

# *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).
  - Analiza pojedinih vrsta grešaka.
    - Greške metode — teorija aproksimacija.
    - Greške u podacima — teorija perturbacija.
  - Uvjetovanost problema.
  - Približno računanje i perturbacije podataka.
  - Mjerenje grešaka — razne norme.
  - 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/](http://web.math.hr/~singer/num_mat/)

Tamo su kompletna **predavanja** od prošle **dvije** godine, a stizat će i **nova** (kako nastaju). Sredit će se ovaj vikend!

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

[http://web.math.hr/~singer/num\\_mat/num\\_mat1.pdf](http://web.math.hr/~singer/num_mat/num_mat1.pdf)

**Skraćena** verzija **skripte** — **2. dio** (drugih **6** tjedana):

[http://web.math.hr/~singer/num\\_mat/num\\_mat2.pdf](http://web.math.hr/~singer/num_mat/num_mat2.pdf)

# Informacije — demonstratori

Kolegij “**Numerička matematika**” ima čak **3 demonstratora**:

- Ervin Duraković, Nika Kenda i Marin Mišur.

Za upute za dogovor i termine demonstratura

- pogledajte **oglase** na oglasnoj ploči i

- **web** stranicu kolegija — pod “**nastava**”.

# 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** (mjerjenja 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 slič **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.



# Analiza grešaka

# Greška metode

Gruba podjela **numeričkih metoda** – prema greškama:

## Egzaktne metode

- ☛ daju **egzaktno** rješenje u **konačnom** broju “koraka” (računskih operacija).

Primjer:

- ☛ Gaussove eliminacije ili LR faktorizacija za linearne sustave.

**Greška** takvih metoda je **nula**, uz egzaktno računanje.

## Približne ili **neegzaktne** metode

- ☛ daju **približno** rješenje problema, u **konačnom** broju “koraka” (računskih operacija).

# Greška metode — približne metode

Mogu biti **egzaktne** — na nekom limesu!

Primjeri:

- zamjena kompliciranog modela jednostavnijim,
- greške diskretizacije (numerička integracija),
- greške odbacivanja/rezanja, konačne iteracije (rješavanje nelinearnih jednačbi)

Analiza grešaka spada u **teoriju aproksimacija**.

Pošteno, to je **standardni** predmet proučavanja **numeričke** matematike, u **širem** smislu

- numerička analiza, funkcionalna analiza, itd.

Time se bavimo **veći** dio kolegija!

# Greške u podacima

Ključno svojstvo **problema** je

- ovisnost **rješenja** o **greškama** ili **perturbacijama** ulaznih podataka.

To spada u **teoriju perturbacije**.

Da bi problem uopće imao smisla, očekujemo

- neku vrstu **neprekidnosti** rješenja,
- ili barem **ograničenu** osjetljivost na perturbacije.

Inače imamo “**loše**” postavljen problem!

Osjetljivost se obično mjeri tzv. **brojem uvjetovanosti** problema (engl. condition number). Može ih biti i više.

# 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:

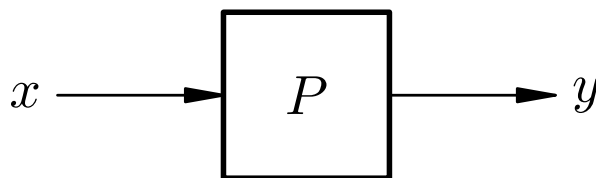
- Koju **točnost rezultata** možemo očekivati
- pri **točnom računanju**
- s (malo) **pomaknutim** — **netočnim podacima**?

# 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



**Problem**  $P$  interpretiramo kao računanje vrijednosti **funkcije**

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

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**:

greška u rezultatu  $\approx$  **uvjetovanost** · 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**!

Više detalja malo kasnije, kad “sredimo” greške zaokruživanja!

## 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 jednačbe oblika  $Ay = x$ , gdje je

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

## 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  $\mathcal{X} = \mathcal{Y} = \mathbb{R}^2$ .

# Približno računanje i perturbacije podataka

# Interpretacija grešaka zaokruživanja

Kod **približnog** računanja — na pr. u aritmetici **računala**, imamo greške **zaokruživanja**

- **spremanjem** ulaznih podataka u algoritam,
- nakon **svake** pojedine aritmetičke operacije.

Ključna stvar za **analizu** tih grešaka je

- svođenje na **teoriju perturbacija**, u smislu
- **egzaktnog** računanja s **perturbiranim** polaznim podacima!

Kako to ide? Ilustracija na IEEE standardu.

## Greške prikaza i aritmetike

Ako je ulazni podatak  $x \in \mathbb{R}$

• unutar raspona brojeva prikazivih u računalu, onda se, umjesto  $x$ , sprema zaokruženi prikazivi broj  $fl(x)$ , tako da vrijedi

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

gdje je

- $\varepsilon$  relativna greška napravljena tim zaokruživanjem,
- a  $u$  je jedinična greška zaokruživanja.

Imamo malu relativnu grešku, a računalo dalje računa

- s perturbiranim polaznim podatkom  $fl(x)$ .

Slična stvar vrijedi i za aritmetičke operacije.

# Zaokruživanje u aritmetici

Osnovna pretpostavka za realnu aritmetiku u računalu:

- za sve četiri osnovne aritmetičke operacije vrijedi ista ocjena greške zaokruživanja kao i za prikaz brojeva.

Isto vrijedi i za neke matematičke funkcije, poput  $\sqrt{\quad}$ , ali ne vrijedi za sve funkcije (na pr. za  $\cos$  i  $\sin$  u okolini nule).

Preciznije: Neka  $\circ$  označava bilo koju operaciju  $+$ ,  $-$ ,  $*$ ,  $/$ . Za prikazive brojeve u dozvoljenom rasponu  $x, y \in \mathcal{F}$ , takve da je i egzaktni rezultat  $x \circ y$  u dozvoljenom rasponu (tj. u  $\mathcal{F}$ ), vrijedi ocjena relativne greške

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

Broj  $\varepsilon$  ovisi o  $x$ ,  $y$ , operaciji  $\circ$  i aritmetici računala.

## Zaokruživanje u aritmetici (nastavak)

Ova ocjena je **ekvivalentna** **idealnom** izvođenju aritmetičkih operacija:

- **egzaktno** izračunaj rezultat operacije  $x \circ y$ ,
- **zaokruži** ga, pri spremanju rezultata u memoriju.

To **ne znači** da računalo **zaista** tako i računa. Naime,

- za  $+$ ,  $-$ ,  $*$  to bi se još i moglo napraviti (egzaktne rezultati imaju konačan binarni prikaz),
- ali kod **dijeljenja** to sigurno ne ide (egzaktan kvocijent može imati beskonačan binarni prikaz).

Dakle, **važno** je samo da dobijemo istu **ocjenu greške** kao u “idealnom” računanju, a **nije važno** kako se stvarno računa!

# Širenje grešaka zaokruživanja

Kad imamo puno operacija — nastaje problem:

- greške se šire i
- treba procijeniti grešku u rezultatu.

kako to napraviti?

Za aritmetiku računala ne vrijedi:

- asocijativnost zbrajanja i množenja,
- distributivnost množenja prema zbrajanju.

Jedino vrijedi:

- komutativnost za zbrajanje i množenje.



# Širenje grešaka zaokruživanja (nastavak)

Za analizu grešaka zaokruživanja ne možemo koristiti nikakva “normalna” pravila za aritmetičke operacije u računalu, jer ti zakoni naprosto ne vrijede.

Stvarna algebarska struktura je izrazito komplicirana i postoje debele knjige na tu temu.

- Vrijede neka “zamjenska” pravila, ali su neupotrebljiva za analizu iole većih proračuna.

Međutim, analiza pojedinih operacija postaje bitno lakša, ako uočimo da:

- greške zaokruživanja u aritmetici računala možemo interpretirati i kao egzaktne operacije, ali na “malo” pogrešnim podacima!

# Širenje grešaka zaokruživanja (nastavak)

Kako? Dovoljno je faktor  $(1 + \varepsilon)$  u ocjeni greške

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

“zalijepiti” na  $x$  i/ili  $y$ . To je isto kao da operand(i) ima(ju) neku relativnu grešku na ulazu u operaciju, a operacija  $\circ$  je egzaktna. Dakle,

- izračunati (ili “zaokruženi”) rezultat jednak je egzaktnom, ali za malo promijenjene (tj. perturbirane) podatke (u relativnom smislu).

Što dobivamo ovom interpretacijom?

- Onda možemo koristiti “normalna” pravila aritmetike za analizu grešaka.

## Širenje grešaka zaokruživanja (nastavak)

Ne zaboravimo još da  $\varepsilon$  ovdje ovisi o  $x$ ,  $y$ , i operaciji  $\circ$ . Kad takvih operacija ima više, pripadne greške obično označavamo nekim indeksom u  $\varepsilon$ .

Na primjer, ako je  $\circ$  **zbrajanje** (+), onda je

$$\begin{aligned} fl(x + y) &= (1 + \varepsilon_{x+y}) (x + y) \\ &= [(1 + \varepsilon_{x+y}) x] + [(1 + \varepsilon_{x+y}) y], \end{aligned}$$

uz  $|\varepsilon_{x+y}| \leq u$ , ako su  $x$ ,  $y$  i  $x + y$  u prikazivom rasponu.

Potpuno ista stvar vrijedi i za **oduzimanje**.

Kod **množenja** i **dijeljenja** možemo birati kojem ulaznom podatku ćemo “zalijepiti” faktor  $(1 + \varepsilon)$ .

# Širenje grešaka zaokruživanja (nastavak)

Za množenje možemo pisati

$$\begin{aligned} fl(x * y) &= (1 + \varepsilon_{x*y}) (x * y) \\ &= [(1 + \varepsilon_{x*y}) x] * y = x * [(1 + \varepsilon_{x*y}) y], \end{aligned}$$

a za dijeljenje

$$\begin{aligned} fl(x/y) &= (1 + \varepsilon_{x/y}) (x/y) \\ &= [(1 + \varepsilon_{x/y}) x] / y = x / [y / (1 + \varepsilon_{x/y})]. \end{aligned}$$

Postoje i druge varijante. Na primjer, da svakom operandu “zalijepimo”  $\sqrt{1 + \varepsilon}$  (odnosno  $1/\sqrt{1 + \varepsilon}$ ), ali to nije naročito važno. Bitno je samo da je izračunati rezultat egzaktan za malo perturbirane podatke.

## Širenje grešaka (bilo kojih)

Zasad **nije vidljivo** koja je točno **korist** od ove interpretacije. Stvar se **bolje** vidi tek kad imamo **više operacija zaredom**.

Međutim, ova ideja s “**malo pogrešnim podacima**” je baš ono što nam **treba** za **analizu širenja grešaka** (i to bez obzira na uzrok grešaka), čim se sjetimo da

- rezultati **ranijih** operacija s nekom **greškom ulaze** u nove operacije.

Naime, **uzroka** grešaka može biti mnogo, ovisno o tome što računamo. Od grešaka **modela** i **metode**, preko grešaka **mjerenja** (u ulaznim podacima), do grešaka **zaokruživanja**.

# Širenje grešaka u aritmetici računala

U aritmetici **računala** postupamo na potpuno **isti** način. Samo treba zgodno iskoristiti onu raniju interpretaciju da je

- **izračunati** (ili “zaokruženi”) rezultat **jednak egzaktnom**, ali za **malo perturbirane** podatke (u relativnom smislu).

A širenje grešaka u egzaktnoj aritmetici znamo.

Svaka aritmetička operacija u računalu samo **povećava** perturbaciju ulaznih podataka za **jedan faktor** oblika  $(1 + \varepsilon)$ , uz ocjenu  $|\varepsilon| \leq u$ , ovisno o tome kojim operandima “zalijepimo” taj faktor.

# Širenje grešaka u aritmetici računala

Na pr., uzmimo da računamo zbroj  $x + y$ , gdje su  $x$  i  $y$  spremljeni u računalu. Znamo da za izračunati rezultat vrijedi

$$\begin{aligned} fl(x + y) &= (1 + \varepsilon_{x+y}) (x + y) \\ &= [(1 + \varepsilon_{x+y}) x] + [(1 + \varepsilon_{x+y}) y], \end{aligned}$$

uz  $|\varepsilon_{x+y}| \leq u$ , ako su  $x$ ,  $y$  i  $x + y$  u prikazivom rasponu.

No,  $x$  i  $y$  već imaju neke greške obzirom na prave egzaktne vrijednosti. I to treba uvrstiti u ovu formulu.

# Natuknice o analizi grešaka

- Bilo koja pojedina **operacija**, **nakon**, recimo prvog čitanja, ili ranijih operacija — tad ide kao ovo gore, ali svagdje imam **dodatni faktor** oblika  $(1 + \varepsilon)$  iz ocjene greške (ako je sve prikazivo u dozvoljenom rasponu).
- Dakle, faktori se “kote” i postoje razne oznake za to, da se lakše čita.
  - oznake:  $\varepsilon$  (s indeksima) — obično za “jedinične” greške (ispod  $u$ ),
  - neko drugo slovo (na pr.  $\eta$ ) s indeksima, za ostale (relativne) greške u analizi.



# Natuknice o analizi grešaka (nastavak)

Pojmovi greške **unaprijed** (forward error), i greške **unatrag** (katkad se zove i **obratna** — backward error).

