

# *Numerička matematika*

## *4. predavanje*

Saša Singer

[singer@math.hr](mailto:singer@math.hr)

[web.math.hr/~singer](http://web.math.hr/~singer)

PMF – Matematički odsjek, Zagreb

# Sadržaj predavanja

- Rješavanje linearnih sustava:
  - Hilbertove matrice.
  - Teorija perturbacije linearnih sustava (nastavak).
  - Uloga reziduala i iterativno poboljšanje rješenja.
  - Struktura LR (LU) faktorizacije.
  - Matrice za koje ne treba pivotiranje.
  - Simetrične pozitivno definitne matrice.
  - Faktorizacija Choleskog.
  - Pivotiranje u faktorizaciji Choleskog.
- Aproksimacija i interpolacija:
  - Uvod u problem aproksimacije (norme, linearnost).

# 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 **tri** godine, a stizat će i **nova** (kako nastaju).

**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 **tri demonstratora**:

- **Anastasia Kruchinina** — termin: **srijeda, 18–20**.
- **Ines Marušić** — termin: **srijeda, 14–16**, uz prethodnu najavu mailom,
- **Melkior Ornik** — termin: **četvrtak, 10–12**.

Demosi lijepo **mole** da im se **najavite** mailom koji dan ranije!

- Njihove mail adrese nađete na **oglasnoj ploči**,
- ili se javite meni.

# Hilbertove matrice

# Hilbertova matrica

**Primjer.** Kod aproksimacije polinomima javljaju se linearni sustavi oblika

$$H_n x = b,$$

gdje je  $H_n$  Hilbertova matrica reda  $n$ , tj.  $(H_n)_{ij} = \frac{1}{i+j-1}$ , ili

$$H_n = \begin{bmatrix} 1 & \frac{1}{2} & \cdots & \frac{1}{n} \\ \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{n+1} \\ \vdots & \vdots & & \vdots \\ \frac{1}{n} & \frac{1}{n+1} & \cdots & \frac{1}{2n-1} \end{bmatrix}.$$

# Hilbertova matrica

Da bismo ispitali **točnost** rješenja, stavimo **desnu** stranu

$$b(i) := \sum_{j=1}^n H_n(i, j) = \sum_{j=1}^n \frac{1}{i+j-1}, \quad i = 1, \dots, n,$$

tako da je egzaktno **rješenje** sustava  $x^T = [1, 1, \dots, 1]$ .

Što možemo očekivati od rješenja takvog sustava?

Pogled na **Frobeniusovu normu** matrice  $A$  kaže da ona **nije naročito velika**, tj.

$$\|H_n\|_F = \sqrt{\sum_{i=1}^n \sum_{j=1}^n \left| \frac{1}{i+j-1} \right|^2} \leq \sqrt{\sum_{i=1}^n \sum_{j=1}^n 1} = n.$$

# Hilbertova matrica

Međutim ... ne treba gledati samo normu matrice!!!

Uvjetovanost Hilbertovih matrica je vrlo visoka:

$n$	$\kappa_2(H_n)$	$n$	$\kappa_2(H_n)$	$n$	$\kappa_2(H_n)$
2	$1.928 \cdot 10^1$	9	$4.932 \cdot 10^{11}$	15	$6.117 \cdot 10^{20}$
3	$5.241 \cdot 10^2$	10	$1.603 \cdot 10^{13}$	16	$2.022 \cdot 10^{22}$
4	$1.551 \cdot 10^4$	11	$5.231 \cdot 10^{14}$	17	$6.697 \cdot 10^{23}$
5	$4.766 \cdot 10^5$	12	$1.713 \cdot 10^{16}$	18	$2.221 \cdot 10^{25}$
6	$1.495 \cdot 10^7$	13	$5.628 \cdot 10^{17}$	19	$7.376 \cdot 10^{26}$
7	$4.754 \cdot 10^8$	14	$1.853 \cdot 10^{19}$	20	$2.452 \cdot 10^{28}$
8	$1.526 \cdot 10^{10}$				



## Hilbertova matrica — $n = 2, 5$

Za sustav s Hilbertovom matricom, u **extended** točnosti, umjesto svih **jedinica** u rješenju, dobivamo:

Red 2

$$x(1) = 1.0000000000000000 \quad x(2) = 1.0000000000000000$$

Red 5

$$x(1) = 1.0000000000000000 \quad x(4) = 0.99999999999999990$$

$$x(2) = 0.9999999999999999 \quad x(5) = 1.00000000000000005$$

$$x(3) = 1.00000000000000007$$

## Hilbertova matrica — $n = 10$

$$x(1) = 1.00000000000003436$$

$$x(6) = 0.9999999294831902$$

$$x(2) = 0.9999999999710395$$

$$x(7) = 1.0000001151701616$$

$$x(3) = 1.0000000006068386$$

$$x(8) = 0.9999998890931838$$

$$x(4) = 0.9999999945453735$$

$$x(9) = 1.0000000580638087$$

$$x(5) = 1.0000000258066880$$

$$x(10) = 0.9999999872591526$$

Uvjetovanost:  $\approx 1.603 \cdot 10^{13}$ .

## Hilbertova matrica — $n = 15$

$$x(1) = 1.00000000005406387 \quad x(9) = 1.0952919444304200$$

$$x(2) = 0.9999999069805858 \quad x(10) = 0.8797820363884070$$

$$x(3) = 1.0000039790948573 \quad x(11) = 1.0994671444236333$$

$$x(4) = 0.9999257525660447 \quad x(12) = 0.9508102511158300$$

$$x(5) = 1.0007543452271621 \quad x(13) = 1.0106027108940050$$

$$x(6) = 0.9953234190795597 \quad x(14) = 1.0012346841153261$$

$$x(7) = 1.0188643674562383 \quad x(15) = 0.9992252029377023$$

$$x(8) = 0.9487142544341838$$

Uvjetovanost:  $\approx 6.117 \cdot 10^{20}$ .

## Hilbertova matrica — $n = 20$

$x(1) =$	1.0000000486333029	$x(11) =$	231.3608002738048500
$x(2) =$	0.9999865995557111	$x(12) =$	-60.5143391625873562
$x(3) =$	1.0008720556363132	$x(13) =$	-57.6674972682886125
$x(4) =$	0.9760210562677670	$x(14) =$	5.1760567992057506
$x(5) =$	1.3512820600312678	$x(15) =$	8.7242780841976215
$x(6) =$	-2.0883247796748707	$x(16) =$	210.1722288687690970
$x(7) =$	18.4001541798146106	$x(17) =$	-413.9544667202651170
$x(8) =$	-63.8982130462650081	$x(18) =$	349.7671855031355400
$x(9) =$	161.8392478869777220	$x(19) =$	-142.9134532513063250
$x(10) =$	-254.7902985140752950	$x(20) =$	25.0584794423327874

Uvjetovanost  $\approx 2.452 \cdot 10^{28}$ .

