Επιστημονικός Υπολογισμός Ι
ΗΥ 343: ΔΙΑΛΕΞΗ 9

Ε. Γαλλόπουλος
Τμήμα Η/Υ & Πληροφορικής
Πανεπιστήμιο Πατρών

ΤΜΗΥΠ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΗΛΕΚΤΡΟΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ & ΠΛΗΡΟΦΟΡΙΚΗΣ

Πανεπιστήμιο Πατρών

Κλασική θεωρία κατάστασης - δείκτης κατάστασης προβλήματος

Ορισμός 3 (Rice' 66). *Έστω γραμμικοί νορμισμένοι*[1] *χώροι* X, Y *που διαθέτουν νόρμα και έστω η απεικόνιση* $f : \Omega \subset X \to Y$, *όπου* Ω *είναι ανοικτό χωρίο. Έστω* x^* *σταθερό και η τιμή* $y^* := f(x^*)$. *Υποθέτουμε ότι τα* x^*, y^* *δεν είναι τα μηδενικά στοιχεία των* X, Y. *Ο ασυμπτωτικός σχετικός δείκτης κατάστασης της απεικόνισης* f *στο* x^* *ως προς μικρές αλλαγές του* x^* *ορίζεται ως*

$$\mathrm{cond}(f; x^*) := \lim_{\delta \to 0} \sup_{\|h\| = \delta} \left\{ \frac{\frac{\|f(x^*+h) - f(x^*)\|}{\|f(x^*)\|}}{\frac{\|h\|}{\|x^*\|}} \right\}$$

εφόσον το όριο υπάρχει.

προσοχή στις νόρμες

Εμείς «απλοποιήσαμε» ως

$$\mathrm{cond}(f; x) := \frac{\|f'(x)\|}{\|f(x)\|} \|x\|$$

προσοχή στην παράγωγο (ιακωβιανό μητρώο της f στο x)

[1] Λέγονται και σταθμισμένοι.

18

- **Προσέξτε ότι αν η συνάρτηση είναι γραμμική, δηλ. f(x+h) = f(x)+f(h) τότε ο αριθμητής (εδώ χρησιμοποιούμε x αντί x^*, απλοποιείται αρκετά:**

