پروژه ۳ درس نرم‌افزار ریاضی

Numerical-Methods

علی تقی زاده:

شماره دانشجویی:

۴۰۳۱۳۰۰۸

---

۲۹ خرداد ۱۴۰۵

سوال ۱: درونیابی لاگرانژ و اسپلاین

در این بخش، هدف یافتن تابعی است که دقیقاً از مجموعه نقاط داده‌شده شامل

$$(۱,۱),\ (۲,۳),\ (۳,۵),\ (۴,۸),\ (۵,۵),\ (۶,۲)$$

عبور کند. برای این منظور از دو روش درونیابی لاگرانژ و اسپلاین مکعبی (Cubic Spline) استفاده شد. همان‌طور که در نتایج خروجی کد پایتون مشخص است، هر دو روش با موفقیت پیاده‌سازی شده و خطای ارزیابی آن‌ها در گره‌های داده‌شده برابر صفر است.

۱. ضابطه توابع به‌دست‌آمده

الف) درونیابی لاگرانژ

از آنجا که ۶ نقطه داده شده است، چندجمله‌ای درونیاب لاگرانژ یک چندجمله‌ای واحد از درجه ۵ خواهد بود. ضابطه دقیق این تابع پس از محاسبه و ساده‌سازی به فرم زیر به دست آمد:

$$P(x) = ۰/۱۷۵x^۵ - ۲/۹۵۸x^۴ + ۱۸/۳۷۵x^۳ - ۵۲/۰۴۱x^۲ + ۶۸/۴۵x - ۳۱/۰ \tag{۱}$$

ب) درونیابی اسپلاین مکعبی

برخلاف روش لاگرانژ، اسپلاین مکعبی از چندجمله‌ای‌های تکه‌ای از درجه ۳ تشکیل شده است که در مرز بازه‌ها، یعنی گره‌ها، دارای مشتقات اول و دوم پیوسته هستند. ضابطه دقیق این توابع برای هر بازه محاسبه شد؛ به عنوان نمونه، ضابطه این تابع در بازه اول $[۱,۲]$ به صورت زیر است:

$$S_۰(x) = -۰/۱۸۶x^۳ + ۰/۵۵۹x^۲ + ۱/۶۲۶x - ۱/۰ \tag{۲}$$

(توابع مربوط به سایر بازه‌ها نیز به طور کامل در خروجی کد درج شده‌اند.)

