

پایداری عددی در مجموع‌یابی مختلط

حسن مجیدیان*

پژوهشگاه علوم انسانی و مطالعات فرهنگی، پژوهشکده دانشنامه‌نگاری، گروه مطالعات میان‌رشته‌ای

مدینه فیروزی، امجد علی‌پناه

دانشگاه کردستان، دانشکده علوم پایه، گروه ریاضی

پذیرش ۹۸/۰۸/۲۸

دریافت ۹۷/۰۷/۲۵

چکیده

عددهای حالت طبیعی و عملی، که محکی برای میزان خطای ناشی از پدیده ازدست‌رفتن رقم‌های بامعنی در محاسبه مجموع عددهای مختلط است، معرفی می‌شوند. رابطه بین این دو عدد و اندازه آنها بررسی می‌شوند. ثابت می‌شود که اگر عدد حالت عملی کوچک باشد، عدد حالت طبیعی نیز کوچک خواهد بود؛ و اگر عدد حالت طبیعی بزرگ باشد، عدد حالت عملی نیز بزرگ خواهد بود. نشان داده می‌شود که عدد حالت عملی برای مجموع‌یابی عددهایی که همگی در یکی از ربع‌های صفحه مختلط قرار دارند، برابر یک است. همچنین یک کران بالا برای عدد حالت طبیعی، وقتی عددها در یک قطاع با زاویه نابیش‌تر از $\pi/2$ قرار دارند، ارائه می‌شود.

واژه‌های کلیدی: مجموع‌یابی مختلط، عدد حالت، پایداری عددی، ازدست‌رفتن رقم‌های بامعنی.

مقدمه

ازدست‌رفتن رقم‌های بامعنی^۱ در تفریق اعداد تقریبی نزدیک به هم، می‌تواند به خطای محاسباتی غیرقابل پذیرشی منجر شود. فرض کنید $\hat{a} = a(1 + \delta a)$ و $\hat{b} = b(1 + \delta b)$ تقریب‌های دو عدد، به ترتیب، a و b با خطاهای نسبی $|\delta a|$ و $|\delta b|$ باشند. در این صورت

$$\left| \frac{(\hat{a} + \hat{b}) - (a + b)}{a + b} \right| = \left| \frac{a\delta a - b\delta b}{a + b} \right| \leq \frac{|a| + |b|}{|a + b|} \max\{|\delta a|, |\delta b|\}.$$

یعنی حاصل ضرب عبارت $\kappa := (|a| + |b|)/|a + b|$ در بزرگ‌ترین خطای نسبی داده‌ها، کران بالایی برای خطای نسبی مجموع است. بنابراین عدد $\kappa \geq 1$ را می‌توان به منزله عدد حالت مجموع در نظر گرفت. مثلاً اگر a و b دو عدد با علامت‌های متفاوت و اندازه‌های بسیار نزدیک به هم باشند، آن‌گاه κ می‌تواند خیلی بزرگ شود و در نتیجه امکان زیاد شدن خطای نسبی مجموع فراهم شود. این همان پدیده از دست‌رفتن رقم‌های بامعنی در تفریق است.

مطلب بالا را می‌توان برای مجموع بیش از دو عدد حقیقی تعمیم داد. به‌طور کلی در محاسبه مجموع عددهای

حقیقی $\sum_{j=1}^n x_j$ عدد حالت بدین‌صورت تعریف می‌شود [۲]:

$$\kappa := \frac{\sum_{j=1}^n |x_j|}{\left| \sum_{j=1}^n x_j \right|}. \quad (1)$$

بدیهی است که $K \geq 1$. هرچه قدر K به ۱ نزدیک‌تر باشد، میزان خطای نسبی مجموع کم‌تر است و اگر عدد حالت بزرگ باشد، خطا «می‌تواند» زیاد باشد. یک راه برای کاهش خطای ناشی از پدیده از دست رفتن رقم‌های بامعنی، جمع کردن عددها به ترتیب نزولی است [۲]، [۶].

