Numerička matematika
2. predavanje

Saša Singer

singer@math.hr

web.math.hr/~singer

PMF – Matematički odjel, Zagreb

Sadržaj predavanja

- Uvodna priča o greškama:
  - Vrste grešaka (ponavljanje).
  - Prikaz brojeva u računalu i greške zaokruživanja.
  - Greške zaokruživanja osnovnih aritmetičkih operacija.
    - Opasno ili “katastrofalno” kraćenje.
  - “Širenje” grešaka (zaokruživanja), stabilni i nestabilni algoritmi.
  - Mjerenje grešaka — razne norme.
  - Uvjetovanost problema.
  - Stabilnost algoritma.
    - Primjeri stabilnih i nestabilnih algoritama.
  - Primjeri grešaka zaokruživanja.

Informacije

Moja web stranica za Numeričku matematiku je

http://web.math.hr/~singer/num_mat/

Skraćena verzija skripte — 1. dio (prvih 7 tjedana):

http://web.math.hr/~singer/num_mat/num_mat1.pdf

Još će se skratiti za par dana!

Skraćena verzija skripte — 2. dio (drugih 7 tjedana):

http://web.math.hr/~singer/num_mat/num_mat2.pdf

Greške (nastavak)

Izvori i vrste grešaka (ponavljanje)

Pri numeričkom rješavanju nekog problema javljaju se četiri vrste grešaka:

- greške modela,
- greške u ulaznim podacima (mjerenjima),
- greške metoda za rješavanje modela,
- greške aritmetike računala.

Za početak, pogledajmo detaljnije zadnju vrstu grešaka koja nastaje zbog "približnog" računanja. To su greške zaokruživanja u

- prikazu brojeva u računalu i
- aritmetici računala.

Prikaz brojeva u računalu

Tipovi brojeva u računalu

U računalu postoje dva bitno različita tipa brojeva:

- cijeli brojevi
- realni brojevi.

Oba skupa su konačni podskupovi odgovarajućih skupova $\mathbb{Z}$ i $\mathbb{R}$ u matematici.

Kao baza za prikaz oba tipa koristi se baza 2.

Cijeli brojevi

Cijeli se brojevi prikazaju korištenjem n bitova — binarnih znamenki, od kojih jedna služi za predznak, a ostalih $n-1$ za znamenke broja.

Matematički gledano,

- aritmetika cijelih brojeva u računalu je modularna aritmetika u prstenu ostataka modulo 2^n, samo je sustav ostataka simetričan oko 0, tj.

$$-2^{n-1}, \ldots, -1, 0, 1, \ldots, 2^{n-1} - 1.$$

- Računalo ne zna izravno operirati s brojevima izvan tog raspona.

Realni brojevi

Realni brojevi r prikazuju se korištenjem mantise m (ili češće, signifikanda) i eksponenta e u obliku

$$r = \pm m \cdot 2^e,$$

pri čemu je e cijeli broj u određenom rasponu, a m racionalni broj za koji vrijedi $1 \leq m < 2$ (tj. mantisa započinje s $1.\ldots$).

- Vodeća jedinica se često ne pamti, pa je mantisa "dulja" za 1 bit tzv. "skriveni bit" (engl. hidden bit).
- Eksponent se prikazuje kao s-bitni cijeli broj, a za mantisu pamti se prvih t znamenki iza binarne točke.
- Po standardu, eksponentu se dodaje "pomak" (engl. bias), da bi eksponent bio nenegativan. Ovo je nebitno za ponašanje aritmetike.

Realni brojevi

Skup svih realnih brojeva prikazivih u računalu je omeđen, a parametriziramo ga duljinom mantise i eksponenta i označavamo s $\mathbb{R}(t, s)$.

mantisa				
$\pm$	m_{-1}	m_{-2}	$\cdots$	m_{-t}

eksponent				
e_{s-1}	e_{s-2}	$\cdots$	e_1	e_0

Ne može se svaki realni broj egzaktno spremiti u računalo.

Ako je broj $x \in \mathbb{R}$ unutar prikazivog raspona i

$$x = \pm \left(\sum_{k=1}^{\infty} b_{-k} 2^{-k} \right) 2^e$$

i mantisa broja ima više od t znamenki, ...

Realni brojevi

... bit će spremljena aproksimacija tog broja $f\ell(x) \in \mathbb{R}(t, s)$ koja se može prikazati kao

$$f\ell(x) = \pm \left(\sum_{k=1}^{t} b^*_{-k} 2^{-k} \right) 2^{e^*}.$$

Slično kao kod decimalne aritmetike

- ako je prva odbačena znamenka 1, broj zaokružujemo nagore,
- a ako je 0, nadolje.

Time smo napravili apsolutnu grešku manju ili jednaku od "pola zadnjeg prikazivog bita", tj. 2^{-t-1+e}.

Relativna greška zaokruživanja

Gledajući relativno, greška je manja ili jednaka

$$\left|\frac{x - f\ell(x)}{x}\right| \leq \frac{2^{-t-1+e}}{2^{-1} \cdot 2^{e}} = 2^{-t},$$

tj. imamo vrlo malu relativnu grešku.

Veličinu 2^{-t} zovemo jedinična greška zaokruživanja (engl. unit roundoff) i uobičajeno označavamo s u.

Za $x \in \mathbb{R}$ unutar prikazivog raspona, umjesto x sprema se zaokruženi broj $f\ell(x) \in \mathbb{R}(t, s)$ i vrijedi

$$f\ell(x) = (1 + \varepsilon)x, \quad |\varepsilon| \leq u,$$

gdje je ε relativna greška napravljena tim zaokruživanjem.

IEEE standard za prikaz brojeva

Prikaz realnih brojeva u računalu zove se prikaz s pomičnim zarezom/točkom (engl. floating point representation), a aritmetika je aritmetika pomičnog zareza/točke (engl. floating point arithmetic).

Veličine s i t prema novom IEEE standardu:

`format`	`32-bitni`	`64-bitni`	`128-bitni`
duljina mantise	23 bita	52 bita	112 bita
duljina eksponenta	8 bitova	11 bitova	15 bitova
jedinična gr. zaokr.	2^{-24}	2^{-53}	2^{-113}
$u \approx$	$5.96 \cdot 10^{-8}$	$1.11 \cdot 10^{-16}$	$9.63 \cdot 10^{-35}$
raspon brojeva $\approx$	$10^{\pm 38}$	$10^{\pm 308}$	$10^{\pm 4932}$

IEEE standard za prikaz brojeva (nastavak)

Većina PC računala (procesora) još ne podržava 128-bitni prikaz i aritmetiku.

Umjesto toga, FPU (Floating–point unit) stvarno koristi

- tzv. tip `extended` iz starog IEEE standarda.

Dio primjera koje ćete vidjeti napravljen je baš u tom tipu!

`format`	`80-bitni`
duljina mantise	64 bita
duljina eksponenta	15 bitova
jedinična gr. zaokr.	2^{-64}
$u \approx$	$5.42 \cdot 10^{-20}$
raspon brojeva $\approx$	$10^{\pm 4932}$

IEEE standard za aritmetiku računala

IEEE standard propisuje i svojstva aritmetike.
Pretpostavka standarda — za osnovne aritmetičke operacije ($\circ$ označava $+$, $-$, $*$, $/$) nad $x, y \in \mathbb{R}(t, s)$ vrijedi

$$fl(x \circ y) = (1 + \varepsilon)\,(x \circ y), \quad |\varepsilon| \leq u,$$

za sve $x, y \in \mathbb{R}(t, s)$ za koje je $x \circ y$ u dozvoljenom rasponu.
Dobiveni rezultat je tada prikaziv, tj. vrijedi $fl(x \circ y) \in \mathbb{R}(t, s)$.
Postoje rezervirani eksponenti koji označavaju "posebno stanje":

- overflow,
- underflow,
- dijeljenje s 0,
- nedozvoljenu operaciju kao što su $0/0$, $\sqrt{-1}$.

Greške i uvjetovanost

Greške — ponavljanje

Pri numeričkom rješavanju nekog problema javljaju se različiti tipovi grešaka:

- greške modela — svođenje realnog problema na neki "matematički" problem,
- greške u ulaznim podacima (mjerenja i sl.),
- greške numeričkih metoda za rješavanje "matematičkog" problema,
- greške "približnog" računanja — obično su to
  - greške zaokruživanja u aritmetici računala.

Greške modela su "izvan" dosega numeričke matematike.

- Spadaju u fiziku, kemiju, biologiju, tehniku, ekonomiju, ...

Greške (nastavak)

Sljedeće tri kategorije (podaci, metoda, računanje) su vezane za "matematički" problem, i

- spadaju u domenu numeričke matematike!

O njima nešto "moramo reći".

Skica numeričkog rješavanja nekog problema sliči algoritmu:

Posebno, ako dozvolimo da, umjesto riječi "algoritam",

- piše i riječ "metoda".

Zamislite da pojam "algoritam" uključuje

- metodu i stvarno računanje rezultata!

Greške (nastavak)

Sve tri vrste grešaka — podaci, metoda, računanje,

- rezultiraju nekom greškom u konačnom rezultatu!

Ta greška nas "zanima".

Uočite da greške u ulaznim podacima možemo gledati

- neovisno o metodi za rješenje problema,
- i tako dolazimo do pojma uvjetovanosti problema.

Za razliku od toga, greške metode i računanja, naravno,

- ovise o metodi, odnosno, algoritmu za rješenje problema.

Uvjetovanost problema

Neformalno rečeno, uvjetovanost problema mjeri

- osjetljivost problema na greške u podacima.

Osnovno svojstvo uvjetovanosti:

- Ne ovisi o konkretnoj numeričkoj metodi za rješenje problema, već samo o problemu.

Svrha uvjetovanosti = daje odgovor na pitanje:

- Koju točnost rezultata možemo očekivati
- pri točnom računanju
- s (malo) pomaknutim — netočnim podacima?

Model problema

Matematički model problema, zovimo ga P:

- za zadani ulaz — podatak $x \in \mathcal{X}$,
- dobivamo izlaz — rezultat $y \in \mathcal{Y}$.

Slikica modela je

$$x \longrightarrow \boxed{P} \longrightarrow y$$

Problem P interpretiramo kao računanje vrijednosti funkcije

$$f : \mathcal{X} \to \mathcal{Y},$$

gdje su $\mathcal{X}$ i $\mathcal{Y}$ odgovarajući matematički objekti. Na primjer, vektorski prostori, a vrlo često su i normirani prostori (treba nam mjera za grešku).

Uvjetovanost problema (nastavak)

Ideja uvjetovanosti:

$$\text{greška u rezultatu} \approx \text{uvjetovanost} \cdot \text{greška u podacima}$$

Ovisi o obje vrijednosti: točnoj x i približnoj $\hat{x}$.

Napomene:

- Obično nas uvjetovanost posebno zanima za male perturbacije (greške, smetnje) podataka.
- Ako je f dovoljno glatka funkcija, možemo koristiti Taylorov razvoj u okolini točnog ulaznog podatka x
- i dobiti procjenu uvjetovanosti preko derivacija!

Primjeri problema (nastavak)

Primjer 1. Računanje sume dva realna broja $x_1, x_2 \in \mathbb{R}$. Tada je

$$f(x_1, x_2) = x_1 + x_2,$$

s tim da je $\mathcal{X} = \mathbb{R}^2$ i $\mathcal{Y} = \mathbb{R}$.

Primjer 2. Računanje produkta dva realna broja $x_1, x_2 \in \mathbb{R}$. Tada je

$$f(x_1, x_2) = x_1 x_2,$$

s tim da je opet $\mathcal{X} = \mathbb{R}^2$ i $\mathcal{Y} = \mathbb{R}$.

Primjeri problema (nastavak)

Primjer 3. Računanje sjecišta pravaca

$$P_1 = \{(y_1, y_2) \in \mathbb{R}^2 \mid a_{11}y_1 + a_{12}y_2 = x_1\},$$
$$P_2 = \{(y_1, y_2) \in \mathbb{R}^2 \mid a_{21}y_1 + a_{22}y_2 = x_1\}.$$

Smatramo da su koeficijenti a_{ij} i x_i, za $i, j = 1, 2$, ulazni podaci.

Ovaj problem pišemo u matričnom zapisu kao linearni sustav od dvije jednadžbe oblika $Ay = x$, gdje je

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \in \mathbb{R}^{2\times 2}, \qquad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2.$$

Primjeri problema (nastavak)

Traženo sjecište je rješenje linearnog sustava $Ay = x$.

Ako pretpostavimo da je matrica A sustava regularna, tj. $\det A \neq 0$, onda je $y = A^{-1}x$. Dakle,

$$f(x) = A^{-1}x,$$

s tim da je $\mathcal{X} = \mathcal{Y} = \mathbb{R}^2$.

Norme i uvjetovanost

Greška na ulazu – što na izlazu?

Zadatak numeričke analize je odrediti vezu između greške na ulazu i greške na izlazu.

Polazni podaci		Izlazni podaci
x_{inp}	Algoritam f $\longrightarrow$	$y_{\mathrm{out}} = f(x_{\mathrm{inp}})$
imaju grešku		imaju grešku
$\Delta x := x_{\mathrm{inp}} - x$		$\Delta y := f(x_{\mathrm{inp}}) - f(x)$

Uzimamo da su $\mathcal{X}$ i $\mathcal{Y}$ (barem) vektorski prostori.

Kako mjeriti grešku?

Kad x_{inp} i $f(x_{\mathrm{inp}})$ nisu brojevi, nego vektori ili matrice, grešku možemo mjeriti:

- po svakoj od komponenata, međutim to je malo previše brojeva,
- kao neku "ukupnu ili najveću" grešku — samo jedan broj i to korištenjem vektorskih/matričnih normi.

Prisjetite se: vektorski prostor na kojem je definirana norma zove se normirani prostor.

Vektorske norme

“Vektorska” norma na vektorskom prostoru V (nad poljem F, gdje je $F = \mathbb{R}$ ili $F = \mathbb{C}$) je

- je svaka funkcija $\| \cdot \| : V \to \mathbb{R}$

koja zadovoljava sljedeća svojstva:

1. $\|x\| \geq 0$, $\forall x \in V$, a jednakost vrijedi ako i samo ako je $x = 0$,
2. $\|\alpha x\| = |\alpha| \, \|x\|, \quad \forall \alpha \in F, \forall x \in V$,
3. $\|x + y\| \leq \|x\| + \|y\|, \quad \forall x, y \in V$
   (nejednakost poznata pod imenom nejednakost trokuta).

