Αριθμητική Ανάλυση Ελάχιστα τετράγωνα

Αναστάσιος Τέφας
tefas@aiia.csd.auth.gr
2310-991932

Ελάχιστα τετράγωνα

Έστω ότι δίνονται τα σημεία (x_i, f_i), $i = 1,\ldots,n$ και θέλουμε να προσδιορίσουμε ένα πολυώνυμο βαθμού 1 της μορφής $y = at + b$, έτσι ώστε το άθροισμα των τετραγώνων:

$$E(a,b) = \sum_{k=1}^{n}\left(f_i - (at_i + b)\right)^2 = \text{ελάχιστο.}$$

Αναγκαία συνθήκη για να ισχύει αυτό είναι:

$$\frac{\partial E(a,b)}{\partial a} = 0, \quad \frac{\partial E(a,b)}{\partial b} = 0,$$

$$\begin{cases} nb + a\sum_{i=1}^{n} t_i = \sum_{i=1}^{n} f_i \\ a\sum_{i=1}^{n} t_i^2 + b\sum_{i=1}^{n} t_i = \sum_{i=1}^{n} f_i \end{cases}$$

Η ορίζουσα των συντελεστών του παραπάνω συστήματος είναι πάντα διάφορη του μηδενός, οπότε το σύστημα έχει μοναδική λύση. Το πολυώνυμο που προκύπτει καλείται *πολυώνυμο 1ου βαθμού ελαχίστων τετραγώνων* και χρησιμοποιείται ευρέως στη στατιστική για τη συσχέτιση μεταξύ δύο ποσοτήτων. Ομοίως, αν θέλουμε να προσδιορίσουμε ένα πολυώνυμο $a_0 + a_1 x + ... + a_m x^m$ βαθμού m, ώστε το άθροισμα των τετραγώνων να είναι ελάχιστο, καλούμαστε να λύσουμε ένα ανάλογο σύστημα $(m+1)\times(m+1)$. Στη συνέχεια θα δούμε πιο αναλυτικά την τεχνική προσέγγισης διακριτών ελαχίστων τετραγώνων καθώς και βέλτιστες προσεγγίσεις σε ευκλείδιους χώρους.

Έχουμε δει στο Κεφάλαιο 3 τι πρέπει να ισχύει για ένα γραμμικό σύστημα ώστε να έχει λύση και προφανώς τα ίδια ισχύουν όταν δεν έχει καμία λύση και είναι αδύνατο. Για παράδειγμα το παρακάτω σύστημα είναι αδύνατο μιας και η $1^η$ με την $3^η$ εξίσωση δεν μπορούν να ικανοποιούνται ταυτόχρονα.

$$x_1 + x_2 = 2$$

$$x_1 - x_2 = 1$$

$$x_1 + x_2 = 3$$

Τι σημαίνει όταν λέμε ότι ένα σύστημα είναι αδύνατο; Μπορεί για παράδειγμα οι συντελεστές να εμπεριέχουν κάποιο σφάλμα και να μας οδηγούν σε αυτό το συμπέρασμα χωρίς να μπορούμε να δώσουμε λύση με τις μεθόδους που βασίζονται στην απαλοιφή κατά Gauss. Αυτό που θα κάνουμε είναι να βρούμε ένα διάνυσμα λύσεων x που είναι ό,τι ***κοντινότερο*** σε λύση.

Αν αυτή την «κοντινότητα» την εκφράσουμε με τη βοήθεια της Ευκλείδιας απόστασης, τότε υπάρχει ένας απλός αλγόριθμος να βρίσκουμε αυτό το x. Αυτή η λύση σε αυτή την περίπτωση είναι η λύση ***ελαχίστων τετραγώνων.***