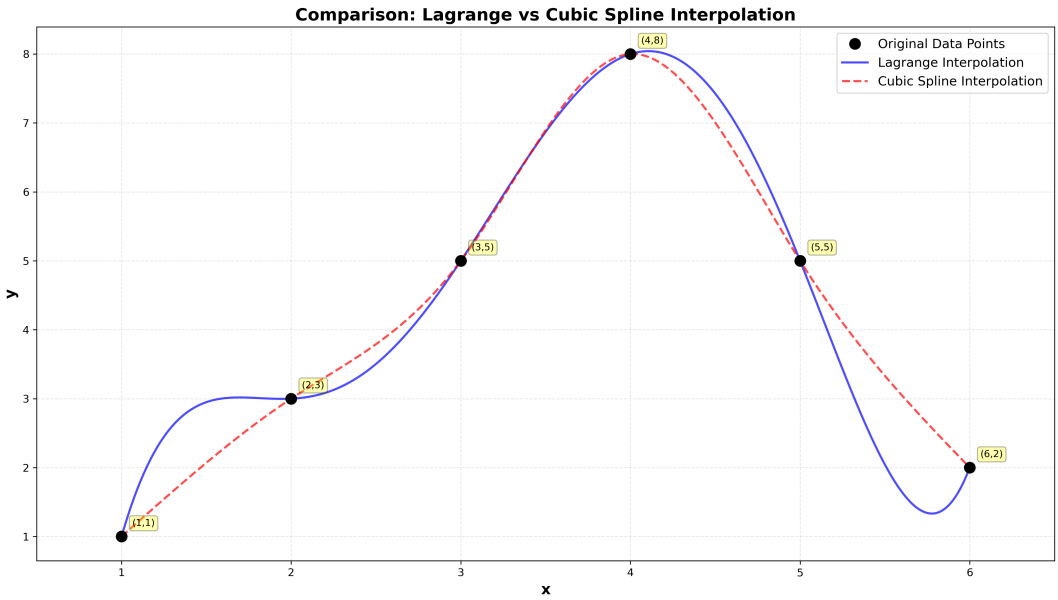
۲. تحلیل نمودار و مقایسه دو روش

با توجه به نمودار رسم‌شده در شکل ۱، تفاوت‌های بنیادین رفتار این دو روش به وضوح قابل مشاهده است:

- **رفتار نوسانی لاگرانژ:** چندجمله‌ای لاگرانژ، یعنی منحنی پیوسته آبی، به دلیل داشتن درجه بالا، یعنی درجه ۵، تمایل به نوسانات شدیدتری بین نقاط دارد. این موضوع به‌خصوص در بازه انتهایی بین $x = ۵$ و $x = ۶$

مشهود است، جایی که منحنی دچار یک افت شدید به سمت پایین شده است. این رفتار نوسانی در لبه‌ها، نمودی از پدیده رونگه (Runge's Phenomenon) در درونیابی چندجمله‌ای است.

- **پایداری اسپلاین:** در مقابل، اسپلاین مکعبی، یعنی خط‌چین قرمز، به دلیل ماهیت تکه‌ای و محدود بودن درجه توابع به عدد ۳، رفتار بسیار نرم‌تر، طبیعی‌تر و پایدارتری از خود نشان می‌دهد. این روش بدون ایجاد نوسانات کاذب بین نقاط، مسیر هموارتری را برای اتصال گره‌ها طی کرده است و برای برازش این نوع داده‌ها انتخاب بهتری محسوب می‌شود.



شکل ۱: مقایسه رفتار توابع درونیابی لاگرانژ و اسپلاین مکعبی بر روی نقاط داده‌شده

سوال ۲: مصورسازی و تحلیل داده‌های زمان اجرای الگوریتم‌ها

در این بخش، زمان اجرای سه الگوریتم مختلف (Alg.1, Alg.2, Alg.3) بر روی مجموعه‌داده‌هایی با اندازه‌های ۱۰۰KB الی ۷۰۰KB مورد ارزیابی قرار گرفته است. داده‌ها به صورت پویا (Dynamic) از فایل Excel خوانده شده‌اند؛ به طوری که با اضافه شدن داده جدید، یعنی سطر مربوط به ۷۰۰KB، کد بدون نیاز به تغییر ساختاری، تمام محاسبات و نمودارها را با موفقیت به‌روزرسانی کرده است.

الف و ب) رسم نمودارها و تحلیل رفتار الگوریتم‌ها

بر اساس داده‌های استخراج‌شده شامل ۷ رکورد کامل، تحلیل‌های آماری و پیچیدگی زمانی الگوریتم‌ها به شرح زیر است:

- **الگوریتم Alg.1:** این الگوریتم با اختلاف، بهینه‌ترین عملکرد را دارد. زمان اجرای آن برای داده ۱۰۰KB برابر ۵۰ms است و در بزرگ‌ترین سایز داده، یعنی ۷۰۰KB، تنها به ۸۰ms می‌رسد. رشد زمان اجرای این الگوریتم به ازای هر ۱۰۰KB تنها ۵ms است که نشان از مقیاس‌پذیری فوق‌العاده بالا و شیب رشد بسیار کند آن دارد. میانگین کل زمان اجرای این الگوریتم ۶۵ms با انحراف معیار بسیار پایین ۱۰ms است.
- **الگوریتم Alg.2:** این الگوریتم دارای سربار اولیه (Initial Overhead) نسبتاً بالایی است، به طوری که کار خود را از ۲۰۰ms برای داده ۱۰۰ کیلوبایتی آغاز می‌کند. با این حال، شیب رشد آن ملایم است و به ازای هر گام، ۲۰ms افزایش می‌یابد. میانگین اجرای این الگوریتم برای کل ۷ رکورد برابر ۲۶۰ms است.
- **الگوریتم Alg.3:** این الگوریتم در سایز کوچک ۱۰۰KB با زمان ۱۰۰ms عملکردی بهتر از Alg.2 دارد، اما نرخ رشد آن بسیار شیب‌دار است و به ازای هر گام، ۱۰۰ms افزایش می‌یابد. این شیب تند باعث می‌شود تا در داده‌های بزرگ، مانند ۷۰۰KB، زمان اجرای آن به ۷۰۰ms برسد که بدترین عملکرد در میان سه الگوریتم است. میانگین کل این الگوریتم ۴۰۰ms با انحراف معیار بالای ۲۰۰ms ثبت شده است.

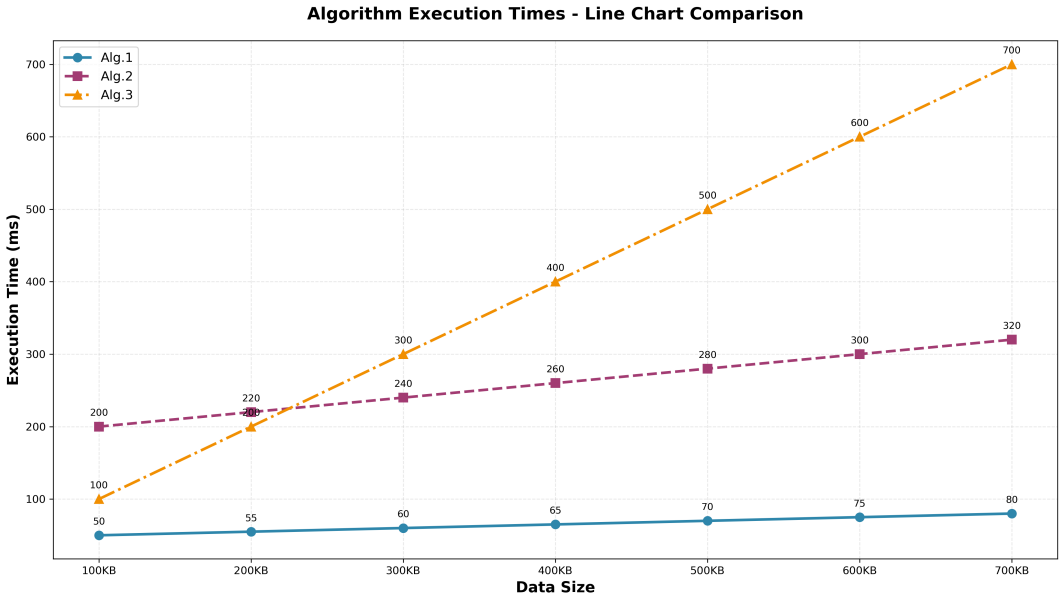
نکته آماری مهم این است که به دلیل رشد کاملاً خطی و یکنواخت زمان‌ها در هر سه الگوریتم، مقادیر میانگین (Mean) و میانه (Median) برای تمامی آن‌ها دقیقاً بر یکدیگر منطبق است.

بررسی نمودارهای ترسیم‌شده

برای تحلیل دقیق‌تر، سه نمودار مختلف به صورت برنامه‌نویسی‌شده استخراج گردیده است.