# Uvjetovanost Hilbertovih matrica

Može se pokazati da za **uvjetovanost** Hilbertove matrice  $H_n$  vrijedi formula

$$\kappa_2(H_n) \approx \frac{(\sqrt{2} + 1)^{4n+4}}{2^{15/4} \sqrt{\pi n}} \quad \text{za } n \rightarrow \infty.$$

Dakle, iako Hilbertove matrice imaju “**idealna**” svojstva,

• **simetrične, pozitivno definitne** (čak **totalno pozitivne** = determinanta **svake** kvadratne podmatrice je **pozitivna**), njihova uvjetovanost **katastrofalno brzo raste!**

“**Krivci**” za to su elementi **inverza**  $H_n^{-1}$ .

## Inverz Hilbertove matrice

Recimo,  $H_5^{-1}$  izgleda ovako:

$$H_5^{-1} = \begin{bmatrix} 25 & -300 & 1050 & -1400 & 630 \\ -300 & 4800 & -18900 & 26880 & -12600 \\ 1050 & -18900 & 79380 & -117600 & 56700 \\ -1400 & 26880 & -117600 & 179200 & -88200 \\ 630 & -12600 & 56700 & -88200 & 44100 \end{bmatrix} .$$

A kako tek izgledaju elementi  $H_{20}^{-1}$ ?

# Inverz Hilbertove matrice

Elementi inverza  $H_n^{-1}$  Hilbertove matrice mogu se eksplicitno izračunati u terminima binomnih koeficijenata

$$(H_n^{-1})_{ij} = (-1)^{i+j} (i + j - 1) \cdot \binom{n+i-1}{n-j} \binom{n+j-1}{n-i} \binom{i+j-2}{i-1}^2.$$

Lako se vidi da ovi elementi vrlo brzo rastu za malo veće  $n$ .

Pogledajte

<http://mathworld.wolfram.com/HilbertMatrix.html>

# Perturbacije linearnih sustava (nastavak)



## Još malo o perturbacijama linearnih sustava

Ocjenu koliko se **najviše** promijenilo rješenje sustava  $Ax = b$ ,

• ako perturbiramo **samo**  $A$  ili **samo**  $b$ ,

• ako perturbiramo **i**  $A$  **i**  $b$ ,

možemo dobiti **direktno** — po **normi** i po **elementima**.

Pretpostavimo da smo perturbirali **samo**  $A$ . Umjesto sustava  $Ax = b$ , tada rješavamo sustav

$$(A + \Delta A)(x + \Delta x) = b.$$

Također, možemo pretpostaviti da za **operatorsku normu perturbacije** vrijedi

$$\|\Delta A\| \leq \varepsilon \|A\|.$$

**Komentar.** Ako je  $\varepsilon$  točnost računanja, tolika perturbacija je **napravljena** već pri **spremanju** elemenata matrice u računalo.

## Perturbacija matrice $A$

Oduzimanjem  $Ax = b$  od  $(A + \Delta A)(x + \Delta x) = b$  dobivamo

$$A \Delta x + \Delta A (x + \Delta x) = 0.$$

Množenjem slijeva s  $A^{-1}$  i sređivanjem dobivamo

$$\Delta x = -A^{-1} \Delta A (x + \Delta x).$$

Uzimanjem norme lijeve i desne strane, a zatim **ocjenjivanjem odozgo**, dobivamo

$$\begin{aligned} \|\Delta x\| &\leq \|A^{-1}\| \|\Delta A\| \|x + \Delta x\| \leq \varepsilon \|A^{-1}\| \|A\| \|x + \Delta x\| \\ &= \varepsilon \kappa(A) (\|x\| + \|\Delta x\|), \end{aligned}$$

pri čemu je  $\kappa(A) = \|A\| \|A^{-1}\|$  **uvjetovanost** matrice  $A$ .

# Perturbacija matrice $A$

Premještanjem na lijevu stranu svih pribrojnika koji sadrže  $\Delta x$  dobivamo

$$(1 - \varepsilon \kappa(A)) \|\Delta x\| \leq \varepsilon \kappa(A) \|x\|.$$

Ako je  $\varepsilon \kappa(A) < 1$ , a to znači i  $\|\Delta A\| \|A^{-1}\| < 1$ , onda je

$$\|\Delta x\| \leq \frac{\varepsilon \kappa(A)}{1 - \varepsilon \kappa(A)} \|x\|,$$

što pokazuje da je **pogreška** u rješenju (relativno, po normi)

- približno **proporcionalna uvjetovanosti** matrice  $A$ .

## Perturbacija vektora $b$

Pretpostavimo sad da, umjesto sustava  $Ax = b$ , rješavamo

$$A(x + \Delta x) = b + \Delta b.$$

Opet, pretpostavljamo da za **operatorsku normu** perturbacije vektora  $b$  vrijedi

$$\|\Delta b\| \leq \varepsilon \|b\|.$$

**Oduzimanjem**  $Ax = b$  od  $A(x + \Delta x) = b + \Delta b$  izlazi

$$A \Delta x = \Delta b.$$

**Množenjem** slijeva s  $A^{-1}$  dobivamo

$$\Delta x = A^{-1} \Delta b.$$

## Perturbacija vektora $b$

Uzimanjem norme lijeve i desne strane, a zatim **ocjenjivanjem odozgo**, dobivamo

$$\begin{aligned}\|\Delta x\| &\leq \|A^{-1}\| \|\Delta b\| \leq \varepsilon \|A^{-1}\| \|b\| = \varepsilon \|A^{-1}\| \|Ax\| \\ &\leq \varepsilon \|A^{-1}\| \|A\| \|x\| = \varepsilon \kappa(A) \|x\|,\end{aligned}$$

što pokazuje da je **pogreška** u rješenju (relativno, po normi)

- ponovno, **proporcionalna uvjetovanosti** matrice  $A$ .

Sada možemo **generalizirati** ove rezultate na slučaj

- kad **perturbiramo** istovremeno i  $A$  i  $b$ .

## Perturbacija matrice $A$ i vektora $b$

**Teorem.** Neka je  $Ax = b$  i

$$(A + \Delta A)(x + \Delta x) = b + \Delta b,$$

gdje je

$$\|\Delta A\| \leq \varepsilon \|E\|, \quad \|\Delta b\| \leq \varepsilon \|f\|,$$

pri čemu je  $E$  neka matrica, a  $f$  neki vektor. Također, neka je

$$\varepsilon \|A^{-1}\| \|E\| < 1.$$

Tada, za  $x \neq 0$ , vrijedi ocjena

$$\frac{\|\Delta x\|}{\|x\|} \leq \frac{\varepsilon}{1 - \varepsilon \|A^{-1}\| \|E\|} \left( \frac{\|A^{-1}\| \|f\|}{\|x\|} + \|A^{-1}\| \|E\| \right).$$

## Perturbacija matrice $A$ i vektora $b$

**Komentar:** Uobičajeno se za  $E$  uzima  $A$ , jer je to **pogreška** koju napravimo spremanjem matrice  $A$  u računalo. Jednako tako, za  $f$  se uzima  $b$ . U tom slučaju je

$$\begin{aligned}\frac{\|\Delta x\|}{\|x\|} &\leq \frac{\varepsilon}{1 - \varepsilon \|A^{-1}\| \|A\|} \left( \frac{\|A^{-1}\| \|b\|}{\|x\|} + \|A^{-1}\| \|A\| \right) \\ &= \frac{\varepsilon}{1 - \varepsilon \kappa(A)} \left( \frac{\|A^{-1}\| \|Ax\|}{\|x\|} + \kappa(A) \right) \\ &\leq \frac{\varepsilon}{1 - \varepsilon \kappa(A)} \left( \frac{\|A^{-1}\| \|A\| \|x\|}{\|x\|} + \kappa(A) \right) \\ &= \frac{2\varepsilon \kappa(A)}{1 - \varepsilon \kappa(A)}.\end{aligned}$$