با توجه به همین تعریف از عدد حالت است که قاعده انتگرال‌گیری

$$\int_a^b f(x)dx \approx \sum_{j=1}^n w_{j,n} f(x_{j,n}) \quad (۲)$$

پایدار خوانده می‌شود هرگاه وزن‌های $w_{j,n}$ همگی هم‌علامت باشند. مثلاً اگر همه وزن‌ها مثبت باشند و تابع نیز در بازه انتگرال‌گیری تغییر علامت ندهد، آن‌گاه

$$\kappa = \frac{\sum_{j=1}^n |w_{j,n} f(x_{j,n})|}{|\sum_{j=1}^n w_{j,n} f(x_{j,n})|} = \frac{\sum_{j=1}^n w_{j,n} |f(x_{j,n})|}{|\sum_{j=1}^n w_{j,n} f(x_{j,n})|} = 1.$$

این تعریف از عدد حالت یک مجموع را می‌توان به‌طور طبیعی به مجموع مختلط $\sum_{j=1}^n z_j$ نیز تعمیم داد:

$$\kappa_{\text{Nat}} = \frac{\sum_{j=1}^n |z_j|}{|\sum_{j=1}^n z_j|}. \quad (۳)$$

پیش‌تر (۳) به‌عنوان "تعمیم عدد حالت حقیقی" مطرح شده است (مثلاً ر.ک. [۱]، [۲]). ولی ما از عنوان "عدد حالت طبیعی" و نماد κ_{Nat} استفاده می‌کنیم چون در مقابل آن "عدد حالت عملی" نیز مطرح می‌شود. از آن‌جاکه مجموع عددهای مختلط با مجموع مجزای قسمت‌های حقیقی و موهومی آنها به‌دست می‌آید، در عمل با عدد حالت (۴) مواجه هستیم:

$$\kappa_{\text{Pra}} = \max \left\{ \frac{\sum_{j=1}^n |x_j|}{|\sum_{j=1}^n x_j|}, \frac{\sum_{j=1}^n |y_j|}{|\sum_{j=1}^n y_j|} \right\}, \quad (۴)$$

که در آن x_j و y_j به ترتیب، قسمت‌های حقیقی و موهومی z_j هستند. توجه کنید که بیشینه خطاهای نسبی در دو مجموع $\sum_{j=1}^n x_j$ و $\sum_{j=1}^n y_j$ محکی برای دقت در محاسبه مجموع مختلط $\sum_{j=1}^n z_j$ است و κ_{Pra} عدد حالت این محک است در صورتی‌که κ_{Nat} عدد حالت خطای نسبی خود مجموع $\sum_{j=1}^n z_j$ است. مفهوم عدد حالت عملی و همچنین نمادهای κ_{Pra} و κ_{Nat} برای نخستین بار در این‌جا معرفی شده‌اند.

مسئله از دست رفتن رقم‌های بامعنی در مجموع‌یابی حقیقی به‌طور مفصل بررسی شده است (مثلاً ن.ک. [۳]، [۴]، [۵]، [۷]، [۹]، [۱۰]، [۱۱]). در مورد مجموع‌یابی مختلط، مقاله اخیر [۱] به بررسی عدد حالت طبیعی یک مجموع‌یابی خاص پرداخته است که درواقع فرمولی برای محاسبه مشتق‌های یک تابع تحلیلی برپایه فرمول کوشی-گورسا است.

در این مقاله، برای نخستین بار، عددهای حالت κ_{Nat} و κ_{Pra} تعریف شده و با توجه به موقعیت z_j ‌ها در صفحه مختلط، به‌طور عام بررسی می‌شوند. برای این کار به‌تعریف یک $(\theta^{(1)}, \theta^{(2)})$ -قطاع و قطاع قائم نیاز است.

به‌سادگی نشان داده می‌شود که اگر z_j ‌ها همگی در یکی از ربع‌های صفحه مختلط واقع شوند، آن‌گاه $\kappa_{\text{Pra}} = 1$. سپس نشان می‌دهیم که هر چه زاویه قطاعی که z_j ‌ها را دربر می‌گیرد کوچک‌تر باشد، κ_{Nat} کران کوچک‌تری خواهد داشت و برای قطاع‌های با زاویه بزرگ‌تر، این عدد می‌تواند بسیار بزرگ شود. به‌طور خاص، یک کران کوچک برای κ_{Nat} وقتی z_j ‌ها در هر قطاع قائمی قرار دارند، ارائه می‌شود. به بیان دقیق‌تر ثابت می‌کنیم که اگر z_j ‌ها همگی در قطاعی با زاویه $\gamma \leq \pi/2$ باشند، آن‌گاه $\kappa_{\text{Nat}} \leq \gamma / \sin \gamma$.