- Gledam algoritam kao preslikavanje: ulaz (domena) u izlaz (kodomena).
- Zanima me greška u rezultatu (kodomenu) — forward.
- To katkad ide, ali je, uglavnom, teško (ili daje loše ocjene). (Primjer - Zlatko, za normu u  $\mathbb{R}^2$ , i još dodaj scaling).
- Lakše je “unatrag” — ista interpretacija kao i za pojedine operacije.

Uočiti da se **akumulacija** faktora  $(1 + \varepsilon)$  **prirodno** radi **unatrag** — inače moram znati grešku za operande (a to je rezultat **unaprijed**).

# Natuknice o analizi grešaka (nastavak)

Postupak “unatrag”:

- Krajnji rezultati algoritma su “egzaktni” ali na ponešto perturbiranim podacima (i napravi se ocjena tih perturbacija u **domeni**).
- A zatim ide matematička **teorija perturbacije**, koja daje ocjene u **kodomeni** (izvod ide za egzaktni račun, pa vrijede normalna pravila).

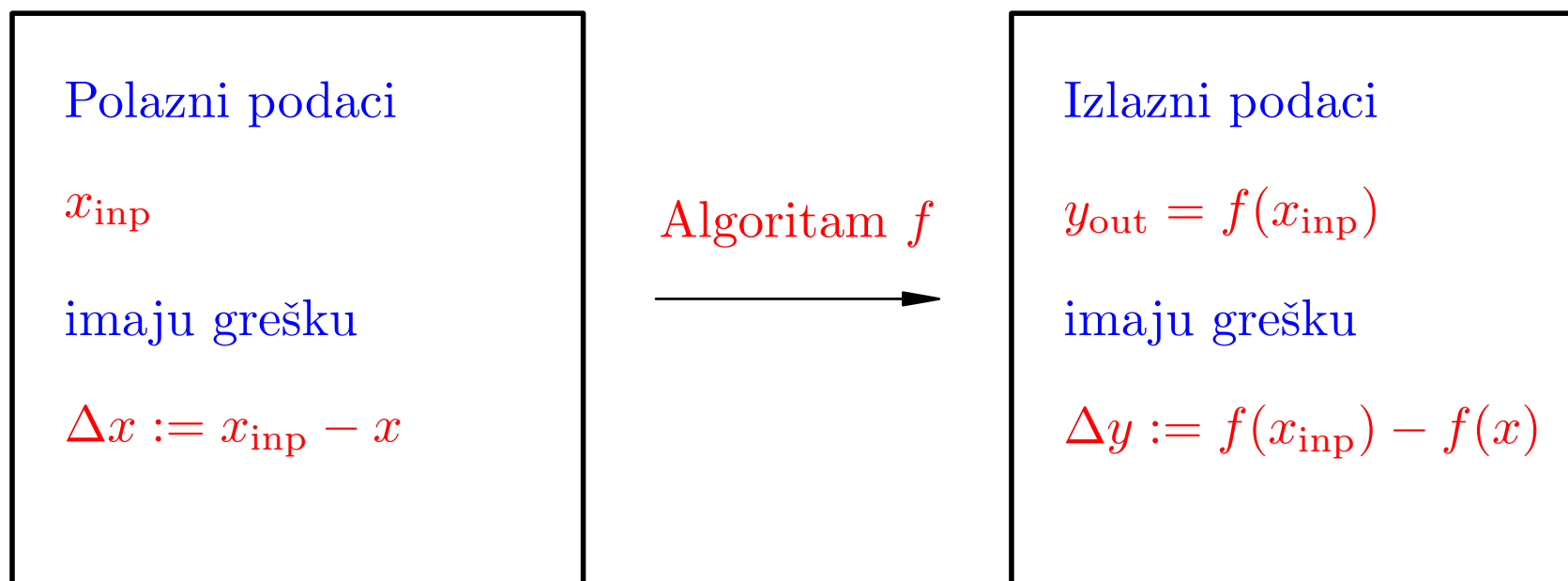
I sad imam pojmove: **stabilno** i **nestabilno** računanje (algoritam) (“prigušivač” ili “pojačalo” grešaka).

- Slikice (skripta NA, Higham).
- Primjeri nestabilnosti — **uklonjivi** i **NEuklonjivi**.

# Norme i uvjetovanost

## Greška na ulazu – što na izlazu?

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



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

# Kako mjeriti grešku?

Kad  $x_{\text{inp}}$  i  $f(x_{\text{inp}})$  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  $\| \cdot \| : V \rightarrow \mathbb{R}$

koja zadovoljava sljedeća svojstva:

1.  $\|x\| \geq 0$ ,  $\forall x \in V$ , a jednakost vrijedi ako i samo ako je  $x = 0$ ,
2.  $\|\alpha x\| = |\alpha| \|x\|$ ,  $\forall \alpha \in F$ ,  $\forall x \in V$ ,
3.  $\|x + y\| \leq \|x\| + \|y\|$ ,  $\forall x, y \in V$   
(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  $\ell_1$  norma  $\|x\|_1 = \sum_{i=1}^n |x_i|$ ,

2. **2-norma** ili  $\ell_2$  norma ili **euklidska norma**

$$\|x\|_2 = (x^* x)^{1/2} = \sqrt{\sum_{i=1}^n |x_i|^2},$$

3.  **$\infty$ -norma** ili  $\ell_\infty$  norma  $\|x\|_\infty = \max_{i=1, \dots, n} |x_i|$ .

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  $\|f\|_1 = \int_a^b |f(t)| dt,$

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

3.  $L_\infty$  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  $\|\cdot\|_a$  i  $\|\cdot\|_b$  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  $\| \cdot \| : \mathbb{C}^{m \times n} \rightarrow \mathbb{R}$  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:

- Maticu  $A$  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

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

- **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”

$$\|A\|_1 = \max_{j=1,\dots,n} \sum_{i=1}^m |a_{ij}|,$$

2. matrična **2-norma**, spektralna norma

$$\|A\|_2 = (\rho(A^*A))^{1/2} = \sigma_{\max}(A),$$

3. matrična  **$\infty$ -norma**, “maksimalna retčana norma”

$$\|A\|_\infty = \max_{i=1,\dots,m} \sum_{j=1}^n |a_{ij}|.$$

$\rho$  je **spektralni radijus** matrice (po aps. vrijednosti maksimalna svojstvena vrijednost), a  $\sigma$  **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

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

- 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:**

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

● **Relativna uvjetovanost:**

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

Problem je **dobro uvjetovan** ako je

●  $\kappa_{\text{rel}}$  što je moguće **manji**, za  $\delta_x \rightarrow 0$ .

## Landauov simbol — red veličine

Neka su  $g, h : \mathbb{R}^n \rightarrow \mathbb{R}^m$  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

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

onda kažemo da je

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

i to pišemo ovako

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



## Landauov simbol (nastavak)

Primjer. Za  $m = n = 1$  je

$$\sin x = \mathcal{O}(x) \quad (x \rightarrow a), \quad \text{za sve } a \in \mathbb{R},$$

$$x^2 + 3x = \mathcal{O}(x) \quad (x \rightarrow 0),$$

$$x^2 - x - 6 = \mathcal{O}(x - 3), \quad (x \rightarrow 3).$$

# Uvjetovanost i Taylorov teorem

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

- Promatramo ponašanje  $f$  za **male** perturbacije  $\Delta x$  u okolini točke  $x$ . Neka je  $\Delta y$  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  $\Delta x$ , **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  $\Delta x$ ).

- 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

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

## Uvjetovanost – primjer

Primjer. **Relativna** uvjetovanost funkcije

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

je

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

što je **veliko** za  $x \approx 1$ .

Pitanje: **Apsolutna** uvjetovanost?

# Primjer uvjetovanosti problema

# Rekurzija za integral

Ispitajmo **uvjetovanost** problema računanja integrala

$$I_n = \int_0^1 \frac{t^n}{t+5} dt$$

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

$$I_k = \int_0^1 t^{k-1} dt - 5I_{k-1} = \frac{1}{k} - 5I_{k-1}, \quad k = 1, 2, \dots, n.$$

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

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

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  $\hat{I}_0$ ,
- rezultat — neka aproksimacija  $\hat{I}_n = f_n(\hat{I}_0)$ .

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

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

Iz definicije integrala:  $I_n$  monotonno padaju po  $n$ , čak

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

Zbrajanjima dobivamo sve manje i manje brojeve!

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

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

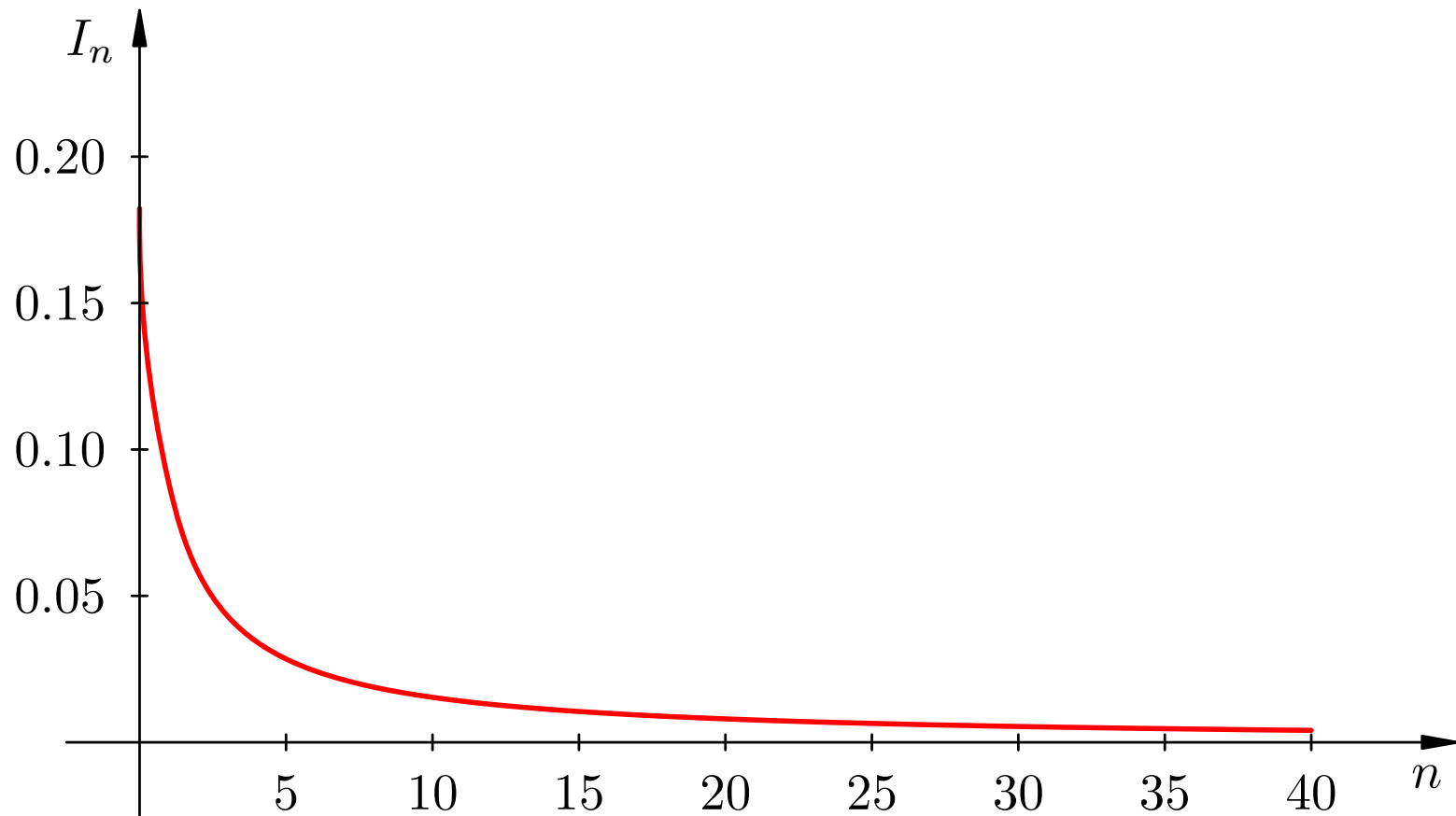
## Rekurzija unaprijed — rezultati

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

Slikice!