## Perturbacija matrice $A$ i vektora $b$

**Dokaz** (skica). Provodi se na sličan način kao za pojedinačne perturbacije matrice  $A$ , odnosno, vektora  $b$ .

Ako od  $(A + \Delta A)(x + \Delta x) = b + \Delta b$  oduzmemo  $Ax = b$ , dobivamo

$$A \Delta x = \Delta b - \Delta A x - \Delta A \Delta x.$$

Množenjem s  $A^{-1}$  slijeva, a zatim korištenjem svojstava **operatorskih normi**, s malo truda, izlazi traženo. ■

Malo kompliciranije, mogu se dobiti i ocjene za perturbacije po **elementima**. Na primjer, uz pretpostavke da je

$$|\Delta A| \leq \varepsilon E, \quad |\Delta b| \leq \varepsilon f,$$

gdje je  $E$  neka matrica, a  $f$  neki vektor (v. Higham, ASNA2).



# Komentar rezultata teorije perturbacija

Uočimo da sve ocjene vrijede

• samo za “dovoljno male” perturbacije matrice  $A$ .

Na primjer, za relativne perturbacije po normi, mora biti

$$\varepsilon \|A^{-1}\| \|E\| < 1, \quad \text{odnosno,} \quad \varepsilon \kappa(A) < 1.$$

Druga relacija se dobiva za  $E = A$ .

U protivnom, ocjena ne vrijedi (nazivnik nula ili krivi znak),

• tj. relativna greška (po normi) može biti po volji velika.

Pitanje. Što kaže obratna analiza grešaka zaokruživanja, tj.

• koje su ocjene na perturbacije?

# Rezultati obratne analize grešaka zaokruživanja

# Obratna ocjena za LR faktorizaciju

**Teorem.** U aritmetici računala računamo LR faktorizaciju zadane matrice  $A$  reda  $n$ . Pretpostavimo da je algoritam uspješno završio,

- bez pojave prevelikih ili premalih brojeva koji nisu prikazivi,
- i bez pokušaja dijeljenja s nulom.

Izračunati trokutasti faktori  $\hat{L}$  i  $\hat{R}$  onda zadovoljavaju

$$\hat{L}\hat{R} = A + \Delta A, \quad |\Delta A| \leq \gamma_n |\hat{L}| |\hat{R}|,$$

gdje je  $\gamma_n$  standardna oznaka za mjeru grešaka zaokruživanja

$$\gamma_n := \frac{nu}{1 - nu}.$$



# Obratna ocjena za rješenje sustava

**Teorem.** U aritmetici računala računamo rješenje linearnog sustava  $Ax = b$ , s matricom  $A$  reda  $n$ .

Uz iste pretpostavke kao u prošlom teoremu, neka su

- $\hat{L}$  i  $\hat{R}$  izračunati trokutasti faktori u LR faktorizaciji matrice  $A$ ,
- i neka je  $\hat{x}$  izračunato rješenje sustava  $Ax = b$ .

Onda postoji perturbacija  $\Delta A$  matrice  $A$  za koju vrijedi

$$(A + \Delta A) \hat{x} = b, \quad |\Delta A| \leq \gamma_{3n} |\hat{L}| |\hat{R}|.$$

Za zaključak o relativnoj grešci, fali nam još

- neka veza između  $|\hat{L}| |\hat{R}|$  i  $|A|$ .

# Put do relativnih ocjena

U idealnom slučaju, željeli bismo da je

$$|\Delta A| \leq u |A|.$$

To bi odgovaralo grešci zaokruživanja koju napravimo

- početnim spremanjem elemenata matrice  $A$  u memoriju računala.

No, to nije realistično. Nad svakim elementom matrice  $A$

- vrši se još najviše  $n$  aritmetičkih operacija.

Zato ne možemo očekivati nešto bolje od ocjene oblika

$$|\Delta A| \leq c_n u |A|,$$

gdje je  $c_n$  “konstanta” reda veličine  $n$ , tj.  $c_n u \approx c \gamma_n$ .

## Relativne ocjene — idealni slučaj

Na primjer, takvu ocjenu **dobivamo** pod uvjetom da  $\hat{L}$  i  $\hat{R}$  zadovoljavaju da je

$$|\hat{L}| |\hat{R}| = |\hat{L}\hat{R}|.$$

To je **idealni** slučaj — i, naravno, **ne vrijedi** uvijek.

Ako to **vrijedi**, onda iz **prvog** teorema izlazi

$$|\hat{L}| |\hat{R}| = |\hat{L}\hat{R}| = |A + \Delta A| \leq |A| + |\Delta A| \leq |A| + \gamma_n |\hat{L}| |\hat{R}|,$$

pa, prebacivanjem članova dobivamo

$$|\hat{L}| |\hat{R}| \leq \frac{1}{1 - \gamma_n} |A|.$$

## Relativne ocjene — idealni slučaj (nastavak)

Ako tu relaciju uvrstimo u drugi teorem, onda izlazi

$$(A + \Delta A) \hat{x} = b, \quad |\Delta A| \leq \frac{\gamma_{3n}}{1 - \gamma_n} |A|,$$

tj. **izračunato** rješenje  $\hat{x}$  ima

- malu obratnu **relativnu** grešku po **komponentama**.

Za koje matrice **vrijedi** “idealno”  $|\hat{L}| |\hat{R}| = |\hat{L}\hat{R}|$ ?

Na primjer, ako LR faktorizacija daje **nenegativne** elemente u faktorima  $L$  i  $R$ , tj. vrijedi  $L, R \geq 0$  (po elementima).

- Takve su tzv. **totalno nenegativne** ili **totalno pozitivne** matrice.

Javljaju se, na primjer, kod **splajn interpolacije** (v. kasnije).

# Što je bitno za stabilnost?

Iz prethodna dva teorema slijedi da stabilnost LR faktorizacije i rješenja linearnog sustava

- ne ovisi o veličini multiplikatora,
- već o veličini elemenata koji se javljaju u matrici  $|\hat{L}| |\hat{R}|$ , relativno obzirom na elementu matrice  $A$ .

Naime, ta matrica  $|\hat{L}| |\hat{R}|$

- može imati male elemente, iako su joj multiplikatori  $m_{ij} = l_{ij}$  veliki,
- ali može imati i velike elemente, a da su joj multiplikatori reda veličine 1.



# Analiza i procjena stabilnosti algoritma

Za lakšu analizu, ne gleda se po svim elementima, već se analizira omjer normi

$$\frac{\| |\hat{L}| |\hat{R}| \|}{\|A\|}.$$

**Bitno:** ovaj omjer ovisi o algoritmu kojim računamo LR faktorizaciju!

Kod LR faktorizacije bez pivotiranja, omjer normi može biti proizvoljno velik. Na primjer, pokažite da je za matricu

$$\begin{bmatrix} \varepsilon & 1 \\ 1 & 1 \end{bmatrix}$$

taj omjer jednak  $\varepsilon^{-1}$ .