ادامه مطالب در این مقاله به‌این صورت سازمان‌دهی شده است. بخش دوم، مفهوم‌ها و تعریف‌های مورد نیاز را دربردارد. در این بخش مفهوم یک $(\theta^{(1)}, \theta^{(2)})$ -قطاع به‌طور خاص معرفی می‌شود. در بخش سوم، ارتباط بین κ_{Nat} و κ_{Pra} در حالت کلی بررسی می‌شود؛ درواقع با چند مثال نشان داده می‌شود که این دو عدد رابطه مقایسه‌ای با

یکدیگر ندارند. ولی وابستگی‌هایی بین اندازه آن دو وجود دارد که اثبات می‌شوند. نتیجه‌های اصلی مقاله درباره کران‌های K_{Pra} و K_{Nat} در بخش چهارم می‌آیند. بخش پایانی نیز به نتیجه‌گیری اختصاص دارد.

پیش‌نیازها

در این بخش به تعریف $(\theta^{(1)}, \theta^{(2)})$ -قطاع و قطاع قائم که در مطالعه K_{Pra} و K_{Nat} اساسی هستند می‌پردازیم. در سراسر مقاله یک موهومی را با نماد i نشان می‌دهیم.

تعریف ۱. فرض کنید $-\pi < \theta^{(1)} < \theta^{(2)} \leq \pi$. در این صورت $(\theta^{(1)}, \theta^{(2)})$ -قطاع، که با نماد $S(\theta^{(1)}, \theta^{(2)})$ نشان داده می‌شود، مجموعه‌ای از عددها در صفحه مختلط است که بدین صورت تعریف می‌شود:

اگر $\theta^{(2)} - \theta^{(1)} \leq \pi$

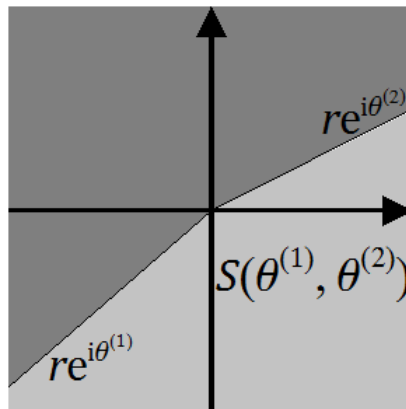
$$S(\theta^{(1)}, \theta^{(2)}) := \{z = re^{i\theta} \mid \theta \in (\theta^{(1)}, \theta^{(2)})\},$$

و اگر $\theta^{(2)} - \theta^{(1)} > \pi$

$$S(\theta^{(1)}, \theta^{(2)}) := \{z = re^{i\theta} \mid \theta \in (-\pi, \theta^{(1)}) \cup (\theta^{(2)}, \pi]\}.$$

زاویه قطاع $S(\theta^{(1)}, \theta^{(2)})$ برابر است با $\theta^{(2)} - \theta^{(1)}$ اگر $\theta^{(2)} - \theta^{(1)} \leq \pi$ و $2\pi - \theta^{(2)} + \theta^{(1)}$ اگر $\theta^{(2)} - \theta^{(1)} > \pi$.

در واقع اگر $re^{i\theta^{(1)}}$ و $re^{i\theta^{(2)}}$ ، $r > 0$ ، دو شعاع باشند، صفحه مختلط را به دوناچه بی‌کران افراز می‌کنند (شکل ۱)، ناحیه کوچک‌تر همان $(\theta^{(1)}, \theta^{(2)})$ -قطاع است که در بالا تعریف شد. زاویه قطاع نیز زاویه نابیش‌تر از π بین دو شعاع است.



شکل ۱. دو ناحیه (بی‌کران) خاکستری کم‌رنگ و پررنگ جدا شده به وسیله شعاع‌های $re^{i\theta^{(1)}}$ و $re^{i\theta^{(2)}}$ ناحیه کم‌رنگ مجموعه $S(\theta^{(1)}, \theta^{(2)})$ را نشان می‌دهد.

تعریف ۲. قطاع با زاویه $\pi/2$ را قائم می‌خوانیم.