Najpoznatije vektorske norme

Kad je $V = \mathbb{R}^n$ ili $V = \mathbb{C}^n$ (kon. dim.), najčešće se koriste sljedeće tri norme:

1. 1-norma ili ℓ_1 norma $\|x\|_1 = \sum_{i=1}^{n} |x_i|$,
2. 2-norma ili ℓ_2 norma ili euklidska norma
$$\|x\|_2 = (x^* x)^{1/2} = \sqrt{\sum_{i=1}^{n} |x_i|^2},$$
3. ∞-norma ili ℓ_∞ norma $\|x\|_\infty = \max_{i=1,\ldots,n} |x_i|$.

Samo je 2-norma izvedena iz skalarnog produkta.

Norme na prostoru funkcija

Definicija vektorskih normi u sebi ne sadrži zahtjev da je vektorski prostor V konačno dimenzionalan.

Na primjer, norme definirane na vektorskom prostoru neprekidnih funkcija na $[a, b]$ (u oznaci $C[a, b]$) definiraju se slično normama na $\mathbb{R}^n$:

1. L_1 norma $\|f\|_1 = \int\limits_a^b |f(t)|\, dt$,

2. L_2 norma $\|f\|_2 = \left(\int\limits_a^b |f(t)|^2\, dt \right)^{1/2}$,

3. L_∞ norma $\|f\|_\infty = \max\{|f(x)| \mid x \in [a, b]\}$.

Ekvivalentnost normi

Može se pokazati da vrijedi sljedeći teorem.

Teorem. Na svakom konačno-dimenzionalnom vektorskom prostoru V sve su norme ekvivalentne, tj. za svake dvije norme $\| \ \|_a$ i $\| \ \|_b$ postoje konstante c i C takve da je

$$c\|v\|_a \leq \|v\|_b \leq C\|v\|_a, \quad \text{za sve } v \in V.$$

Na primjer,

$$\|x\|_\infty \leq \|x\|_2 \leq \sqrt{n}\|x\|_\infty$$

za sve $x \in \mathbb{R}^n$.

Razlika između teorije i prakse — kad je n ogroman.

Matrične norme

Zamijenimo li u definiciji vektorske norme formalno vektor matricom, dobivamo matričnu normu.

Matrična norma je svaka funkcija $\|\cdot\| : \mathbb{C}^{m\times n} \to \mathbb{R}$ koja zadovoljava sljedeća svojstva:

1. $\|A\| \geq 0$, $\forall A \in \mathbb{C}^{m\times n}$, a jednakost vrijedi ako i samo ako je $A = 0$,
2. $\|\alpha A\| = |\alpha|\,\|A\|$, $\forall \alpha \in \mathbb{R}$, $\forall A \in \mathbb{C}^{m\times n}$,
3. $\|A + B\| \leq \|A\| + \|B\|$, $\forall A, B \in \mathbb{C}^{m\times n}$.

Tome se često dodaje zahtjev konzistentnosti

4. $\|AB\| \leq \|A\|\,\|B\|$

kad god je matrični produkt AB definiran.

Matrične norme (nastavak)

Matrične norme nastaju na dva načina:

- Matricu A promatramo kao vektor s $m \times n$ elemenata i za taj vektor koristimo odgovarajuću vektorsku normu.

  Najpoznatija takva norma odgovara vektorskoj 2-normi i zove se euklidska, Frobeniusova, Hilbert–Schmidtova, ili Schurova norma

$$\|A\|_F = (\mathrm{tr}(A^* A))^{1/2} = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2}.$$

- operatorske norme:

$$\|A\| = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|} \quad (\text{ ili } \ = \max_{\|x\|=1} \|Ax\|).$$

Matrične norme (nastavak)

Uvrštavanjem odgovarajućih vektorskih normi, dobivamo

1. matrična 1-norma, “maksimalna stupčana norma”
$$\|A\|_1 = \max_{j=1,\ldots,n} \sum_{i=1}^{m} |a_{ij}|,$$
2. matrična 2-norma, spektralna norma
$$\|A\|_2 = (\rho(A^*A))^{1/2} = \sigma_{\max}(A),$$
3. matrična ∞-norma, “maksimalna retčana norma”
$$\|A\|_\infty = \max_{i=1,\ldots,m} \sum_{j=1}^{n} |a_{ij}|.$$

ρ je spektralni radijus matrice (po aps. vrijednosti maksimalna svojstvena vrijednost), a σ singularna vrijednost matrice.

Matrične norme (nastavak)

Svojstva:

- I matrične norme nisu međusobno neovisne (slično kao vektorske norme) — ekvivalentnost.
- Matrična 2-norma se teško računa pa se uobičajeno procjenjuje korištenjem ostalih normi.
- Za svaku operatorsku normu vrijedi

$$\|Ay\| \leq \|A\| \, \|y\|,$$

za svaki vektor y, što se često koristi kod ocjena. Formula direktno izlazi iz definicije operatorske norme.

Mjerenje grešaka i uvjetovanost

Relativna/apsolutna uvjetovanost problema mjeri koliko je problem osjetljiv na odgovarajuće promjene polaznih podataka.

- Apsolutna greška: $\|\Delta x\|$, $\|\Delta y\|$, (svaka norma u svom prostoru), gdje je

$$\Delta x = x - \hat{x}, \qquad \Delta y = y - \hat{y}.$$

- Apsolutna uvjetovanost:

$$\kappa_{\text{abs}}(x) := \frac{\|\Delta y\|}{\|\Delta x\|}.$$

Veza s derivacijom!

Mjerenje grešaka i uvjetovanost (nastavak)

U praksi se češće koristi relativna mjera za grešku (na primjer, zbog aritmetike računala).

- Relativna greška:

$$\delta_x := \frac{\|\Delta x\|}{\|x\|}, \qquad \delta_y := \frac{\|\Delta y\|}{\|y\|}.$$

- Relativna uvjetovanost:

$$\kappa_{\text{rel}}(x) := \frac{\|\delta_y\|}{\|\delta_x\|}.$$

Problem je dobro uvjetovan ako je

- κ_{rel} što je moguće manji, za $\delta_x \to 0$.

Landauov simbol — red veličine

Neka su $g, h : \mathbb{R}^n \to \mathbb{R}^m$ funkcije, $\|\cdot\|_{\mathbb{R}^n}$ i $\|\cdot\|_{\mathbb{R}^m}$ norme i neka je $x_0 \in \mathbb{R}^n$.

Ako postoje konstante $C > 0$ i $\delta > 0$ takve da za sve x vrijedi

$$\|x - x_0\|_{\mathbb{R}^n} \leq \delta \quad \Longrightarrow \quad \|g(x)\|_{\mathbb{R}^m} \leq C\|h(x)\|_{\mathbb{R}^m},$$

onda kažemo da je

"funkcija g reda $\mathcal{O}$ od h, za x koji teži prema x_0"

i to pišemo ovako

$$g(x) = \mathcal{O}(h(x)) \quad (x \to x_0).$$

Landauov simbol (nastavak)

Primjer. Za $m = n = 1$ je

$$\begin{aligned} \sin x &= \mathcal{O}(x) \quad (x \to a), \quad \text{za sve } a \in \mathbb{R}, \\ x^2 + 3x &= \mathcal{O}(x) \quad (x \to 0), \\ x^2 - x - 6 &= \mathcal{O}(x - 3), \quad (x \to 3). \end{aligned}$$

Uvjetovanost i Taylorov teorem

Istražimo uvjetovanost problema za funkciju $f : \mathbb{R} \to \mathbb{R}$.

- Promatramo ponašanje f za male perturbacije Δx u okolini točke x. Neka je Δy pripadna perturbacija funkcijske vrijednosti $y = f(x)$, tj. $f(x + \Delta x) = y + \Delta y$.
- Neka je f još dva puta neprekidno derivabilna. Korištenjem Taylorovog polinoma stupnja 1 dobivamo

$$\begin{aligned}\Delta y &= f(x + \Delta x) - f(x)\\ &= f'(x)\Delta x + \frac{f''(x + \theta\Delta x)}{2!}(\Delta x)^2, \quad \theta \in (0, 1).\end{aligned}$$

Uvjetovanost i Taylorov teorem (nastavak)

- Za male perturbacije Δx, apsolutni oblik ove relacije je

$$\Delta y = f'(x)\,\Delta x + O((\Delta x)^2),$$

odakle slijedi da je $f'(x)$ ili $|f'(x)|$ apsolutna uvjetovanost funkcije f (za male Δx).

- Ako je $x \neq 0$ i $y \neq 0$, onda joj je relativna forma

$$\frac{\Delta y}{y} = \frac{x f'(x)}{f(x)}\,\frac{\Delta x}{x} + O\left(\left(\frac{\Delta x}{x}\right)^2\right),$$ pa relativnu uvjetovanost funkcije f možemo definirati kao

$$\kappa_{\text{rel}}(x) = (\operatorname{cond} f)(x) := \left|\frac{x f'(x)}{f(x)}\right|.$$

Uvjetovanost – primjer

Primjer. Relativna uvjetovanost funkcije

$$f(x) = \ln x,$$

je

$$(\operatorname{cond} f)(x) = \left| \frac{x f'(x)}{f(x)} \right| = \left| \frac{1}{\ln x} \right|,$$

što je veliko za $x \approx 1$.

Pitanje: Apsolutna uvjetovanost?

Primjer uvjetovanosti problema

Rekurzija za integral

Ispitajmo uvjetovanost problema računanja integrala

$$I_n = \int\limits_0^1 \frac{t^n}{t+5}\, dt$$

za fiksni prirodni broj n.

U ovom obliku, problem je napisan kao preslikavanje iz $\mathbb{N}$ u $\mathbb{R}$ i ne “paše” ranijem pojmu problema.

- Domena nije $\mathbb{R}$, nego $\mathbb{N}$ (diskretan skup), pa nema smisla govoriti o neprekidnosti, derivabilnosti i sl.

Zato prvo transformiramo problem.

Rekurzija za integral (nastavak)

Nađimo vezu između I_k i I_{k-1}, s tim da I_0 znamo izračunati

$$I_0 = \int_0^1 \frac{1}{t+5}\,dt = \ln(t+5)\,\Big|_0^1 = \ln\frac{6}{5}.$$

Za početak, očito vrijedi da je

$$\frac{t}{t+5} = 1 - \frac{5}{t+5},$$

Množenjem obje strane s t^{k-1} dobivamo

$$\frac{t^k}{t+5} = t^{k-1} - 5\,\frac{t^{k-1}}{t+5}.$$

Rekurzija za integral (nastavak)

Na kraju, integracijom na segmentu $[0, 1]$ izlazi

$$I_k = \int\limits_0^1 t^{k-1}\,dt - 5I_{k-1} = \frac{1}{k} - 5I_{k-1}, \quad k = 1, 2, \ldots, n.$$

Dakle, I_k je rješenje (linearne, nehomogene) diferencijske jednadžbe

$$y_k = -5y_{k-1} + \frac{1}{k}, \quad k = 1, 2, \ldots,$$

uz početni uvjet $y_0 = I_0$.

Rekurzija za integral (nastavak)

Varijacija početnog uvjeta definira niz funkcija f_k, $y_k = f_k(y_0)$.

Zanima nas relativna uvjetovanost funkcije f_n u točki $y_0 = I_0$, u ovisnosti o $n \in \mathbb{N}$.

- I_0 nije egzaktno prikaziv,
- umjesto I_0 spremi se aproksimacija $\widehat{I}_0$,
- rezultat — neka aproksimacija $\widehat{I}_n = f_n(\widehat{I}_0)$.

Indukcijom se lako dokaže da vrijedi

$$y_n = f_n(y_0) = (-5)^n y_0 + p_n,$$

gdje je p_n ovisi samo o nehomogenim članovima rekurzije, ali ne i o y_0.

Rekurzija za integral (nastavak)

Relativna uvjetovanost je

$$(\operatorname{cond} f_n)(y_0) = \left| \frac{y_0 f_n'(y_0)}{y_n} \right| = \left| \frac{y_0 (-5)^n}{y_n} \right|.$$

Iz definicije integrala: I_n monotono padaju po n, čak

$$\lim_{n\to\infty} I_n = 0.$$

Zbrajanjima dobivamo sve manje i manje brojeve!

$$(\operatorname{cond} f_n)(I_0) = \frac{I_0 \cdot 5^n}{I_n} > \frac{I_0 \cdot 5^n}{I_0} = 5^n.$$

f_n je vrlo loše uvjetovana u $y_0 = I_0$, i to tim gore kad n raste.

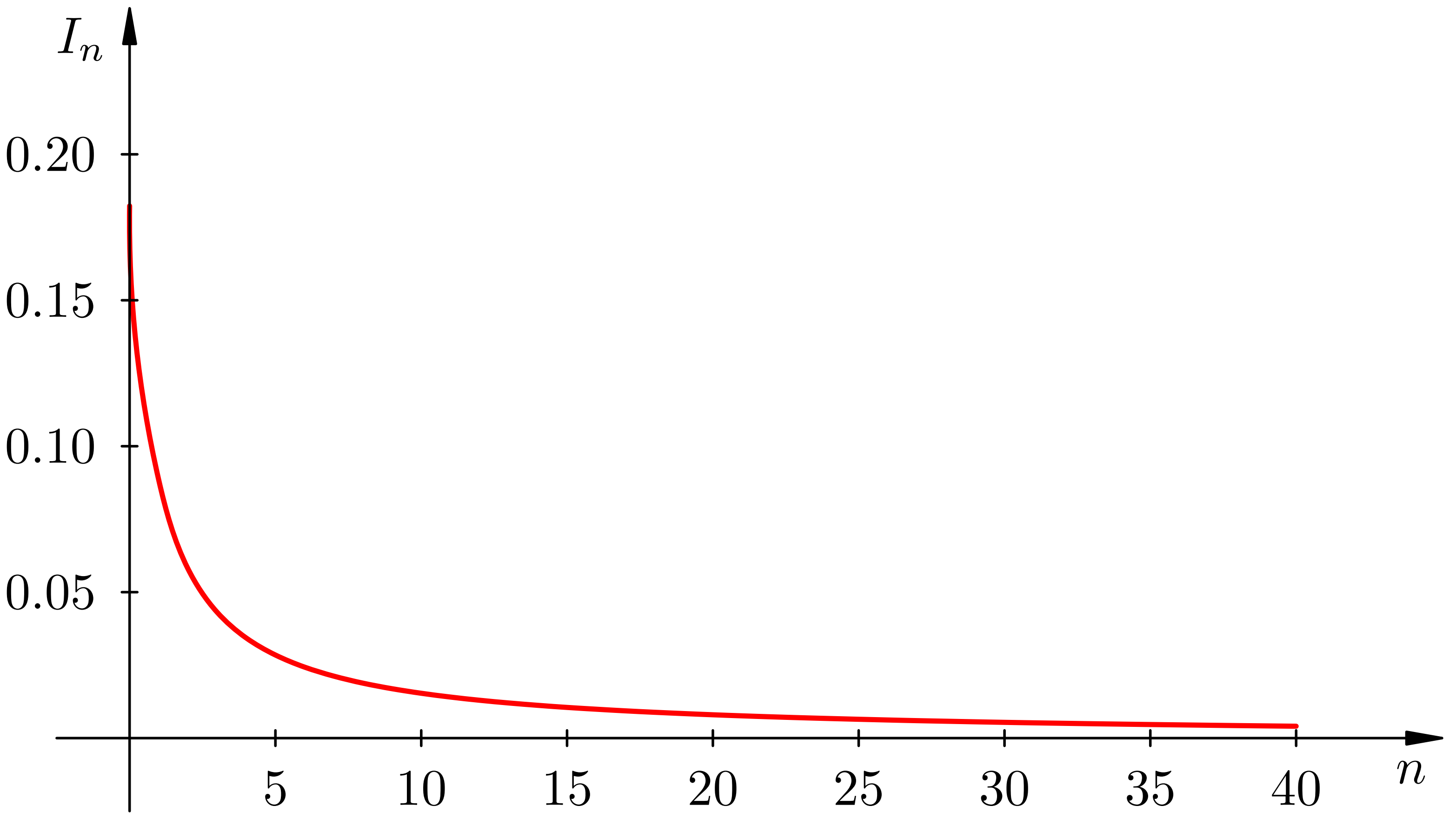
Rekurzija unaprijed — rezultati

Pitanje: Kako se loša uvjetovanost vidi, kad stvarno računamo $f_n(I_0)$?