# Stabilnost parcijalnog pivotiranja

Kod **parcijalnog** pivotiranja znamo da vrijedi

$$|\ell_{ij}| \leq 1 \quad \text{za sve } i \geq j.$$

Kad uvrstimo  $m_{ik} = \ell_{ik}$  u formule **transformacije**

$$a_{ij}^{(k+1)} = a_{ij}^{(k)} - m_{ik}a_{kj}^{(k)},$$

indukcijom po koracima eliminacije, dobivamo da vrijedi

$$|r_{ij}| \leq 2^{i-1} \max_{k \leq i} |a_{kj}|.$$

Dakle, kod **parcijalnog** pivotiranja

•  $L$  je **malen**, a  $R$  je **ograđen relativno** obzirom na  $A$ .

# Ocjena stabilnosti preko faktora rasta

Tradicionalno, **obratna** analiza greške izražava se preko **faktora rasta** (engl. growth factor)

$$\rho_n = \frac{\max_{i,j,k} |a_{ij}^{(k)}|}{\max_{i,j} |a_{ij}|}.$$

U procesu **Gaussovih** eliminacija, očito vrijedi da je

$$|r_{ij}| = |a_{ij}^{(i)}| \leq \rho_n \max_{i,j} |a_{ij}|.$$

što daje ogradu za **R**, **relativno** obzirom na **A**.

Može se naći i precizna veza između **omjera normi** i **faktora rasta**.

# Obratna ocjena za sustav preko faktora rasta

**Teorem** (Wilkinson). Neka je  $A$  regularna kvadratna matrica reda  $n$  i neka je  $\hat{x}$  izračunato rješenje sustava  $Ax = b$

- Gaussovima eliminacijama s parcijalnim pivotiranjem u aritmetici računala.

Tada vrijedi

$$(A + \Delta A) \hat{x} = b, \quad \|\Delta A\|_{\infty} \leq n^2 \gamma_{3n} \rho_n \|A\|_{\infty}. \quad \blacksquare$$

U prethodnom teoremu, pretpostavka da koristimo parcijalno pivotiranje nije nužna.

Naime, isto vrijedi i za Gaussove eliminacije bez pivotiranja, samo s malo drugačijom konstantom.

Naravno, faktor rasta može biti puno veći!

# Rezidual i iterativno poboljšanje rješenja

## Rezidual približnog rješenja

Kad rješenje sustava  $Ax = b$  računamo približno (računalom),

☛ umjesto pravog rješenja  $x$ , dobivamo približno rješenje  $\hat{x}$ .

Veličinu

$$r = b - A\hat{x},$$

zovemo **rezidual** izračunatog rješenja  $\hat{x}$ .

**Napomena.** Egzaktni rezidual pravog rješenja  $x$  je  $r = 0$ !

Međutim, ako je (egzaktni) rezidual

☛ **velik**, onda sigurno **nismo blizu** pravom rješenju,

☛ ali rezidual može biti **malen**, a da izračunato rješenje sustava nije **ni blizu** pravom.

## Izračunati rezidual

Primjer. Gledamo **izračunato** rješenje  $\hat{x}$  linearnog sustava

$$H_{20}x = b$$

s desnom stranom takvom da je  $x^T = [1, 1, \dots, 1]$ .

Kad računamo u **extended** točnosti,

• **izračunati** rezidual  $\hat{r} = b - A\hat{x}$  je **nula-vektor** (kraćenje), a komponente rješenja  $\hat{x}$  su bile u **stotinama**.

Ovo ponašanje je u **skladu** s **teorijom perturbacija**, koja

- **garantira mali** rezidual  $r$ , za iole razumne perturbacije,
- a rješenje može biti **katastrofalno**, ako je **uvjetovanost** matrice  $A$  **velika**.

# Uloga reziduala — poboljšanje točnosti

Reziduali se mogu iskoristiti za poboljšavanje netočnog rješenja linearnog sustava.

To se obično provodi u tri koraka — može i iterativno.

- Izračuna se rezidual  $r = b - A\hat{x}$ , pri čemu je  $\hat{x}$  izračunato (ili približno) rješenje sustava.
- Riješi se sustav  $Ad = r$ , gdje je  $d$  korekcija.
- Korekcija se doda izračunatom rješenju

$$y = \hat{x} + d,$$

što bi trebalo dati bolje rješenje  $y$ .

Postupak se može ponoviti s  $y$ , umjesto  $\hat{x}$ .



# Računanje reziduala — mora u većoj točnosti

Ovo ima smisla **samo** ako se **prvi** korak

- računanje reziduala  $r = b - A\hat{x}$
- radi u **većoj** točnosti od **one** u kojoj je **izračunat**  $\hat{x}$ .

To je nužno zbog **kraćenja**, tako da

- **izračunati**  $\hat{r}$  ima **dovoljnu** relativnu točnost.

Preostala **dva** koraka standardno se rade

- u “**običnoj**” točnosti, kao za  $\hat{x}$  (to je **ideja!**).

Tipično se računanje reziduala radi u

- **dvostrukoj** točnosti — jedinična greška reda veličine  $u^2$ , obzirom na **jednostruku**.

Na pr. **double**, prema **single**.

**Kad ne treba pivotirati  
u LR faktorizaciji?**

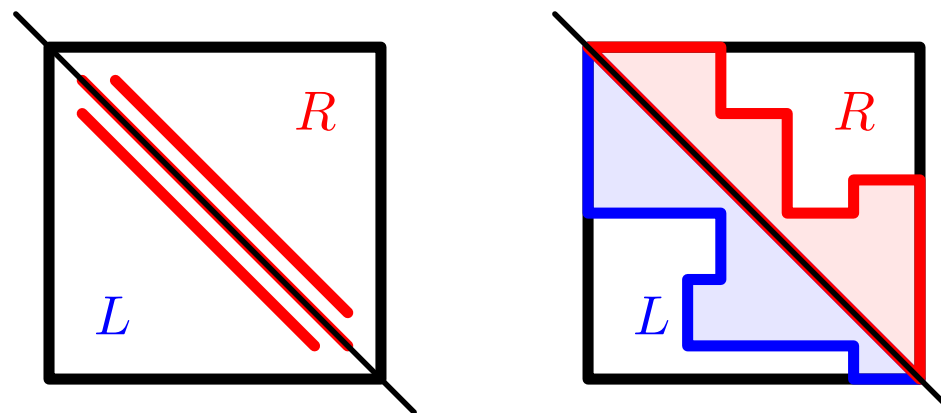
# Struktura LR faktorizacije

Ako matrica  $A$  koja ulazi u LR faktorizaciju ima nekakvu **strukturu**, pitanje je kad će se ta struktura **sačuvati**.

To je **posebno bitno** za

- sustave gdje je  $A$  takva da se **bitna** informacija o njoj može spremiti u **bitno manje** od  $n^2$  elemenata.

Ako **ne pivotiramo**, onda se čuvaju, recimo, sljedeće forme:



Prva su **vrpčaste** matrice, a druga su “rupe **udesno i nadolje**”.

# Kad ne moramo pivotirati?

Dakle, zgodno je znati kad **ne treba** pivotirati, a da

- imamo **garantiranu stabilnost** algoritma **eliminacija**, odnosno, **LR** faktorizacije.

