دانشگاه صنعتی امیرکبیر
دانشکده ریاضی و علوم کامپیوتر

پایان‌نامه کارشناسی ارشد

رشته ریاضی کاربردی(گرایش آنالیز عددی)

روشهای عددی محاسبه ریشه های ماتریسی

استاد راهنما:
دکترمصطفی شمسی
استاد مشاور:
دکتر مهدی دهقان
نگارنده:
امیر صادقی

تیر ۱۳۸۷

بسمه تعالی

فرم اطلاعات پایان نامه

کارشناسی ارشد و دکترا

(پلی تکنیک تهران)

تاریخ :

پیوست :

دانشگاه صنعتی امیر کبیر

معاونت پژوهشی

نام و نام خانوادگی :امیر صادقی **دانشجوی آزا د** ☐ **بورسیه** ☐ **معادل** ☐

شمارهٔ دانشجوئی : 85113051 **دانشکده :** ریاضی و علوم کامپیوتر **رشتهٔ تحصیلی : ریاضی**

نام و نام خانوادگی استاد راهنما : دکتر مصطفی شمسی

عنوان پایان نامه به فارسی: روشهای عددی محاسبه ریشه های ماتریسی

عنوان پایان نامه به انگلیسی: Numerical methods for computing matrix roots

نوع پروژه : **کارشناسی ارشد :** ■ **دکتـــــری** ☐ **کاربردی** ☐ **بنیادی** ☐ **توسعه ای** ☐ **نظری** ■

تاریخ شروع : اسفند ۸۶ **تاریخ خاتمه :** ت یر 78 **تعداد واحد :** ۶ واحد

سازمان تأمین کننده اعتبار :

واژه های کلیدی به فارسی: توابع ماتریسی- ریشه های ماتریسی- بلوک های ژوردان- روش شور- روش نیوتن

واژه های کلیدی به انگلیسی : **matrix function- matrix roots- Jordan block- Schur method- newton method**

نظرها و پیشنهادها به منظور بهبود فعالیتهای پژوهشی دانشگاه :

استاد راهنما : دکتر مصطفی شمسی

دانشجو : امیر صادقی

امضاء استاد راهنما : **تاریخ :**

نسخهٔ ۱ : معاونت پژوهشی

نسخهٔ ۲: کتابخانه و به انضمام دوجلد پایان نامه به منظور تسویه حساب با کتابخانه و مرکز اسناد و مدارک علمی

چکیده

به بیان ساده ریشه pام یک ماتریس $\mathbf{A}$ عبارتست از ماتریسی مانند $\mathbf{X}$ بطوری که $\mathbf{X}^p = \mathbf{A}$. بلافاصله پس از تعریف ریشه pام ماتریسی وجود یا عدم وجود و تعداد ریشه ها مطرح می‌شود. در صورتی که ریشه pام ماتریسی منحصر به فرد نباشد، آنگاه ریشه pام اصلی مورد توجه قرار می‌گیرد. در این پایان نامه با بررسی موارد فوق به برخی کاربردهای ریشه های ماتریسی در حل معادلات دیفرانسیل، در مدل های مارکوف و تعیین میانگین هندسی ماتریسی اشاره می‌شود. همچنین روشهای نیوتن و شور برای محاسبه ریشه pام ماتریسی بررسی می شود و در مورد همگرایی وپایداری این روشها بطور مختصر توضیح داده می‌شود و با مثالهای عددی کارایی و نقاط ضعف آنها گزارش می‌شود.

فهرست

فصل اول

تعاریف و مقدمات

در این فصل به مقدمات و تعاریف اولیه می‌پردازیم.

فضای برد و پوچی ماتریس و رتبه ماتریس

برای هر ماتریس $m \times n$ مانند $\mathbf{A}$ دو زیر فضای بسیار مهم به نام های برد[1] $\mathbf{A}$ و پوچی[2] $\mathbf{A}$ وجود دارد که به ترتیب با نمادهای $R(\mathbf{A})$ و $N(\mathbf{A})$ نمایش داده می‌شوند و به صورت زیر تعریف می‌شوند:

$$R(\mathbf{A}) = \left\{ b \in \mathbb{R}^m : b = \mathbf{A}x \ \text{ for } \ x \in \mathbb{R}^n \right\}$$

$$N(\mathbf{A}) = \left\{ x \in \mathbb{R}^n : \mathbf{A}x = 0 \right\}$$

فرض کنیم S زیر فضای $\mathbb{R}^m$ باشد. زیر فضای $S^{\perp}$ تعریف شده با

$$S^{\perp} = \left\{ y \in \mathbb{R}^m : \forall x \in S \ \ y^T x = 0 \right\}$$

را زیر فضای متعامد S گویند. نشان داده می‌شود

---

[1] Range

[2] Nullity

$$N(\mathbf{A}) = R(\mathbf{A}^T)^{\perp}$$
$$R(\mathbf{A})^{\perp} = N(\mathbf{A}^T)$$

بعد $N(\mathbf{A})$ پوچی $\mathbf{A}$ گفته می‌شود و با $\mathrm{null}(\mathbf{A})$ نمایش داده می‌شود.

فرض کنیم $\mathbf{A}$ ماتریسی $m \times n$ باشد. آنگاه زیر فضای تولید شده با بردارهای سطری $\mathbf{A}$، فضای سطری $\mathbf{A}$ گفته می‌شود. زیر فضای تولید شده توسط بردارهای ستونی $\mathbf{A}$، فضای ستونی $\mathbf{A}$ نامیده می‌شود.