Slikice!

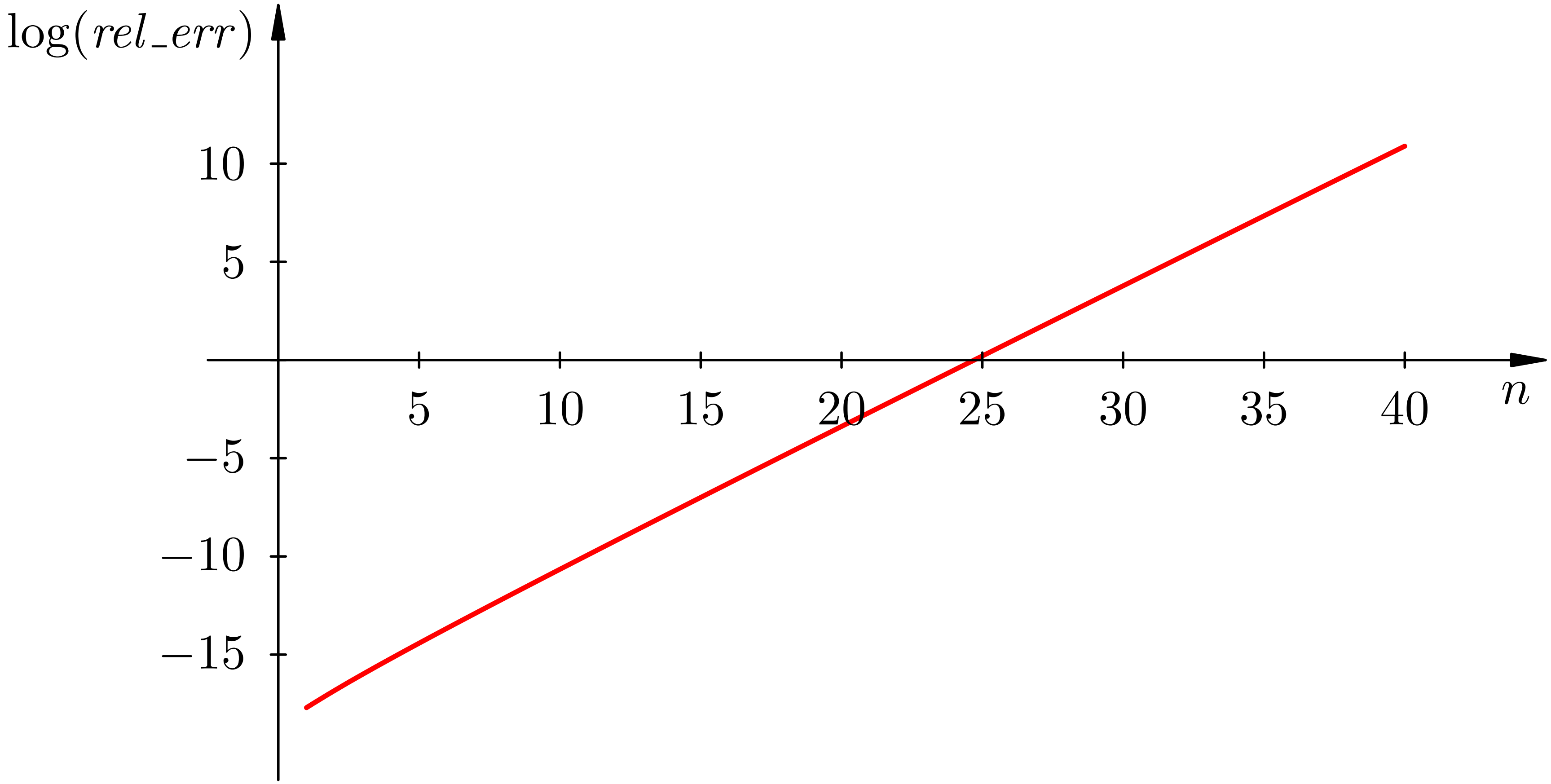
Pokaži program i rezultate!

Točne vrijednosti integrala



egzaktne/točne vrijednosti integrala I_n

Rekurzija unaprijed za I_n



(log$_{10}$) relativne greške izračunate vrijednosti integrala I_n rekurzijom unaprijed

Rekurzija za integral (nastavak)

Može li se loša uvjetovanost izbjeći?

- Može — okretanjem rekurzije.

Treba uzeti neki $\nu > n$ i "silazno" računati

$$y_{k-1} = \frac{1}{5}\left(\frac{1}{k} - y_k\right), \qquad k = \nu, \nu - 1, \ldots, n + 1.$$

Problem: kako izračunati početnu vrijednost y_ν.

Nova rekurzija definira niz funkcija g_k, koje vežu y_n i y_ν, uz $\nu > n$, tj.

$$y_n = g_n(y_\nu).$$

Rekurzija za integral (nastavak)

Relativna uvjetovanost za g_n

$$(\operatorname{cond} g_n)(y_\nu) = \left| \frac{y_\nu \, (-1/5)^{\nu-n}}{y_n} \right|, \quad \nu > n.$$

Za $y_\nu = I_\nu$, je $y_n = I_n$, a iz monotonosti I_n slijedi

$$(\operatorname{cond} g_n)(I_\nu) = \frac{I_\nu}{I_n} \cdot \left(\frac{1}{5}\right)^{\nu-n} < \left(\frac{1}{5}\right)^{\nu-n}, \quad \nu > n,$$

što je ispod 1, tj. greške se prigušuju.

Rekurzija za integral (nastavak)

Ako je $\widehat{I}_\nu$ neka aproksimacija za I_ν, onda za relativne perturbacije vrijedi

$$\left|\frac{\widehat{I}_n - I_n}{I_n}\right| = (\operatorname{cond} g_n)(I_\nu) \cdot \left|\frac{\widehat{I}_\nu - I_\nu}{I_\nu}\right| < \left(\frac{1}{5}\right)^{\nu-n} \cdot \left|\frac{\widehat{I}_\nu - I_\nu}{I_\nu}\right|.$$

Zbog linearnosti funkcije g_n, ova relacija vrijedi za bilo kakve perturbacije, a ne samo male.

- Početna vrijednost $\widehat{I}_\nu$ uopće ne mora biti blizu prave I_ν.
- Možemo uzeti $\widehat{I}_\nu = 0$, čime smo napravili relativnu grešku od 100% u početnoj vrijednosti ...

Rekurzija za integral (nastavak)

- ... a još uvijek dobivamo $\widehat{I}_n$ s relativnom greškom

$$\left|\frac{\widehat{I}_n - I_n}{I_n}\right| < \left(\frac{1}{5}\right)^{\nu-n}, \quad \nu > n.$$

- Povoljnim izborom ν, ocjenu na desnoj strani možemo napraviti po volji malom — ispod tražene točnosti ε.
- Dovoljno uzeti $\nu \geq n + \dfrac{\log(1/\varepsilon)}{\log 5}$, i $\widehat{I}_\nu = 0$ i računamo vrijednosti

$$\widehat{I}_{k-1} = \frac{1}{5}\left(\frac{1}{k} - \widehat{I}_k\right), \quad k = \nu, \nu-1, \ldots, n+1.$$

Rekurzija unatrag — rezultati

Pitanje: Kako se dobra uvjetovanost vidi, kad stvarno računamo $g_n(I_\nu)$?

Pokaži program i rezultate za $\varepsilon = 10^{-19}$!

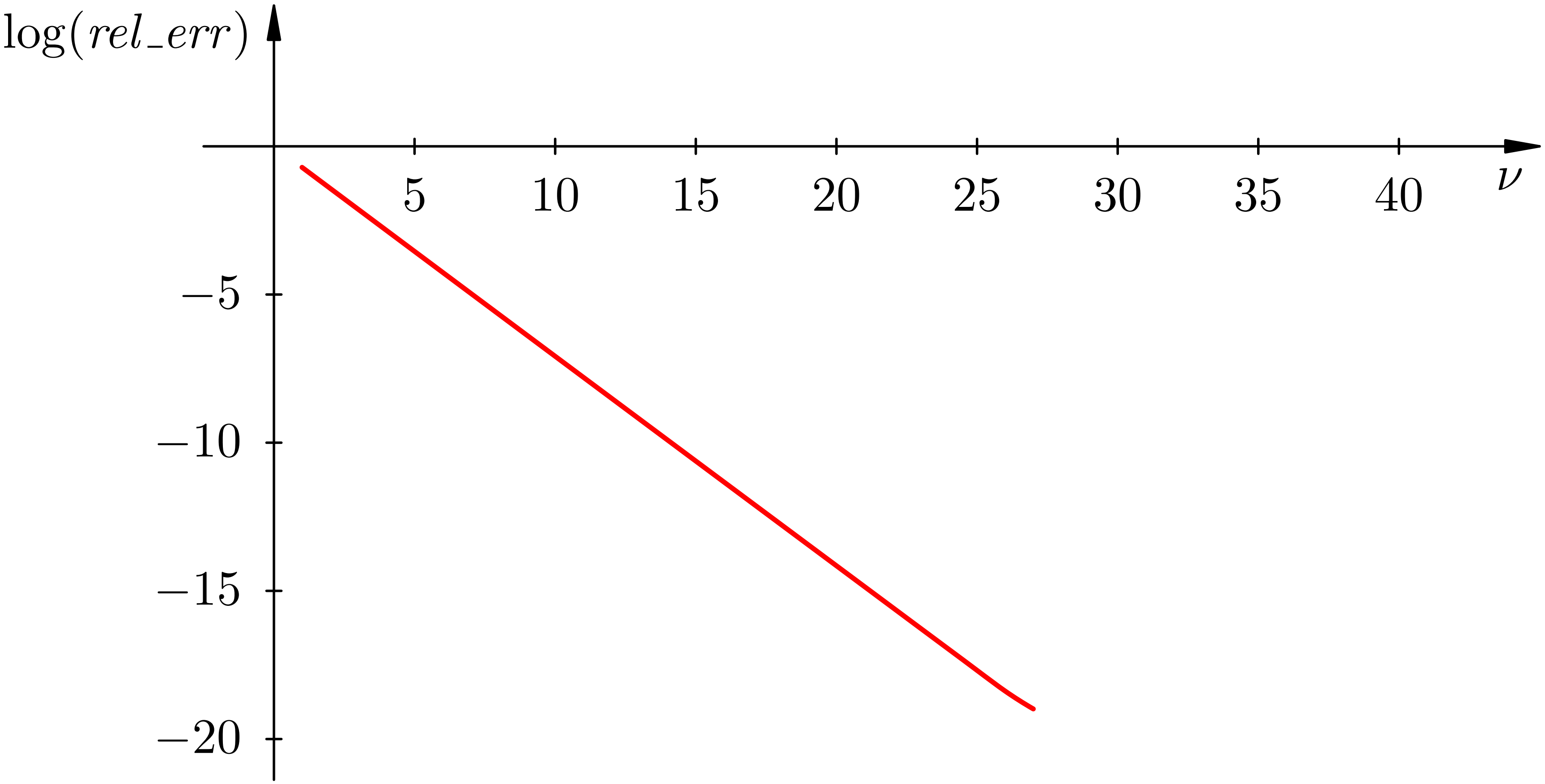
- Za ovaj ε dobijemo

$$\nu \geq n + \frac{\log(1/\varepsilon)}{\log 5} \approx n + 28.$$

  Dakle, “silazno” računamo 28 vrijednosti.

- Stvarna početna vrijednost je $\widehat{I}_\nu = 0$.

Rekurzija unazad za I_{40} — ovisno o startu ν



($\log_{10}$) relativne greške izračunate vrijednosti integrala I_{40} obratnom rekurzijom za $I_{40+\nu} = 0$

Primjer grešaka zaokruživanja

Primjer rasprostiranja grešaka

Primjer. Vrijednost

$$f_n(x) = (x - n)^{10}, \quad n = 0, \ldots, 10,$$

računamo u aritmetici računala u okolini točke n.

Primijetite da je graf funkcije $(x - n)^{10}$ translatirani graf funkcije x^{10} za n jedinica udesno.

Funkcijsku vrijednost funkcija f_n možemo izračunati na više načina koji su matematički ekvivalentni, ali nisu numerički jednaki.

Primjer rasprostiranja grešaka (nastavak)

Računat ćemo:

- translacijom grafa funkcije x^{10} za n jedinica udesno,
- korištenjem binomne formule

$$(x - n)^{10} = \sum_{k=0}^{10} \binom{10}{k} x^k (-n)^{10-k},$$

  s tim da polinom na desnoj strani računamo Hornerovom shemom.

Odgovor: U okolini točke n je $(x - n)^{10}$ mali broj. Članovi u sumi na desnoj strani su alternirajući po predznaku i rastu s porastom n.

Primjer rasprostiranja grešaka (nastavak)

Kako izgledaju grafovi?

- zeleno — graf dobiven translacijom,
- crveno — korištenjem binomne formule.