**Odgovor.** Postoje tipovi matrica kod kojih **ne moramo** pivotirati. Na primjer, to su:

- strogo **dijagonalno dominantne** matrice po **stupcima**, tj. matrice za koje vrijedi

$$|a_{jj}| > \sum_{\substack{i=1 \\ i \neq j}}^n |a_{ij}|,$$

- **dijagonalno dominantne** matrice po **recima** ( $i \leftrightarrow j$ ),
- **simetrične pozitivno definitne** matrice (v. malo kasnije).

## Kad ne moramo pivotirati?

Za **dijagonalno dominantne** matrice po **stupcima**, treba samo pokazati da iza **prvog koraka** eliminacije **ostaju** dijagonalno dominantne po stupcima.

1. **Zaključak.**  $a_{11} \neq 0$  i maksimalan po apsolutnoj vrijednosti u 1. stupcu, pa sigurno možemo napraviti 1. korak eliminacije.

Dobivamo matricu  $A^{(2)}$  oblika

$$A^{(2)} = \begin{bmatrix} a_{11} & a_1 \\ 0 & S \end{bmatrix},$$

pri čemu je  $S$  **regularna** (dokaz korištenjem determinanti).

2. **Korak.** Moramo pokazati da je matrica  $S$  **dijagonalno dominantna** po stupcima.

## Kad ne moramo pivotirati?

Za  $j = 2, \dots, n$  vrijedi

$$\sum_{\substack{i=2 \\ i \neq j}}^n |a_{ij}^{(2)}| = \sum_{\substack{i=2 \\ i \neq j}}^n \left| a_{ij} - \frac{a_{i1} a_{1j}}{a_{11}} \right| \leq \sum_{\substack{i=2 \\ i \neq j}}^n |a_{ij}| + \left| \frac{a_{1j}}{a_{11}} \right| \sum_{\substack{i=2 \\ i \neq j}}^n |a_{i1}|$$

(dijagonalna dominantnost obje sume)

$$< (|a_{jj}| - |a_{1j}|) + \left| \frac{a_{1j}}{a_{11}} \right| (|a_{11}| - |a_{j1}|)$$

$$= |a_{jj}| - \left| \frac{a_{1j}}{a_{11}} a_{j1} \right| \quad (\text{koristimo } |a| - |b| \leq |a - b|)$$

$$\leq \left| a_{jj} - \frac{a_{1j}}{a_{11}} a_{j1} \right| = |a_{jj}^{(2)}|,$$

što pokazuje da je i  $A^{(2)}$  dijagonalno dominantna po stupcima.

## Dijagonalno dominantne matrice — preciznije

Za kompleksnu matricu  $A \in \mathbb{C}^{n \times n}$  kažemo da je **dijagonalno dominantna** po **stupcima** ako vrijedi

$$|a_{jj}| \geq \sum_{\substack{i=1 \\ i \neq j}}^n |a_{ij}|, \quad j = 1, \dots, n.$$

Ako vrijedi **stroga** nejednakost ( $>$ ), za **sve**  $j = 1, \dots, n$ , onda kažemo da je  $A$  **strogo** dijagonalno dominantna po **stupcima**.

Matrica  $A$  je (**strogo**) dijagonalno dominantna po **recima**, ako je  $A^*$  (**strogo**) dijagonalno dominantna po **stupcima** ( $i \leftrightarrow j$ ).

U oba slučaja, **Gaussove eliminacije** i **LR faktorizacija** su

🔴 savršeno **stabilne** i **bez pivotiranja**.

# GE i LR za dijagonalno dominantne matrice

**Teorem** (Wilkinson). Neka je  $A$  kompleksna regularna kvadratna matrica reda  $n$ .

- Ako je  $A$  dijagonalno dominantna po recima ili stupcima, tada  $A$  ima LR faktorizaciju bez pivotiranja i za faktor rasta vrijedi  $\rho_n \leq 2$ .
- Ako je  $A$  dijagonalno dominantna po stupcima, tada je  $|l_{ij}| \leq 1$  za sve  $i, j$ , u LR faktorizaciji bez pivotiranja.

To znači da parcijalno pivotiranje ne radi nikakve zamjene redaka (najveći element je već na dijagonali). ■

**Napomena.** Regularnost samo osigurava da dijagonalni elementi ne smiju biti nula, jer dozvoljavamo  $\geq$ .

**Dokaz.** Sličan prethodnom (v. skripta ili Higham, ASNA2).



# Simetrične pozitivno definitne matrice

# Simetrične pozitivno definitne matrice

Za simetrične/hermitske pozitivno definitne matrice radi se “simetrizirana varijanta” LR faktorizacije

- jer je 2 puta brža nego obična LR faktorizacija,
- čuva strukturu matrice  $A$  — čak i kad računamo u aritmetici računala, množenjem faktora uvijek dobivamo simetričnu matricu.

“Simetrizirana LR” faktorizacija zove se faktorizacija Choleskog.

Prisjećanje. Matrica je hermitska ako je

$$A = A^*.$$

Ako je  $A \in \mathbb{R}^{n \times n}$ , onda je hermitska matrica isto što i simetrična, tj.  $*$  =  $T$ .

# Simetrične pozitivno definitne matrice

Pozitivna definitnost matrice se ne vidi odmah. Uobičajeno se unaprijed, iz prirode problema zna da je neka matrica pozitivno definitna.

Matrica  $A \in \mathbb{F}^{n \times n}$  je pozitivno definitna ako je

$$x^* Ax > 0, \quad \text{za svaki } x \in \mathbb{F}^n, \quad x \neq 0.$$

Ekvivalentni uvjeti za pozitivnu definitnost:

☛ sve svojstvene vrijednosti od  $A$  su pozitivne, tj. vrijedi

$$\lambda_k(A) > 0, \quad \text{za } k = 1, \dots, n,$$

gdje  $\lambda_k$  označava  $k$ -tu najveću svojstvenu vrijednost;

# Simetrične pozitivno definitne matrice

Ekvivalentni uvjeti (nastavak):

☛ sve vodeće glavne minore od  $A$  su pozitivne, tj. vrijedi

$$\det(A_k) > 0, \quad \text{za } k = 1, \dots, n,$$

gdje je  $A_k = A(1 : k, 1 : k)$  vodeća glavna podmatrica od  $A$  reda  $k$ .

**Posljedica.** Sve vodeće glavne podmatrice  $A_k$  su regularne, za  $k = 1, \dots, n$ .

**Digresija.** Katkad se puno lakše vidi da neka matrica nije pozitivno definitna. Pokažite da nisu pozitivno definitne matrice koje

☛ na dijagonali imaju negativan element ili nulu.

## Pozitivna definitnost i simetrija

Za **kompleksne** matrice  $A \in \mathbb{C}^{n \times n}$ , može se pokazati da vrijedi

●  $A$  je **pozitivno definitna**  $\implies A$  je **hermitska** ( $A = A^*$ ).

Za **realne** matrice  $A \in \mathbb{R}^{n \times n}$  to **ne mora** vrijediti, tj.

● **pozitivno definitna** matrica **ne mora** biti **simetrična** (može biti i  $A \neq A^T$ ).