به عنوان مثال، محورهای حقیقی و موهومی، صفحه مختلط را به چهار قطاع قائم تقسیم می‌کنند.

ارتباط بین K_{Pra} و K_{Nat}

در این بخش، وابستگی‌هایی بین اندازه K_{Pra} و K_{Nat} را ثابت می‌کنیم. ولی ابتدا با ارائه چند مثال، نشان می‌دهیم که بین این دو عدد هیچ رابطه مقایسه‌ای وجود ندارد.

مثال ۱. فرض کنید $z_1 = 1 + i$ و $z_2 = -2 - 4i$ در این صورت

$$|z_1| = \sqrt{2}, \quad |z_2| = 2\sqrt{5}, \quad |z_1 + z_2| = \sqrt{10}.$$

بنابر تعریف داریم

$$\kappa_{\text{Nat}} = \frac{\sqrt{2} + 2\sqrt{5}}{\sqrt{10}} \approx 1.86.$$

از طرفی

$$\begin{aligned} \kappa_{\text{Pra}} &= \max \left\{ \frac{\sum_{j=1}^2 |x_j|}{|\sum_{j=1}^2 x_j|}, \frac{\sum_{j=1}^2 |y_j|}{|\sum_{j=1}^2 y_j|} \right\} = \max \left\{ \frac{|1| + |-2|}{|1 - 2|}, \frac{|1| + |-4|}{|1 - 4|} \right\} \\ &= \max\{3, 5/3\} = 3. \end{aligned}$$

بنابراین $\kappa_{\text{Nat}} < \kappa_{\text{Pra}}$ **مثال ۲.** فرض کنید $z_1 = -2i$ و $z_2 = 2$ در این صورت

$$|z_1| = |z_2| = 2, \quad |z_1 + z_2| = 2\sqrt{2}.$$

بنابر تعریف داریم

$$\kappa_{\text{Nat}} = \frac{2 + 2}{2\sqrt{2}} = \sqrt{2}.$$

از طرفی

$$\begin{aligned} \kappa_{\text{Pra}} &= \max \left\{ \frac{\sum_{j=1}^2 |x_j|}{|\sum_{j=1}^2 x_j|}, \frac{\sum_{j=1}^2 |y_j|}{|\sum_{j=1}^2 y_j|} \right\} = \max \left\{ \frac{|2| + |0|}{|2 + 0|}, \frac{|0| + |-2|}{|0 - 2|} \right\} \\ &= \max\{1, 1\} = 1. \end{aligned}$$

پس $\kappa_{\text{Nat}} > \kappa_{\text{Pra}}$ **مثال ۳.** فرض کنید $z_1 = z_2 = 1 + i$ در این صورت

$$|z_1| = |z_2| = \sqrt{2}, \quad |z_1 + z_2| = 2\sqrt{2}.$$

بنابر تعریف داریم

$$\kappa_{\text{Nat}} = \frac{\sqrt{2} + \sqrt{2}}{2\sqrt{2}} = 1.$$

از طرفی

$$\kappa_{\text{Pra}} = \max \left\{ \frac{\sum_{j=1}^2 |x_j|}{|\sum_{j=1}^2 x_j|}, \frac{\sum_{j=1}^2 |y_j|}{|\sum_{j=1}^2 y_j|} \right\} = \max \left\{ \frac{|1| + |1|}{|1 + 1|}, \frac{|1| + |1|}{|1 + 1|} \right\} = \max\{1, 1\} = 1$$

پس $\kappa_{\text{Nat}} = \kappa_{\text{Pra}}$

بنابراین، باتوجه به مثال‌های ۱-۳ می‌توان گفت بین κ_{Pra} و κ_{Nat} هیچ رابطه مقایسه‌ای وجود ندارد. با این حال نتیجه زیر برقرار است که از آن، وابستگی‌هایی بین اندازه این دو عدد به دست می‌آید.