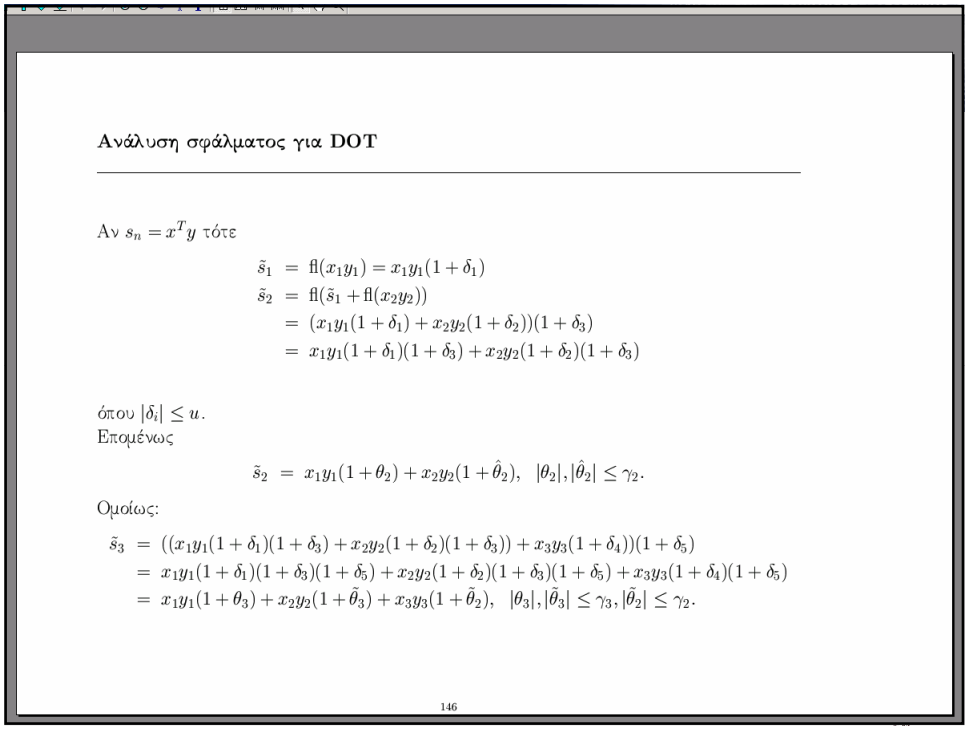
$$\frac{\|f(h)\|}{\|f(x)\|}$$

- **«υπενθυμίζουμε» ότι η νόρμα συνάρτησης (όπως και του μητρώου) ορίζεται ως**

$$\|f\| = \sup_{h \neq 0} \frac{\|f(h)\|}{\|h\|}$$

Οπότε για γραμμικές συναρτήσεις ο δ.κ. μπορεί να υπολογιστεί ως

$$\text{cond}(f;x) = \|f\| \frac{\|x\|}{\|f(x)\|}$$



Ανάλυση σφάλματος για DOT

Αν $s_n = x^T y$ τότε

$$\begin{aligned}
\tilde{s}_1 &= \text{fl}(x_1 y_1) = x_1 y_1 (1+\delta_1) \\
\tilde{s}_2 &= \text{fl}(\tilde{s}_1 + \text{fl}(x_2 y_2)) \\
&= (x_1 y_1 (1+\delta_1) + x_2 y_2 (1+\delta_2))(1+\delta_3) \\
&= x_1 y_1 (1+\delta_1)(1+\delta_3) + x_2 y_2 (1+\delta_2)(1+\delta_3)
\end{aligned}$$

όπου $|\delta_i| \leq u$.
Επομένως

$$\tilde{s}_2 = x_1 y_1 (1+\theta_2) + x_2 y_2 (1+\hat{\theta}_2), \quad |\theta_2|, |\hat{\theta}_2| \leq \gamma_2.$$

Ομοίως:

$$\begin{aligned}
\tilde{s}_3 &= ((x_1 y_1 (1+\delta_1)(1+\delta_3) + x_2 y_2 (1+\delta_2)(1+\delta_3)) + x_3 y_3 (1+\delta_4))(1+\delta_5) \\
&= x_1 y_1 (1+\delta_1)(1+\delta_3)(1+\delta_5) + x_2 y_2 (1+\delta_2)(1+\delta_3)(1+\delta_5) + x_3 y_3 (1+\delta_4)(1+\delta_5) \\
&= x_1 y_1 (1+\theta_3) + x_2 y_2 (1+\tilde{\theta}_3) + x_3 y_3 (1+\tilde{\theta}_2), \quad |\theta_3|, |\tilde{\theta}_3| \leq \gamma_3, |\tilde{\theta}_2| \leq \gamma_2.
\end{aligned}$$

146

Μπορούμε να φράξουμε κατευθείαν το εμπρός σφάλμα

Έχουμε

$$\begin{aligned}\tilde{s}_n &= x_1y_1\prod_{\substack{j=1\\ j\neq 2}}^{n+1}(1+\delta_j)+x_2y_2\prod_{j=2}^{n+1}(1+\delta_j)+\\ &\quad ...x_3y_3\prod_{j=3}^{n+1}(1+\delta_j)+\cdots+x_ny_n\prod_{j=n}^{n+1}(1+\delta_j).\end{aligned}$$

Από το γνωστό λήμμα:

$$\begin{aligned}\tilde{s}_n &= x_1y_1(1+\theta_n)+x_2y_2(1+\hat{\theta}_n)+\\ &\quad ...x_3y_3(1+\theta_{n-1})+\cdots+x_ny_n(1+\theta_2).\end{aligned}$$

Μπορούμε επομένως να συμπεράνουμε ότι:

$$\begin{aligned}|\tilde{s}_n-s_n| &\leq (|x_1y_1|+\cdots+|x_n||y_n|)\gamma_n\\ &\leq |x|^\top|y|\gamma_n\end{aligned}$$

Αυτό μας δίνει μια ένδειξη για το πώς μπορεί να συσσωρευθεί το σφάλμα!