رتبه ماتریس $\mathbf{A}$ بعد فضای ستونی $\mathbf{A}$ می‌باشد و با $\mathrm{rank}(\mathbf{A})$ نمایش داده می‌شود. ماتریس $\mathbf{A} \in \mathbb{R}^{n\times n}$ نامنفرد است اگر $\mathrm{rank}(\mathbf{A}) = n$ در غیر این صورت $\mathbf{A}$ منفرد است. یک ماتریس $\mathbf{A} \in \mathbb{R}^{m\times n}$ دارای رتبه کامل ستونی است اگر دارای ستون های مستقل خطی باشد. بطور مشابه رتبه کامل سطری تعریف می‌شود. یک ماتریس رتبه کامل گفته می‌شود اگر دارای رتبه سطری کامل یا رتبه ستونی کامل باشد. اگر $\mathbf{A}$ ماتریسی $m \times n$ باشد، رتبه آن دارای ویژگی های زیر است:

$$\begin{aligned}
&\mathrm{rank}(\mathbf{A}) = \mathrm{rank}(\mathbf{A}^T) \\
&\mathrm{rank}(\mathbf{A}) + \mathrm{null}(\mathbf{A}) = n \\
&\mathrm{rank}(\mathbf{AB}) \geq \mathrm{rank}(\mathbf{A})\mathrm{rank}(\mathbf{B}) - n \\
&\mathrm{rank}(\mathbf{AB}) \leq \min\left\{\mathrm{rank}(\mathbf{A}), \mathrm{rank}(\mathbf{B})\right\} \\
&\mathrm{rank}(\mathbf{A}+\mathbf{B}) \leq \mathrm{rank}(\mathbf{A}) + \mathrm{rank}(\mathbf{B})
\end{aligned}$$

طیف و شعاع طیفی

ماتریس $\mathbf{A} \in \mathbb{C}^{n\times n}$ مفروض است. بردار $x \in \mathbb{C}^n$ یک بردار ویژه و $\lambda \in \mathbb{C}$ یک مقدار ویژه $\mathbf{A}$ هستند اگر

$$\mathbf{A}x = \lambda x, \quad x \neq 0$$

همچنین وقتی x بردار ویژه متناظر با مقدار ویژه λ باشد، آنگاه برای هر $\alpha \neq 0$ بردار αx نیز یک بردار ویژه برای همان مقدار ویژه می‌باشد.

چند جمله**ای** مشخصه[۳] ماتریس $\mathbf{A}$ بصورت زیر تعریف می‌شود:

$$p_n(z) = \det(\mathbf{A} - z\mathbf{I})$$

---

[۳] Characteristic polynomial

توجه داریم مطابق با تعریف فوق، $\lambda \in \mathbb{C}$ یک مقدار ویژه ماتریس $\mathbf{A}$ است اگر و فقط اگر $p_n\left(\lambda\right) = 0$.

قضیه (۱.۱) (کیلی همیلتون)[۴]: یک ماتریس مربعی در چندجمله‌ای مشخصه اش صدق می‌کند. به عبارت دیگر اگر $\mathbf{A}$ یک ماتریس $n \times n$ و $p_n\left(\lambda\right)$ چند جمله‌ای مشخصه آن باشد، آنگاه $p_n\left(\mathbf{A}\right)$ برابر ماتریس صفر است.

برهان: به [6] مراجعه شود.■

همچنین چند جمله‌ای که اولاً از کمترین درجه است و ثانیاً بازای ماتریس $\mathbf{A}$ برابر صفر شود، چند جمله‌ای مینیمال[۵] نامیده می‌شود و با m نمایش داده می‌شود.

اگربرای $i = 1,\cdots,n$ ، λ_i مقادیر ویژه ماتریس $\mathbf{A}$ باشد، آنگاه مجموعه مقادیر ویژه ماتریس $\mathbf{A}$ را طیف[۶] $\mathbf{A}$ گوییم و با $\mathrm{spec}(\mathbf{A})$ نمایش داده می‌شود. به عبارت دیگرداریم:

$$\mathrm{spec}(\mathbf{A}) = \left\{\lambda_i : \det(\mathbf{A} - \lambda_i \mathbf{I}) = 0\right\}$$

همچنین شعاع طیفی[۷] ماتریس $\mathbf{A}$ را بصورت زیر تعریف می‌کنیم:

$$\rho(\mathbf{A}) = \max_{1 \le i \le n} \left\{\left|\lambda_i\right| : \det(\mathbf{A} - \lambda_i \mathbf{I}) = 0\right\}$$

مقادیر منفرد

فرض کنیم $\mathbf{A} \in \mathbb{C}^{m \times n}$ ماتریسی باشد که در آن $m \ge n$ است. آنگاه مقادیر ویژه ماتریس $n \times n$ متقارن $\mathbf{A}^T\mathbf{A}$ حقیقی و مثبت می‌باشند. فرض کنیم این مقادیر ویژه با $\sigma_i^{\,2}$ نمایش داده شوند بطوری که

$$\sigma_1^{\,2} \ge \sigma_2^{\,2} \ge \cdots \ge \sigma_n^{\,2}$$

در این صورت $\sigma_1, \sigma_2, \cdots, \sigma_n$ مقادیر منفرد[۸] $\mathbf{A}$ نامیده می‌شوند.