۱. نمودار خطی (Line Chart)

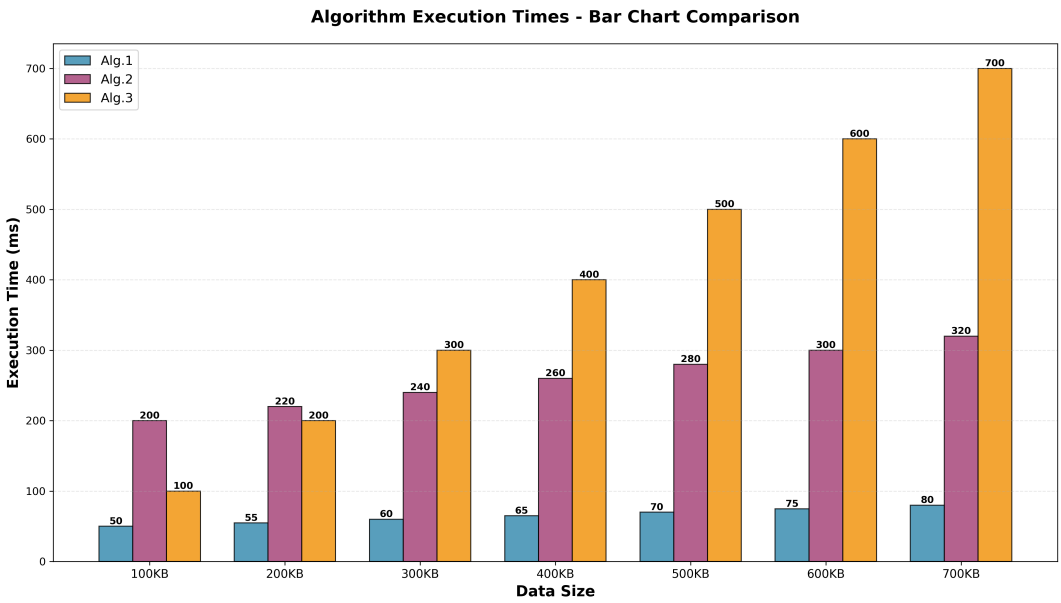
در شکل ۲ روند رشد خطی هر سه الگوریتم قابل رؤیت است. مهم‌ترین نقطه این نمودار، تقاطع خطوط Alg.2 و Alg.3 در سایز ۲۰۰KB است. این نقطه نشان می‌دهد که برای داده‌های کوچک‌تر از ۲۰۰ کیلوبایت، الگوریتم ۳ انتخاب بهتری است؛ اما برای داده‌های بزرگ‌تر از آن، الگوریتم ۲، با وجود سربار اولیه بیشتر، به دلیل شیب ملایم‌تر، کاملاً برتری می‌یابد.



شکل ۲: نمودار خطی مقایسه زمان اجرای الگوریتم‌ها بر حسب اندازه داده، با احتساب داده ۷۰۰KB

۲. نمودار میله‌ای (Bar Chart)

شکل ۳ مقایسه نقطه‌به‌نقطه الگوریتم‌ها را ارائه می‌دهد. در این نمودار به وضوح دیده می‌شود که ستون‌های مربوط به Alg.3 با چه سرعتی نسبت به دو الگوریتم دیگر در حال افزایش هستند و در انتهای نمودار، اختلاف فاحشی با سایرین ایجاد کرده‌اند.