قضیه ۱. اگر $z_j = x_j + iy_j$ ، $j = 1, \dots, n$ آن‌گاه

$$\frac{\sqrt{2}}{2} \min \left\{ \frac{\sum_{j=1}^n |x_j|}{|\sum_{j=1}^n x_j|}, \frac{\sum_{j=1}^n |y_j|}{|\sum_{j=1}^n y_j|} \right\} \leq \kappa_{\text{Nat}} \leq \frac{\sum_{j=1}^n |x_j|}{|\sum_{j=1}^n x_j|} + \frac{\sum_{j=1}^n |y_j|}{|\sum_{j=1}^n y_j|}. \quad (5)$$

برهان: داریم

$$|\sum_{j=1}^n z_j|^2 = (\sum_{j=1}^n x_j)^2 + (\sum_{j=1}^n y_j)^2 \leq 2 \max \left\{ (\sum_{j=1}^n x_j)^2, (\sum_{j=1}^n y_j)^2 \right\}. \quad (6)$$

از طرفی

$$\sum_{j=1}^n (x_j^2 + y_j^2)^{1/2} \geq \sum_{j=1}^n |x_j|, \quad \sum_{j=1}^n (x_j^2 + y_j^2)^{1/2} \geq \sum_{j=1}^n |y_j|. \quad (7)$$

از (۶) و (۷) نتیجه می‌شود

$$\begin{aligned}\kappa_{\text{Nat}} &= \frac{\sum_{j=1}^n (x_j^2 + y_j^2)^{1/2}}{\left(\left(\sum_{j=1}^n x_j \right)^2 + \left(\sum_{j=1}^n y_j \right)^2 \right)^{1/2}} \geq \frac{\sum_{j=1}^n (x_j^2 + y_j^2)^{1/2}}{\sqrt{2} \max\{|\sum_{j=1}^n x_j|, |\sum_{j=1}^n y_j|\}} \\ &\geq \frac{\sqrt{2}}{2} \min \left\{ \frac{\sum_{j=1}^n |x_j|}{|\sum_{j=1}^n x_j|}, \frac{\sum_{j=1}^n |y_j|}{|\sum_{j=1}^n y_j|} \right\}.\end{aligned}$$

قسمت دوم (نابرابری سمت راست (۵)) نیز از نابرابری مثلثی

$$(x_j + y_j)^{1/2} \leq |x_j| + |y_j|$$

و نابرابری‌های زیر به‌دست می‌آید:

$$\left(\sum_{j=1}^n x_j \right)^2 + \left(\sum_{j=1}^n y_j \right)^2 \geq \left(\sum_{j=1}^n x_j \right)^2, \quad \left(\sum_{j=1}^n x_j \right)^2 + \left(\sum_{j=1}^n y_j \right)^2 \geq \left(\sum_{j=1}^n y_j \right)^2.$$

از رابطه (۵) نتیجه‌های زیر به‌دست می‌آیند. به‌دلیل ساده بودن برهان‌ها، از آوردن آنها خودداری کرده‌ایم.

نتیجه ۱. اگر $\kappa_{\text{Pra}} \leq U$ ، آن‌گاه $\kappa_{\text{Nat}} \leq 2U$.

گزاره اخیر نشان می‌دهد که اگر عدد حالت عملی کوچک باشد، آن‌گاه عدد حالت طبیعی نیز کوچک است. عکس این مطلب در حالت کلی درست نیست. مثلاً برای دو عدد $z_1 = -1 + (1 + \varepsilon)i$ و $z_2 = -1 - i$ ، به ازای $\varepsilon > 0$ بسیار کوچک، عدد حالت عملی برابر است با $1 + 2\varepsilon^{-1}$ که می‌تواند خیلی بزرگ باشد. ولی

$$\kappa_{\text{Nat}} = \frac{\sqrt{2} + \sqrt{1 + (1 + \varepsilon)^2}}{\sqrt{4 + \varepsilon^2}} \rightarrow \sqrt{2},$$

وقتی $\varepsilon \rightarrow 0$.

نتیجه ۲. اگر $L \leq \kappa_{\text{Nat}}$ ، آن‌گاه $L/2 \leq \kappa_{\text{Pra}}$.

گزاره اخیر نشان می‌دهد که اگر عدد حالت طبیعی بزرگ باشد، آن‌گاه عدد حالت عملی نیز بزرگ است. چنان‌که در مثال اخیر (ذیل نتیجه ۱) دیدیم، عکس این مطلب در حالت کلی درست نیست.

نتیجه ۳. اگر z ها همگی در یک ربع صفحه مختلط واقع شوند، آن‌گاه $\kappa_{\text{Nat}} \leq 2$.