---

[۴] Cayley-Hamilton Theorem

[۵] Minimal polynomial

[۶] Spectrum

[۷] Spectral radius

[۸] Singular valué

بعنوان مثال اگر $\mathbf{A}=\begin{pmatrix}0&1\\2&2\end{pmatrix}$، آنگاه $\mathbf{A}^T\mathbf{A}=\begin{pmatrix}4&4\\4&5\end{pmatrix}$ و مقادیر ویژه $\mathbf{A}^T\mathbf{A}$ عبارتند از $\left(\frac{9\pm\sqrt{65}}{2}\right)$. پس مقادیر منفرد $\mathbf{A}$ عبارتند از:

$$\sigma_1=\sqrt{\left(\frac{9+\sqrt{65}}{2}\right)}\quad,\quad\sigma_2=\sqrt{\left(\frac{9-\sqrt{65}}{2}\right)}$$

ضرب و جمع کرونکر

تعریف: فرض کنیم $\mathbf{A}\in\mathbb{C}^{m\times n}$ و $\mathbf{B}\in\mathbb{C}^{p\times q}$. حاصلضرب کرونکر[۹] دو ماتریس $\mathbf{A}$ و $\mathbf{B}$ ماتریس $\left(\mathbf{A}\otimes\mathbf{B}\right)\in\mathbb{C}^{mp\times nq}$ می‌باشد و بصورت زیر تعریف می‌شود:

$$\mathbf{A}\otimes\mathbf{B}=\begin{bmatrix}a_{11}\mathbf{B}&a_{12}\mathbf{B}&\cdots&a_{1n}\mathbf{B}\\a_{21}\mathbf{B}&a_{22}\mathbf{B}&\cdots&a_{2n}\mathbf{B}\\\vdots&\vdots&\ddots&\vdots\\a_{m1}\mathbf{B}&a_{m2}\mathbf{B}&\cdots&a_{mn}\mathbf{B}\end{bmatrix}$$

حاصلضرب کرونکر دارای ویژگیهای زیر است:

$$\begin{aligned}
&\left(\mathbf{A}\otimes\mathbf{B}\right)\neq\left(\mathbf{B}\otimes\mathbf{A}\right)\\
&\mathbf{A}\otimes\left(\mathbf{B}\otimes\mathbf{C}\right)=\left(\mathbf{A}\otimes\mathbf{B}\right)\otimes\mathbf{C}\\
&\mathbf{A}\otimes\left(\mathbf{B}+\mathbf{C}\right)=\left(\mathbf{A}\otimes\mathbf{B}\right)+\left(\mathbf{A}\otimes\mathbf{C}\right)\\
&\left(\alpha_{\mathbf{A}}\mathbf{A}\otimes\alpha_{\mathbf{B}}\mathbf{B}\right)=\alpha_{\mathbf{A}}\alpha_{\mathbf{B}}.\left(\mathbf{A}\otimes\mathbf{B}\right)\\
&\left(\mathbf{A}\otimes\mathbf{B}\right)^T=\left(\mathbf{A}^T\otimes\mathbf{B}^T\right)\\
&\left(\mathbf{A}\otimes\mathbf{B}\right)\left(\mathbf{C}\otimes\mathbf{D}\right)=\mathbf{AC}\otimes\mathbf{BD}\\
&\left(\mathbf{A}\otimes\mathbf{B}\right)^{-1}=\left(\mathbf{A}^{-1}\otimes\mathbf{B}^{-1}\right)\\
&rank\left(\mathbf{A}\otimes\mathbf{B}\right)=rank(\mathbf{A}).rank(\mathbf{B})\\
&Tr\left(\mathbf{A}\otimes\mathbf{B}\right)=Tr(\mathbf{A}).Tr(\mathbf{B})\\
&\det\left(\mathbf{A}\otimes\mathbf{B}\right)=\det(\mathbf{A})^{rank(\mathbf{B})}.\det(\mathbf{B})^{rank(\mathbf{A})}
\end{aligned}$$

[۹] Kroneker product

تعریف: فرض کنیم $\mathbf{A} \in \mathbb{C}^{n\times n}$ و $\mathbf{B} \in \mathbb{C}^{m\times m}$. آنگاه جمع کرونکر[۱۰] دو ماتریس $\mathbf{A}$ و $\mathbf{B}$ ماتریس $\left(\mathbf{A} \oplus \mathbf{B}\right) \in \mathbb{R}^{mn\times mn}$ می‌باشد و بصورت $\mathbf{A} \oplus \mathbf{B} = \left(\mathbf{I}_m \otimes \mathbf{A}\right) + \left(\mathbf{B} \otimes \mathbf{I}_n\right)$ تعریف می‌شود.

توجه داریم که $\mathbf{A} \oplus \mathbf{B} \neq \mathbf{B} \oplus \mathbf{A}$.