Χρησιμοποιήσαμε τους κανόνες διάδοσης σφάλματος και με εμπρός ανάλυση βρήκαμε φράγμα για το εμπρός σφάλμα

147

Να δούμε τι μπορεί να συμπεράνουμε χρησιμοποιώντας «πίσω ανάλυση»

Έχουμε

$$\begin{aligned}\tilde{s}_n &= x_1y_1\prod_{\substack{j=1\\ j\neq 2}}^{n+1}(1+\delta_j)+x_2y_2\prod_{j=2}^{n+1}(1+\delta_j)+\\ &\quad ...x_3y_3\prod_{j=3}^{n+1}(1+\delta_j)+\cdots+x_ny_n\prod_{j=n}^{n+1}(1+\delta_j).\end{aligned}$$

Από το γνωστό λήμμα:

$$\begin{aligned}\tilde{s}_n &= x_1y_1(1+\theta_n)+x_2y_2(1+\hat{\theta}_n)+\\ &\quad ...x_3y_3(1+\theta_{n-1})+\cdots+x_ny_n(1+\theta_2).\end{aligned}$$

Το υπολογισθέν DOT είναι το ακριβές εσωτερικό γινόμενο για στοιχεία $x_1,...,x_n,y_1(1+\theta_n),y_2(1+\hat{\theta}_n),...,y_n(1+\theta_2)$, όπου για τα θ ισχύει το φράγμα

$$|\theta_j|\leq\frac{ju}{1-ju}=\gamma_j.$$

147

Αποδείξαμε ότι

$$\text{fl}(x^T y) = (x + \Delta x)^T y = x^T (y + \Delta y),$$
$$\text{όπου } |\Delta x| \leq \gamma_n |x|, |\Delta y| \leq \gamma_n |y|.$$

Δηλαδή το υπολογισμένο DOT είναι το ίδιο με το ακριβές DOT για στοιχεία εισόδου $x + \Delta x, y$. Επίσης

$$|\Delta x| \leq \gamma_n |x| \Rightarrow \|\Delta x\|_\infty \leq \gamma_n \|x\|_\infty$$

άρα το σχετικό πίσω σφάλμα είναι φραγμένο ως εξής

$$\frac{\|\Delta x\|_\infty}{\|x\|_\infty} \leq \gamma_n$$

και μπορούμε να θέσουμε για δείκτη κατάστασης του αλγορίθμου το $\text{cond}(f_{\text{prog}}) = \frac{\gamma_n}{\mathbf{u}}$.

ο υπολογισμός του εσωτερικού γινομένου είναι προς τα πίσω ευσταθής.

Το ακριβές φράγμα εξαρτάται από τη σειρά υπολογισμού. Το παραπάνω φράγμα υπολογίσθηκε σύμφωνα με το συνηθισμένο τρόπο υπολογισμού (από τα αριστερά προς τα δεξιά).

Χρησιμοποιήσαμε τους κανόνες διάδοσης σφάλματος και αποδείξαμε ότι ο αλγόριθμος είναι πίσω ευσταθής.
Δηλαδή είχαμε επιτυχημένη χρήση της πίσω ανάλυσης σφάλματος

45

Προσέξτε: Αν για όλα τα στοιχεία ισχύει $x_i y_i \geq 0$, τότε $|x|^\top |y| = |x^\top y| = |s_n|$ επομένως:

$$\frac{|\tilde{s}_n - s_n|}{|s_n|} \leq \gamma_n$$

Έχουμε δηλαδή μια σαφή ένδειξη για το μέγιστο σχετικό σφάλμα του υπολογισμού!

Θα δούμε στη συνέχεια ότι η παρακάτω παρατήρηση είναι εξίσου σημαντική:

Το υπολογισθέν DOT είναι το ακριβές εσωτερικό γινόμενο για στοιχεία $x_1, ..., x_n, y_1(1+\theta_n), y_2(1+\hat{\theta}_n), ..., y_n(1+\theta_2)$, όπου για τα θ ισχύει το φράγμα

$$|\theta_j| \leq \frac{ju}{1-ju} = \gamma_j.$$

148

Κατάσταση του μαθηματικού προβλήματος «εσωτερικό γινόμενο» (DOT)

$$f([x;y]) := x^T y, \quad x, y \in \mathbb{R}^n$$

Θέτουμε $X := [x;y] \in \mathbb{R}^{2n}$ άρα

$$\text{cond}(f;X) = \frac{\|X\|}{\|x^T y\|} \| \frac{\partial f}{\partial X}|_{[x;y]} \|.$$

και

$$\frac{\partial f}{\partial X}|_{[x;y]} = [y;x]^\top \in \mathbb{R}^{1\times 2n}$$