$$\begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$$

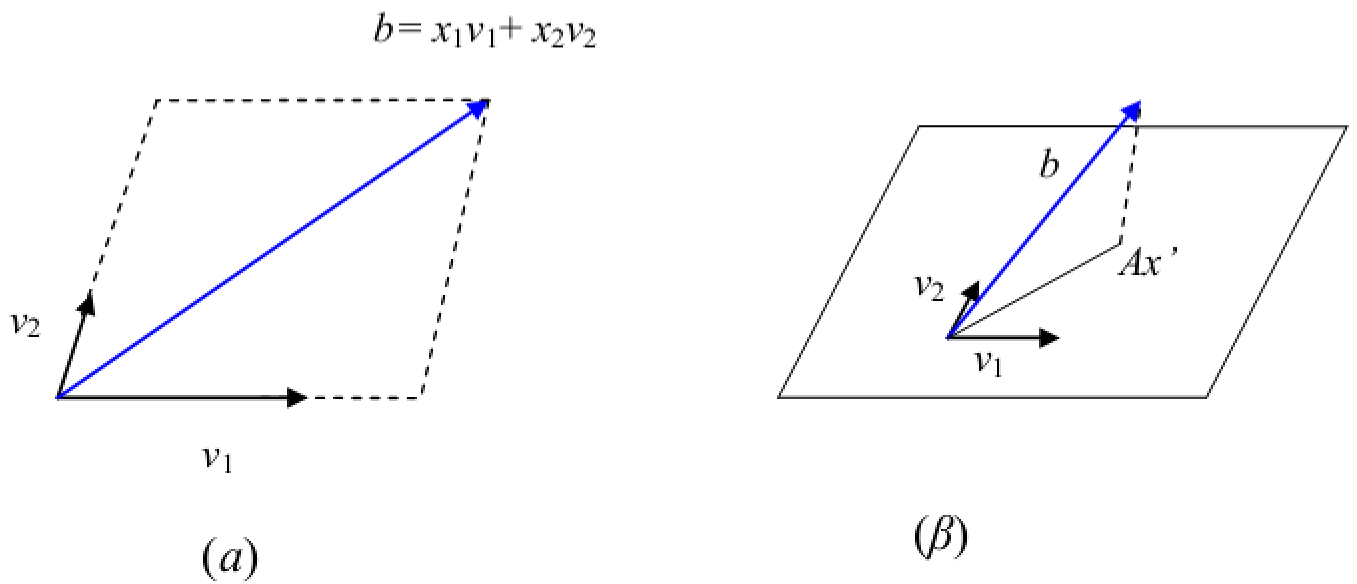
το οποίο μπορεί να γραφεί ως

$$x_1 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$$

Γενικά κάθε $m \times n$ σύστημα μπορεί να γραφεί σε αυτή τη μορφή σαν μία εξίσωση διανυσμάτων:

$$x_1 v_1 + x_2 v_2 + \cdots + x_n v_n = b$$

Στο παράδειγμά μας, προσπαθούμε να εκφράσουμε το b σαν έναν γραμμικό συνδυασμό από τρισδιάστατα διανύσματα. Ο συνδυασμός αυτός δημιουργεί μία επιφάνεια στο R^3 και έχει λύση μόνο αν το διάνυσμα b κείται σε αυτή την επιφάνεια.



Σχήμα 1: Γεωμετρική λύση ενός συστήματος 3 εξισώσεων με 2 αγνώστους. (α) Για να έχει λύση το διάνυσμα b θα πρέπει να είναι γραμμικός συνδυασμός των v_1 και v_2. (β) Αν το b είναι εκτός επιφάνειας δεν θα υπάρχει λύση. Η λύση ελαχίστων τετραγώνων x' δίνει ένα διάνυσμα Ax' σαν λύση που είναι το κοντινότερο στο b σε σχέση με την Ευκλείδια απόσταση.

Με βάση το σχήμα 1(β), μπορούμε να υπολογίσουμε μία προσεγγιστική λύση που να είναι η κοντινότερη σε σχέση με την ευκλείδια απόσταση. Αυτό το διάνυσμα Ax' έχει την εξής ιδιότητα: η διαφορά b-Ax' είναι κάθετη στην επιφάνεια $\{Ax | x \in R^n\}$. Θα εκμεταλλευτούμε αυτή την ιδιότητα για να υπολογίσουμε το x', την λύση ελαχίστων τετραγώνων.

Γνωρίζουμε ότι το b-Ax' είναι κάθετο στην επιφάνεια $\{Ax | x \in R^n\}$. Επομένως:

$$(Ax)^T(b - Ax') = 0, \forall x \in R^n \Rightarrow$$
$$x^T A^T(b - Ax') = 0, \forall x \in R^n$$

το οποίο σημαίνει ότι το διάνυσμα $A^T(b - Ax')$ είναι κάθετο σε κάθε διάνυσμα $x \in R^n$, καθώς και στον εαυτό του. Αυτό μπορεί να συμβεί μόνο όταν:

$$A^T(b - Ax') = 0$$

Αυτή η εξίσωση δίνει ένα σύστημα εξισώσεων για την παραγωγή της λύσης ελαχίστων τετραγώνων x', οι οποίες ονομάζονται ***κανονικές εξισώσεις***.