بطور مثال فرض کنیم $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ و $\mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}$ باشد. آنگاه خواهیم داشت:

$$\begin{aligned}
\mathbf{A} \oplus \mathbf{B} &= \left(\mathbf{I}_3 \otimes \mathbf{A}\right) + \left(\mathbf{B} \otimes \mathbf{I}_2\right) \\
&= \begin{bmatrix} a_{11} & & & a_{12} & & \\ & a_{11} & & & a_{12} & \\ & & a_{11} & & & a_{12} \\ a_{21} & & & a_{22} & & \\ & a_{21} & & & a_{22} & \\ & & a_{21} & & & a_{22} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & b_{13} & & & \\ b_{21} & b_{22} & b_{23} & & & \\ b_{31} & b_{32} & b_{33} & & & \\ & & & b_{11} & b_{12} & b_{13} \\ & & & b_{21} & b_{22} & b_{23} \\ & & & b_{31} & b_{32} & b_{33} \end{bmatrix} \\
&= \begin{bmatrix} a_{11}+b_{11} & b_{12} & b_{13} & a_{12} & & \\ b_{21} & a_{11}+b_{22} & b_{23} & & a_{12} & \\ b_{31} & b_{32} & a_{11}+b_{33} & & & a_{12} \\ a_{21} & & & a_{22}+b_{11} & b_{12} & b_{13} \\ & a_{21} & & b_{21} & a_{22}+b_{22} & b_{23} \\ & & a_{21} & b_{31} & b_{32} & a_{33}+b_{22} \end{bmatrix}
\end{aligned}$$

عملگر $\mathrm{Vec}(\mathbf{A})$

برای هر ماتریس $\mathbf{A} = (a_{ij}) \in \mathbb{C}^{m\times n}$ عملگر $\mathrm{Vec}(\mathbf{A}) \in \mathbb{C}^{mn}$ را بصورت زیر تعریف می‌کنیم:

$$\mathrm{Vec}(\mathbf{A}) = \left(a_{11}, \cdots, a_{m1}, a_{12}, \cdots, a_{m2}, \cdots, a_{1n}, \cdots, a_{mn}\right)^T$$

عملگر $Vec(\mathbf{A})$ نیز دارای ویژگیهای زیر است:

---

[۱۰] Kroneker sum

$$
\begin{aligned}
&\mathrm{Vec}(\mathbf{AXB}) = (\mathbf{B}^T \otimes \mathbf{A}).\mathrm{Vec}(\mathbf{X})\\
&\mathrm{Tr}(\mathbf{A}^T\mathbf{B}) = \mathrm{Vec}(\mathbf{A})^T\mathrm{Vec}(\mathbf{B})\\
&\mathrm{Vec}(\mathbf{A}+\mathbf{B}) = \mathrm{Vec}(\mathbf{A}) + \mathrm{Vec}(\mathbf{B})\\
&\mathrm{Vec}(\alpha\mathbf{A}) = \alpha.\mathrm{Vec}(\mathbf{A})
\end{aligned}
$$

نرم برداری

نرم برداری تابعی بصورت $\left\|\cdot\right\| : \mathbb{C}^n \to \mathbb{R}$ است که در خواص زیر صدق می‌کند:

$$
\begin{aligned}
&\forall x \in \mathbb{C}^n : x \neq 0 \Rightarrow \left\|x\right\| > 0, \quad \left(\left\|x\right\| = 0 \Leftrightarrow x = 0\right),\\
&\forall \alpha \in \mathbb{C}, x \in \mathbb{C}^n : \left\|\alpha x\right\| = \left|\alpha\right|.\left\|x\right\|,\\
&\forall x, y \in \mathbb{C}^n : \left\|x+y\right\| \leq \left\|x\right\| + \left\|y\right\|.
\end{aligned}
$$

از نرم های بسیار مفید $p-$نرم یا نرم‌های هولدر[۱۱] است که بصورت زیر تعریف می‌شود:

$$
\left\|x\right\|_p = \left(\sum_{i=1}^{n}\left|x_i\right|^p\right)^{1/p}, \qquad p \geq 1
$$

سه مورد از $p-$نرمهایی که کاربرد وسیعی دارند عبارتند از نرم یک، نرم دو یا اقلیدسی[۱۲] و نرم بینهایت که بصورت زیر تعریف می‌شوند:

$$
\begin{aligned}
\left\|x\right\|_1 &= \sum_{i=1}^{n}\left|x_i\right|\\
\left\|x\right\|_2 &= \left(\sum_{i=1}^{n}\left|x_i\right|^2\right)^{1/2}\\
\left\|x\right\|_\infty &= \max_{1\leq i\leq n}\left|x_i\right|
\end{aligned}
$$

یک ویژگی مهم $p-$نرمهای برداری صدق کردن در نامساوی هولدر است:

$$
\left|x^T y\right| \leq \left\|x\right\|_p \left\|y\right\|_q
$$

که در آن $\frac{1}{p} + \frac{1}{q} = 1$. همچنین حالت خاص نامساوی هولدرکه با قرار دادن $p = q = 2$ به دست می‌آید و به نامساوی شوارتز[۱۳] معروف است، بصورت زیر است:

---

[۱۱] Holder

[۱۲] Euclidean Norm

$$\left|x^T y\right| \le \left(\sum_{i=1}^{n}\left|x_i\right|^2\right)^{1/2}\left(\sum_{i=1}^{n}\left|y_i\right|^2\right)^{1/2}$$

همه $p-$نرمهای برداری در فضاهای با بعد متناهی هم ارزند، به این معنی که ثابت های $0 < \alpha$ و $0 < \beta$ موجودند بطوریکه:

$$\alpha\left\|x\right\|_p \le \left\|x\right\|_q \le \beta\left\|x\right\|_p,$$

هم ارزی $p-$نرمهای یک، اقلیدسی و بی‌نهایت به صورت زیر می‌باشد:

$$\begin{aligned}&\left\|x\right\|_2 \le \left\|x\right\|_1 \le \sqrt{n}\left\|x\right\|_2\\&\left\|x\right\|_\infty \le \left\|x\right\|_2 \le \sqrt{n}\left\|x\right\|_\infty\\&\left\|x\right\|_\infty \le \left\|x\right\|_1 \le n\left\|x\right\|_\infty\end{aligned}$$