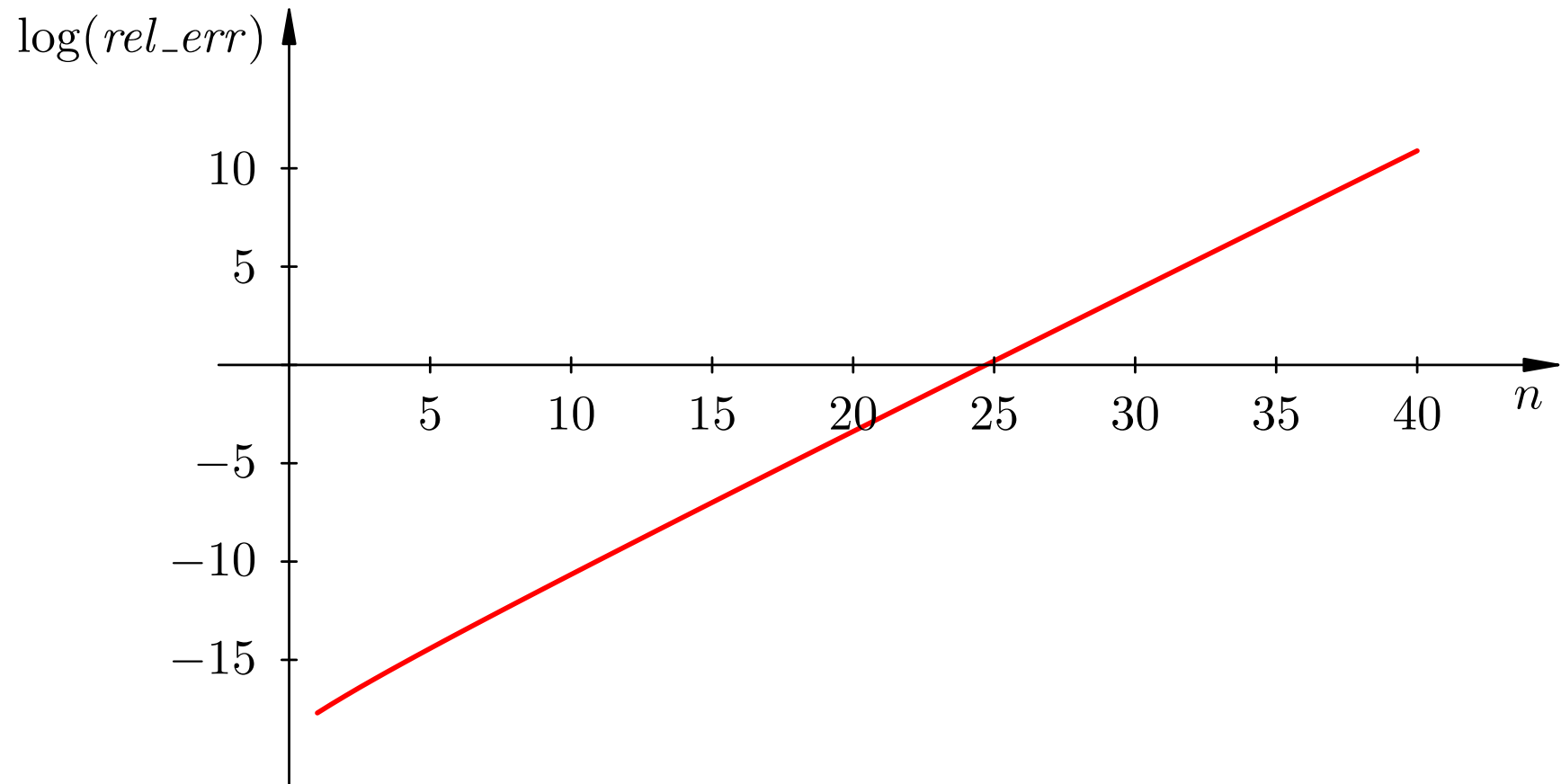
Pokaži program i rezultate!

# 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

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

**Problem:** kako izračunati početnu vrijednost  $y_\nu$ .

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

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

## Rekurzija za integral (nastavak)

Relativna uvjetovanost za  $g_n$

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

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

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

što je ispod 1, tj. greške se prigušuju.

## Rekurzija za integral (nastavak)

Ako je  $\hat{I}_\nu$  neka aproksimacija za  $I_\nu$ , onda za **relativne perturbacije** vrijedi

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

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

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



## Rekurzija za integral (nastavak)

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

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

- Povoljnim izborom  $\nu$ , ocjenu na desnoj strani možemo napraviti **po volji malom** — ispod tražene točnosti  $\varepsilon$ .

- Dovoljno uzeti  $\nu \geq n + \frac{\log(1/\varepsilon)}{\log 5}$ , i  $\hat{I}_\nu = 0$  i računamo vrijednosti

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

## Rekurzija unatrag — rezultati

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

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

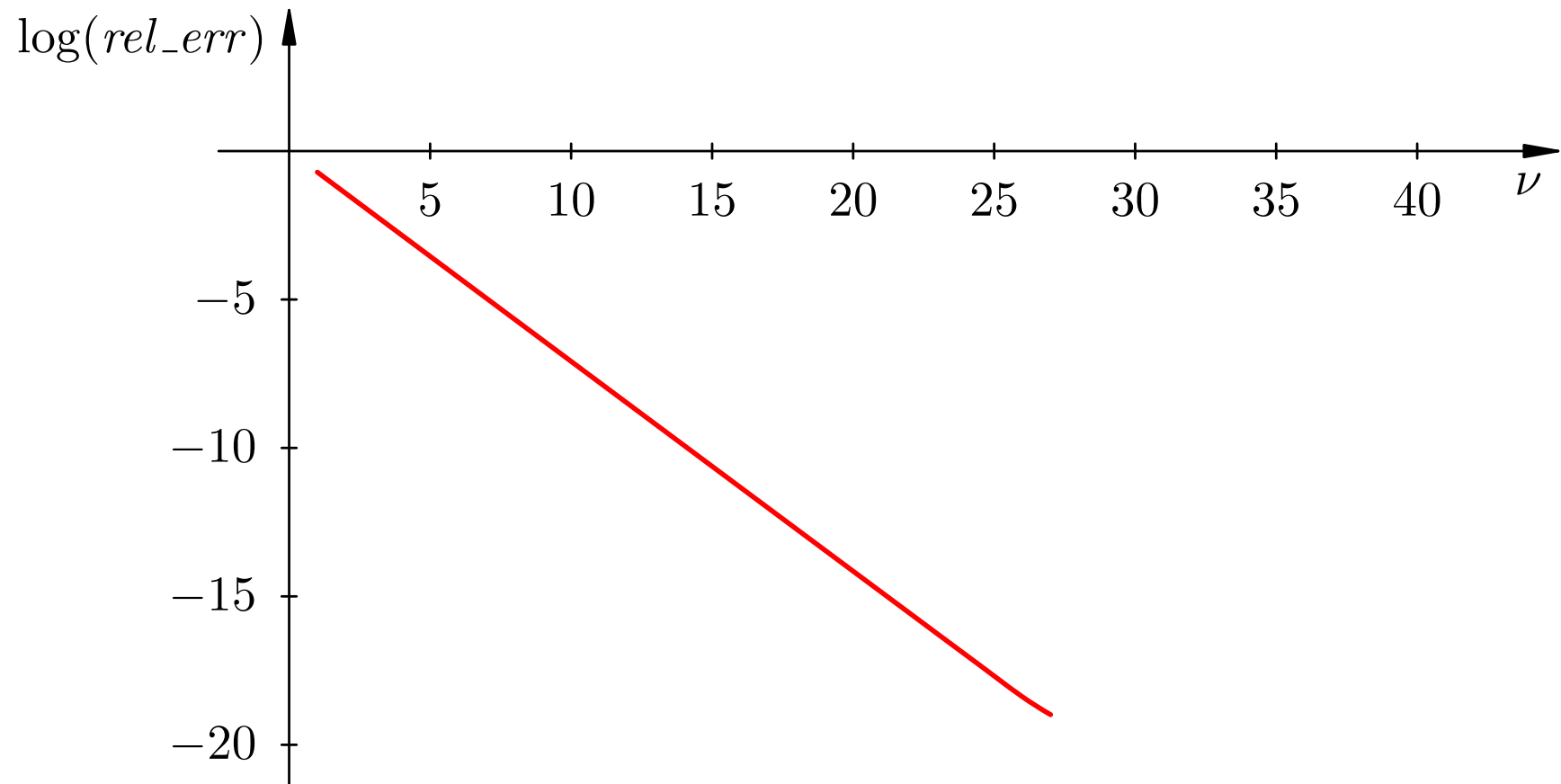
🔴 Za ovaj  $\varepsilon$  dobijemo

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

Dakle, “**silazno**” računamo **28** vrijednosti.

🔴 Stvarna početna vrijednost je  $\hat{I}_\nu = 0$ .

## Rekurzija unazad za $I_{40}$ — ovisno o startu $\nu$



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

# Primjer grešaka zaokruživanja

# Primjer rasprostiranja grešaka

Primjer. Vrijednost

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

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.

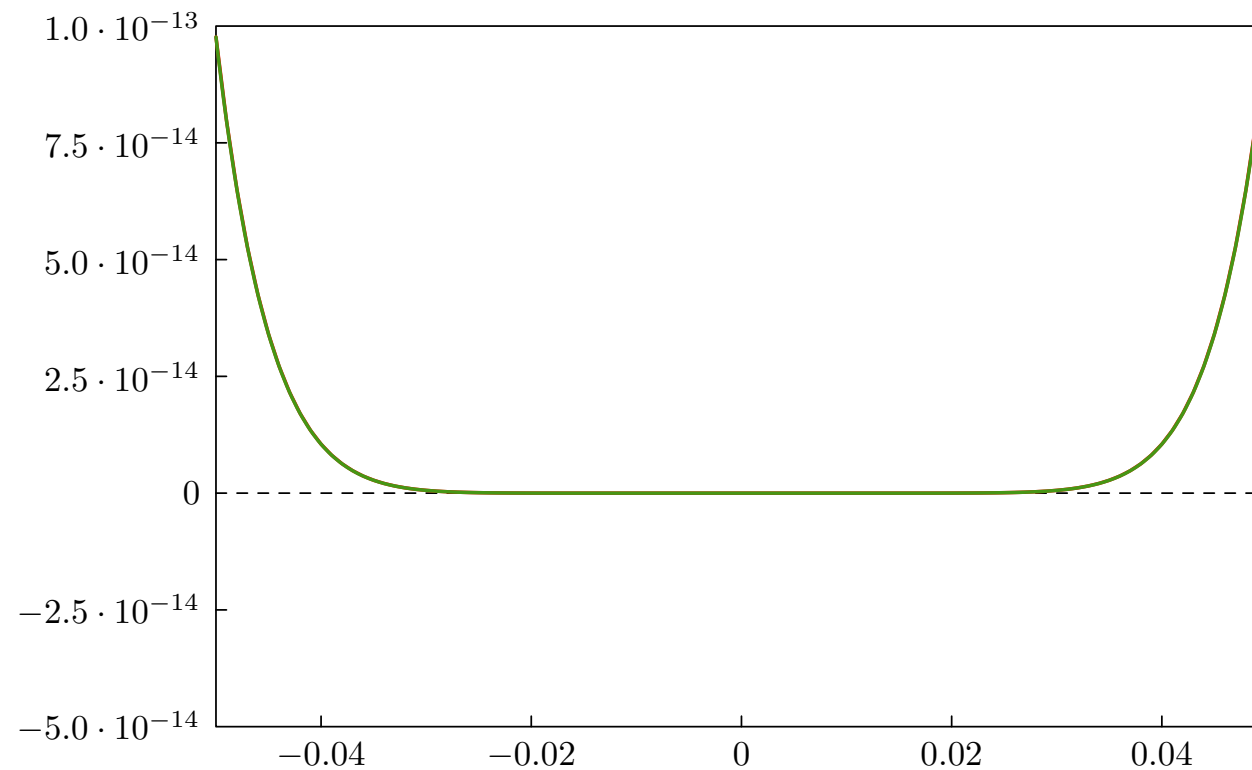
Za svaki  $n$  crtamo dvije slike grafa funkcije  $f_n$ :

- 📍 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)