Međutim, u **numerici** se vrlo često koristi “**stroža**” varijanta pojma — koja, po **definiciji**, uključuje i **simetriju**:

● **Realna** matrica  $A \in \mathbb{R}^{n \times n}$  **pozitivno definitna** ako je **simetrična** i za svaki  $x \in \mathbb{R}^n$ ,  $x \neq 0$ , vrijedi  $x^T A x > 0$ .

Da ne bude zabune, u nastavku, koristimo “**strožu**” definiciju! U tom slučaju, originalni pojam (**bez** simetrije) katkad se zove samo “**pozitivnost**” matrice  $A$ .

# LR Faktorizacija za sim. poz. def. matrice

Iz ekvivalentnog uvjeta odmah izlazi da se za

- hermitsku/simetričnu pozitivno definitnu matricu uvijek može napraviti
- LR faktorizacija bez pivotiranja (v. Teorem o LR)!

# Simetrizacija — faktorizacija Choleskog

Tvrdnja. Za hermitsku/simetričnu pozitivno definitnu matricu  $A$ , LR faktorizaciju možemo napisati u obliku

$$A = LDL^*.$$

Ta faktorizacija se obično zove  $LDL^*$  faktorizacija.

Kako se dobije ova faktorizacija?

U LR faktorizaciji matrice  $A$ , faktor  $R$  se rastavi na

$$R = DM^*,$$

gdje je

- $D$  dijagonalna,
- a  $M^*$  gornja trokutasta s jedinicama na dijagonali.

# Faktorizacija Choleskog

Da se dobije takva faktorizacija,

- dijagonalni elementi  $R$  stave se na dijagonalu od  $D$ ,
- svaki redak u  $R$  podijeli se s dijagonalnim elementom u tom retku da se dobije  $M^*$ .

Dakle,

$$A = LDM^*, \quad M \text{ donjetrokutasta, regularna.}$$

Zbog hermitičnosti/simetrije vrijedi

$$A = A^* = (LDM^*)^* = MDL^*,$$

pa je  $LDM^* = MDL^* \dots$



# Faktorizacija Choleskog

(nastavak) ...

$$LDM^* = MDL^*.$$

Množenjem s lijeva s  $L^{-1}$  i zdesna s  $L^{-*}$  dobivamo

$$DM^*L^{-*} = L^{-1}MD.$$

Na lijevoj strani imamo produkt **gornjetrokutastih** matrica, a na desnoj strani **donjetrokutastih**, pa su ti produkti **dijagonalne matrice**.

Te dijagonalne matrice su **jednake**  $D$  (jer i  $M$  i  $L$  imaju na dijagonali **jedinice**), pa je

$$L^{-1}MD = D \implies MD = LD \implies M = L.$$

# Faktorizacija Choleskog

Nadalje,  $D$  ima **pozitivne elemente**, jer bi inače postojao vektor  $x$  takav da je

$$(Ax, x) = (LDL^*x, x) = (DL^*x, L^*x) := (Dy, y) \leq 0.$$

Dakle,  $D$  možemo **rastaviti** na

$$D = \Delta \cdot \Delta$$

gdje je  $\Delta$  **dijagonalna** i  $\Delta_{ii} = \sqrt{D_{ii}}$ .

Tada  $LDL^*$  faktorizaciju možemo napisati u obliku:

$$A = LDL^* = (L\Delta)(\Delta L^*) = (L\Delta)(\Delta^*L^*) = (L\Delta)(L\Delta)^*.$$

# Faktorizacija Choleskog

Uz oznaku  $R := (L\Delta)^*$  dobivamo faktorizaciju Choleskog

$$A = R^*R.$$

**Digresija.** Mnogi slovom  $L$  označavaju  $L := L\Delta$ , pa se u literaturi faktorizacija Choleskog može naći napisana kao

$$A = LL^*.$$

**Oprez:** taj  $L$  nema jedinice na dijagonali!

Kad znamo da postoji, Faktorizacija Choleskog se može i direktno izvesti, znajući da je  $A = R^*R$ .

# Algoritam

Ograničimo se na **realni** slučaj. Tada je

$$a_{ij} = \sum_{k=1}^i r_{ki} r_{kj}, \quad i \leq j,$$

pa dobivamo sljedeću **rekurziju** za elemente:

za  $j = 1, \dots, n$  :

$$r_{ij} = \left( a_{ij} - \sum_{k=1}^{i-1} r_{ki} r_{kj} \right) / r_{ii}, \quad i = 1, \dots, j-1,$$

$$r_{jj} = \left( a_{jj} - \sum_{k=1}^{j-1} r_{kj}^2 \right)^{1/2}.$$

Za  $j = 1$  računamo samo  $r_{11}$ .

# Algoritam

Greške zaokruživanja  $\Rightarrow$  moguć negativan izraz pod korijenom.

## Faktorizacija Choleskog

```
za j = 1 do n radi {  
    /* Nađi j-ti stupac od R */  
    /* Supstitucija unaprijed iznad dijagonale */  
    za i = 1 do j - 1 radi {  
        sum = A[i, j];  
        za k = 1 do i - 1 radi {  
            sum = sum - R[k, i] * R[k, j];  
        };  
        R[i, j] = sum / R[i, i];  
    };  
};
```

# Algoritam

```
    /* Dijagonalni element */  
    sum = A[j, j];  
    za k = 1 do j - 1 radi {  
        sum = sum - R[k, j]**2;  
    };  
    ako je sum > 0.0 onda {  
        R[j, j] = sqrt(sum)  
    };  
    inače  
        /* Matrica nije pozitivno definitna, STOP */  
};
```

# Komentar na algoritam

Uočimo da:

- se po prethodnoj rekurziji matrica  $R$  generira **stupac po stupac**, dok se u LR faktorizaciji  $R$  generirala **redak po redak**, a  $L$  **stupac po stupac**;
- ovo je tzv. *jik* varijanta algoritma, a naziv dolazi od **poretka petlji**, uz prirodno imenovanje indeksa;
- ovo **nije** jedina varijanta za realizaciju algoritma.

**Složenost algoritma** (broj aritmetičkih operacija) je približno jednaka

$$OP(n) \sim \frac{1}{3} n^3,$$

što je približno **polovina** složenosti LR faktorizacije.

## Rješenje linearnog sustava

Kad imamo faktorizaciju Choleskog  $A = R^T R$ , onda se rješenje linearnog sustava  $Ax = b$  svodi na dva rješavanja trokutastih sustava

$$R^T y = b, \quad Rx = y,$$

koje lako rješavamo:

● sustav  $R^T y = b$  — supstitucijom unaprijed

$$y_1 = b_1 / r_{11}$$

$$y_i = \left( b_i - \sum_{j=1}^{i-1} r_{ji} y_j \right) / r_{ii}, \quad i = 2, \dots, n,$$



# Rješenje linearnog sustava

- sustav  $Rx = y$  — supstitucijom unatrag

$$x_n = y_n / r_{nn}$$

$$x_i = \left( y_i - \sum_{j=i+1}^n r_{ij} x_j \right) / r_{ii}, \quad i = n - 1, \dots, 1.$$

Za **razliku** od LR faktorizacije, ovdje u obje supstitucije imamo dijeljenja.