همچنین برای $q > p$ می توان نشان داد:

$$\left\|x\right\|_p \le \left\|x\right\|_q \le n^{\left(\frac{1}{q}-\frac{1}{p}\right)}\left\|x\right\|_p.$$

نرم ماتریسی

تعریف: نرم ماتریسی تابع $\left\|\cdot\right\| : \mathbb{C}^{m\times n} \to \mathbb{R}$ است که در خواص زیر صدق می‌کند:

$$\begin{aligned}&\forall \mathbf{A} \in \mathbb{C}^{m\times n} : \left\|\mathbf{A}\right\| > 0, \quad \left(\left\|\mathbf{A}\right\| = 0 \Leftrightarrow \mathbf{A} = 0\right),\\&\forall \alpha \in \mathbb{R}, \mathbf{A} \in \mathbb{C}^{m\times n} : \left\|\alpha\mathbf{A}\right\| = \left|\alpha\right|.\left\|\mathbf{A}\right\|,\\&\forall \mathbf{A},\mathbf{B} \in \mathbb{C}^{m\times n} : \left\|\mathbf{A}+\mathbf{B}\right\| \le \left\|\mathbf{A}\right\| + \left\|\mathbf{B}\right\|.\end{aligned}$$

با در نظر گرفتن ماتریس $\mathbf{A}$ و نرم برداری $\left\|\cdot\right\|$ می‌توان نرم القایی یا $p-$نرم ماتریسی را بصورت زیر تعریف کرد:

$$\left\|\mathbf{A}\right\|_p = \max_{x\neq 0}\frac{\left\|\mathbf{A}x\right\|_p}{\left\|x\right\|_p}$$

---

[۱۳] Schwarz Inequality

استفاده از فرمول فوق برای محاسبه $p-$نرم یک ماتریس مشکل می‌باشد. اما برای نرم‌های معمول فرمول‌های جایگزین ساده‌تری به صورت زیر وجود دارد:

$$\|\mathbf{A}\|_1 = \max_{1\le j\le n}\sum_{i=1}^{m}|a_{ij}|$$
$$\|\mathbf{A}\|_\infty = \max_{1\le i\le m}\sum_{j=1}^{n}|a_{ij}|$$
$$\|\mathbf{A}\|_2 = \left(\rho(\mathbf{A}^*\mathbf{A})\right)^{\frac{1}{2}} = \sigma_{\max}(\mathbf{A})$$

یکی از ویژگیهای بسیار مفید نرم القایی[۱۴] ماتریسی این است که

$$\forall x \neq 0 \ : \ \|\mathbf{A}x\|_p \le \|\mathbf{A}\|_p . \|x\|_p$$

لازم به ذکر است یک نرم ماتریسی مانند $\|\cdot\|_M$ و نرم برداری $\|\cdot\|_V$ سازگارند، اگر داشته باشیم:

$$\|\mathbf{A}x\|_V \le \|\mathbf{A}\|_M . \|x\|_V$$

یکی از نرمهای ماتریسی سازگار[۱۵] با نرم برداری دو، نرم فروبنیوس[۱۶] است و بصورت زیر تعریف می‌شود:

$$\|\mathbf{A}\|_F = \left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2\right)^{\frac{1}{2}} = \left(Tr(\mathbf{A}^*\mathbf{A})\right)^{\frac{1}{2}} = \left(\sum_{i=1}^{n}\sigma^2{}_i\right)^{\frac{1}{2}}$$

همچنین یک نرم ماتریسی سازگار گفته می‌شود اگر برای دو ماتریس $\mathbf{A}$ و $\mathbf{B}$ داشته باشیم:

$$\|\mathbf{AB}\| \le \|\mathbf{A}\|.\|\mathbf{B}\|$$

لازم به ذکر است که در چنین حالتی باید حاصل ضرب $\mathbf{AB}$ تعریف شده باشد. به عنوان دو نمونه از نرم های سازگار می توان به نرم الحاقی و نرم فروبنیوس اشاره کرد. و نیز به عنوان مثالی از یک نرم ناسازگار می‌توان به ماکسیمم نرم اشاره کرد

$$\|\mathbf{A}\|_{\max} = \max_{i,j}|a_{ij}|$$

بهترین کرانی که برای هر $\mathbf{A}\in\mathbb{C}^{m\times n}$ و $\mathbf{B}\in\mathbb{C}^{n\times p}$ بدست می‌آید برابرست با:

---

۱۴ Subordinate matrix norm

۱۵ Consistent

۱۶ Frobenius norm

$$\left\|\mathbf{AB}\right\|_{\max} \leq n.\left\|\mathbf{A}\right\|_{\max}.\left\|\mathbf{B}\right\|_{\max}$$

شبیه نرم برداری، هم ارزی دو نرم ماتریسی به صورت زیر تعریف می‌شود.

اگر $\left\|\cdot\right\|_{\nu}$ و $\left\|\cdot\right\|_{\mu}$ دو نرم ماتریسی باشند و $\alpha > 0$ و $\beta > 0$ وجود داشته باشند بطوری که

$$\alpha\left\|\mathbf{A}\right\|_{\mu} \leq \left\|\mathbf{A}\right\|_{\nu} \leq \beta\left\|\mathbf{A}\right\|_{\mu}$$

آنگاه این دو نرم باهم معادلند.