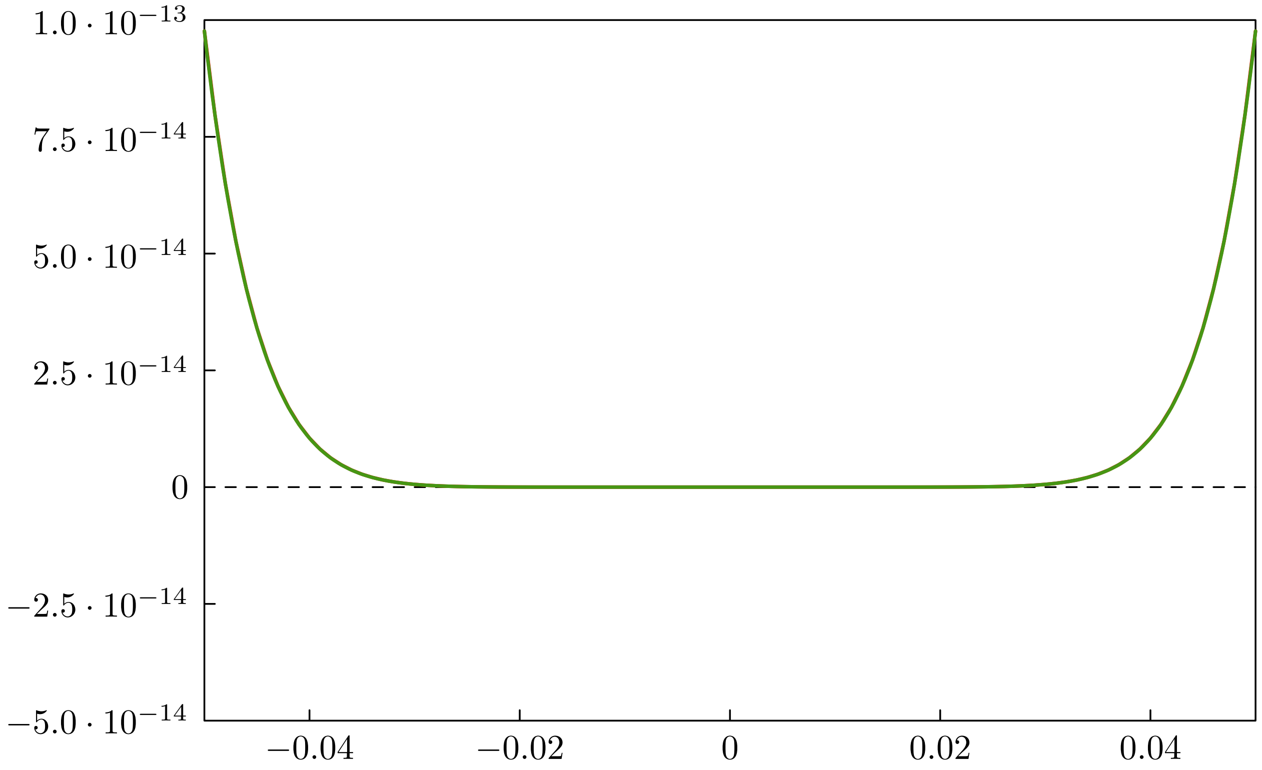
Za svaki n crtamo dvije slike grafa funkcije f_n:

- na intervalu $[n - 0.05, n + 0.05]$,
- na intervalu $[n - r, n + r]$, gdje je r odabran tako da ovaj interval sadrži numeričke nultočke od f_n.

Obratite pažnju na skale po x i y!

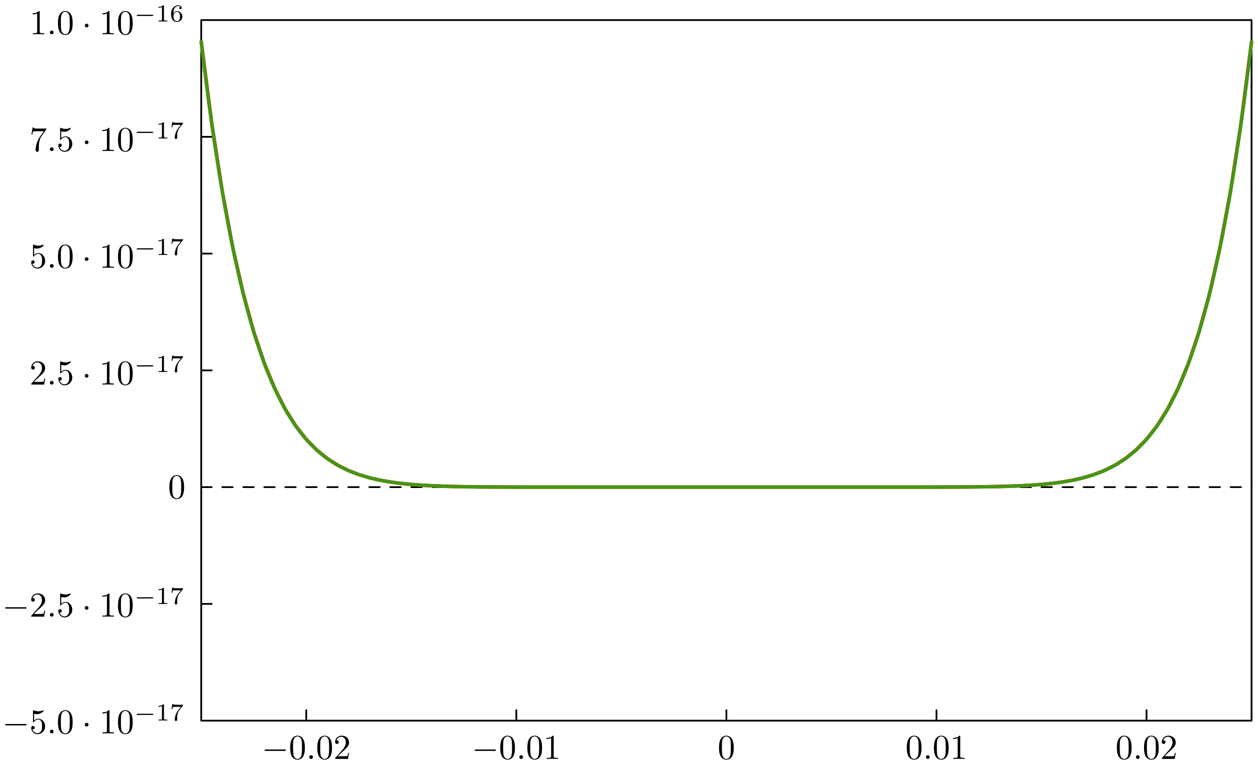
Primjer rasprostiranja grešaka — $n = 0$ *(1)*

$$(x - 0)^{10} = x^{10}$$



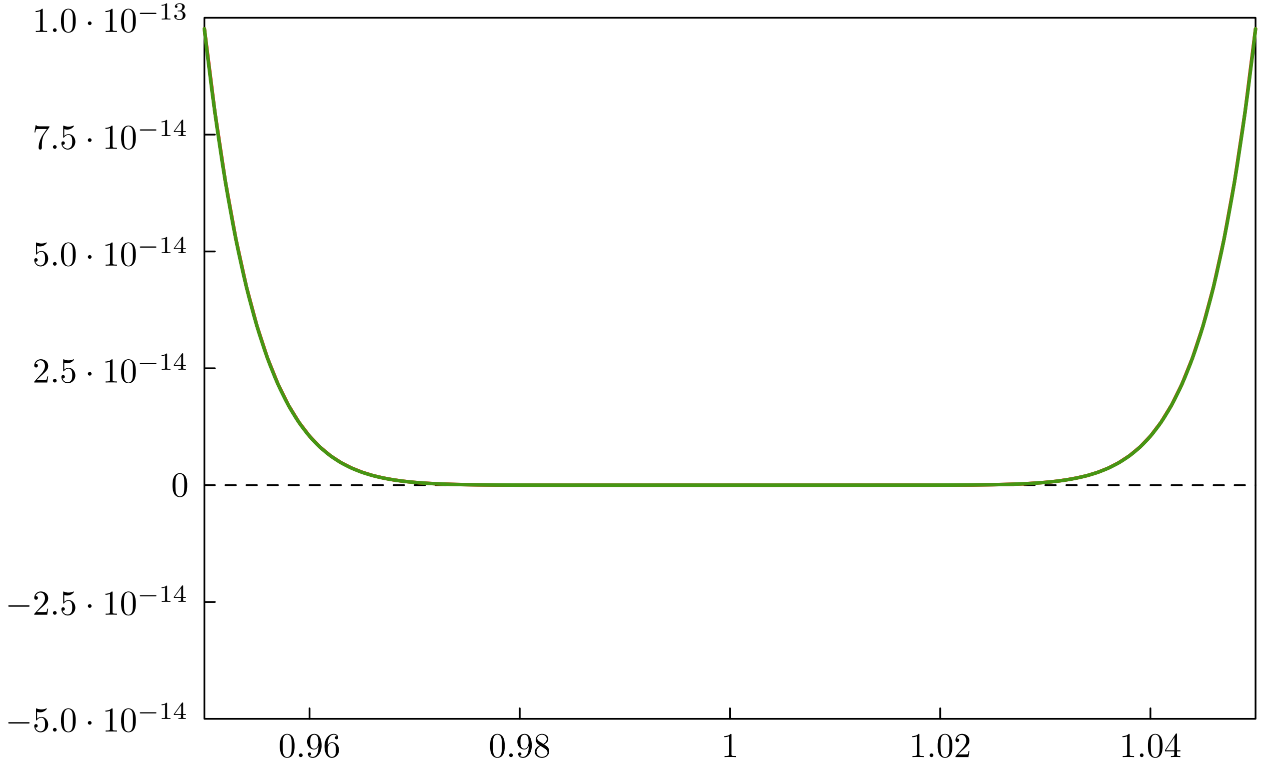
Primjer rasprostiranja grešaka — $n = 0$ *(2)*

$$(x - 0)^{10} = x^{10}$$



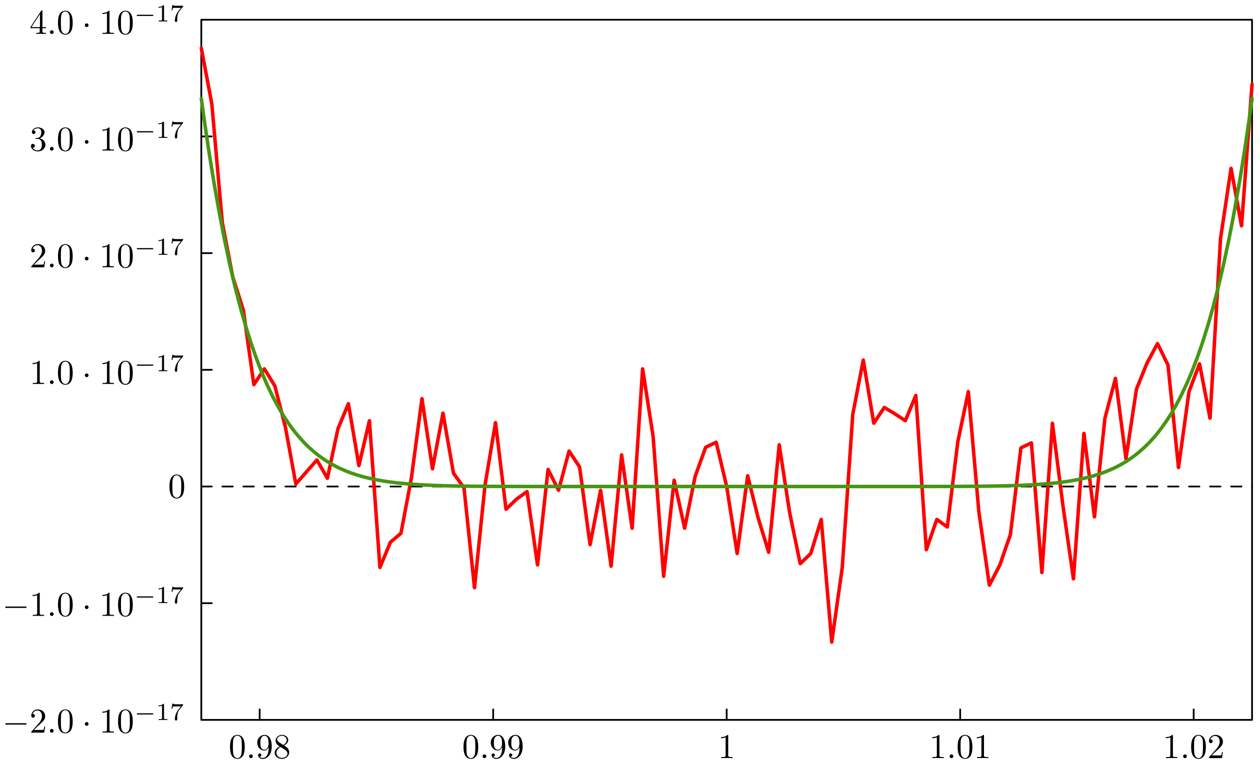
Primjer rasprostiranja grešaka — $n = 1$ (1)

$$(x-1)^{10} = x^{10} - 10x^9 + 45x^8 - 120x^7 + 210x^6 - 252x^5 + 210x^4 - 120x^3 + 45x^2 - 10x + 1$$