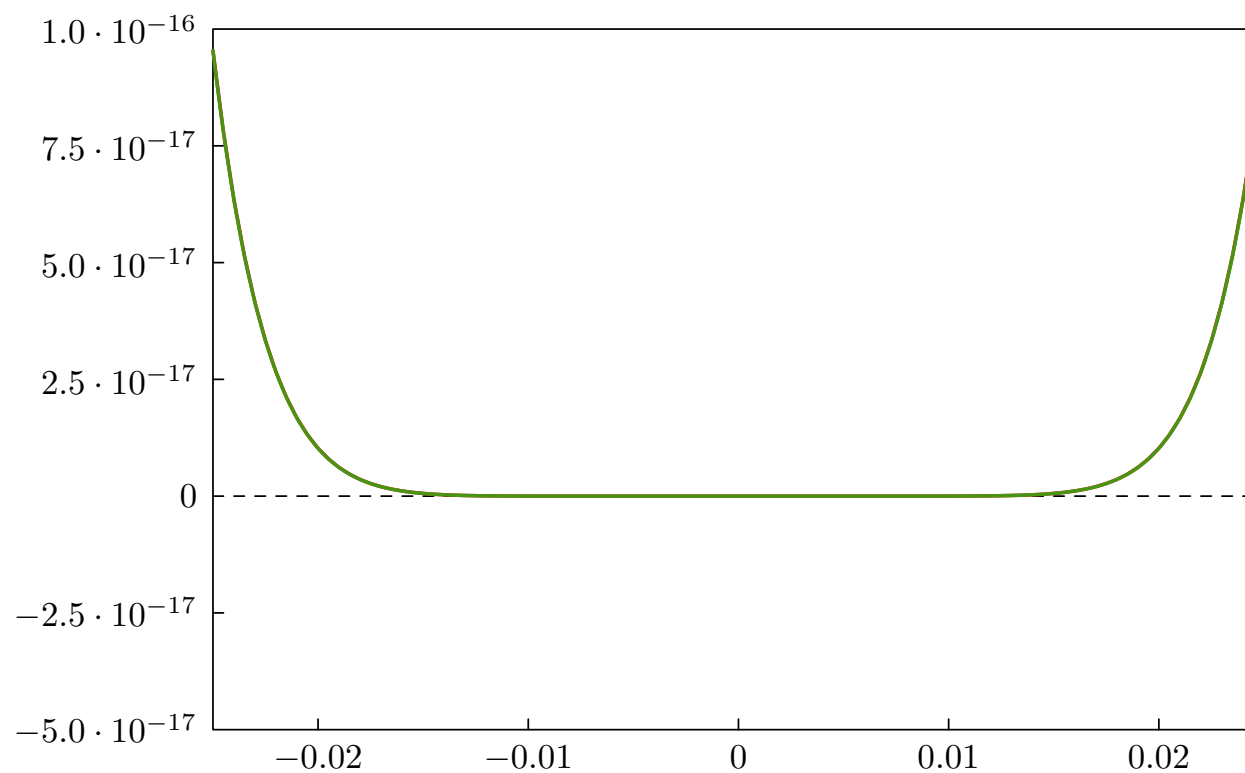
$$(x - 0)^{10} = x^{10}$$





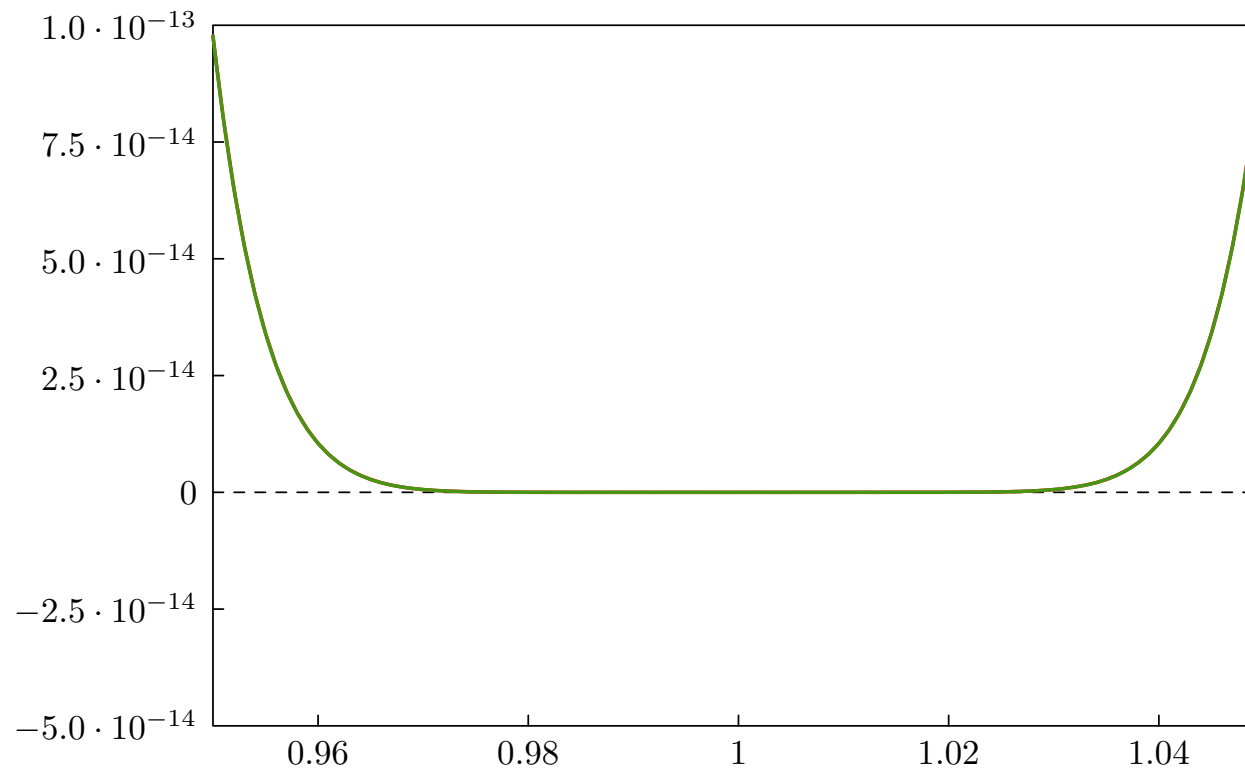
## Primjer rasprostiranja grešaka — $n = 0$ (2)

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



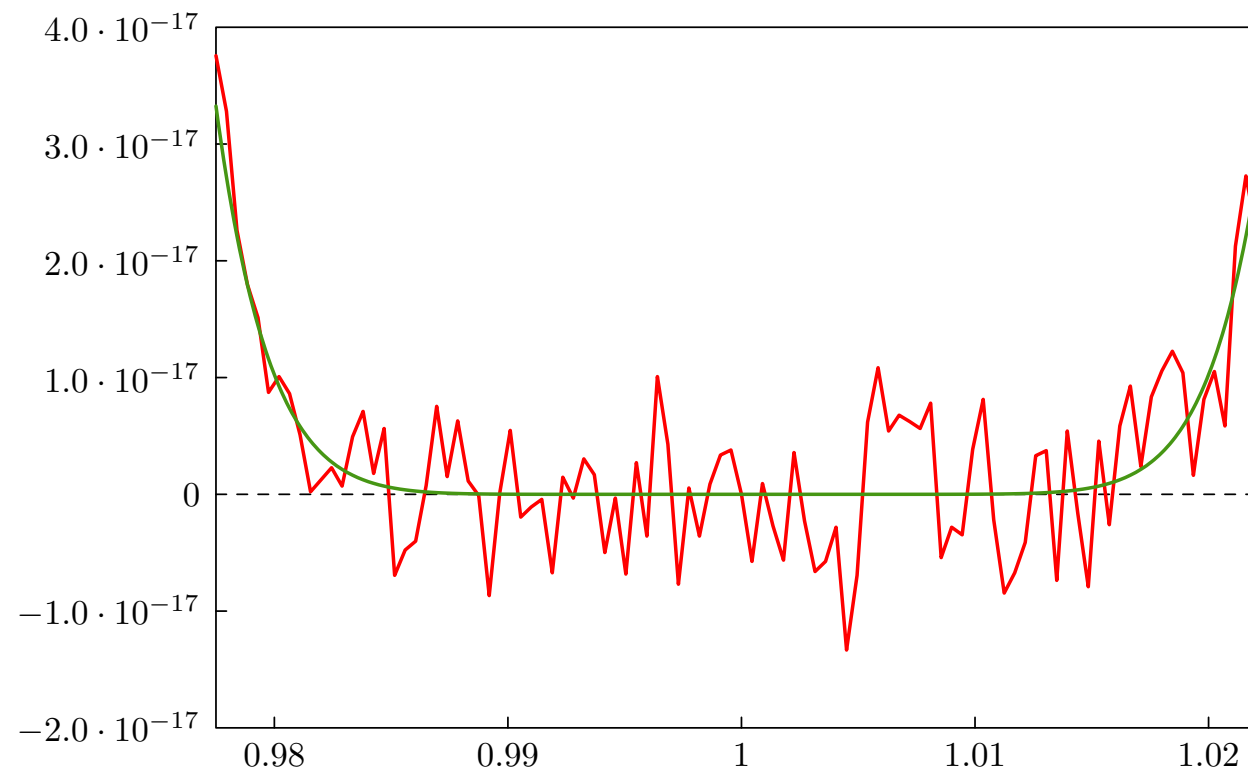
# Primjer rasprostiranja grešaka — $n = 1$ (1)

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



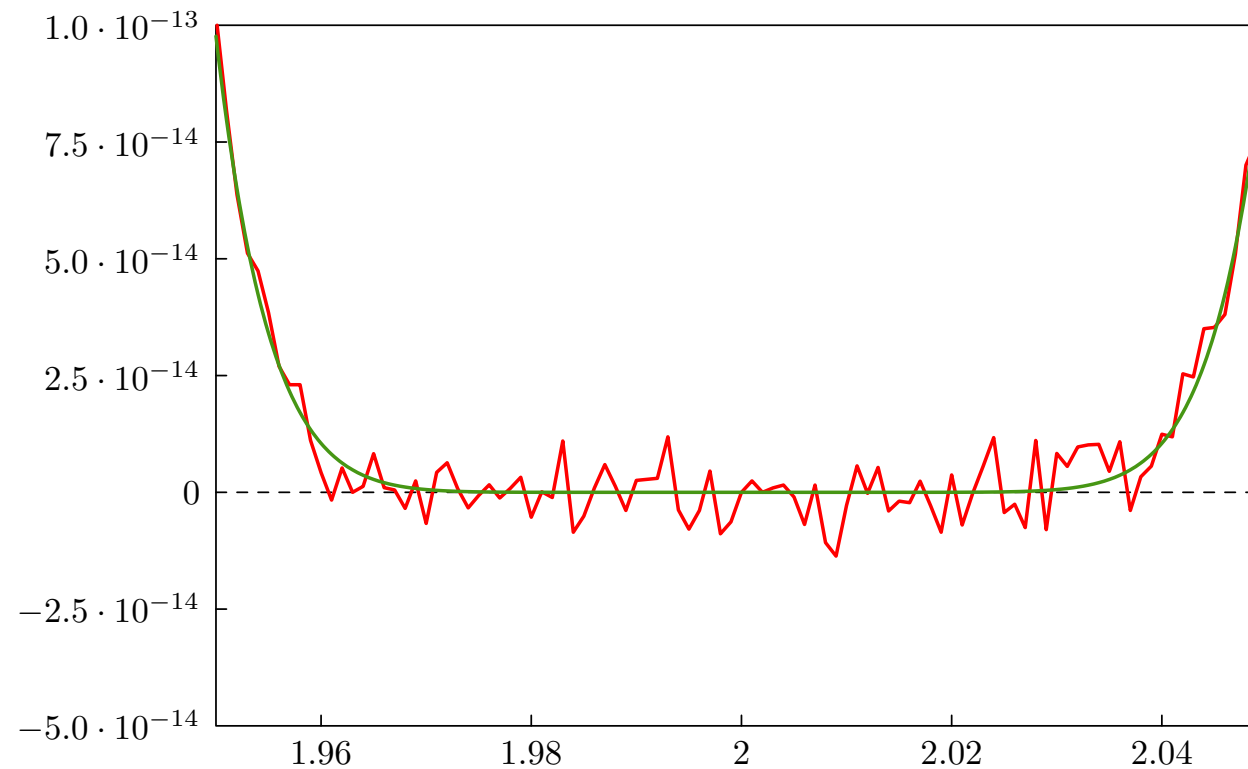
## Primjer rasprostiranja grešaka — $n = 1$ (2)

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



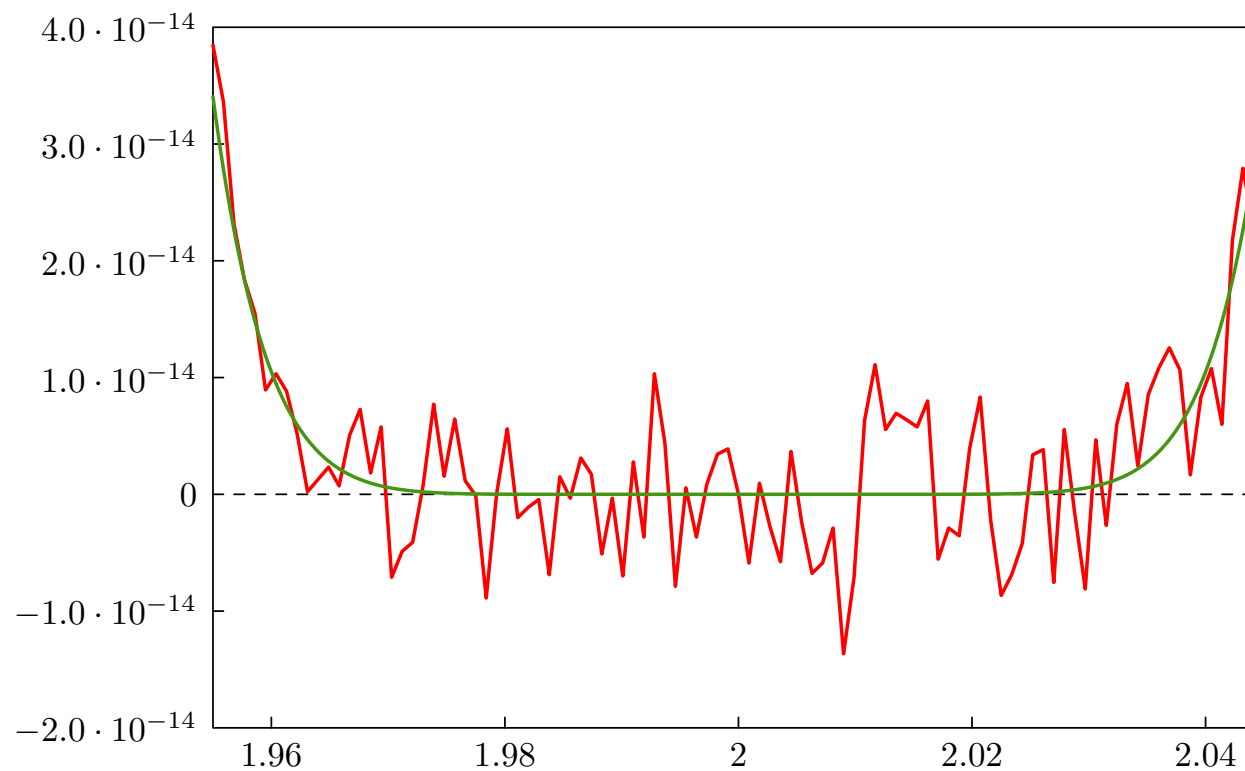
# Primjer rasprostiranja grešaka — $n = 2$ (1)