Zbog toga se često koristi  $LDL^T$  oblik faktorizacije:

- u algoritmu **nema** računanja **drugih korijena**;
- rješavaju se **tri** linearna sustava:

$$Lz = b, \quad Dy = z, \quad L^T x = y.$$

- $L$  ima **jediničnu** dijagonalu, pa štedimo  $n$  **dijeljenja**.

## Može li $LDL^T$ za simetrične matrice?

Može li se  $LDL^T$  faktorizacija napraviti za bilo koju simetričnu matricu  $A$  (uz dozvolu da matrica  $D$  ima i negativne elemente)?

To ne vrijedi! Kontraprimjer:

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

Pomaže li simetrična permutacija redaka/stupaca? Ne!

Poopćenje na indefinitne matrice dobivamo tako da

• dozvolimo dijagonalne blokove reda 2 u matrici  $D$ .

# Pivotiranje u faktorizaciji Choleskog

I kod faktorizacije Choleskog možemo koristiti pivotiranje.

- Da bismo očuvali simetriju radne matrice, pivotiranje mora biti “simetrično”, tj. radimo istovremene zamjene redaka i stupaca u  $A$

$$A \rightarrow P^T A P,$$

gdje je  $P$  matrica permutacije.

- “Simetrična zamjena”  $\Rightarrow$  dijagonalni element zamjenjuje mjesto s dijagonalnim!
- Standardni izbor pivotnog elementa u  $k$ -tom koraku je

$$a_{rr}^{(k)} = \max_{k \leq i \leq n} a_{ii}^{(k)},$$

što odgovara potpunom pivotiranju u Gaussovima eliminacijama.

# Pivotiranje u faktORIZACIJI Choleskog

Ovim postupkom dobivamo faktORIZACIJU Choleskog

$$P^T A P = R^T R,$$

a za elemente matrice  $R$  vrijedi

$$r_{kk}^2 \geq \sum_{i=k}^j r_{ij}^2, \quad j = k + 1, \dots, n, \quad k = 1, \dots, n.$$

Posebno, to znači da  $R$  ima nerastuću dijagonalu

$$r_{11} \geq \dots \geq r_{nn} > 0.$$

**Napomena.** Usporedite kasnije s rezultatom za QR faktORIZACIJU s pivotiranjem stupaca.

# Aproksimacija i interpolacija

# Općenito o problemu aproksimacije

Što je problem **aproksimacije**?

**Poznate** su **neke** informacije o funkciji  $f$ , definiranoj na nekom podskupu  $X \subseteq \mathbb{R}$ .

Na osnovu tih **informacija**, želimo funkciju  $f$

- **zamijeniti** nekom **drugom** funkcijom  $\varphi$  na skupu  $X$ , ili na još **većem** skupu,
- tako da su funkcije  $f$  i  $\varphi$  **bliske** u nekom **smislu**.

Skup  $X$  je najčešće:

- **interval** oblika  $[a, b]$  (koji može biti i **neograničen**), ili
- **diskretni skup** točaka.

**Pitanje**: Zašto uopće želimo **zamjenu**  $f \mapsto \varphi$ ?

# Oblici problema aproksimacije

Problem aproksimacije javlja se u dva bitno različita oblika.

Prvi oblik: Znamo funkciju  $f$  (analitički ili slično),

• ali je njezina forma prekomplikirana za računanje.

U tom slučaju,

• izaberemo neke informacije o  $f$  i

• po nekom kriteriju odredimo aproksimacijsku funkciju  $\varphi$ .

Prednosti ovog oblika problema aproksimacije:

• Možemo birati informacije o  $f$  koje ćemo koristiti.

• Jednako tako, možemo ocijeniti grešku dobivene aproksimacije  $\varphi$ , obzirom na prave vrijednosti funkcije  $f$ .

# Oblici problema aproksimacije (nastavak)

Drugi oblik: Ne znamo funkciju  $f$ ,

- već samo neke informacije o njoj,
- na primjer, vrijednosti na nekom (diskretnom) skupu točaka.

Zamjenska funkcija  $\varphi$  određuje se iz raspoloživih informacija.

- Osim samih podataka (poznate vrijednosti),
- ove informacije mogu uključivati i očekivani oblik ponašanja tih podataka (tj. funkcije  $\varphi$ ).

Mane ovog oblika problema aproksimacije:

- Ne može se napraviti ocjena pogreške,
- bez dodatnih informacija o nepoznatoj funkciji  $f$ .



# Prvi oblik problema — primjene

Prvi oblik se obično koristi u teoriji

- za razvoj numeričkih metoda na bazi aproksimacije.

Na primjer, za numeričko

- integriranje funkcija (integriramo aproksimaciju),
- rješavanje diferencijalnih jednažbi.

Praktični primjer:

- programska biblioteka za računanje raznih elemenatnih funkcija (`exp`, `sin`, `cos`, `sqrt` i sl),

Traži se maksimalna brzina i puna točnost, na razini osnovne greške zaokruživanja.

Realizacija standardno ide racionalnim aproksimacijama.

## Drugi oblik problema — primjene

Drugi oblik problema se vrlo često javlja u praksi.

Na primjer,

- kod mjerenja nekih veličina (rezultat je “tablica”),
- osim izmjerenih podataka, pokušavamo aproksimirati i podatke koji se nalaze “između” izmjerenih točaka.

To je ključna svrha ovakve aproksimacije!

Napomena. Kod mjerenja se javljaju i greške mjerenja.

- Zato postoje posebne tehnike — vrste aproksimacija, za “ublažavanje” tako nastalih grešaka.

Na primjer, metoda najmanjih kvadrata.

## Izbor aproksimacijske funkcije $\varphi$

Aproksimacijska funkcija  $\varphi$  bira se:

- prema **prirodi modela** — izbor dolazi iz **problema**,
- ali tako da bude relativno **jednostavna** za **računanje**.

Obično se **prvo fiksira** (izabere) neki **skup** funkcija  $\mathcal{F}$ .

- **Onda** se traži “**najbolja**” aproksimacija  $\varphi$  iz tog skupa  $\mathcal{F}$ .

Skup  $\mathcal{F}$  može biti **vektorski prostor**, ali ne mora.

Za **praktično** računanje, funkcija  $\varphi$  obično ovisi

- o nekom **konačnom** broju **parametara**  $a_k$ ,  $k = 0, \dots, m$ ,
- koje treba **odrediti** po nekom **kriteriju** aproksimacije.

**Ideja:** **Sve moguće** vrijednosti ovih  $m + 1$  parametara određuju skup **svih** “**dozvoljenih**” funkcija  $\mathcal{F}$ .

# Parametrizacija aproksimacijske funkcije $\varphi$

Kad funkciju  $\varphi$  zapišemo u obliku

$$\varphi(x) = \varphi(x; a_0, a_1, \dots, a_m),$$

kao funkciju koja ovisi o parametrima  $a_k$ , onda kažemo da smo

- da smo izabrali opći oblik aproksimacijske funkcije  $\varphi$  (u odnosu na skup  $\mathcal{F}$ ).

Prema obliku ovisnosti o parametrima, aproksimacijske funkcije možemo grubo podijeliti na:

- linearne aproksimacijske funkcije,
- nelinearne aproksimacijske funkcije.

Koje su bitne razlike između ove dvije grupe?