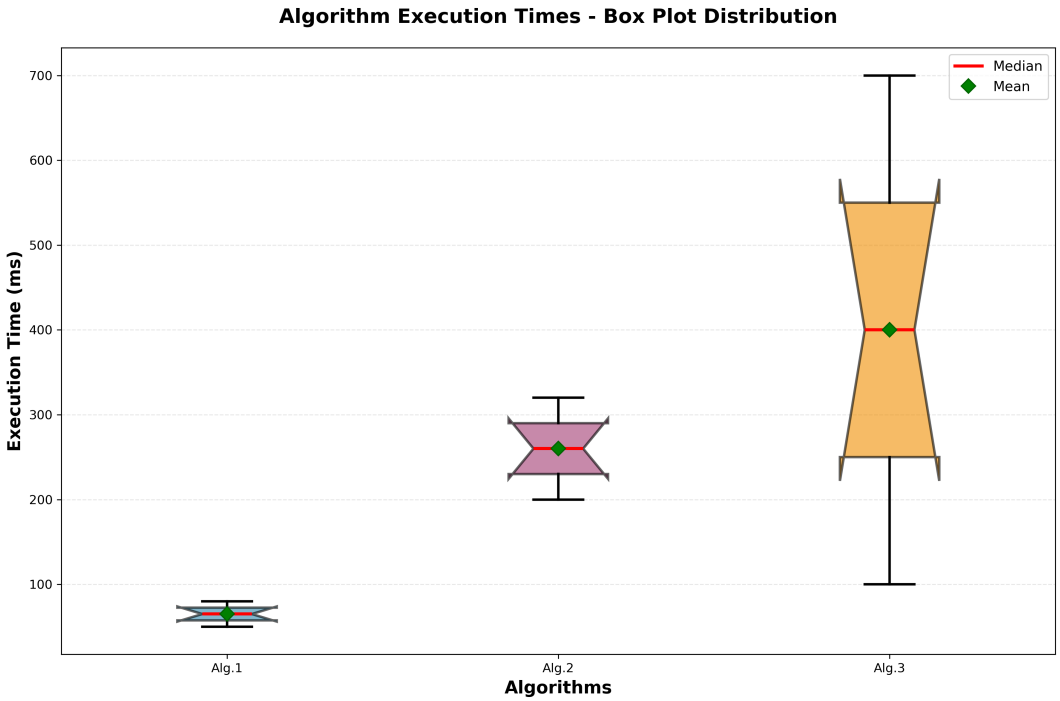


شکل ۳: نمودار میله‌ای زمان اجرای الگوریتم‌ها به تفکیک سایز داده‌ها

۳. نمودار جعبه‌ای (Box Plot)

شکل ۴ برای تحلیل پراکندگی، یعنی واریانس داده‌ها، رسم شده است. فشردگی جعبه مربوط به Alg.1 نشان‌دهنده پایداری زمانی این الگوریتم فارغ از اندازه داده است. در مقابل، کشیدگی شدید جعبه Alg.3 نشانگر حساسیت بالا و ناپایداری

شدید این الگوریتم در برابر افزایش حجم داده ورودی است. منطبق بودن نشانگر میانگین، یعنی لوزی سبز، بر خط میانه، یعنی خط قرمز، در هر سه جعبه، تأییدکننده توزیع متقارن و رشد خطی داده‌هاست.



شکل ۴: نمودار جعبه‌ای توزیع پراکندگی زمان اجرای الگوریتم‌ها

ج) محاسبه میانگین زمان اجرای الگوریتم ۲ در بازه ۱۰۰ تا ۶۰۰ کیلوبایت

طبق خواسته دقیق سوال، در این بخش صرفاً داده‌های مربوط به الگوریتم ۲ برای اندازه‌های ۱۰۰KB تا ۶۰۰KB، بدون احتساب سطر جدید ۷۰۰ کیلوبایت، در نظر گرفته شده است.

مقادیر استخراج‌شده از فایل برابرند با:

$$۲۰۰, \quad ۲۲۰, \quad ۲۴۰, \quad ۲۶۰, \quad ۲۸۰, \quad ۳۰۰$$

محاسبه میانگین (Average) به صورت زیر است:

$$\text{Average} = \frac{۲۰۰ + ۲۲۰ + ۲۴۰ + ۲۶۰ + ۲۸۰ + ۳۰۰}{۶} = \frac{۱۵۰۰}{۶} = ۲۵۰/۰\,\text{ms}$$

این مقدار توسط اسکریپت پایتون نیز به صورت دقیق محاسبه شده و با نتیجه تحلیلی فوق کاملاً مطابقت دارد.