άρα

$$\text{cond}(f;X) = \frac{\|[x;y]\|}{|x^T y|} \|[y;x]\|.$$

Θεωρώντας $\|[x;y]\| = \|[y;x]\|$ έχουμε

$$\text{cond}(f;X) \;=\; \frac{\|[x;y]\|^2}{|x^T y|}.$$

Η κατάσταση εξαρτάται από το $|x^T y|$, δηλ. από το $\cos(x,y)$. Το φράγμα θα είναι μεγάλο (οπότε και το δυνάμει σφάλμα μεγάλο) αν $\cos(x,y) \approx 0$ ενώ $\|x\|, \|y\| \gg 0$.

Σφάλμα sAXPY (Εύκολη ανάλυση)

Ξεκινάμε από το σφάλμα κάθε στοιχείου του $z \leftarrow y + \alpha x$:

$$\begin{aligned} \text{fl}(\zeta_i) &= \text{fl}(\eta_i + \text{fl}(\alpha \cdot \xi_i)) \\ &= (\eta_i + \alpha\xi_i(1+\delta_{1,i}))(1+\delta_{2,i}) \\ &= \eta_i(1+\delta_{2,i}) + \alpha\xi_i(1+\delta_{1,i})(1+\delta_{2,i}) \\ &= \eta_i(1+\delta_{2,i}) + \alpha\xi_i(1+\theta_{2,i}), \end{aligned}$$

όπου $|\delta_{i,j}| \le u$ και $|\theta_{2,i}| \le \gamma_2 = \frac{2u}{1-2u}$.

Από την παραπάνω ανάλυση εύκολα προκύπτει ότι χρειαζόμαστε για να υπολογίσουμε φράγματα για το πίσω καθώς και το εμπρος σφάλμα.

195

Εμπρος σφάλμα

$$\begin{aligned} |\zeta_i - \tilde{\zeta}_i| &= |\delta_{1,i}\eta_i + \theta_{2,i}\alpha\xi_i| \\ &\leq u|\eta_i| + \gamma_2|\alpha\xi_i| \\ &\leq u|\eta_i| + 2u|\alpha\xi_i| + O(u^2) \\ |z - \tilde{z}| &\leq u(|y| + 2|\alpha||x|) + O(u^2) \end{aligned}$$

- Φράξαμε το «απόλυτο εμπρός σφάλμα» για κάθε στοιχείο.
- Αν θέλουμε μεγαλύτερη σαφήνεια, χρησιμοποιούμε το παρακάτω τέχνασμα:
  Υποθέτουμε ότι $n\mathbf{u} \leq 0.01$. Τότε $1/(1 - n\mathbf{u}) < 1.0101 < 1.02$

  $$\gamma_n = \frac{n\mathbf{u}}{1 - n\mathbf{u}} < 1.02n\mathbf{u}$$

  επομένως

  $$|\zeta_i - \tilde{\zeta}_i| < \mathbf{u}|\eta_i| + 1.02 \times 2\mathbf{u}|\alpha\xi_i| < 2.04(|\eta_i| + |\alpha\xi_i|)\mathbf{u}$$

  και στη νόρμα μεγίστου:

  $$\|z - \tilde{z}\|_\infty \leq 2.04(\|y\|_\infty + |\alpha|\|x\|_\infty)\mathbf{u}$$

Πίσω σφάλμα

Είδαμε ότι:

$$\mathrm{fl}(\zeta_i) = \eta_i(1 + \delta_{2,i}) + \alpha\xi_i(1 + \theta_{2,i}),$$

όπου $|\delta_{i,j}| \leq u$ και $|\theta_{2,i}| \leq \gamma_2 = \frac{2u}{1-2u}$.
Επομένως, το υπολογισμένο $\tilde{z}$ μπορεί να θεωρηθεί σαν το ακριβές sAXPY:

$$\tilde{z} = \tilde{y} + \alpha\tilde{x}$$

όπου

$$\tilde{y} = y + \Delta y, \qquad \tilde{x} = x + \Theta x,$$

όπου

$$\Delta = \begin{pmatrix} \delta_{2,1} & 0 & \cdots & 0 \\ 0 & \delta_{2,2} & \cdots & \vdots \\ \ddots & \cdots & \ddots & \vdots \\ 0 & \cdots & 0 & \delta_{2,n} \end{pmatrix}, \Theta = \begin{pmatrix} \theta_{2,1} & 0 & \cdots & 0 \\ 0 & \theta_{2,2} & \cdots & \vdots \\ \ddots & \cdots & \ddots & \vdots \\ 0 & \cdots & 0 & \theta_{2,n} \end{pmatrix}$$