این نتیجه با گزاره بدیهی ۱ در مورد عدد حالت عملی، قابل‌مقایسه است.

گزاره ۱. اگر z ها همگی در یک ربع صفحه مختلط واقع شوند، آن‌گاه $\kappa_{\text{Pra}} = 1$.

در بخش بعد، کرانی برای κ_{Nat} وقتی z ها همگی در یک قطاع قائم قرار می‌گیرند، ارائه می‌دهیم. بنابراین تعمیمی از نتیجه ۳ داریم.

محک‌هایی برای اندازه عدد حالت طبیعی

فرم قطبی عددهای مختلط،

$$z_j = r_j e^{i\theta_j}, \quad r_j > 0, \quad \theta_j \in (-\pi, \pi],$$

را در نظر بگیرید. بنابه فرمول اوایلر داریم

$$|\sum_{j=1}^n z_j|^2 = \left(\sum_{j=1}^n r_j \cos \theta_j \right)^2 + \left(\sum_{j=1}^n r_j \sin \theta_j \right)^2. \quad (8)$$

در نتیجه

$$\kappa_{\text{Nat}}^2 = \frac{\left(\sum_{j=1}^n r_j \right)^2}{\left(\sum_{j=1}^n r_j \cos \theta_j \right)^2 + \left(\sum_{j=1}^n r_j \sin \theta_j \right)^2} = \frac{\left(\sum_{j=1}^n r_j \right)^2}{A}, \quad (9)$$

که در آن

$$A := \sum_{j=1}^n r_j^2 + 2 \sum_{j \leq k} r_j r_k \cos(\theta_j - \theta_k).$$

رابطه (۹) نقش $\cos(\theta_j - \theta_k)$ به‌ازای r_j های ثابت را نشان می‌دهد. یعنی اگر طول عددها ثابت باشند، زاویه قطاعی که آنها را دربرمی‌گیرد در مقدار عدد حالت طبیعی نقش کلیدی بازی می‌کند. هرچه زاویه قطاع کوچک‌تر باشد، مقدار $\cos(\theta_j - \theta_k)$ ، به‌ازای هر j و k ($j \leq k$)، به ۱ نزدیک‌تر است و در نتیجه A به بیشینه خود نزدیک‌تر می‌شود و κ_{Nat} کوچک‌تر می‌شود. در نهایت اگر زاویه قطاع به صفر میل کند، $\cos(\theta_j - \theta_k) \rightarrow 1$ ، و در نتیجه

$$\kappa_{\text{Nat}} \rightarrow \frac{\sum_{j=1}^n r_j}{\left(\sum_{j=1}^n r_j^2 + 2 \sum_{j \leq k} r_j r_k\right)^{1/2}} = \frac{\sum_{j=1}^n r_j}{\left|\sum_{j=1}^n r_j\right|} = 1.$$

مطلب اخیر را به‌طور مستقیم نیز می‌توان نشان داد. فرض کنید $z_j = r_j e^{i\theta}$ ، به‌ازای $\theta \in (-\pi, \pi]$. در این صورت

$$\kappa_{\text{Nat}} = \frac{\sum_{j=1}^n r_j}{\left|e^{i\theta} \sum_{j=1}^n r_j\right|} = 1.$$

در حالت ساده $n = 2$

$$A = r_1^2 + r_2^2 + 2r_1 r_2 \cos(\theta_1 - \theta_2).$$

مقدار A به عدد $\cos(\theta_1 - \theta_2)$ وابسته است که هر چه به -1 نزدیک‌تر باشد (برای r_1 و r_2 ثابت)، مقدار A کم‌تر می‌شود. در ادامه، گزاره‌هایی را درباره κ_{Nat} ، در این حالت خاص به‌دست می‌آوریم.