سوال ۳: محاسبه عددی انتگرال و مقایسه روش‌ها

در این بخش، هدف محاسبه مقدار انتگرال معین تابع $f(x) = e^{x^۲}$ در بازه $[۰, ۱]$ است. این انتگرال دارای پاسخ تحلیلی با توابع مقدماتی نیست و باید از طریق روش‌های عددی تقریب زده شود. مقدار مرجع و بسیار دقیق این انتگرال برابر با ۱٫۴۶۲۶۵۱۷۴۵۹۰۷ محاسبه شده است که از آن برای ارزیابی خطای مطلق روش‌های مختلف استفاده می‌کنیم.

الف) بررسی روش‌های نیوتن-کوتس، ذوزنقه‌ای و سیمپسون

روش‌های ذوزنقه‌ای (Trapezoidal) و سیمپسون (Simpson) از دسته فرمول‌های نیوتن-کوتس هستند که بازه انتگرال‌گیری را به n زیربازه مساوی با گام h تقسیم می‌کنند.

- **روش ذوزنقه‌ای مرکب:** این روش تابع را با خطوط راست، یعنی چندجمله‌ای درجه ۱، تقریب می‌زند. از نظر تئوری، خطای برشی این روش از مرتبه $O(h^۲)$ است. داده‌های خروجی این موضوع را کاملاً اثبات می‌کنند؛ به عنوان مثال، وقتی تعداد بازه‌ها (n) از ۱۶ به ۳۲، یعنی ۲ برابر، افزایش می‌یابد، خطای مطلق از $۱٫۷۷ \times ۱۰^{-۳}$ به $۴٫۴۲ \times ۱۰^{-۴}$ کاهش می‌یابد. نسبت این دو خطا تقریباً برابر ۴ است ($۲^۲ = ۴$) که تاییدکننده مرتبه خطای تقریب است. این روش در $n = ۱۲۸$ به خطای $۲٫۷۷ \times ۱۰^{-۵}$ رسیده است که برای کارهای دقیق، نیازمند محاسبات سنگین‌تری است.
- **روش سیمپسون مرکب:** این روش از سهمی‌ها، یعنی چندجمله‌ای درجه ۲، برای تقریب استفاده می‌کند و خطای تئوریک آن از مرتبه $O(h^۴)$ است. با بررسی داده‌ها می‌بینیم با دو برابر شدن n از ۱۶ به ۳۲، خطا از $۴٫۵۸ \times ۱۰^{-۶}$ به $۲٫۸۸ \times ۱۰^{-۷}$ می‌رسد. نسبت این کاهش خطا تقریباً برابر ۱۶ است ($۲^۴ = ۱۶$). این روش به دلیل مرتبه همگرایی بالاتر، با همان تعداد گره $n = ۱۲۸$، به خطای بسیار ناچیز $۱٫۱۳ \times ۱۰^{-۹}$ دست یافته است که عملکردی بسیار بهتر از روش ذوزنقه‌ای دارد.

ب) بررسی انتگرال‌گیری گوسی، گاوس-لژاندر

برخلاف روش‌های نیوتن-کوتس که از نقاط با فواصل یکسان استفاده می‌کنند، روش گاوس-لژاندر (Gauss-Legendre) هم نقاط، یعنی گره‌ها، و هم وزن‌ها را به گونه‌ای بهینه می‌کند که بالاترین دقت ممکن به دست آید. یک فرمول n نقطه‌ای گوسی می‌تواند چندجمله‌ای‌هایی تا درجه $۲n - ۱$ را به طور دقیق انتگرال‌گیری کند.