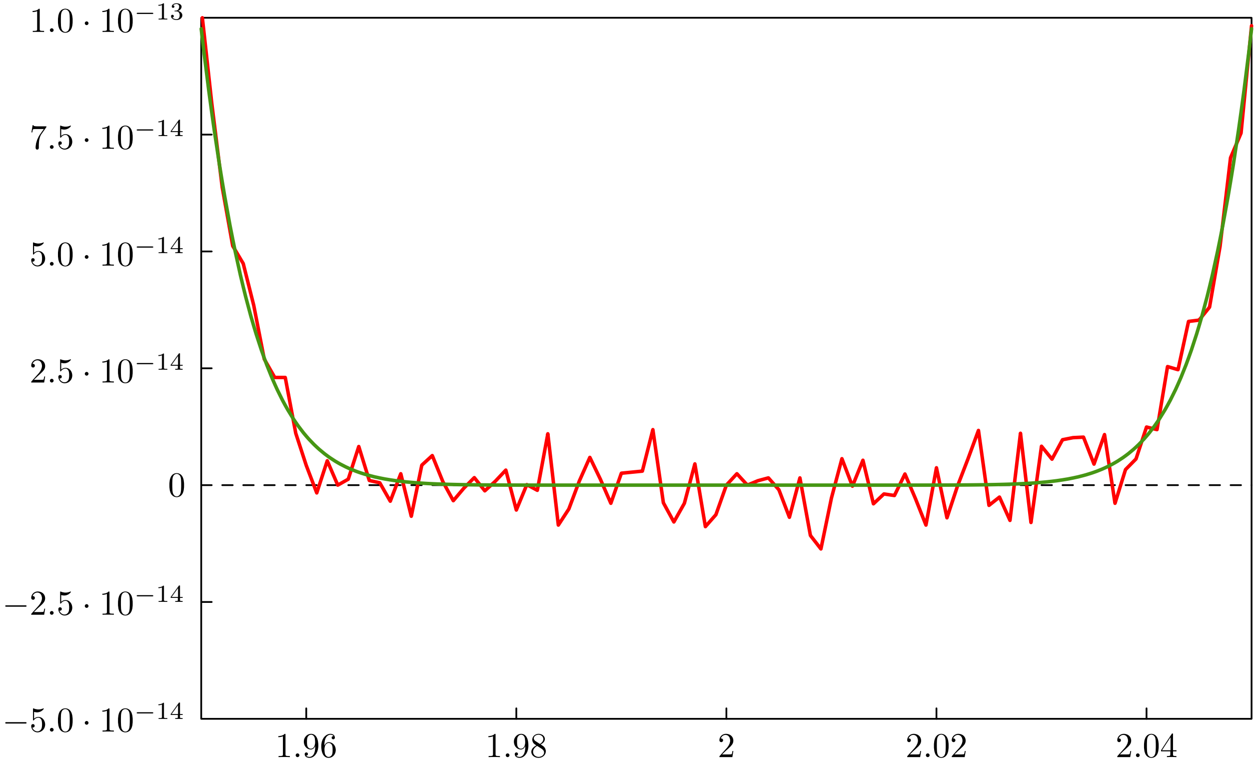
Primjer rasprostiranja grešaka — $n = 1$ (2)

$$(x-1)^{10} = x^{10} - 10x^9 + 45x^8 - 120x^7 + 210x^6 - 252x^5$$
$$+ 210x^4 - 120x^3 + 45x^2 - 10x + 1$$



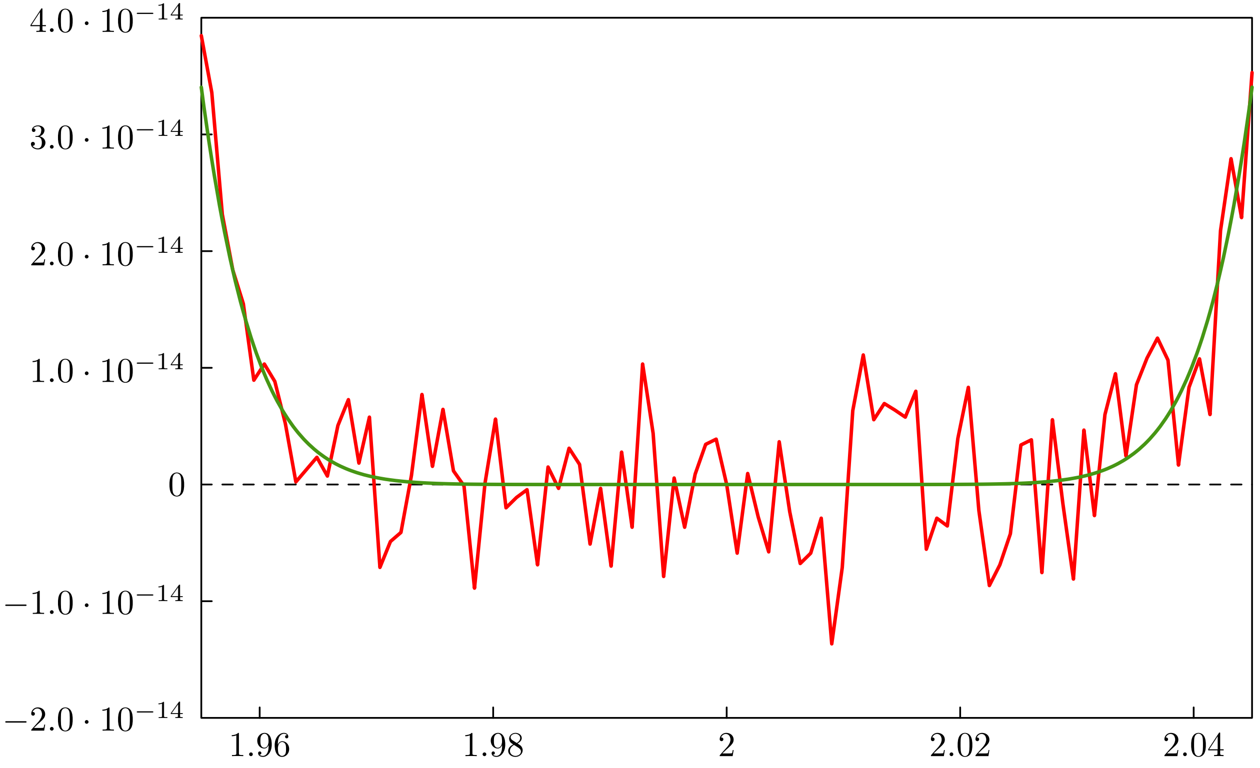
Primjer rasprostiranja grešaka — $n = 2$ (1)

$$(x-2)^{10} = x^{10} - 20x^9 + 180x^8 - 960x^7 + 3360x^6 - 8064x^5 \\ + 13440x^4 - 15360x^3 + 11520x^2 - 5120x + 1024$$



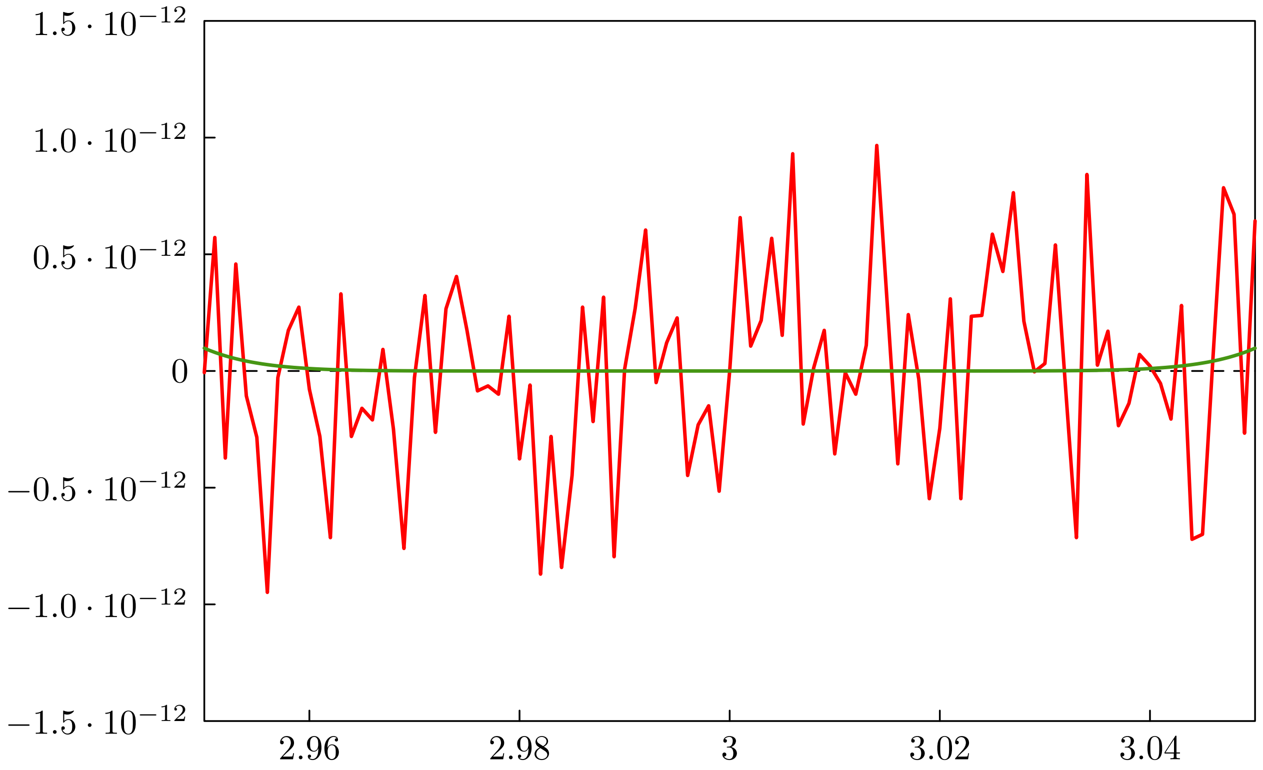
Primjer rasprostiranja grešaka — $n = 2$ (2)

$$(x-2)^{10} = x^{10} - 20x^9 + 180x^8 - 960x^7 + 3360x^6 - 8064x^5$$
$$+ 13440x^4 - 15360x^3 + 11520x^2 - 5120x + 1024$$



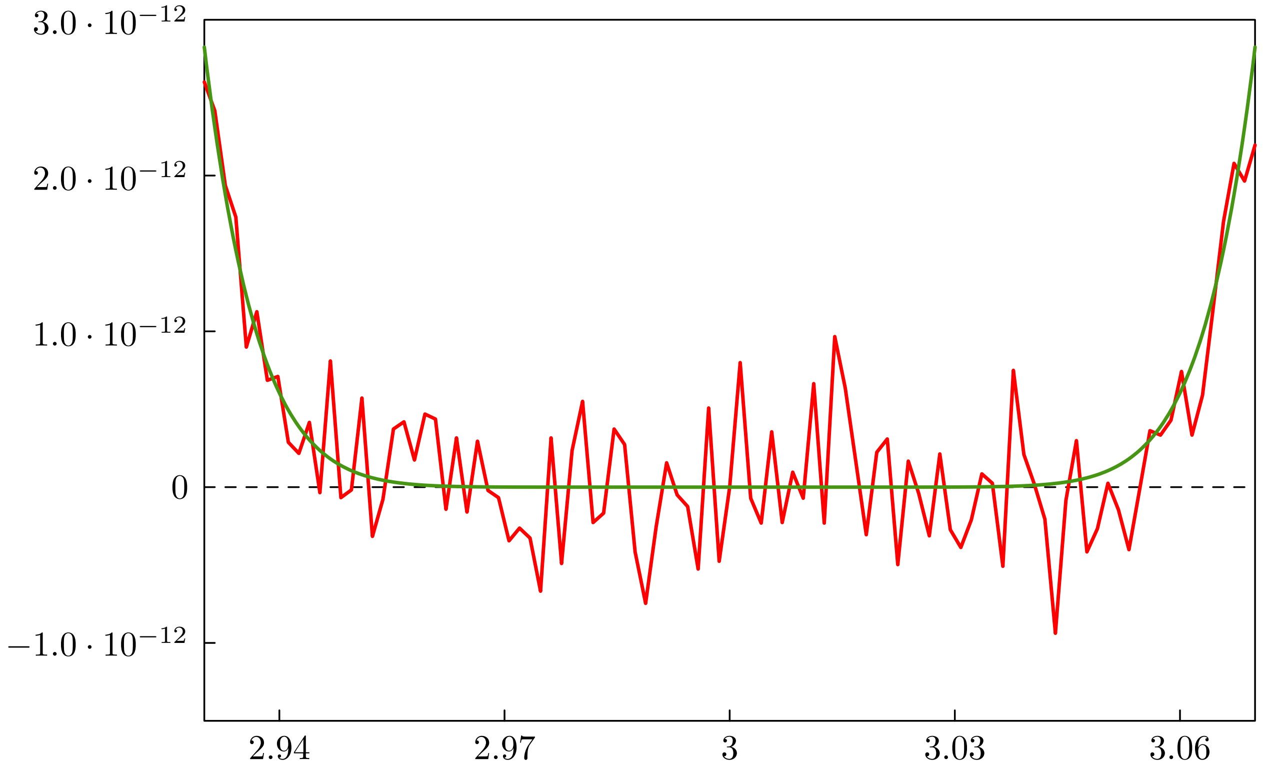
Primjer rasprostiranja grešaka — $n = 3$ (1)

$$(x-3)^{10} = x^{10} - 30x^9 + 405x^8 - 3240x^7 + 17010x^6 - 61236x^5 + 153090x^4 - 262440x^3 + 295245x^2 - 196830x + 59049$$



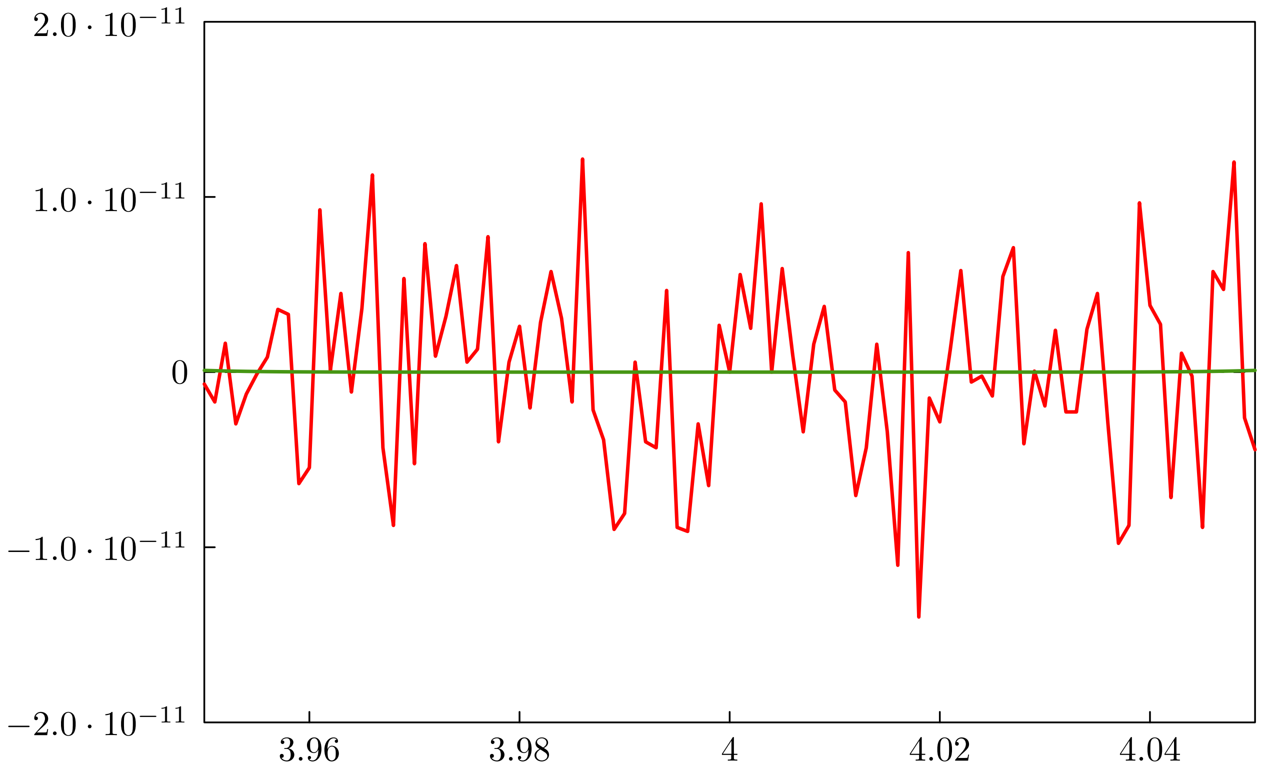
Primjer rasprostiranja grešaka — $n = 3$ *(2)*

$$(x-3)^{10} = x^{10} - 30x^9 + 405x^8 - 3240x^7 + 17010x^6 - 61236x^5$$
$$+ 153090x^4 - 262440x^3 + 295245x^2 - 196830x + 59049$$