Παρατήρηση: Οι Δ, Θ είναι διαγώνιοι με τα διαγώνια στοιχεία τους φραγμένο από u και γ_2 αντίστοιχα. Τότε

$$\|\Delta\| \leq u, \|\Theta\| \leq \gamma_2$$

για οποιαδήποτε από τις γνωστές νόρμες.
Επομένως μπορούμε να πούμε ότι για τη συνάρτηση $f : \mathbb{R}^{2n+1} \to \mathbb{R}^n$,

$$f([x; y; \alpha]) = y + \alpha x$$

έχουμε

$$f([x + \Theta x; y + \Delta y; \alpha]) = f_{\text{prog}}([x; y : \alpha])$$

επομένως αν ονομάσουμε $X = [x; y; \alpha]$, $\tilde{X} = [x + \Theta x; y + \Delta y; \alpha]$, τότε

$$\begin{aligned} \|\tilde{X} - X\| &= \|[\Theta x; \Delta y; 0]\| \\ &= \| \begin{pmatrix} \Theta & & \\ & \Delta & \\ & & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ \alpha \end{pmatrix} \| \leq \| \begin{pmatrix} \Theta & & \\ & \Delta & \\ & & 0 \end{pmatrix} \| \|X\| \end{aligned}$$

επομένως

$$\frac{\|\tilde{X} - X\|}{\|X\|} < 2.04u$$

Επομένως, μπορούμε να πούμε ότι

$$\text{cond}(f_{\text{prog}}) \leq 2.04.$$

Ασυμπτωτικός δείκτης κατάστασης

Θα χρησιμοποιήσουμε τον ορισμό. Θυμηθείτε ότι η συνάρτηση sAXPY είναι $f : \mathbb{R}^{2n+1} \to \mathbb{R}^n$, επομένως το Ιακωβιανό μητρώο

$$f'(X) \in \mathbb{R}^{n \times (2n+1)}$$

όπου $X = [x; y; \alpha] \in \mathbb{R}^{2n+1}$. Ειδικότερα, προσέξτε ποιά είναι τα στοιχεία του Ιακωβιανού μητρώου:

$$f'(X) = \left(\begin{array}{cccc|ccc|c} \alpha & 0 & & & 1 & 0 & & \xi_1 \\ 0 & \alpha & & & 0 & 1 & & \xi_2 \\ & & \ddots & & & & \ddots & \vdots \\ & & & \alpha & & & 1 & \xi_n \end{array} \right) = [\alpha I, I, x] \in \mathbb{R}^{n \times (2n+1)}$$

Επομένως

$$\text{cond}(f; X) = \frac{\|f'(X)\|}{\|z\|} \|X\|$$

όπου

$$\|f'(X)\|_\infty = \max_j \{1 + |\alpha| + |\xi_j|\} = 1 + |\alpha| + \|x\|_\infty$$

$$\|X\|_\infty = \max\{\|x\|_\infty, \|y\|_\infty, |\alpha|\}$$

200

Συγκεντρώνοντας τα αποτελέσματα:

$$\mathrm{cond}(f;X) = \frac{\|f'(X)\|}{\|z\|}\|X\|$$

όπου

$$\|f'(X)\|_\infty = \max_j\{1+|\alpha|+|\xi_j|\} = 1+|\alpha|+\|x\|_\infty, \qquad \|X\|_\infty = \max\{\|x\|_\infty, \|y\|_\infty, |\alpha|\}$$

και

$$\mathrm{cond}(f_{\mathrm{prog}}) < 2.04$$

βλέπουμε πως το εμπρός σχετικό σφάλμα μπορεί να γίνει μεγάλο αν η νόρμα του αποτελέσματος $\|z\|$ είναι μικρή.