Κανονικές Εξισώσεις

Δοθέντος του αδύνατου συστήματος

$$Ax = b$$

λύνουμε το σύστημα

$$A^T A x' = A^T b$$

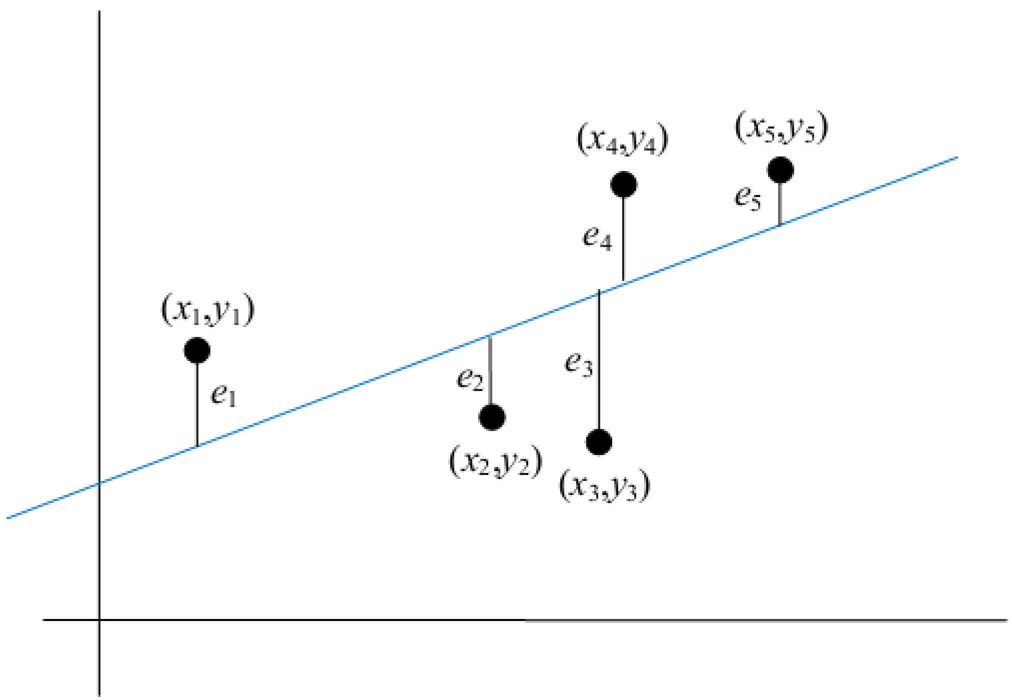
για την λύση ελαχίστων τετραγώνων.

Παράδειγμα

Παράδειγμα 6.1 Βρείτε την λύση ελαχίστων τετραγώνων για το σύστημα:

$$\begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$$

Έστω (x_1,y_1), (x_2,y_2),…, (x_m,y_m) ένα σύνολο σημείων στο επίπεδο (δεδομένα). Δοθείσης μίας οικογένειας μοντέλων (π.χ. $y=a+bx$) βρίσκουμε το καλύτερο στιγμιότυπο αυτής της οικογένειας ώστε να ταιριάζει καλύτερα στα δεδομένα με βάση τη νόρμα 2 (ευκλείδια απόσταση). Η ιδέα των ελαχίστων τετραγώνων είναι η μέτρηση της διαφοράς του μοντέλου από τα πραγματικά δεδομένα με το τετράγωνο του σφάλματος και επιλογή του κατάλληλου μοντέλου ώστε να ελαχιστοποιούνται αυτές οι ποσότητες. Αυτό φαίνεται στο παρακάτω σχήμα:



Σχήμα 2: Η καλύτερη γραμμή είναι αυτή που ελαχιστοποιεί την ποσότητα $e_1^2 + e_2^2 + e_3^2 + e_4^2 + e_5^2$ σε σχέση με όλες τις γραμμές $y=a+bx$.

Παράδειγμα 6.2 Ποια είναι η καλύτερη γραμμή που ταιριάζει στα δεδομένα (1,2), (-1,1) και (1,3).

Με βάση και το προηγούμενο παράδειγμα μπορούμε να δώσουμε την παρακάτω διαδικασία ώστε να λύνουμε ταιριάσματα ελαχίστων τετραγώνων:

Δοθέντος ενός συνόλου σημείων (x_1,y_1), (x_2,y_2),…, (x_m,y_m):

1. Επιλέγουμε ένα παραμετροποιημένο μοντέλο (γραμμή, πολυώνυμο κτλ.) που θα χρησιμοποιηθεί για το ταίριασμα των δεδομένων.
2. Αντικατάσταση των δεδομένων στο μοντέλο. Έτσι παίρνουμε ένα γραμμικό σύστημα για την εύρεση των παραμέτρων.
3. Επίλυση των αντίστοιχων κανονικών εξισώσεων και του συστήματος που προκύπτει.

Παράδειγμα 6.3 Ποια είναι η καλύτερη γραμμή και ποια η καλύτερη τετραγωνική καμπύλη που ταιριάζει στα δεδομένα (-1,1), (0,0), (1,0) και (2,-2).