کم‌ترین مقدار A ، یعنی بدترین عدد حالت، زمانی رخ می‌دهد که زاویه بین z_1 و z_2 ، برابر با π باشد. در این صورت

$$\kappa_{\text{Nat}}^2 = \frac{r_1^2 + r_2^2 + 2r_1 r_2}{r_1^2 + r_2^2 - 2r_1 r_2}. \quad (10)$$

حال کافی است مقادیر r_1 و r_2 را با هم مقایسه کنیم. مقدار r_1 بالاخره ضریبی از r_2 است، یعنی $\alpha > 0$ چنان وجود دارد که $r_1 = \alpha r_2$. در نتیجه رابطه (۱۰) بدین صورت ساده می‌شود:

$$\kappa_{\text{Nat}}^2 = \frac{(\alpha + 1)^2}{(\alpha - 1)^2}.$$

سه رفتار مجانبی را برای α در نظر می‌گیریم:

- اگر $\alpha \rightarrow 1$ ، یعنی z_1 به قرینه z_2 نسبت به مبدا میل کند، آن‌گاه $\kappa_{\text{Nat}} \rightarrow \infty$.
- اگر $\alpha \rightarrow 0$ ، یعنی طول z_1 در قیاس با z_2 خیلی کوچک باشد، آن‌گاه $\kappa_{\text{Nat}} \rightarrow 1$.
- اگر $\alpha \rightarrow \infty$ ، یعنی طول z_1 در قیاس با z_2 خیلی بزرگ باشد، آن‌گاه $\kappa_{\text{Nat}} \rightarrow 1$.

اگر زاویه بین z_1 و z_2 قائمه باشد، آن‌گاه

$$\kappa_{\text{Nat}}^2 = \frac{(r_1 + r_2)^2}{r_1^2 + r_2^2}.$$

در نتیجه

$$\kappa_{\text{Nat}}^2 = \frac{(1 + \alpha)^2}{1 + \alpha^2} = 1 + \frac{2\alpha}{1 + \alpha^2} \leq 2.$$

در حالت کلی $n \geq 2$ ، با تغییری در لم ۶.۳ مرجع [۸]، می‌توان کرانی برای عدد حالت طبیعی مربوط به z_j های

واقع در یک قطاع قائم یا کوچک‌تر به‌دست آورد.

قضیه ۲. فرض کنید z_1, \dots, z_n در قطاع $S(\theta^{(1)}, \theta^{(2)})$ با زاویه γ نابیش‌تر از $\pi/2$ قرار گرفته باشند. در این صورت

$$\kappa_{\text{Nat}} \leq \frac{\gamma}{\sin \gamma}.$$

برهان: صورت قطبی $z_j = r_j e^{i\theta_j}$ ، به‌ازای $-\pi < \theta_j \leq \pi$ ، را در نظر بگیرید. اگر θ شناسه عددی درون $S(\theta^{(1)}, \theta^{(2)})$ باشد، آن‌گاه $\cos(\theta_j - \theta) > 0$ ، برای هر $j = 1, \dots, n$ در نتیجه

$$|\sum_{j=1}^n z_j| = |\sum_{j=1}^n e^{-i\theta} z_j| \geq \Re \sum_{j=1}^n r_j e^{i(\theta_j - \theta)} = \sum_{j=1}^n r_j \cos(\theta_j - \theta). \quad (11)$$

اگر $\theta^{(2)} - \theta^{(1)} \leq \pi$ ، آن‌گاه طبق تعریف $\gamma = \theta^{(2)} - \theta^{(1)}$ و داریم

$$\int_{\theta^{(1)}}^{\theta^{(2)}} \cos(\theta_j - \theta) d\theta = \sin(\gamma - \beta_j) + \sin \beta_j. \quad (12)$$

که در آن $\beta_j := \theta_j - \theta^{(1)}$. عبارت سمت راست برای (۱۲)، تابعی مقعر از β_j روی $(0, \gamma)$ است از این‌رو، کمینه خود را در یکی از نقطه‌های انتهایی آن می‌گیرد. بنابراین

$$\sin(\gamma - \beta_j) + \sin \beta_j \geq \sin \gamma. \quad (13)$$

با انتگرال‌گیری از دوطرف نابرابری (۱۱) برحسب θ روی $(\theta^{(1)}, \theta^{(2)})$ ، و با توجه به (۱۲) و (۱۳)، نتیجه حاصل می‌شود.