قضیه(۲.۱): فرض کنیم $\mathbf{A} \in \mathbb{C}^{m\times n}$ باشد. در این صورت داریم:

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\left\|\mathbf{A}\right\|_{\infty} \leq \left\|\mathbf{A}\right\|_{2} \leq \sqrt{m}\left\|\mathbf{A}\right\|_{\infty}\\
&\left\|\mathbf{A}\right\|_{2} \leq \left\|\mathbf{A}\right\|_{F} \leq \sqrt{n}\left\|\mathbf{A}\right\|_{2}\\
&\frac{1}{\sqrt{m}}\left\|\mathbf{A}\right\|_{1} \leq \left\|\mathbf{A}\right\|_{2} \leq \sqrt{n}\left\|\mathbf{A}\right\|_{1}\\
&\left\|\mathbf{A}\right\|_{2} \leq \sqrt{\left\|\mathbf{A}\right\|_{1}\left\|\mathbf{A}\right\|_{\infty}}
\end{aligned}$$

برهان: به [9] مراجعه شود. ■

عدد حالت

عدد حالت[۱۷] ماتریس $\mathbf{A}$ متناظر با p- نرم $\left\|\cdot\right\|_{p}$ را با $\kappa_p(\mathbf{A})$ یا $\mathrm{Cond}_p(\mathbf{A})$ نمایش می‌دهند و بصورت زیر تعریف می‌شود:

$$\mathrm{Cond}_p\left(\mathbf{A}\right) = \left\|\mathbf{A}\right\|_{p}.\left\|\mathbf{A}^{-1}\right\|_{p}$$

که در آن معمولاً p یک، دو یا بینهایت در نظر گرفته می‌شود. قرارداد می‌کنیم اگرعدد حالت را به صورت $\mathrm{Cond}\left(\mathbf{A}\right)$ نوشتیم، منظور عدد حالت تحت نرم دو یعنی $\mathrm{Cond}_2\left(\mathbf{A}\right)$ باشد.

عدد حالت می‌تواند در تعیین بدحالت بودن و نزدیک به منفرد بودن یک ماتریس استفاده شود. بطوری که اگر عدد حالت بزرگ باشد این ماتریس یک ماتریس نزدیک به منفرد است. عدد حالت یک ماتریس دارای ویژگی های مهم زیر می‌باشد:

---

[۱۷] Condition number

- تحت هر p – نرم ماتریسی داریم $\mathrm{Cond}_p(\mathbf{A}) \ge 1$.
- اگر $\mathbf{A}$ یک ماتریس متعامد باشد، آنگاه عدد حالت $\mathbf{A}$ تحت نرم دو برابر یک خواهد بود.
- $\mathrm{Cond}(\mathbf{A}^T\mathbf{A}) = (\mathrm{Cond}(\mathbf{A}))^2$.
- $\mathrm{Cond}(\mathbf{A}^T) = \mathrm{Cond}(\mathbf{A})$.
- $\mathrm{Cond}(\mathbf{AB}) \le \mathrm{Cond}(\mathbf{A})\mathrm{Cond}(\mathbf{B})$.
- برای هر اسکالر α داریم $\mathrm{Cond}(\alpha\mathbf{A}) = \mathrm{Cond}(\mathbf{A})$.
- اگر $\lambda_1,\cdots,\lambda_n$ مقادیر ویژه ماتریس $\mathbf{A}$ باشند، بطوری که $|\lambda|_1 \ge |\lambda_n| \ge \cdots \ge |\lambda_n|$ آنگاه $\mathrm{Cond}(\mathbf{A}) \ge |\lambda|_1 / |\lambda_n|$.
- $\mathrm{Cond}(\mathbf{A}) = \sigma_1 / \sigma_n$ بطوری که $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_n$ مقادیر منفرد $\mathbf{A}$ هستند.

همگرایی دنباله ماتریسی

یک دنباله از ماتریس های $\mathbf{A}^{(1)} = (a_{ij}^{(1)}), \mathbf{A}^{(2)} = (a_{ij}^{(2)}),\cdots$ به ماتریس $\mathbf{A} = (a_{ij})$ همگراست اگر

$$\lim_{k\to\infty} a_{ij}^{(k)} = a_{ij} \quad i,j = 1,\cdots,n$$

و اگر دنباله ماتریسی $\{\mathbf{A}^{(k)}\}$ به $\mathbf{A}$ همگرا باشد، می‌نویسیم

$$\lim_{k\to\infty} \mathbf{A}^{(k)} = \mathbf{A}$$

در قضیه زیر شرط لازم و کافی برای همگرایی یک دنباله ماتریسی بیان می‌شود.

قضیه (۳.۱): **دنباله‌ای** از ماتریس های $\mathbf{A}^{(1)},\mathbf{A}^{(2)},\cdots$ به ماتریس $\mathbf{A}$ همگرا هستند اگر و فقط اگر

$$\lim_{k\to\infty} \left\|\mathbf{A} - \mathbf{A}^{(k)}\right\| = 0$$

که $\|\cdot\|$ هر نرم ماتریسی می‌تواند باشد.

برهان: به [6] مراجعه شود.■

همچنین نمادگذاری $\sum_{k=1}^{\infty}\mathbf{A}_k$ به معنی $\lim_{k\to\infty}\sum_{i=1}^{k}\mathbf{A}_i$ می‌باشد. شرایط همگرایی یک سری ماتریسی در قضیه زیر بیان شده است.