201

Γενικά σχόλια

- **Η αυστηρή χρήση της πίσω ανάλυσης δεν είναι πάντα εύκολη**
  - **αποτυγχάνει ακόμα και σε απλές περιπώσεις**
    - **μη γραμμικοί υπολογισμοί**
    - **υπολογισμοί με «πολλά αποτελέσματα» σε σύγκριση με το πλήθος των στοιχείων εισόδου**
  - **έχει όμως φανεί πάρα πολύ χρήσιμη σε σημαντικούς αλγορίθμους της γραμμικής άλγεβρας που μπορούν να αποδειχτούν πίσω ευσταθείς**
    - **επίλυση συστημάτων με QR, εύρεση ιδιοτιμών με QR, «σχεδόν» επίλυση συστημάτων με LU**
    - **υπολογισμοί με πολυώνυμα**
    - **οπότε το σφάλμα εξαρτάται αποκλειστικά από το δείκτη κατάστασης του προβλήματος**

Σφάλμα εξωτερικού γινομένου

Έστω ο υπολογισμός

$$\mathrm{fl}(C) = \mathrm{fl}(ab^\top), a, b \in \mathbb{R}^n$$

άρα

$$\mathrm{fl}(\gamma_{ij}) = \alpha_i \beta_j (1 + \delta_{ij})$$

που σημαίνει ότι

$$\mathrm{fl}(C) = ab^\top + E$$

όπου $e_{ij} = \alpha_i \beta_j \delta_{ij}$. Αν είχαμε πίσω ευστάθεια θα υπήρχαν $\tilde{a}, \tilde{b}$ κοντά στα a, b ώστε

$$\mathrm{fl}(C) = \tilde{a}\tilde{b}^\top = (a + \Delta a)(b + \Delta b)^\top$$

Τότε θα ίσχυε

$$E = \overbrace{a\Delta b^\top + \Delta a\, b^\top + \Delta a \Delta b^\top}^{\text{τάξη} \approx 2} = [a, \Delta a] \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{bmatrix} b^\top \\ \Delta b^\top \end{bmatrix}$$

Ισχύουν όμως ότι:

$$\begin{aligned} \mathrm{rank}(XY) &\leq \min(\mathrm{rank}(X), \mathrm{rank}(Y)), \\ \mathrm{rank}(X) + \mathrm{rank}(Y) &\leq \mathrm{rank}(X) + \mathrm{rank}(Y)) \end{aligned}$$

Επομένως,

$$\mathrm{rank}(E) \leq \mathrm{rank}\left[\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}\right] = 2$$

Επομένως, αν ο αλγόριθμος ήταν πίσω ευσταθής θα ίσχυε ότι

$$n = \mathrm{rank}(E) \leq 2$$

πράγμα που είναι γενικά αδύνατο.

Η ανανέωση 1ης τάξης δεν είναι πίσω ευσταθής

Εμπρός σφάλμα

Επομένως η πίσω ανάλυση δεν μπορεί να μας βοηθήσει. Προσέξτε όμως ότι το εμπρός σφάλμα μπορεί να εκτιμηθεί άμεσα!

$$|\mathrm{fl}(\gamma_{ij}) - \alpha_i\beta_j| = |\alpha_i\beta_j\delta_{ij}| \leq |\alpha_i\beta_j|u$$

άρα

$$\frac{|\mathrm{fl}(\gamma_{ij}) - \alpha_i\beta_j|}{|\alpha_i\beta_j|} \leq u$$

Σχετικά με το σφάλμα στην πράξη MV

Δύο τρόποι υπολογισμού του MV $y = Ax$:

Κατά γραμμές: $\eta_i = a_{i,:}^\top x, i = 1:n$

Κατά στήλες: $\eta = \sum_{i=1}^n \xi_i a_{:,i}$

Δίνουν οι δύο τρόποι το ίδιο σφάλμα;

Και οι δύο μέθοδοι οδηγούν ακριβώς στα ίδια σφάλματα στρογγύλευσης

$$\tilde{\eta}_i = (a_{i,:} + \Delta a_{i,:})^\top x, \quad |\Delta a_{i,:}| \leq \gamma_n |a_{i,:}|.$$

Επομένως έχουμε πίσω ευστάθεια γιατί:

$$\tilde{y} = (A + \Delta A)x, \quad |\Delta A| \leq \gamma_n |A|$$

Από την άλλη, μπορούμε να υπολογίσουμε άμεσα για το εμπρός σφάλμα ότι:

$$|y - \tilde{y}| \leq \gamma_n |A||x|.$$