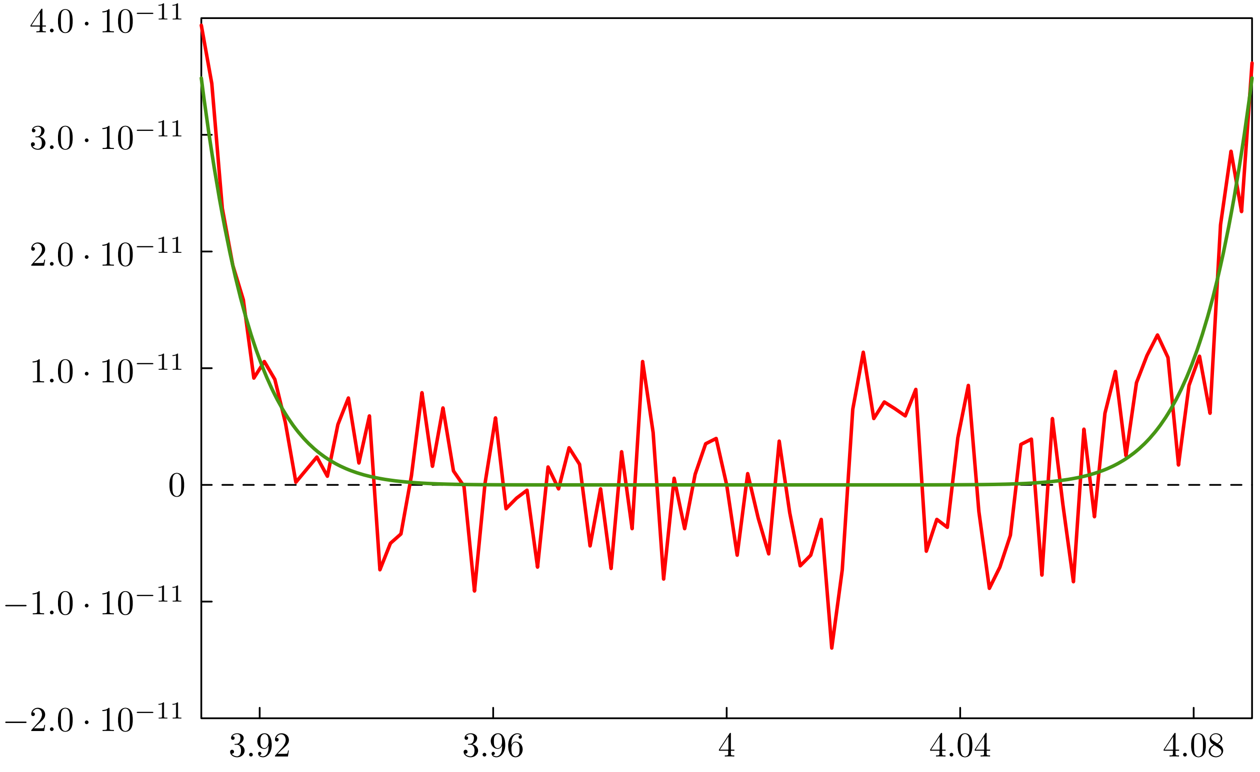
Primjer rasprostiranja grešaka — $n = 4$ (1)

$$\begin{aligned}(x-4)^{10} = x^{10} &- 40x^9 + 720x^8 - 7680x^7 + 53760x^6 \\ &- 258048x^5 + 860160x^4 - 1966080x^3 \\ &+ 2949120x^2 - 2621440x + 1048576\end{aligned}$$



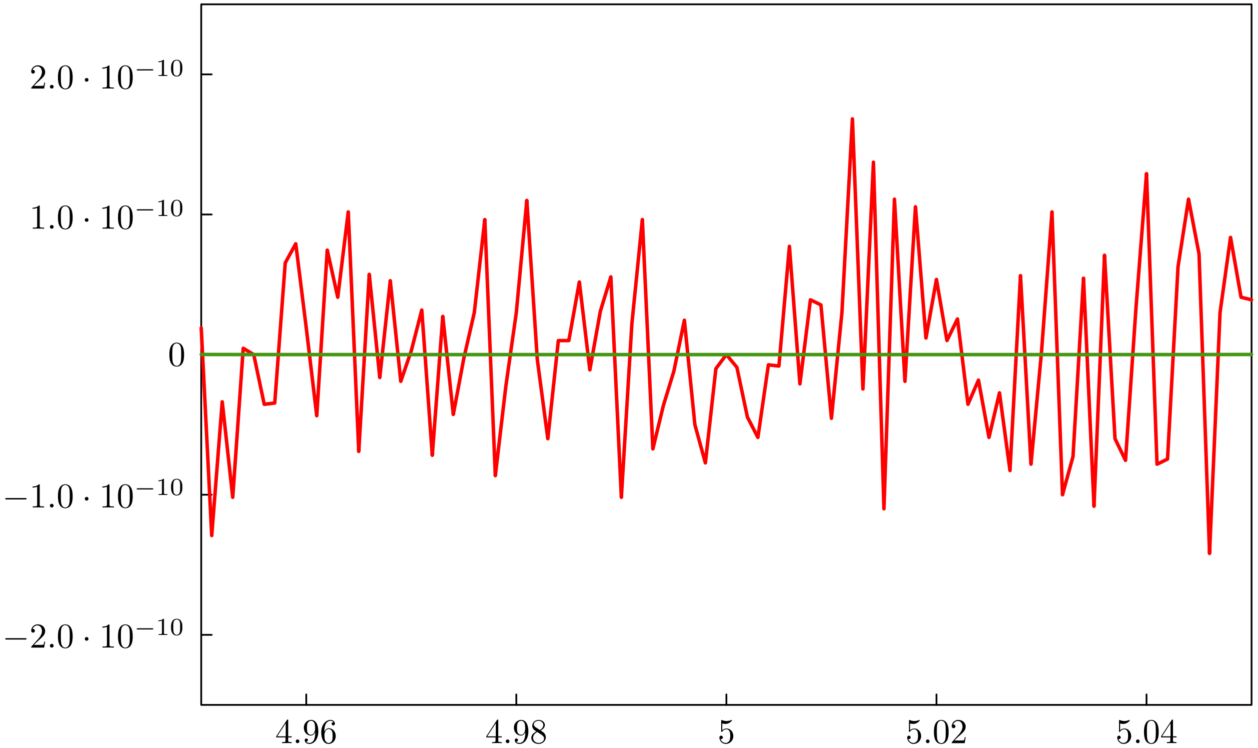
Primjer rasprostiranja grešaka — $n = 4$ *(2)*

$$
\begin{aligned}
(x-4)^{10} = x^{10} &- 40x^9 + 720x^8 - 7680x^7 + 53760x^6 \\
&- 258048x^5 + 860160x^4 - 1966080x^3 \\
&+ 2949120x^2 - 2621440x + 1048576
\end{aligned}
$$



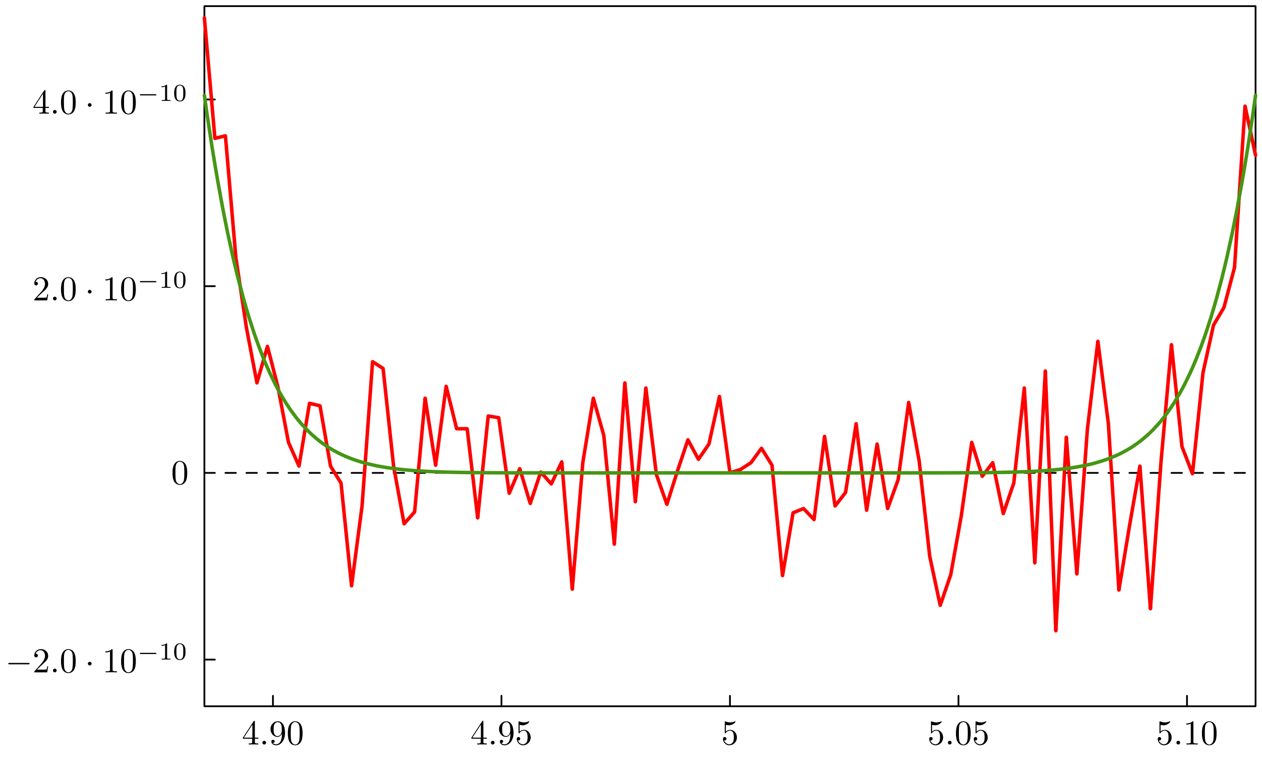
Primjer rasprostiranja grešaka — $n = 5$ (1)

$$
\begin{aligned}
(x-5)^{10} = x^{10} &- 50x^9 + 1125x^8 - 15000x^7 + 131250x^6 \\
&- 787500x^5 + 3281250x^4 - 9375000x^3 \\
&+ 17578125x^2 - 19531250x + 9765625
\end{aligned}
$$



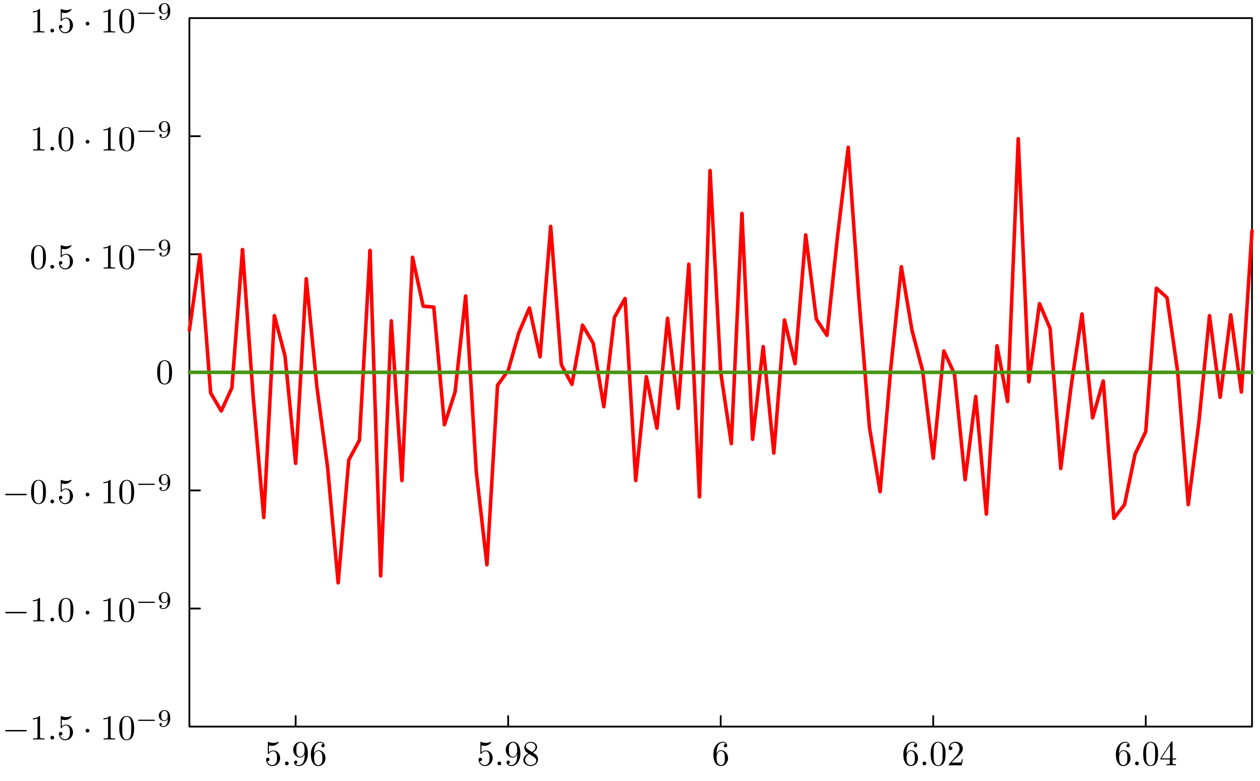
Primjer rasprostiranja grešaka — $n = 5$ *(2)*

$$
\begin{aligned}
(x-5)^{10} = x^{10} &- 50x^9 + 1125x^8 - 15000x^7 + 131250x^6 \\
&- 787500x^5 + 3281250x^4 - 9375000x^3 \\
&+ 17578125x^2 - 19531250x + 9765625
\end{aligned}
$$