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



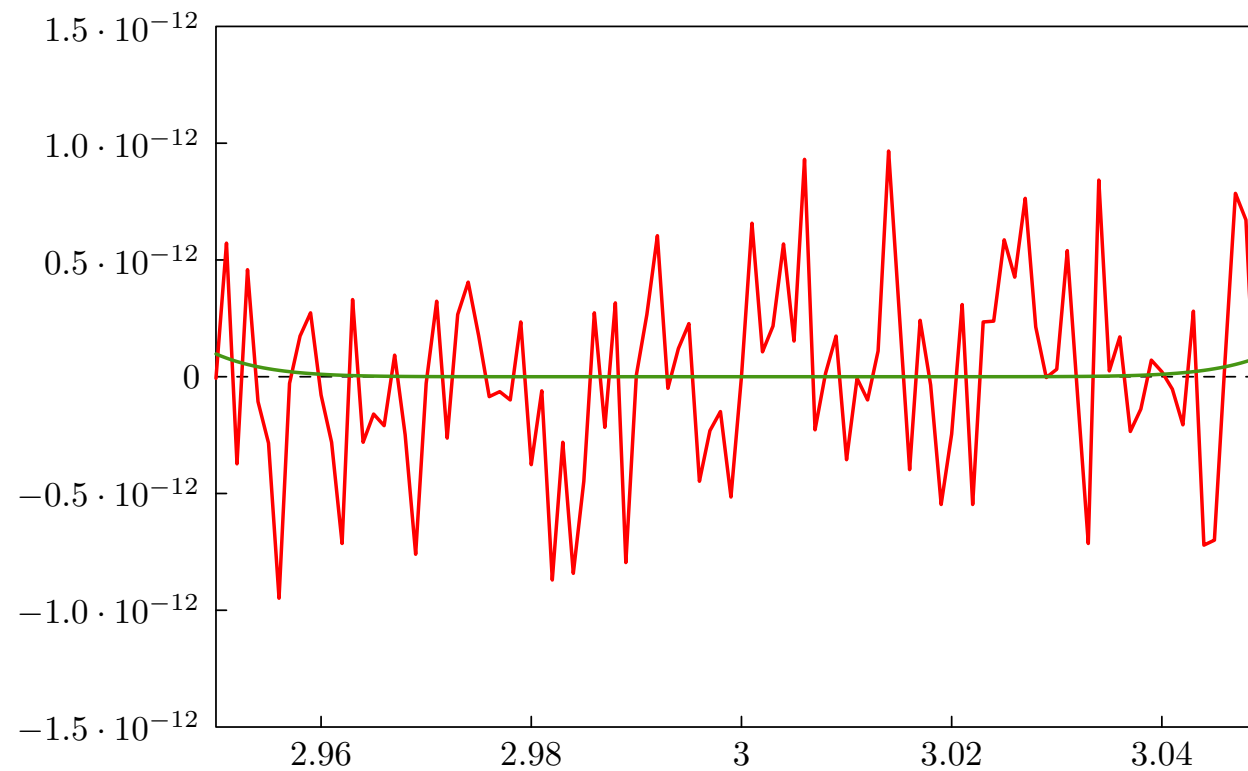
## Primjer rasprostiranja grešaka — $n = 2$ (2)

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



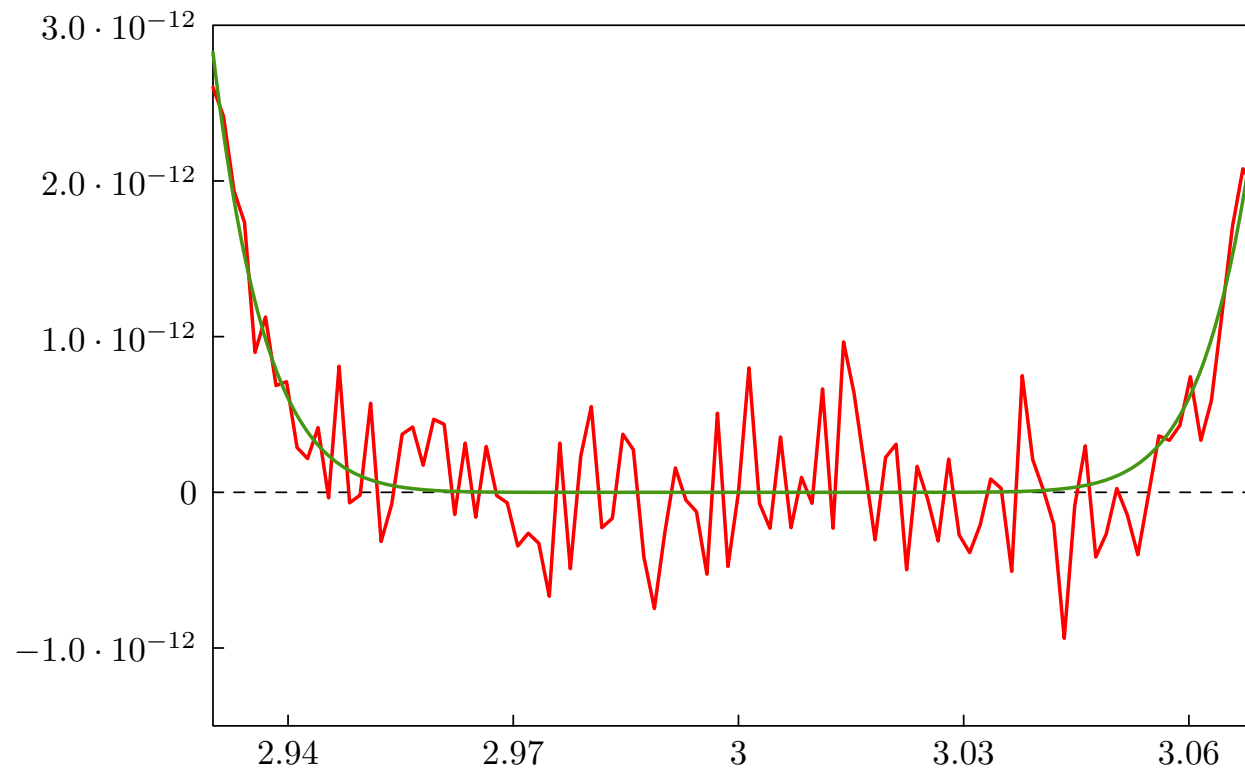
# Primjer rasprostiranja grešaka — $n = 3$ (1)

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



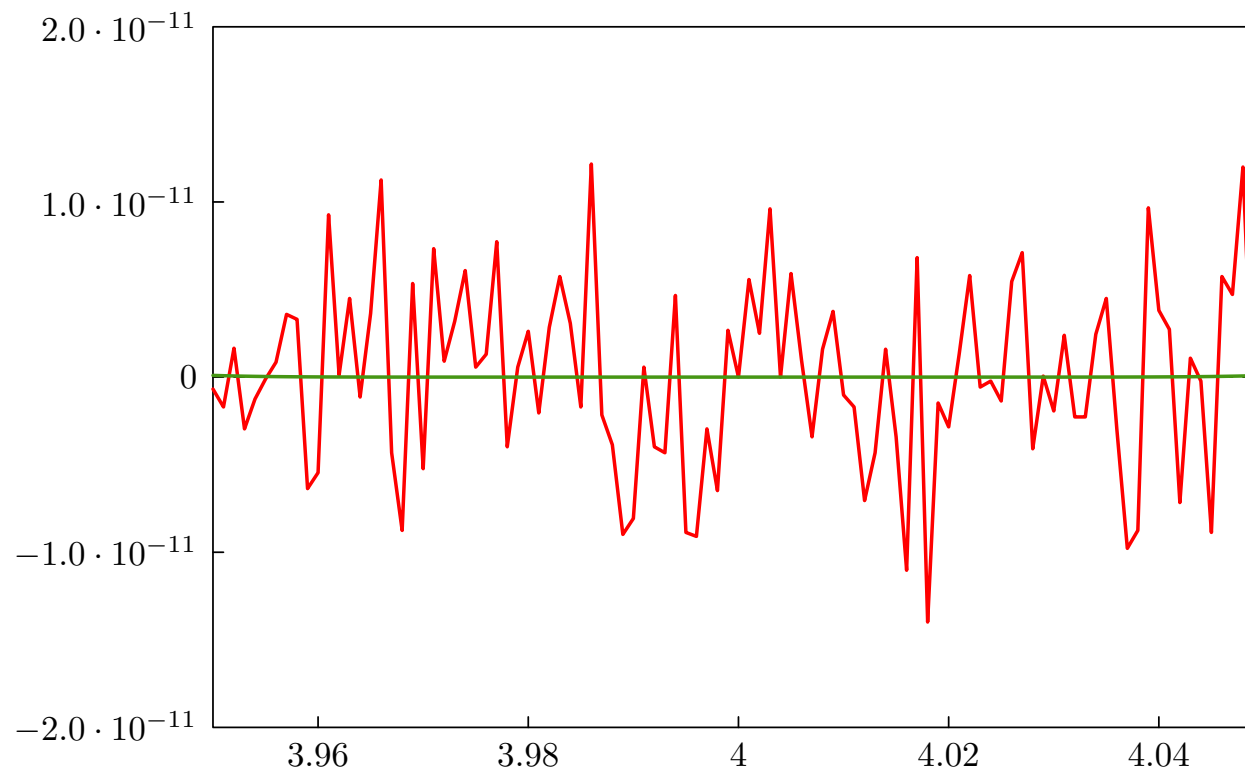
## Primjer rasprostiranja grešaka — $n = 3$ (2)

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



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

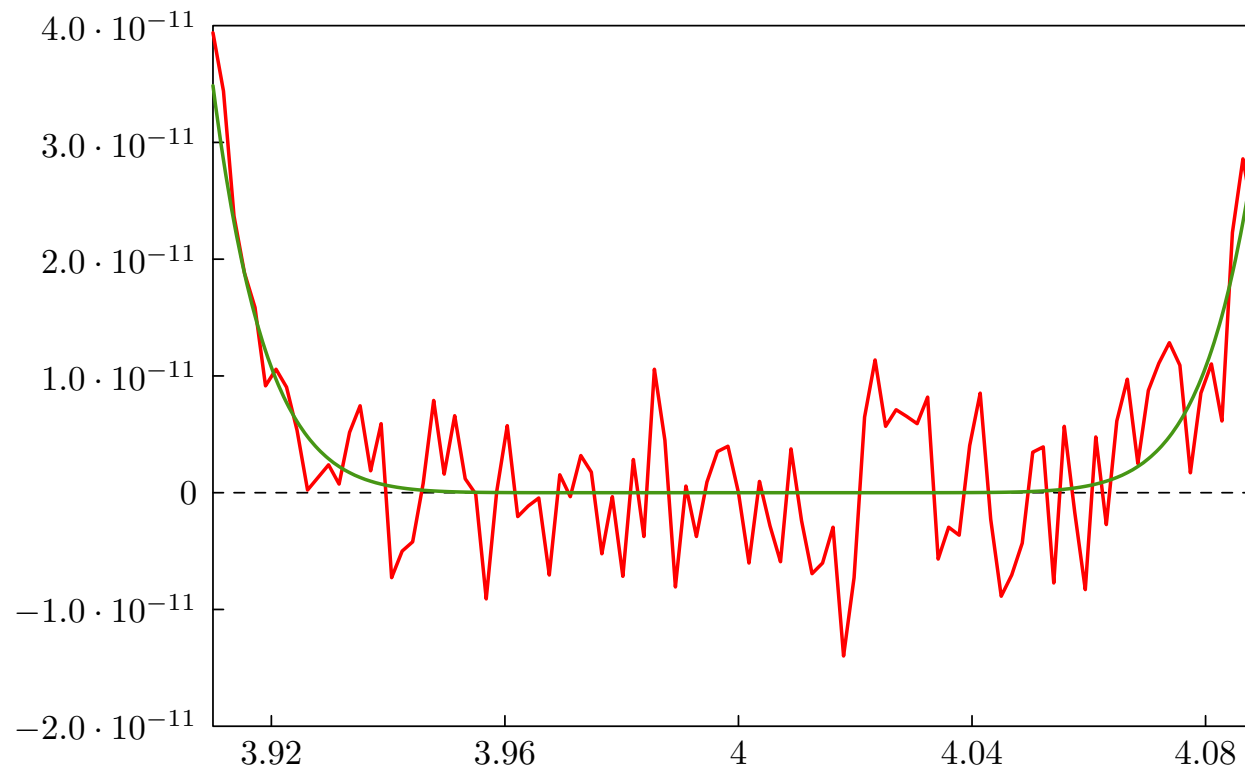
$$\begin{aligned}(x - 4)^{10} = & x^{10} - 40x^9 + 720x^8 - 7680x^7 + 53760x^6 \\ & - 258048x^5 + 860160x^4 - 1966080x^3 \\ & + 2949120x^2 - 2621440x + 1048576\end{aligned}$$