به‌طور مشابه اگر $\theta^{(2)} - \theta^{(1)} > \pi$ ، آن‌گاه طبق تعریف $\gamma = 2\pi - \theta^{(2)} + \theta^{(1)}$ و داریم

$$\left(\int_{-\pi}^{\theta^{(1)}} + \int_{\theta^{(2)}}^{\pi} \right) \cos(\theta_j - \theta) d\theta = \sin(\gamma + \beta_j) - \sin \beta_j. \quad (14)$$

که در آن $\beta_j := \theta_j - \theta^{(1)}$. عبارت سمت راست برای (۱۴) روی $(-\gamma_1, 0) \cup (2\pi - \gamma, 2\pi - \gamma_1]$ ، تابعی مقعر از β_j است، که در آن $\gamma_1 = \theta^{(1)} + \pi$. از این‌رو، کمینه خود را در یکی از نقطه‌های انتهایی آن می‌گیرد. بنابراین

$$\sin(\gamma + \beta_j) - \sin \beta_j \geq \sin \gamma. \quad (15)$$

با انتگرال‌گیری از دوطرف نابرابری (۱۱) برحسب θ روی $(-\pi, \theta^{(1)}) \cup (\theta^{(2)}, \pi]$ ، و با توجه به (۱۴) و (۱۵)، نتیجه حاصل می‌شود.

نتیجه ۴. برای مجموع‌یابی عددی که در یک قطاع قائم قرار دارند، داریم

$$\kappa_{\text{Nat}} \leq \frac{\pi}{2}.$$

نتیجه‌گیری

در این مقاله عدد حالت طبیعی و عملی برای مجموع‌یابی مختلط تعریف شدند. نشان دادیم که مقدار زاویه قطاعی که عددها را دربرمی‌گیرد، در اندازه عدد حالت طبیعی نقش کلیدی دارد: هرچه مقدار زاویه کم‌تر، عدد حالت طبیعی کوچک‌تر. در حالت خاص اگر قطاع قائم باشد یا زاویه‌ای کم‌تر از $\pi/2$ داشته باشد، یک کران بالا برای عدد حالت طبیعی ارائه کردیم. بررسی عدد حالت عملی بسیار ساده‌تر است. در واقع اگر عددها در یکی از ربع‌های صفحه مختلط قرار گیرند، عدد حالت عملی برابر ۱ خواهد بود. در باره رابطه بین عدد حالت طبیعی و عملی نیز نتیجه‌هایی به‌دست آمد. نشان دادیم که بین این دو عدد هیچ رابطه مقایسه‌ای وجود ندارد. همچنین ثابت کردیم اگر عدد حالت عملی کوچک باشد، آن‌گاه عدد حالت طبیعی نیز کوچک است، و اگر عدد حالت طبیعی بزرگ باشد، آن‌گاه عدد حالت عملی نیز بزرگ است.

منابع

1. Bornemann F., "Accuracy and stability of computing high-order derivatives of analytic functions by Cauchy integrals", Found. Comput. Math., 11 (2011) 1-63.
2. Higham N. J., "Accuracy and Stability of Numerical Algorithms", 2nd edn., SIAM (2002).

3. Higham N. J., "The accuracy of floating point summation", *SIAM J. Sci. Comput.*, 14 (1993) 783-799.
4. Jeannerod C-P., Siegfried M. R., "On relative errors of floating-point operations: optimal bounds and applications", *Math. Comput.*, 87 (2018) 803-819.
5. Lange M., Siegfried M. R., "Error estimates for the summation of real numbers with application to floating-point summation", *BIT Numer. Math.*, 57 (2017) 927-941.
6. Ogita T., Rump S. M., Oishi S., "Accurate sum and dot product", *SIAM J. Sci. Comput.*, 26 (2005) 1955-1988.
7. Pan V. Y., Murphy B., Qian G., Rosholt R. E., "A new error-free floating-point summation algorithm", *Comput. Math. Appl.*, 57 (2009) 560-564.
8. Rudin W., "Real and Complex Analysis", 3rd edn., Mc Graw-Hill, New York (1987).
9. Rump S. M., "Ultimately fast accurate summation", *SIAM J. Sci. Comput.*, 31 (2009) 3466-3502.
10. Zhu Y-K., Wayne B. H., "Algorithm 908: Online exact summation of floating-point streams", *ACM Trans. Math. Softw.*, 37 (2010) 37.
11. Zhu Y-K., Wayne B.H., "Correct rounding and a hybrid approach to exact floating-point summation", *SIAM J. Sci. Comput.*, 31 (2009) 2981-3001.