Βέλτιστες προσεγγίσεις

Ορισμός 6.1.1 Εστω X ένας πραγματικός διανυσματικός χώρος. Μία απεικόνιση $(.,.): X \times X \rightarrow \boldsymbol{R}$ καλείται ***εσωτερικό γινόμενο*** στο X, αν ισχύουν:

- $(x + y , z) = (x,y) + (x , z)$ για κάθε $x,y,z \in X$
- $(\lambda x , y) = \lambda (x , y)$ για κάθε $x,y \in X$, $\lambda \in \boldsymbol{R}$
- $(x , y) = (y , x)$ για κάθε $x,y \in X$
- $(x , x) > 0$ για κάθε $x \in X\text{-}\{0\}$.

Ένας πραγματικός διανυσματικός χώρος πεπερασμένης διάστασης, στον οποίο έχει ορισθεί ένα εσωτερικό γινόμενο, καλείται ***ευκλείδειος χώρος***. Αν ισχύει $(x , y) = 0$, θα λέμε ότι τα x,y είναι κάθετα μεταξύ τους ***κάθετα***, ***ή ορθογώνια***. Η ποσότητα $||x|| = \sqrt{(x,x)}$ καλείται ***νόρμα*** του στοιχείου x και η ποσότητα $d(x,y) = || x - y ||$ καλείται ***απόσταση*** μεταξύ των στοιχείων x και y.

Παραδείγματα:

(1) Στο διανυσματικό χώρο $C[a,b]$ των συνεχών συναρτήσεων στο κλειστό διάστημα $[a,b]$, η απεικόνιση:

$$(f,g)=\int_a^b f(x)g(x)dx$$

είναι ένα εσωτερικό γινόμενο.

(2) Στο διανυσματικό χώρο $\boldsymbol{R}^n = \{x=(x_1,\ldots,x_n): x_i \in \boldsymbol{R}\}$, η απεικόνιση:

$$(x,y)=\sum_{k=1}^{n} x_i y_i$$

είναι ένα εσωτερικό γινόμενο.

Ορισμός 6.1.2 Εστω X ένας ευκλείδιος διανυσματικός χώρος με νόρμα $\|.\|$, $x \in X$ και έστω Y ένα υποσύνολο του X. Ένα στοιχείο $y \in Y$ για το οποίο ισχύει:

$$||x\text{-}y|| \le ||x\text{-}z||, \text{ για κάθε } z \in Y,$$

καλείται ***βέλτιστη προσέγγιση*** του X από το Y. Εναλλακτικά μπορεί να γράψει κάποιος ότι $(x,z) = (y,z)$, για κάθε $z \in Y$.

Θεώρημα 6.1.1 Εστω X ένας ευκλείδιος διανυσματικός χώρος με νόρμα $\|.\|$, $x \in X$ και Y ένας υπόχωρος του X. Ένα στοιχείο $y \in Y$ είναι βέλτιστη προσέγγιση του x από το y, αν και μόνον αν ισχύει:

$$\text{για κάθε } z \in Y, \quad (x-y,z) = 0. \qquad (6.1)$$

Αν λοιπόν ο υπόχωρος *Y* είναι πεπερασμένης διάστασης *n* και $\{s_1, \ldots, s_n\}$ είναι μία βάση του, τότε κάθε στοιχείο *z* του *Y* μπορεί να γραφεί ως γραμμικός συνδυασμός των στοιχείων της βάσης ως εξής:

$$z = \sum_{k=1}^{n} z_k \ s_k \ ,$$

οπότε η σχέση (6.1) ικανοποιείται αν και μόνον αν:

$$(x - y, s_k) = 0, k = 1, \ldots, n$$

Συνεπώς, αν $y = \sum_{k=1}^{n} y_k \ s_k$, όπου y_k είναι άγνωστοι συντελεστές, έχουμε:

$$\begin{pmatrix} (s_1,s_1) & (s_1,s_2) & \cdots & (s_1,s_n) \\ (s_2,s_1) & (s_2,s_2) & \cdots & (s_2,s_n) \\ \vdots & \vdots & \ddots & \vdots \\ (s_n,s_1) & (s_n,s_1) & \cdots & (s_n,s_n) \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} (x,s_1) \\ (x,s_2) \\ \vdots \\ (x,s_n) \end{pmatrix}. \qquad (6.2)$$

Παράδειγμα 6.1 Να προσδιορισθεί το πολυώνυμο 2ου βαθμού, το οποίο είναι η βέλτιστη προσέγγιση της συνάρτησης $f(x) = sin(\pi x)$ στο διάστημα [-1,1], στο χώρο των πολυωνύμων 2ου βαθμού, ως προς το σύνηθες εσωτερικό γινόμενο του χώρου (βλέπε παράδειγμα 1). Υπολογίστε το σφάλμα της βέλτιστης προσέγγισης.