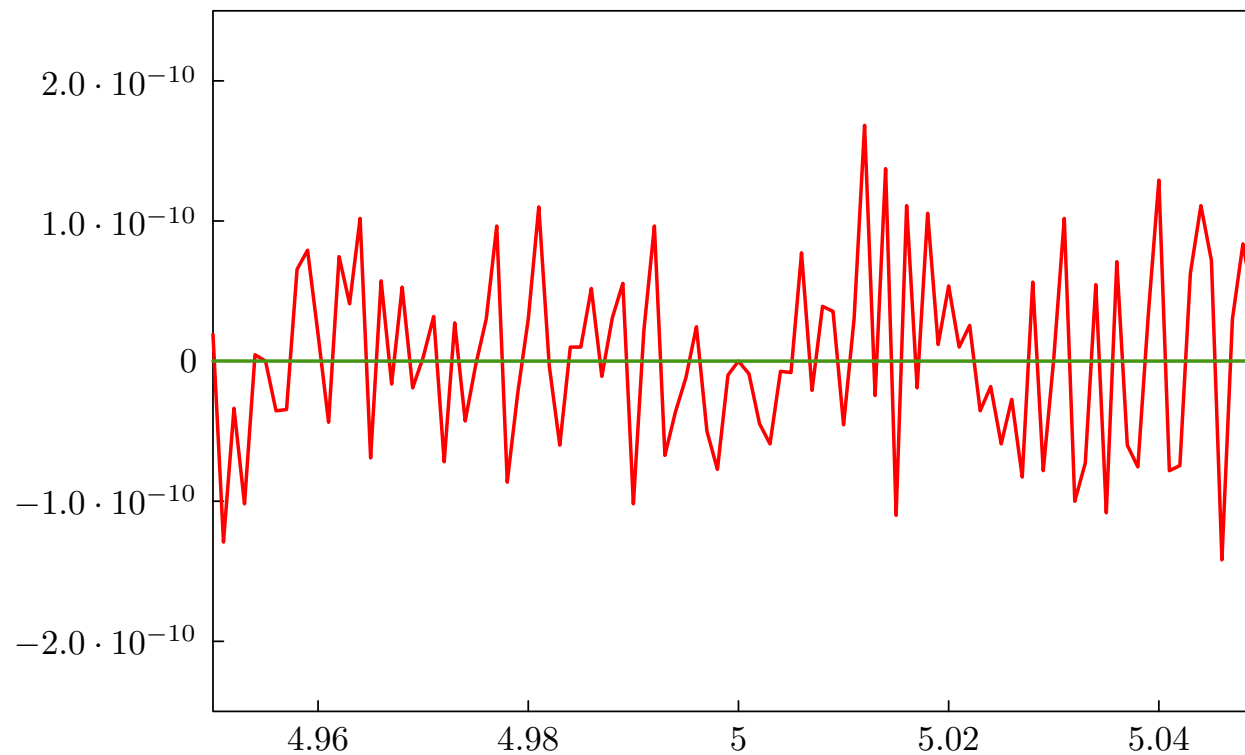
## Primjer rasprostiranja grešaka — $n = 4$ (2)

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



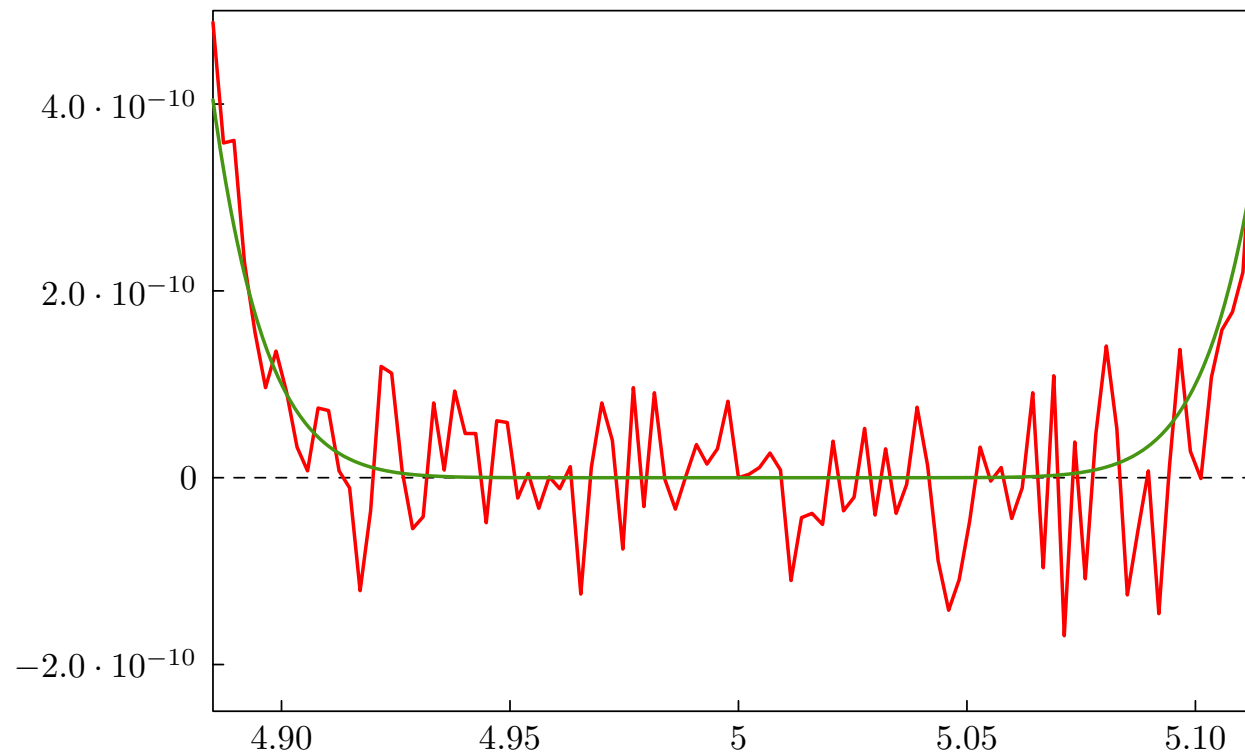
## Primjer rasprostiranja grešaka — $n = 5$ (1)

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



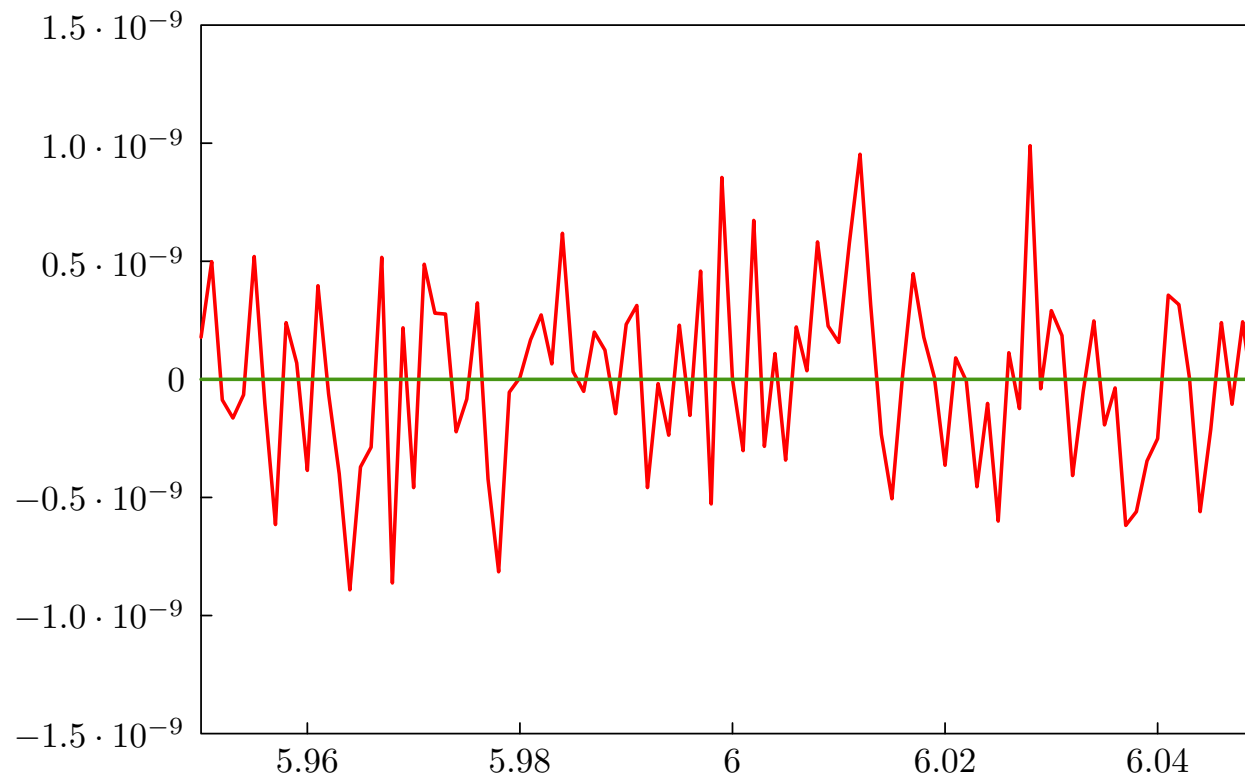
## Primjer rasprostiranja grešaka — $n = 5$ (2)

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



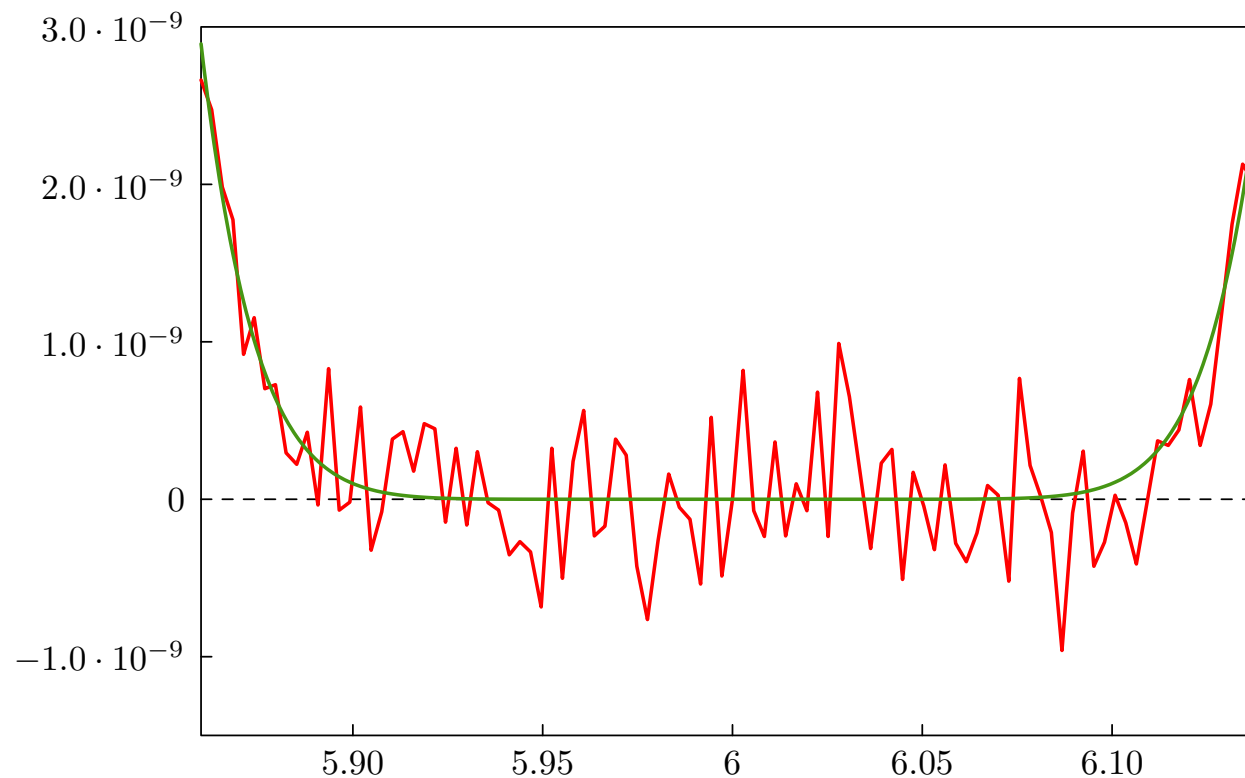
# Primjer rasprostiranja grešaka — $n = 6$ (1)

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



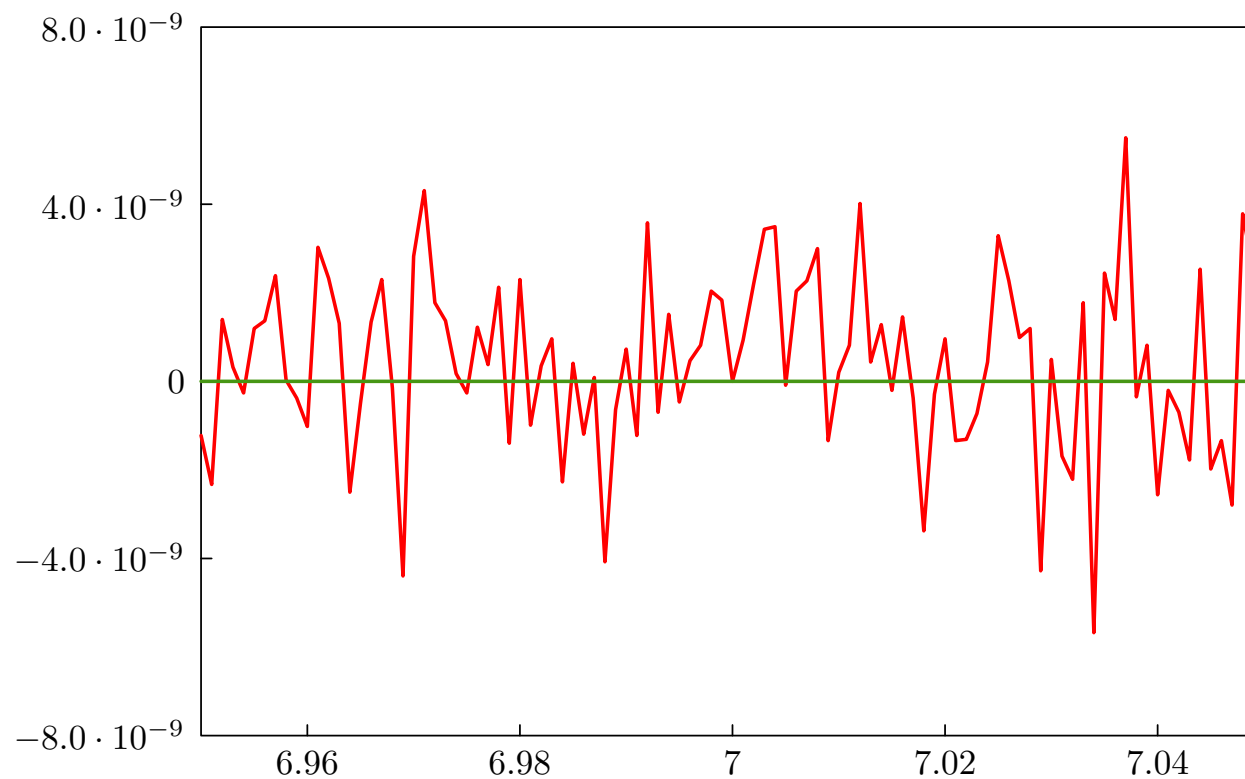
## Primjer rasprostiranja grešaka — $n = 6$ (2)