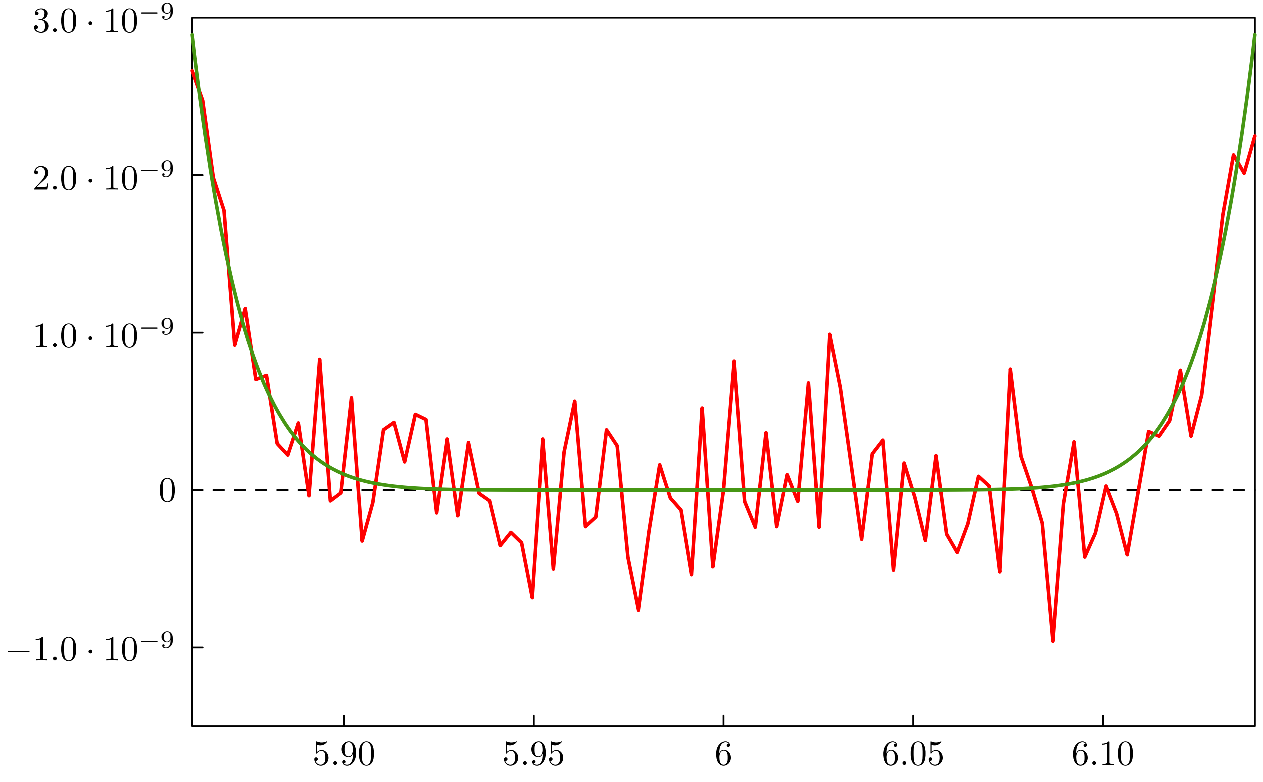
Primjer rasprostiranja grešaka — $n = 6$ *(1)*

$$
\begin{aligned}
(x-6)^{10} = x^{10} &- 60x^9 + 1620x^8 - 25920x^7 + 272160x^6 \\
&- 1959552x^5 + 9797760x^4 - 33592320x^3 \\
&+ 75582720x^2 - 100776960x^1 + 60466176
\end{aligned}
$$



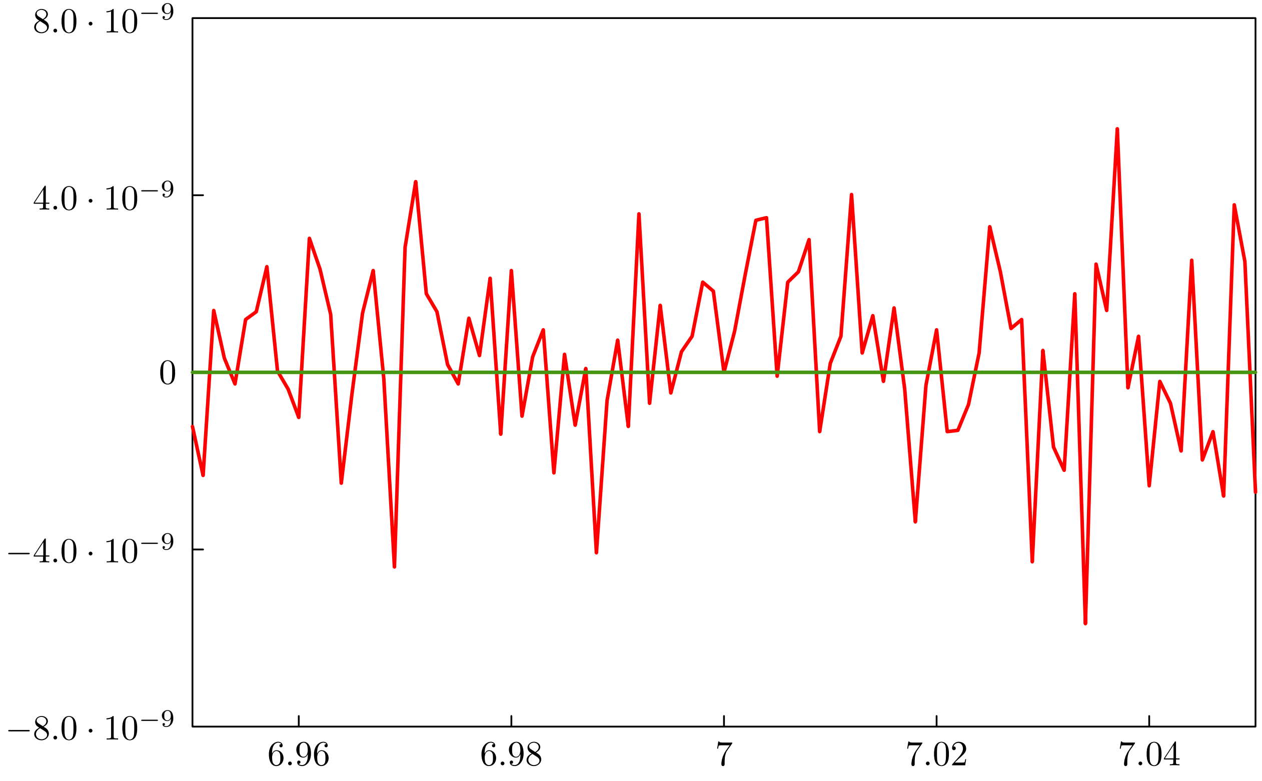
Primjer rasprostiranja grešaka — $n = 6$ (2)

$$(x-6)^{10} = x^{10} - 60x^9 + 1620x^8 - 25920x^7 + 272160x^6 - 1959552x^5 + 9797760x^4 - 33592320x^3 + 75582720x^2 - 100776960x^1 + 60466176$$



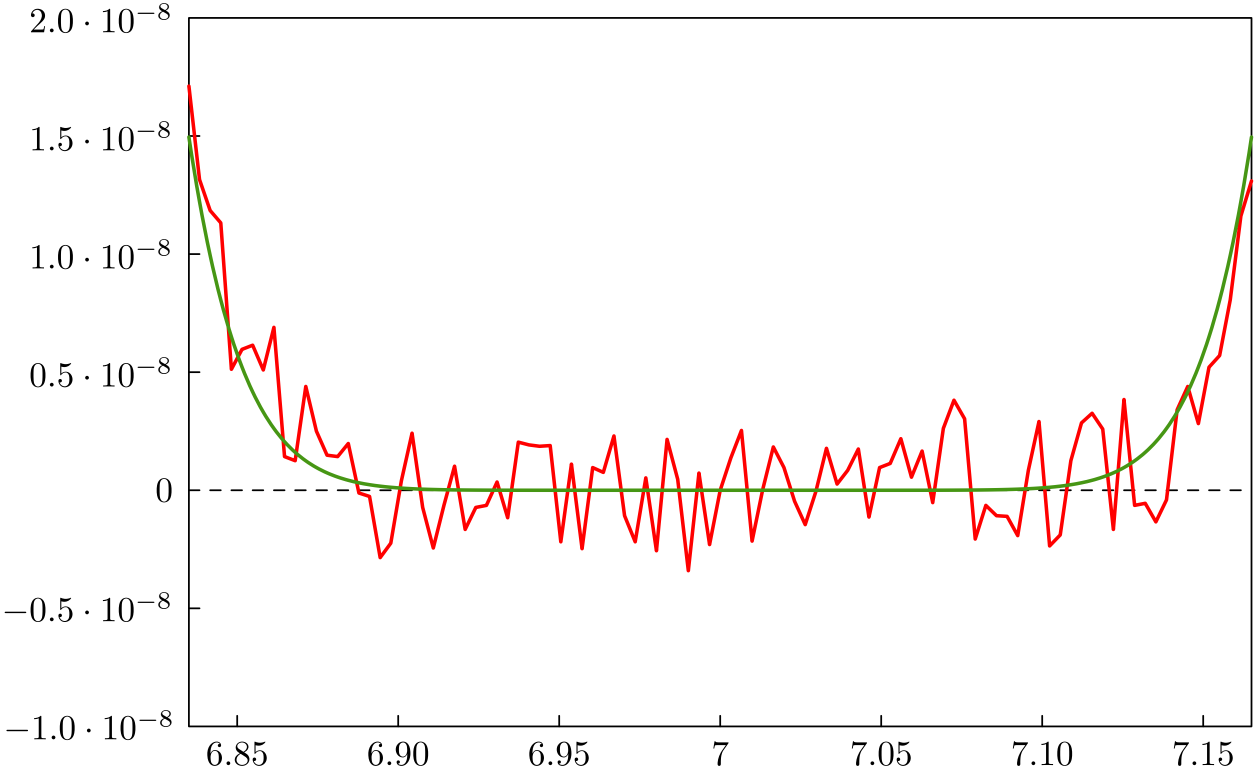
Primjer rasprostiranja grešaka — $n = 7$ (1)

$$
\begin{aligned}
(x-7)^{10} = x^{10} &- 70x^9 + 2205x^8 - 41160x^7 + 504210x^6 \\
&- 4235364x^5 + 24706290x^4 - 98825160x^3 \\
&+ 259416045x^2 - 403536070x^1 + 282475249
\end{aligned}
$$



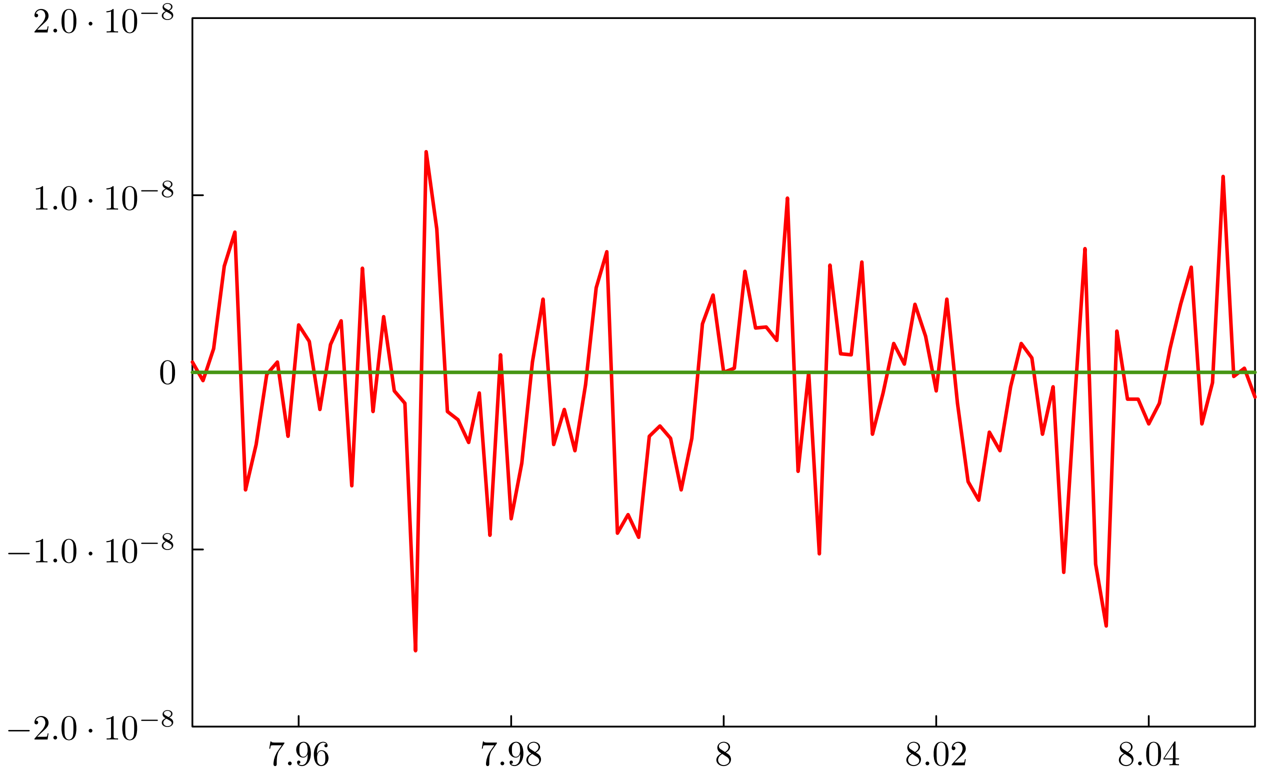
Primjer rasprostiranja grešaka — $n = 7$ *(2)*

$$
\begin{aligned}
(x-7)^{10} = x^{10} &- 70x^9 + 2205x^8 - 41160x^7 + 504210x^6 \\
&- 4235364x^5 + 24706290x^4 - 98825160x^3 \\
&+ 259416045x^2 - 403536070x^1 + 282475249
\end{aligned}
$$