قضیه (٤.١): سری ماتریسی $\sum_{k=1}^{\infty} \mathbf{A}_k$ همگراست اگر سری $\sum_{k=1}^{\infty} \left\| \mathbf{A}_k \right\|$ برای بعضی نرم های ماتریسی همگرا باشد.

برهان: به [6] مراجعه شود. ■

دستگاه اعداد ممیز سیار

یک عدد ممیز سیار نرمال شده غیر صفر مانند x در مبنای دو دارای فرم زیر است:

$$
\begin{aligned}
x &= \left(-1\right)^s d_1.d_2d_3\cdots d_t 2^e \\
x &= \pm d_1.d_2\cdots d_t 2^e \\
x &= \pm r 2^e
\end{aligned}
$$

وقتی e نما یا توان[۱۸]، r ارقام معنی دار[۱۹]، t دقت[۲۰]، $\left(-1\right)^s$ علامت و $\left(.d_2d_3\cdots d_t\right)$ قسمت اعشاری[۲۱] نامیده می‌شوند. توجه داریم که t متناهی است و

$$
\begin{aligned}
d_1 &= 1 \\
d_i &= 0 \ or \ 1 \qquad 2 \le i \le t
\end{aligned}
$$

در قضیه زیر خطای نمایش عدد حقیقی x در سیستم ممیز سیار را ارائه می‌دهیم.

قضیه (٥.١): فرض کنیم $fl(x)$ نمایش ممیز سیار عدد حقیقی x باشد. آنگاه

$$
\frac{\left|fl(x) - x\right|}{\left|x\right|} \le \mu = \begin{cases} \dfrac{1}{2}\beta^{1-t} & \text{for rounding} \\ \beta^{1-t} & \text{for chopping} \end{cases}
$$

برهان: به [6] مراجعه شود. ■

تعریف: عدد μ در قضیه فوق دقت ماشین یا خطای گرد کردن[۲۲]نامیده می‌شود و کوچکترین عدد ممیز سیار مثبتی است که $fl(1+\mu) > 1$.

---

[۱۸] Exponent
[۱۹] Significant
[۲۰] Precision
[۲۱] Fraction

در اغلب ماشینها μ از مرتبه 10^{-16} و 10^{-6} به ترتیب برای دقت مضاعف و معمولی می باشد.

قضیه (۶.۱): فرض کنیم x و y نمایش دو عدد ممیز سیار باشد و $fl(x+y)$، $fl(x-y)$، $fl(x.y)$ و $fl(x/y)$ به ترتیب نمایشگر محاسبه جمع، تفریق، ضرب و تقسیم باشند. آنگاه

- $fl(x \pm y) = (x \pm y)(1+\delta)$ وقتی $|\delta| \le \mu$.
- $fl(x.y) = (x.y)(1+\delta)$ وقتی $|\delta| \le \mu$.
- اگر $y \neq 0$ آنگاه $fl(x/y) = (x/y)(1+\delta)$ وقتی $|\delta| \le \mu$.

در کامپیوترهایی که از IEEE استاندارد استفاده می‌کنند، قانون زیر ممکن است درست باشد:

- $fl(x+y) = x(1+\delta_1) + y(1+\delta_2)$ وقتی $|\delta_1| \le \mu$ و $|\delta_2| \le \mu$.

برهان: به [6] مراجعه شود. ■

در قضایای زیر جمع و ضرب n عدد ممیز سیار را نمایش می‌دهیم.

قضیه(۷.۱): فرض کنیم $x_1,\cdots,x_n$، n عدد ممیز سیار باشند. آنگاه

$$\begin{aligned} & fl(x_1 + x_2 + \cdots + x_n) - (x_1 + x_2 + \cdots + x_n) \\ & \approx x_1(\delta_1 + \delta_2 + \cdots + \delta_n) + x_2(\delta_2 + \cdots + \delta_n) + \cdots + x_n\delta_n \end{aligned}$$

وقتی برای هر $i = 1,\cdots,n$ داشته باشیم $|\delta_i| \le \mu$.

برهان: به مرجع [6] مراجعه شود. ■

قضیه(۸.۱): فرض کنیم $x_1,\cdots,x_n$، n عدد ممیز سیار باشند. آنگاه

$$fl(x_1 \times x_2 \times \cdots \times x_n) \approx (1+\varepsilon)\prod_{i=1}^{n} x_i$$

وقتی $\left|(1+\delta_2)(1+\delta_3)\cdots(1+\delta_n) - 1\right| = \varepsilon$ و برای هر $i = 1,\cdots,n$ داشته باشیم $|\delta_i| \le \mu$.