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



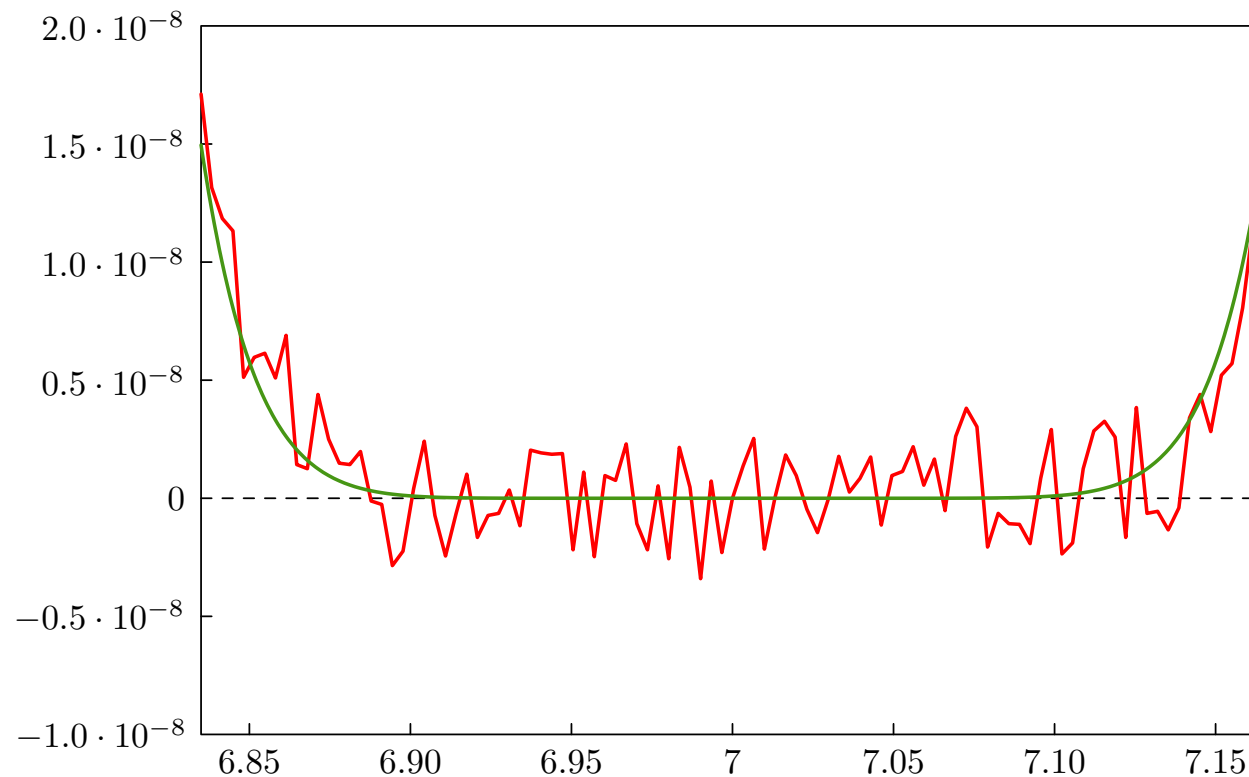
# Primjer rasprostiranja grešaka — $n = 7$ (1)

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



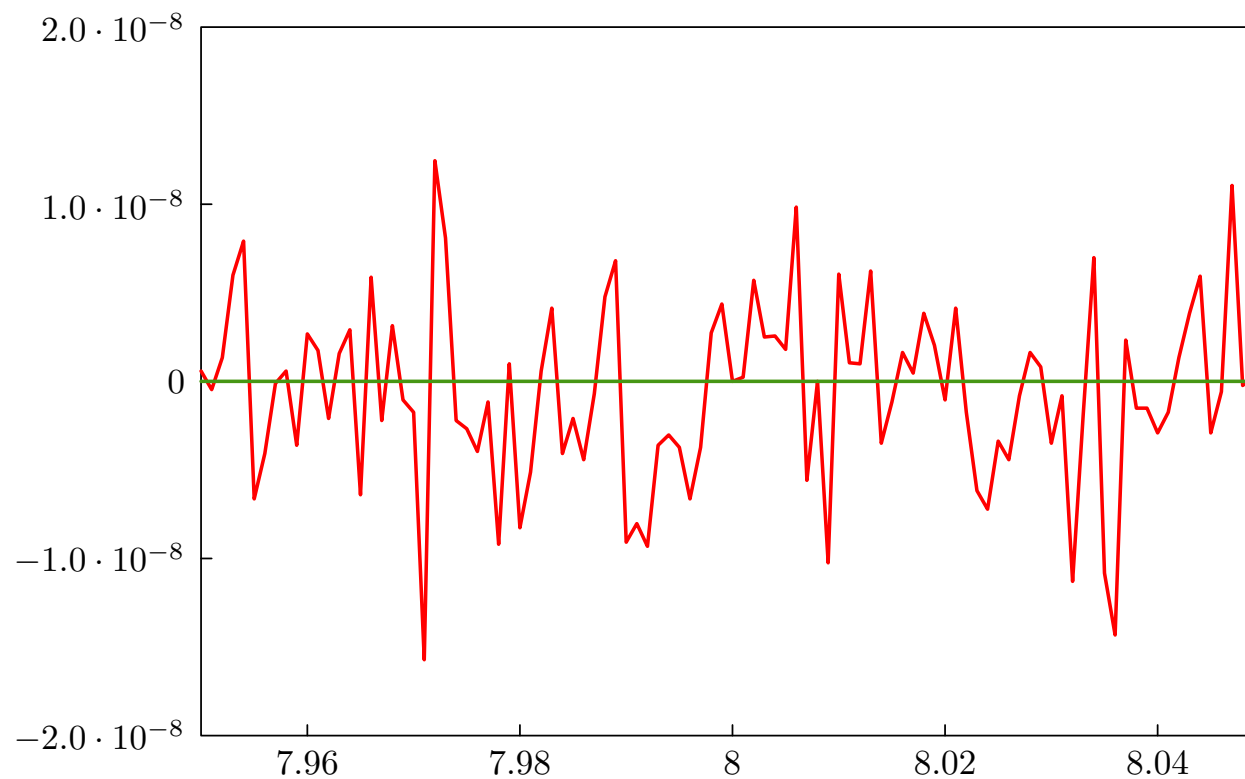
## Primjer rasprostiranja grešaka — $n = 7$ (2)

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



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

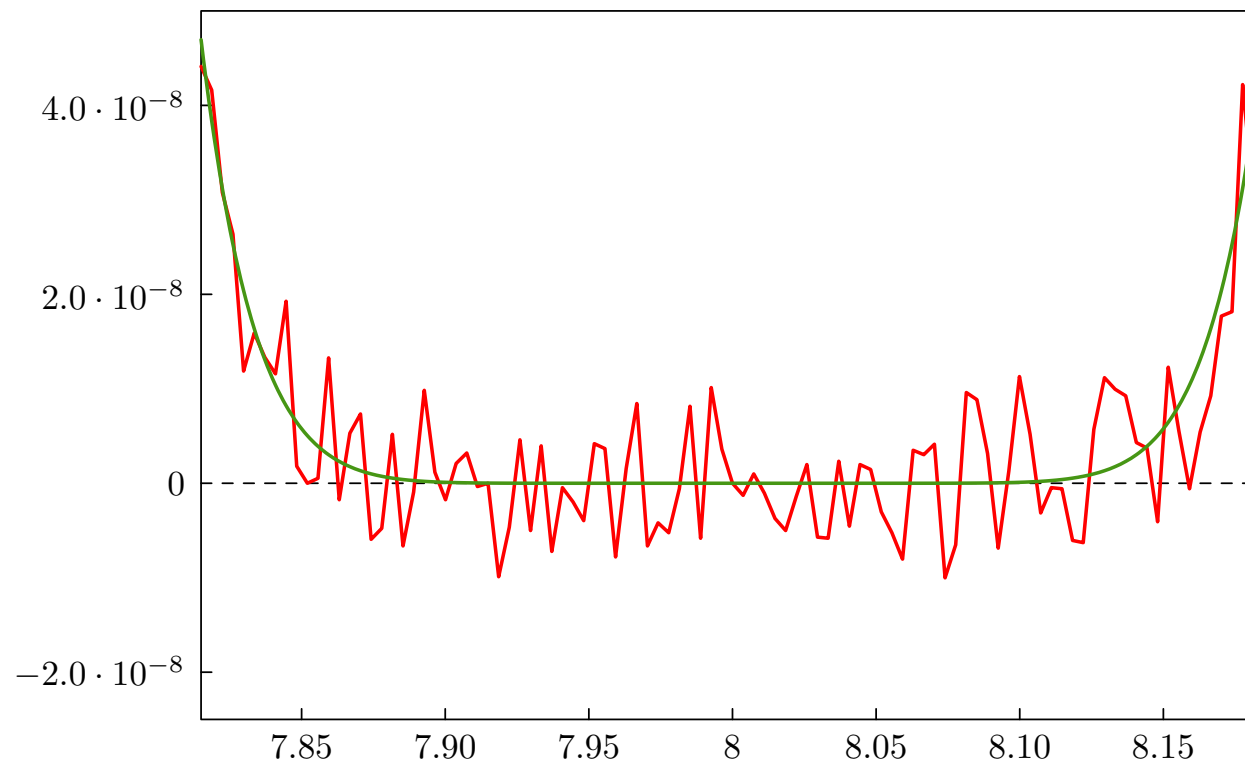
$$\begin{aligned}(x - 8)^{10} = & x^{10} - 80x^9 + 2880x^8 - 61440x^7 + 860160x^6 \\ & - 8257536x^5 + 55050240x^4 - 251658240x^3 \\ & + 754974720x^2 - 1342177280x^1 + 1073741824\end{aligned}$$





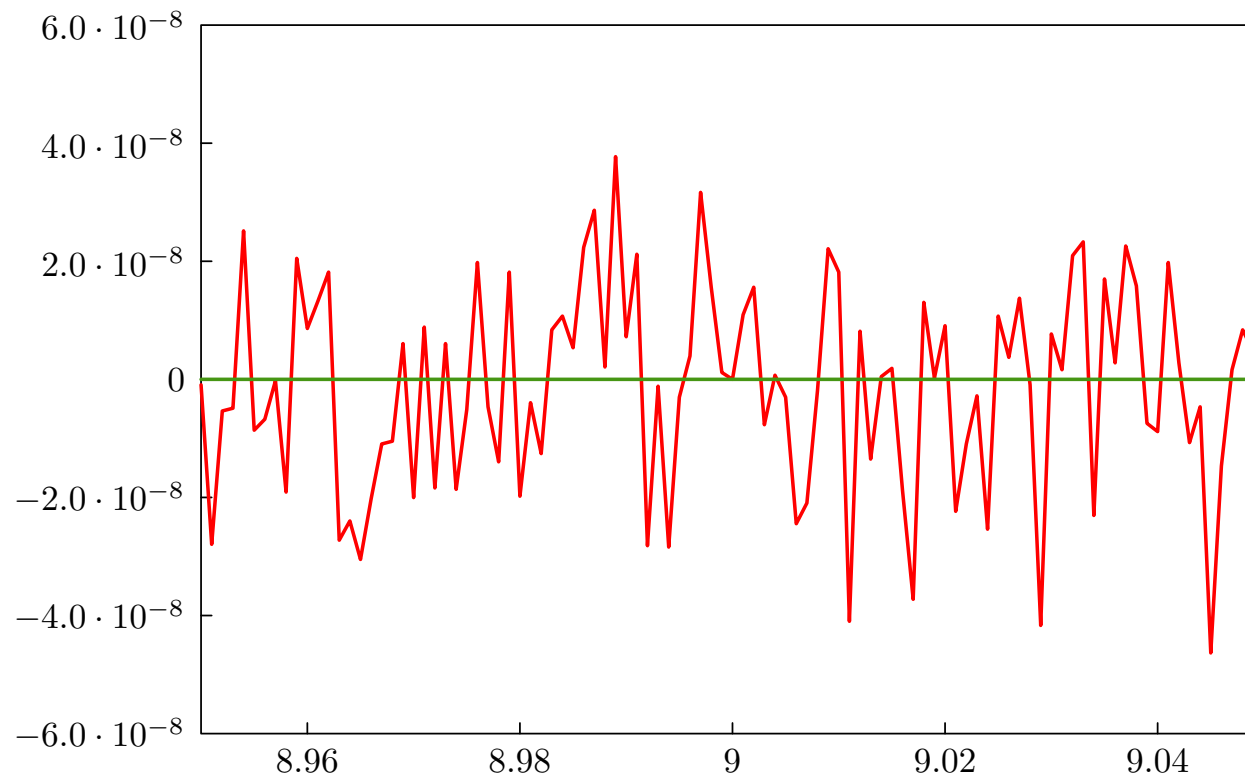
## Primjer rasprostiranja grešaka — $n = 8$ (2)