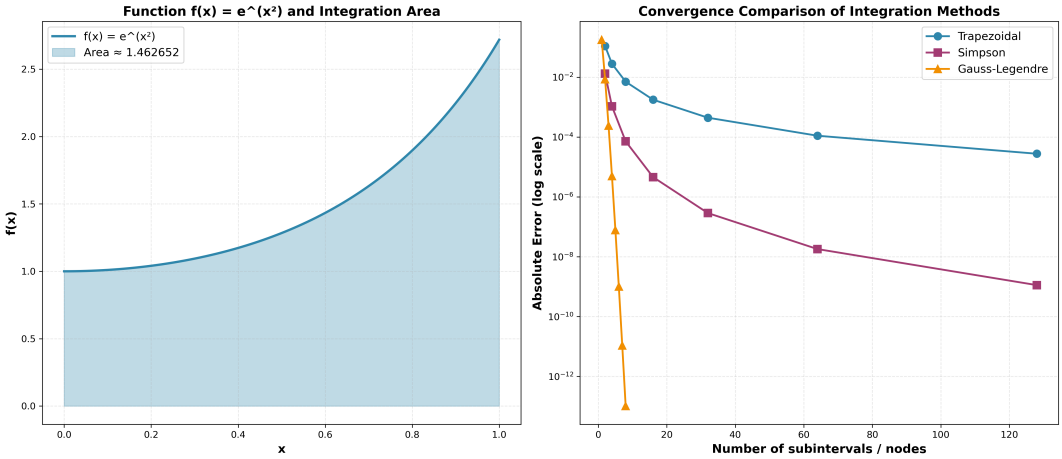
نتایج اجرای این روش شگفت‌انگیز است. از آنجا که تابع $f(x) = e^{x^۲}$ تابعی بی‌نهایت بار مشتق‌پذیر و هموار (Smooth) است، روش گوسی اصطلاحاً دارای «همگرایی طیفی» (Spectral Convergence) برای آن است.

همان‌طور که در خروجی‌ها پیداست، این روش تنها با استفاده از **۸ نقطه ارزیابی** به خطای $۱٫۰۲ \times ۱۰^{-۱۳}$ رسیده است که در حد دقت ماشین (Machine Precision) است. این در حالی است که روش‌های قبلی حتی با ۱۲۸ نقطه ارزیابی نتوانستند به این دقت نزدیک شوند.

پ) تحلیل نمودارها و نتیجه‌گیری نهایی

برای درک بهتر رفتار الگوریتم‌ها، دو نمودار رسم شده است. این نمودارها در شکل ۵ نمایش داده شده‌اند.

- **نمودار سمت چپ:** مساحت زیر نمودار تابع $f(x) = e^{x^2}$ را نشان می‌دهد که تحدب رو به بالای آن مشهود است.
- **نمودار سمت راست، نمودار همگرایی:** این نمودار که در مقیاس لگاریتمی رسم شده است، حیاتی‌ترین بخش تحلیل است. محور افقی تعداد نقاط یا زیربازه‌ها و محور عمودی خطای مطلق است. شیب هر خط نشان‌دهنده سرعت همگرایی آن روش است. خط آبی، یعنی ذوزنقه، شیب ملایمی دارد. خط بنفش، یعنی سیمپسون، شیب تندتری دارد. اما خط نارنجی، یعنی گاوس–لژاندر، سقوطی تقریباً عمودی را تجربه می‌کند که نشان می‌دهد با افزودن تنها چند نقطه، خطا به صورت نمایی (Exponential) کاهش می‌یابد.



شکل ۵: چپ: مساحت زیر منحنی تابع. راست: نمودار لگاریتمی مقایسه سرعت همگرایی روش‌های انتگرال‌گیری

نتیجه‌گیری: برای توابع هموار نظیر آنچه در این مسئله بررسی شد، روش گاوس–لژاندر با اختلاف بسیار زیاد، کارآمدترین (Most Efficient) روش است؛ زیرا با کمترین تعداد فراخوانی تابع، یعنی کمترین هزینه محاسباتی، بیشترین دقت را به ارمغان می‌آورد. با این حال، اگر داده‌ها به صورت جدولی با فواصل یکسان داده شده باشند، نه به صورت تابع تحلیلی، استفاده از روش سیمپسون بهترین گزینه در دسترس خواهد بود.