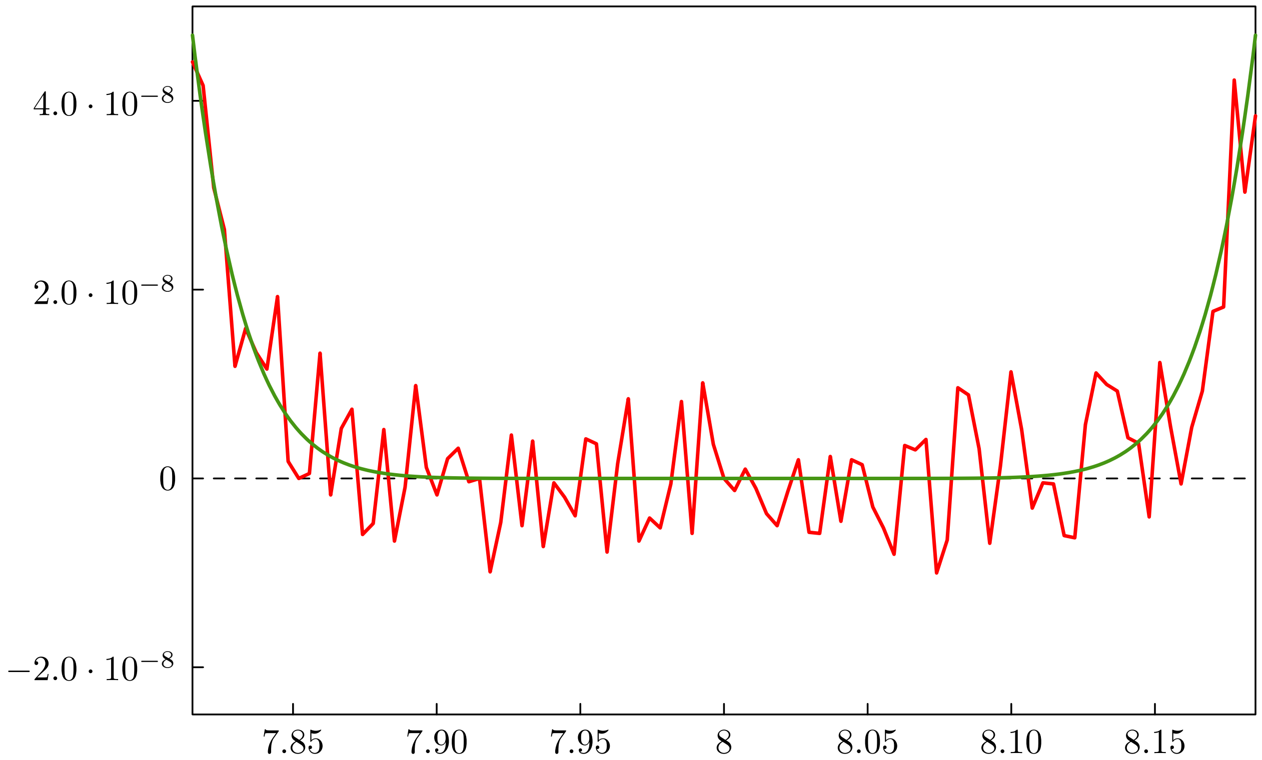
Primjer rasprostiranja grešaka — $n = 8$ (1)

$$
\begin{aligned}
(x-8)^{10} = x^{10} &- 80x^9 + 2880x^8 - 61440x^7 + 860160x^6 \\
&- 8257536x^5 + 55050240x^4 - 251658240x^3 \\
&+ 754974720x^2 - 1342177280x^1 + 1073741824
\end{aligned}
$$



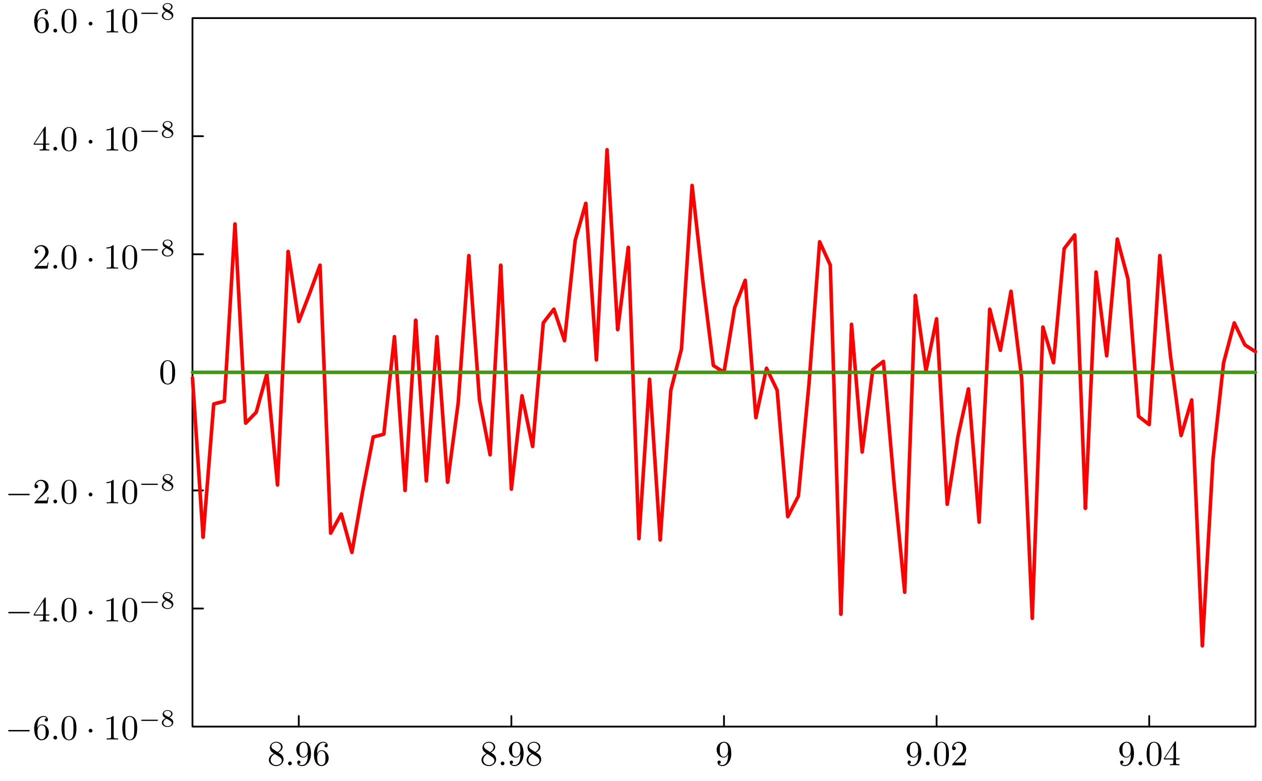
Primjer rasprostiranja grešaka — $n = 8$ (2)

$$(x-8)^{10} = x^{10} - 80x^9 + 2880x^8 - 61440x^7 + 860160x^6 \\ - 8257536x^5 + 55050240x^4 - 251658240x^3 \\ + 754974720x^2 - 1342177280x^1 + 1073741824$$



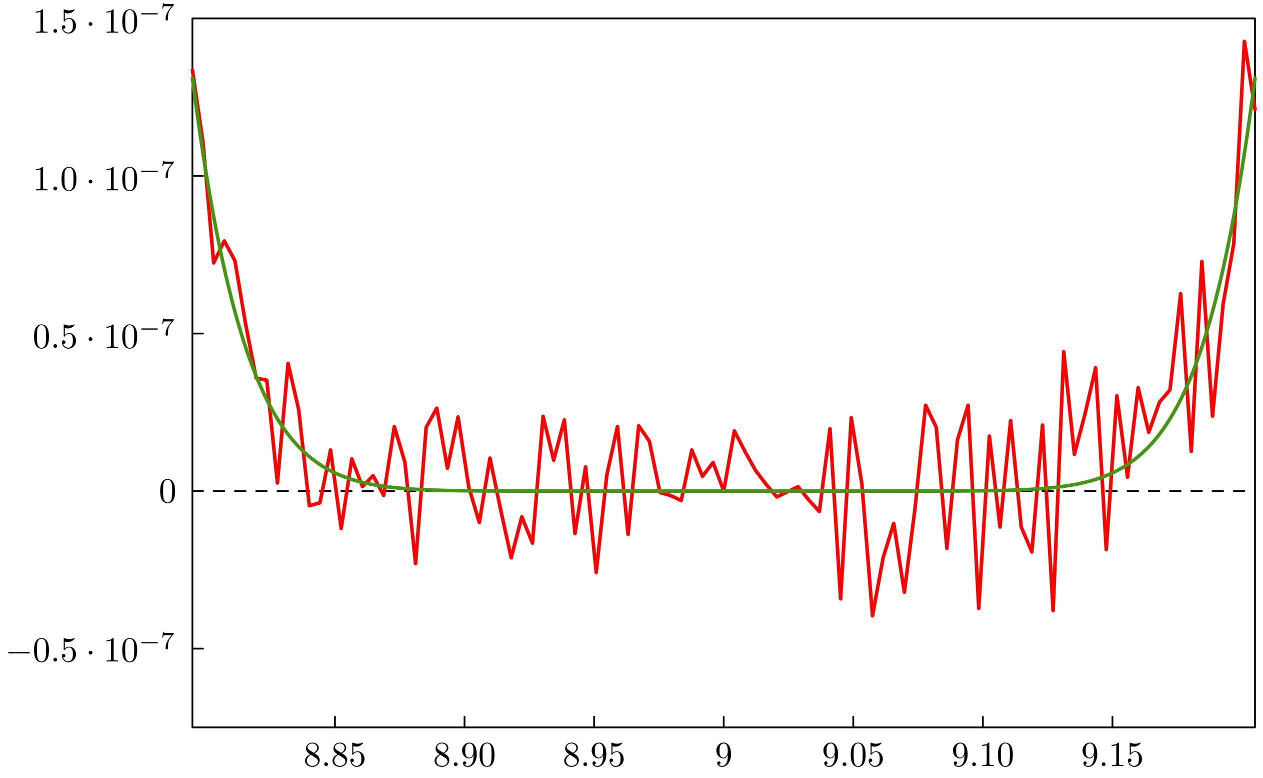
Primjer rasprostiranja grešaka — $n = 9$ (1)

$$
\begin{aligned}
(x-9)^{10} = x^{10} &- 90x^9 + 3645x^8 - 87480x^7 + 1377810x^6 \\
&- 14880348x^5 + 111602610x^4 - 573956280x^3 \\
&+ 1937102445x^2 - 3874204890x^1 + 3486784401
\end{aligned}
$$



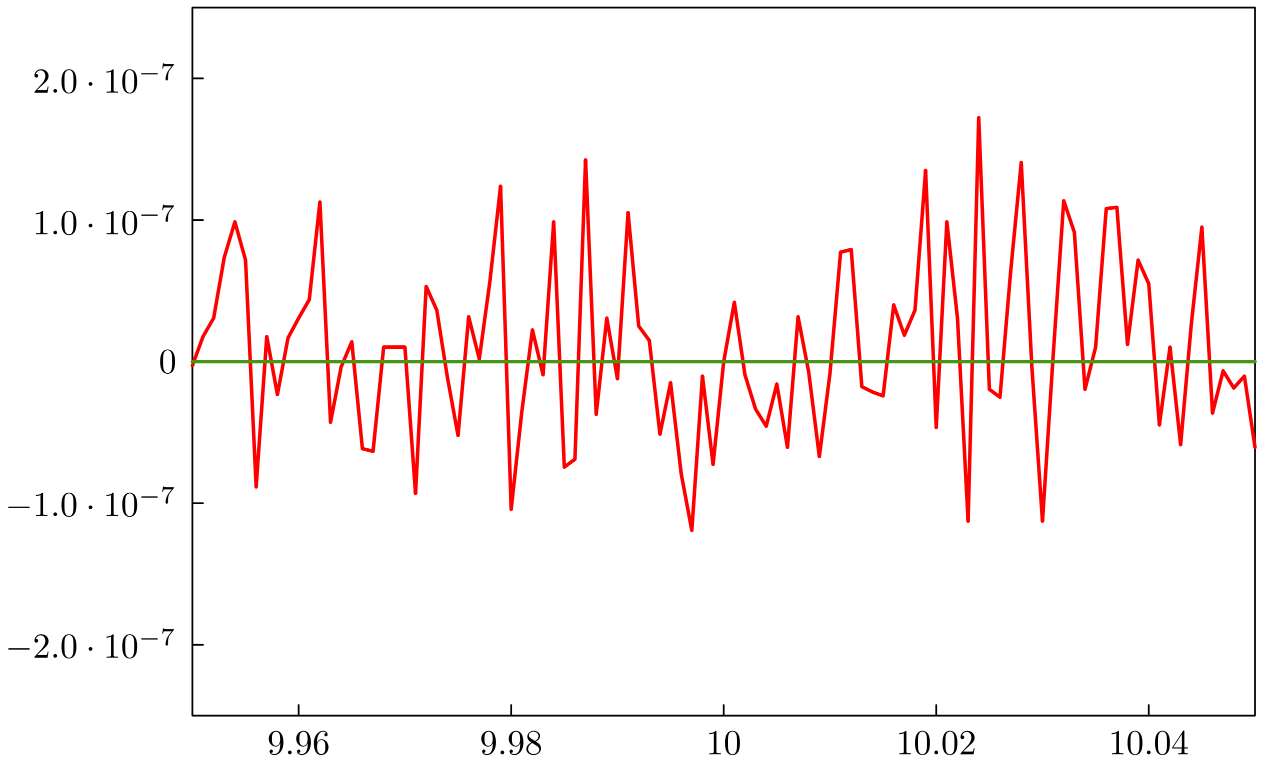
Primjer rasprostiranja grešaka — $n = 9$ (2)

$$
\begin{aligned}
(x-9)^{10} = x^{10} &- 90x^9 + 3645x^8 - 87480x^7 + 1377810x^6 \\
&- 14880348x^5 + 111602610x^4 - 573956280x^3 \\
&+ 1937102445x^2 - 3874204890x^1 + 3486784401
\end{aligned}
$$



Primjer rasprostiranja grešaka — $n = 10$ *(1)*

$$
\begin{aligned}
(x-10)^{10} = x^{10} &- 100x^9 + 4500x^8 - 120000x^7 + 2100000x^6 \\
&- 25200000x^5 + 210000000x^4 - 1200000000x^3 \\
&+ 4500000000x^2 - 10000000000x^1 + 10000000000
\end{aligned}
$$



Primjer rasprostiranja grešaka — $n = 10$ (2)

$$(x-10)^{10} = x^{10} - 100x^9 + 4500x^8 - 120000x^7 + 2100000x^6 - 25200000x^5 + 210000000x^4 - 1200000000x^3 + 4500000000x^2 - 10000000000x^1 + 10000000000$$