντοποίηση SVD έχουν ιδιαίτερη σημασία. Αυτά είναι τα προβλήματα όπου ο πίνακας $A$ έχει ταξη $< n$, πιθανότατα $m < n$ έτσι ώστε το σύστηματων εξισώσεων να είναι υποκαθορισμένο. Τέτοια προβλήματα δεν έχουν

μοναδική λύση εκτός και αν κάποιο έχει κάποια επιπρόσθετη υπόθεση, όπως το $x$ να έχει όσο πιο μικρή νόρμα γίνεται. Μια περαιτέρω περιπλκή είναι ότι η σωστή λύση εξαρτάται από την τάξη του πίνακα $A$ και το να καθορίζουμε την τάξη π΄νακα ενώ έχουμε σφάλματα από στρογγυλοποίηση δεν είναι ποτέ μια τετριμμένη περίπτωση.

Γι' αυτό τα προβλήματα ελαχίστων τετραγώνων χαμηλής τάξης δεν είναι μια ενδιαφέρουσα υποκατηγορία προβλημάτων ελαχίστων τετραγώνων αλλά είναι θεμελιωθώς διαφορετική. Αφού ο ορισμός μιας λύσης είναι καινούργιος, δεν υπάρχει λόγος ένας αλγόριθμος ο οποίος είναι ευσταθής για προβλήματα πλήρους τάξης να πρέπει να είναι επίσης ευσταθής και για προβλήματα χαμηλής τάξης. Για την ακρίβεια, οι μόνοι ευσταθείς αλγόριθμοι για προβλήματα χαμηλής τάξης είναι αυτοί που βασίζονται στην παραγοντοποίηση SVD. Μια εναλλακτική μέθοδος είναι η τριγωνοποίηση householder με χρήση οδηγού-στήλης η οποία μέθοδος είναι ευσταθής για σχεδόν όλα τα προβλήματα.

**Παράδειγμα στο Matlab**

Θέλουμε να βρούμε τη λύση *x* του προβλήματος *Ax=b.*

Θέτουμε για *m=100* , *n=15* και *t=(0:m-1)'/(m-1)*

*A=[A t.^(i-1)]* και *y=exp(sin(4*t))*

Κανονικοποιούμε με *b=b/2006.787453080206*

Λύνουμε πρώτα με τη μέθοδο SVD και στη συνέχεια με την QR.

Θα συγκρίνουμε τους 2 αλγόριθμους με βάση το αποτέλεσμα

- SVD

Το αποτέλεσμα είναι $x_{15} = 0.99999998230471$

- QR

Το αποτέλεσμα είναι $x_{15} = 1.00000031528723$

| *y* | *x* | |
|---|---|---|
| $1.0$ | $1.1\times10^{5}$ | *b* |
| $2.3\times10^{10}$ | $3.2\times10^{10}$ | *A* |

## *Σχόλια*

Χάρη στην κανονικοποίηση η σωστή απάντηση θα ήταν $x_{15} = 1$. Οπότε όταν εφαρμόζουμε την QR παραγοντοποίηση έχουμε ένα σχετικό σφάλμα της τάξης $3\times10^{-7}$. Ο υπολογισμός αυτός έγινε με αριθμητική διπλής ακρίβειας με $\varepsilon_{\mu\eta\chi\alpha\nu\acute{\eta}\varsigma} \approx 10^{-16}$. Αυτό σημαίνει ότι τα σφάλματα στρογγυλοποίησης έχουν αυξηθεί κατά ένα παράγοντα της τάξης $10^{9}$. Ο δείκτης κατάστασης του *x* όταν

διαταράσσουμε τον A είναι της τάξης του $10^{10}$. Άρα η ανακρίβεια του $x_{15}$ μπορεί να εξηγηθεί λόγω κακής κατάστασης και όχι λόγω αστάθειας. Ο αλγόριθμος αυτός είναι προς τα πίσω ευσταθής. Το αποτέλεσμα που παίρνουμε όταν εφαρμόζουμε την SVD παραγοντοποίηση είναι πολύ πιο ακριβές και κατά συνέπεια ο αλγόριθμος αυτός είναι ευσταθής.

# ΒΙΒΛΙΟΓΡΑΦΙΑ

[1] Numerical Linear Algebra, Lloyd N. Trefethen-David Bau III, Siam
[2] Σημειώσεις Ανάλυση Πινάκων, Ιωάννης Β. Μαρουλάς
[3] Αριθμητική ανάλυση με εφαρμογές σε Matlab και Mathematica, Παπαγεωργίου-Τσίτουρας, Εκδόσεις Συμεών
[4] Εισαγωγή στην αριθμητική ανάλυση, Χρυσοβέρης-Μπακόπουλος, Εκδόσεις Συμεών