برهان: به مرجع [6] مراجعه شود. ■

لم (۱.۱): اگر $|\delta_i| \le \mu$ و برای $i = 1,\cdots,n$ داشته باشیم $\rho_i = \pm 1$ و نیز $n\mu < 1$، آنگاه

---

[۲۲] Roundoff error

$$\prod_{i=1}^{n}\left(1+\delta_{i}\right)^{\rho_i}=1+\theta_{n}$$

وقتی

$$\left|\theta_{n}\right| \leq \frac{n\mu}{1-n\mu}=\gamma_{n}$$

برهان: به [12] مراجعه شود. ■

لم(۲.۱): برای هر عدد صحیح مثبت k فرض کنیم θ_k مقدار کرانی باشد که $\left|\theta_{k}\right| \leq \gamma_{k}=\dfrac{k\mu}{1-k\mu}$. بنا براین روابط زیر درست هستند:

$$\left(1+\delta_{k}\right)\left(1+\delta_{j}\right)=\left(1+\delta_{k+j}\right)$$

$$\frac{\left(1+\delta_{k}\right)}{\left(1+\delta_{j}\right)}=\begin{cases}1+\delta_{k+j} & (j \leq k) \\ 1+\delta_{k+2j} & (j>k)\end{cases}$$

$$\gamma_{k}\gamma_{j} \leq \gamma_{\min(k,j)}$$

$$i\gamma_{k} \leq \gamma_{ik}$$

$$\gamma_{k}+u \leq \gamma_{k+1}$$

$$\gamma_{k}+\gamma_{j}+\gamma_{kj} \leq \gamma_{k+j}$$

برهان: به [12] مراجعه شود. ■

در بعضی تحلیل های خطا، استفاده از نمادگذاری فوق می تواند بسیار پرزحمت باشد. بنابراین برای رسیدن به دقت بیشتر در جملات γ_k می‌توان از نمادگذاری مفید زیر استفاده کرد:

$$\tilde{\gamma}_{k}=\frac{cku}{1-cku}$$

که در آن c یک ثابت صحیح کوچک می باشد. اکنون می‌توان کران خطایی برای محاسبات ماتریس ممیز سیار بدست آورد.

قضیه(۹.۱): فرض کنیم $\left|\mathbf{M}\right|=\left(\left|m_{ij}\right|\right)$ و نیز $\mathbf{A}$ و $\mathbf{B}$ دو ماتریس ممیز سیار و c یک عدد ممیز سیار باشند. آنگاه

- $fl\left(c\mathbf{A}\right)=c\mathbf{A}+\mathbf{E}$ که در آن $\left|\mathbf{E}\right|\le\mu\left|c\mathbf{A}\right|$.
- $fl\left(\mathbf{A}+\mathbf{B}\right)=\left(\mathbf{A}+\mathbf{B}\right)+\mathbf{E}$ که در آن $\left|\mathbf{E}\right|\le\mu\left(\left|\mathbf{A}+\mathbf{B}\right|\right)$.

اگر $\mathbf{A}$ و $\mathbf{B}$ ماتریس هایی باشند که ضرب آنهاتعریف شده باشد، آنگاه

- $fl\left(\mathbf{AB}\right)=\left(\mathbf{AB}\right)+\mathbf{E}$ که در آن $\left|\mathbf{E}\right|\le n\mu\left|\mathbf{A}\right|\left|\mathbf{B}\right|+O\left(\mu^2\right)$.

برهان: به [6] مراجعه شود. ■

توجه: در قضیه فوق عبارت $O\left(\mu^2\right)$ مرتبه خطا را بیان می‌کند و اگر c یک ثابت باشد، $c\mu^2$ کرانی برای آن می‌باشد.

توجه: چون $\left\|\left|\mathbf{E}\right|\right\|\le\left\|\mathbf{E}\right\|$، می‌توانیم معادله آخر را بصورت $fl\left(\mathbf{AB}\right)=\left(\mathbf{AB}\right)+\mathbf{E}$ وقتی $\left\|\mathbf{E}\right\|\le\left\|\left|\mathbf{E}\right|\right\|\le n\mu\left\|\left|\mathbf{A}\right|\right\|\left\|\left|\mathbf{B}\right|\right\|+O\left(\mu^2\right)$ نوشت. به ویژه برای نرم های $\left\|\cdot\right\|_2$ و $\left\|\cdot\right\|_\infty$ خواهیم داشت:

$$\left\|\mathbf{E}\right\|_\infty\le n\mu\left\|\mathbf{A}\right\|_\infty\left\|\mathbf{B}\right\|_\infty+O\left(\mu^2\right)$$

$$\left\|\mathbf{E}\right\|_2\le n^2\mu\left\|\mathbf{A}\right\|_2\left\|\mathbf{B}\right\|_2+O\left(\mu^2\right)$$

قضیه (۱۰.۱): $fl\left(\mathbf{AB}\right)=\left(\mathbf{AB}\right)+\mathbf{E}$ وقتی $\left\|\mathbf{E}\right\|_2\le n^2\mu\left\|\mathbf{A}\right\|_2\left\|\mathbf{B}\right\|_2+O\left(\mu^2\right)$.

برهان: به [6] مراجعه شود.■

نتیجه (ضرب ماتریس و بردار): اگر b یک بردار باشد، داریم:

$$fl\left(\mathbf{A}b\right)=\left(\mathbf{A}b\right)+e$$

وقتی $\left\|e\right\|_2\le n^2\mu\left\|\mathbf{A}\right\|_2\left\|\mathbf{B}\right\|_2+O\left(\mu^2\right)$.

نتیجه (ضرب ماتریس و یک ماتریس متعامد): فرض کنیم $\mathbf{A}\in\mathbb{R}^{n\times n}$ و $\mathbf{Q}\in\mathbb{R}^{n\times n}$ یک ماتریس متعامد باشد. آنگاه

$$fl\left(\mathbf{QA}\right)=\mathbf{Q}\left(\mathbf{A}+\mathbf{E}\right)$$

وقتی $\left\|\mathbf{E}\right\|_2\le O\left(\mu\right)\left\|\mathbf{A}\right\|_2$.