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



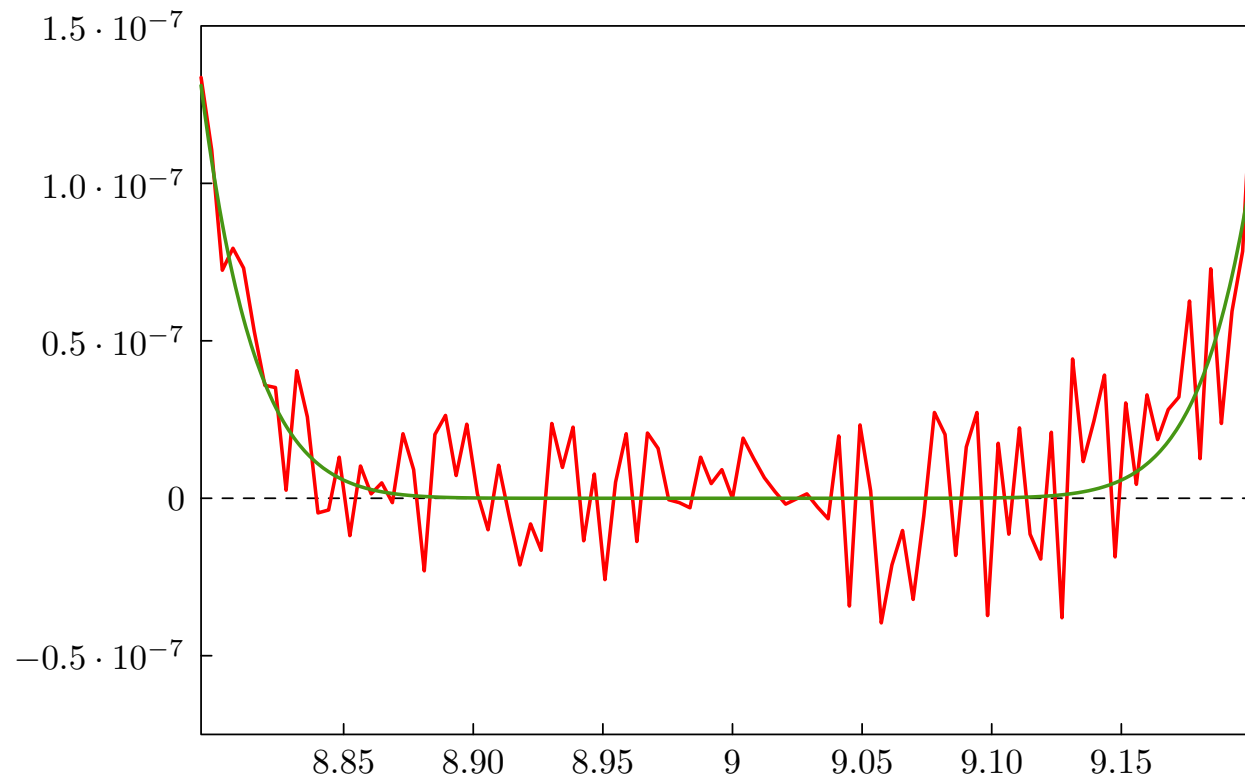
# Primjer rasprostiranja grešaka — $n = 9$ (1)

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



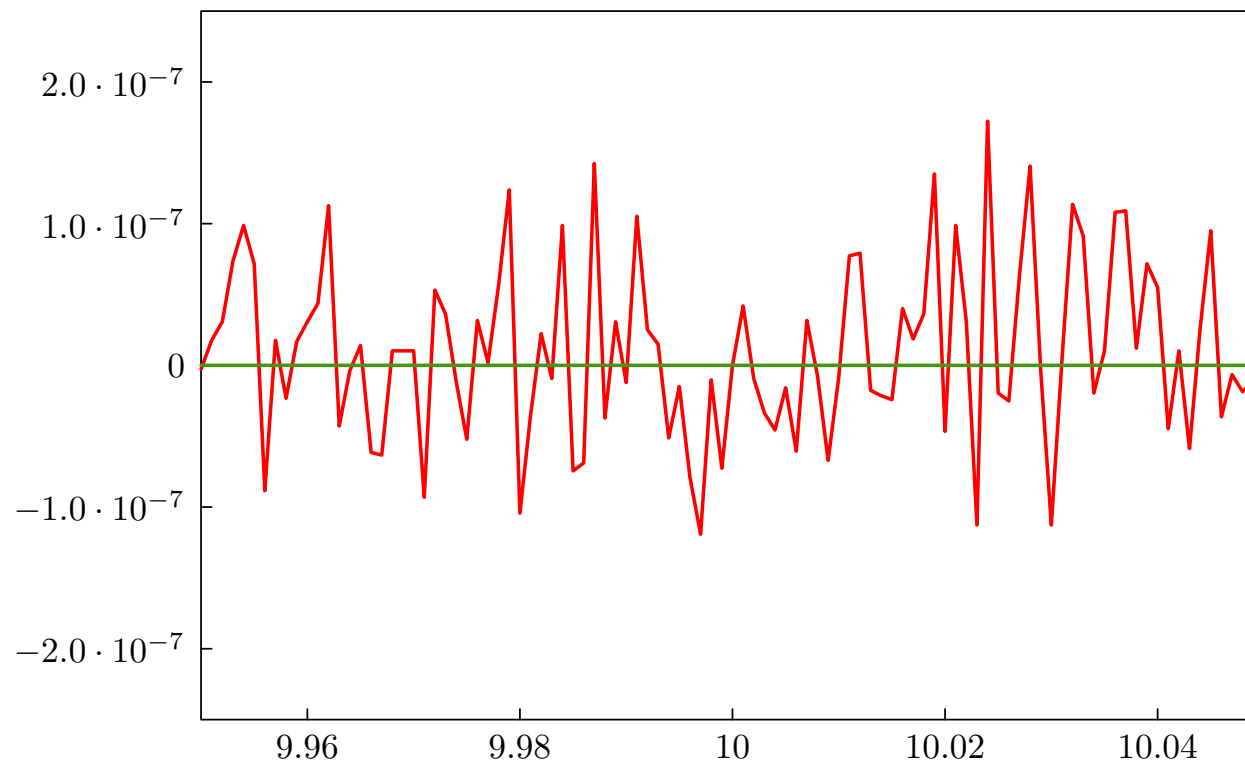
## Primjer rasprostiranja grešaka — $n = 9$ (2)

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



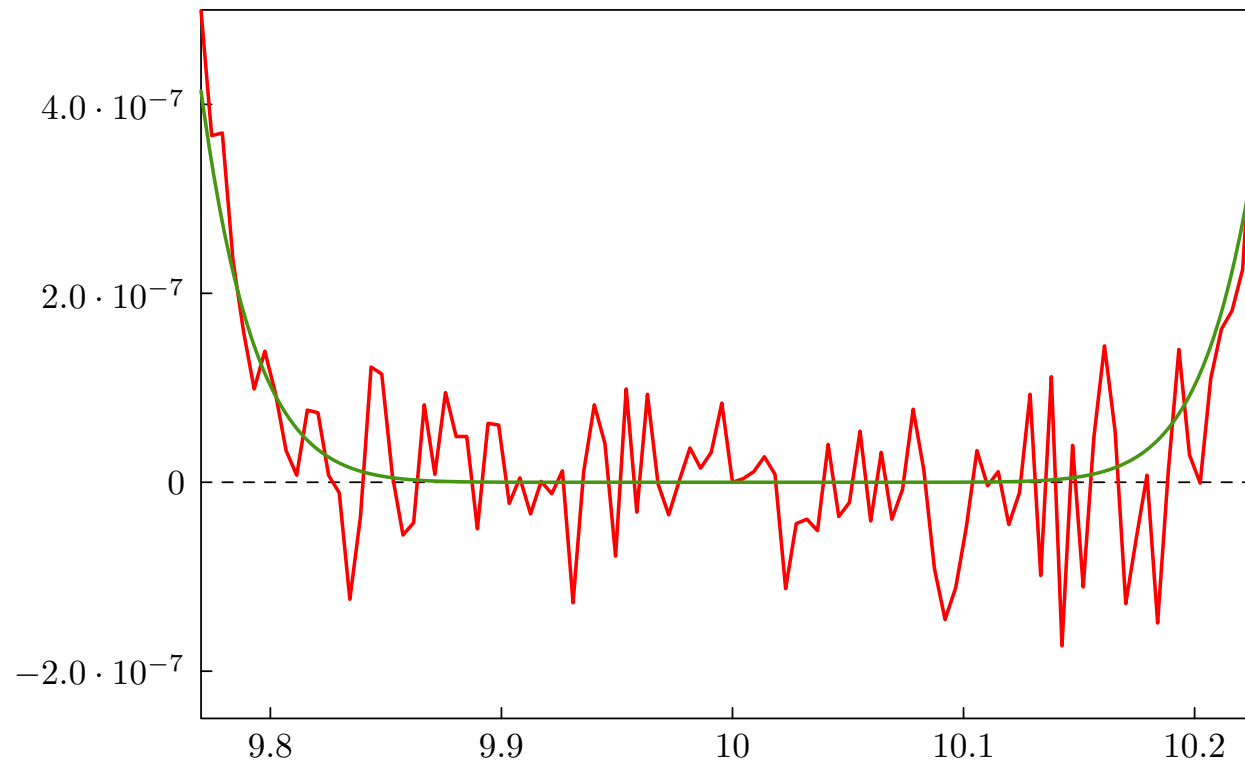
# Primjer rasprostiranja grešaka — $n = 10$ (1)

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



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

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



# Primjeri izbjegavanja kraćenja

# Kvadratna jednadžba

Uzmimo da treba riješiti (realnu) kvadratnu jednadžbu

$$ax^2 + bx + c = 0,$$

gdje su  $a$ ,  $b$  i  $c$  zadani, i vrijedi  $a \neq 0$ .

Matematički gledano, problem je lagan: imamo 2 rješenja

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

Numerički gledano, problem je mnogo izazovniji:

- ni uspješno računanje po ovoj formuli,
- ni točnost izračunatih korijena,

ne možemo uzeti “zdravo za gotovo”.

## Kvadratna jednadžba — problem

Primjer:  $x^2 - 56x + 1 = 0$ . U aritmetici s 5 decimala dobijemo

$$x_1 = \frac{56 - \sqrt{3132}}{2} = \frac{56 - 55.964}{2} = 0.018000,$$

$$x_2 = \frac{56 + \sqrt{3132}}{2} = \frac{56 + 55.964}{2} = 55.982.$$

Točna rješenja su

$$x_1 = 0.0178628\dots \quad \text{i} \quad x_2 = 55.982137\dots$$

**Manji** od ova dva korijena —  $x_1$ , ima **samo dvije** točne znamenke (**kraćenje**).



# Kvadratna jednadžba — popravak

Prvo izračunamo **većeg** po apsolutnoj vrijednosti, po formuli

$$x_2 = \frac{-(b + \text{sign}(b)\sqrt{b^2 - 4ac})}{2a},$$

a **manjeg** po apsolutnoj vrijednosti, izračunamo iz

$$x_1 \cdot x_2 = \frac{c}{a}$$

(Vieta), tj. formula za  $x_1$  je

$$x_1 = \frac{c}{x_2 a}.$$

Opasnog **kraćenja** za  $x_1$  više **nema!**

# Kvadratna jednadžba (nastavak)

Ovo je bila samo **jedna**, od (barem) **tri** “opasne” točke za računanje. Preostale **dvije** su:

- “kvadriranje” pod korijenom — mogućnost za **overflow**.  
Rješenje — “skaliranjem”.
- **oduzimanje** u diskriminanti (**kraćenje**) — **nema** jednostavnog rješenja.
  - To je odraz **nestabilnosti** problema, jer tad imamo **dva bliska korijena** koji su **osjetljivi** na male **perturbacije** koeficijenata jednadžbe.
  - Na primjer, pomak  $c$  = pomak grafa “**gore–dolje**”.

# Neki primjeri izbjegavanja kraćenja

Primjer. Treba izračunati

$$y = \sqrt{x + \delta} - \sqrt{x},$$

gdje su  $x$  i  $\delta$  zadani ulazni podaci, s tim da je  $x > 0$ ,

• a  $|\delta|$  vrlo mali broj.

U ovoj formuli, očito, dolazi do velike greške zbog kraćenja — zaokruživanje korijena prije oduzimanja.

Ako formulu “deracionaliziramo” u oblik

$$y = \frac{\delta}{\sqrt{x + \delta} + \sqrt{x}},$$

problema više nema!

# Neki primjeri izbjegavanja kraćenja

Primjer. Treba izračunati

$$y = \cos(x + \delta) - \cos x,$$

gdje su  $x$  i  $\delta$  zadani **ulazni** podaci, s tim da je  $|\cos x|$  razumno velik,

• a  $|\delta|$  **vrlo mali** broj.

Opet, dolazi do **velike greške** zbog **kraćenja**.

Ako formulu napišmo u “**produktnom**” obliku

$$y = -2 \sin \frac{\delta}{2} \sin \left( x + \frac{\delta}{2} \right),$$

problema više **nema!**