# Linearne aproksimacijske funkcije

Opći oblik **linearne** aproksimacijske funkcije je

$$\varphi(x) = a_0\varphi_0(x) + a_1\varphi_1(x) + \dots + a_m\varphi_m(x),$$

gdje su  $\varphi_0, \dots, \varphi_m$  **poznate** funkcije koje **znamo** računati.

Linearnost u **ovisnosti** o parametrima znači:

- traženi parametri su **koeficijenti** u **linearnoj kombinaciji poznatih** funkcija.

Velika **prednost**: **Određivanje** parametara  $a_k$  obično vodi na “**linearne**” probleme (koji su **lakše** rješivi od **nelinearnih**):

- sustave linearnih** jednadžbi, ili
- linearne** probleme **optimizacije**.

# Linearne aproksimacijske funkcije (nastavak)

Standardni model za linearni oblik aproksimacije:

- skup “dozvoljenih” funkcija  $\mathcal{F}$  je vektorski prostor, a
- funkcije  $\varphi_0, \dots, \varphi_m$  su neka baza u tom prostoru.

Unaprijed se bira (fiksira):

- vektorski prostor  $\mathcal{F}$ , odgovarajuće dimenzije  $m + 1$ ,
- baza  $\varphi_0, \dots, \varphi_m$  u  $\mathcal{F}$ .

Napomena. Kod približnog numeričkog računanja,

- “dobar” izbor baze je ključan za stabilnost postupka
- i točnost izračunatih vrijednosti parametara aproksimacijske funkcije  $\varphi$ .

# Primjer 1 — polinomi

Nekoliko primjera najčešće korištenih vektorskih prostora  $\mathcal{F}$ .

**Polinomi.** Uzimamo  $\mathcal{F} = \mathcal{P}_m$ , gdje je  $\mathcal{P}_m$  vektorski prostor polinoma stupnja  $\leq m$ .

**Standardni** izbor baze je  $\varphi_k(x) = x^k$ , za  $k = 0, \dots, m$ , tj.

$$\varphi(x) = a_0 + a_1x + \dots + a_mx^m.$$

**Nije** nužno da  $\varphi$  zapisujemo u bazi  $\{1, x, \dots, x^m\}$ . Upravo suprotno, vrlo često je neka druga baza **bitno bolja**.

- Na primjer,  $\{1, (x - x_0), (x - x_0)(x - x_1), \dots\}$ , gdje su  $x_0, x_1, \dots$  **zadane** točke (v. kod interpolacije).
- **Ortogonalni** polinomi, obzirom na **pogodno** izabrani skalarni produkt (v. kod najmanjih kvadrata).

## Primjer 2 — trigonometrijski polinomi

Trigonometrijski “polinomi”. Za funkcije  $\varphi_k$  uzima se prvih  $m + 1$  funkcija iz skupa

$$\{ 1, \cos x, \sin x, \cos 2x, \sin 2x, \dots \}.$$

Koriste se za aproksimaciju **periodičkih** funkcija na intervalu **perioda** — ovdje, recimo,  $[0, 2\pi]$ .

● **Primjena** je, na primjer, u **obradi i modeliranju signala**.

Varijacije u izboru **baze**:

● Koristi se **dodatni faktor** u argumentu **sinusa i kosinusa** ( $x \mapsto \lambda x$ ) — koji služi za **kontrolu perioda**.

● Ponekad se biraju **samo parne** ili **samo neparne** funkcije iz ovog skupa.



## Primjer 3 — polinomni splajnovi

**Polinomni splajnovi.** To su funkcije koje su “po dijelovima” **polinomi**. Ako su zadane točke  $x_0 < \dots < x_n$ , onda se **splajn** funkcija na svakom **podintervalu**

- svodi na **polinom** određenog fiksnog (**niskog**) stupnja,

$$\varphi \Big|_{[x_{k-1}, x_k]} = p_k, \quad k = 1, 2, \dots, n,$$

a  $p_k$  su **polinomi** najčešće stupnjeva 1, 2, 3 ili 5.

U točkama  $x_i$  obično se zahtijeva da funkcija  $\varphi$  zadovoljava još

- i “**uvjete ljepljenja**” vrijednosti **funkcije** i nekih njezinih **derivacija**, ili nekih **aproksimacija** za te **derivacije**.

Splajnovi se često koriste zbog dobrih **ocjena greške** aproksimacije i **kontrole oblika** aproksimacijske funkcije.

# Nelinearne aproksimacijske funkcije

Nelinearne aproksimacijske funkcije  $\varphi$

$$\varphi(x) = \varphi(x; a_0, a_1, \dots, a_m),$$

imaju nelinearnu ovisnosti o parametrima aproksimacijske funkcije  $a_0, \dots, a_m$ .

Pripadni skup ‘dozvoljenih’ funkcija  $\mathcal{F}$  najčešće

● nije vektorski prostor.

Određivanje parametara  $a_k$ , općenito, vodi na “nelinearne” probleme:

● sustave nelinearnih jednažbi, ili

● nelinearne probleme optimizacije.

## Primjer 4 — eksponencijalne funkcije

Par **primjera** najčešće korištenih oblika **nelinearnih** aproksimacijskih funkcija.

**Eksponencijalne aproksimacije.** Imaju oblik **linearne kombinacije eksponencijalnih** funkcija s **parametrima** u eksponentu:

$$\varphi(x) = c_0 e^{b_0 x} + c_1 e^{b_1 x} + \dots + c_r e^{b_r x},$$

Broj **nezavisnih** parametara je  $m + 1 = 2r + 2$ .

Opisuju, na primjer,

- procese **rasta** i **odumiranja** u raznim **populacijama**,
- s primjenom u **biologiji**, **ekonomiji** i **medicini**.

## Primjer 5 — racionalne funkcije

Racionalne funkcije. Imaju opći oblik

$$\varphi(x) = \frac{b_0 + b_1x + \cdots + b_r x^r}{c_0 + c_1x + \cdots + c_s x^s},$$

i  $m + 1 = r + s + 1$  nezavisnih parametara, a ne  $r + s + 2$ , kako formalno piše.

Objašnjenje. Razlomci se mogu proširivati,

- ako su  $b_i, c_i$  parametri, onda su to i  $tb_i, tc_i$ , za  $t \neq 0$ ;
- uvijek možemo fiksirati jedan od koeficijenata  $b_i$  ili  $c_i$ , a koji je to — obično slijedi iz prirode modela.

Ovako definirane racionalne funkcije imaju mnogo bolja svojstva aproksimacije nego polinomi, a pripadna teorija je relativno nova.

# Kriteriji aproksimacije — interpolacija

Interpolacija je zahtjev da se funkcije  $f$  i  $\varphi$  podudaraju na nekom konačnom skupu točaka.

- Te točke nazivamo **čvorovi interpolacije**.
- Zahtjevu se može, ali i ne mora, **dodati** zahtjev da se u čvorovima, **osim funkcijskih vrijednosti, poklapaju i vrijednosti nekih derivacija**.

U **najjednostavnijem** obliku interpolacije, kad se podudaraju samo funkcijske vrijednosti, od podataka o funkciji  $f$

- koristi se **samo informacija** o njenoj vrijednosti na skupu od  $n + 1$  točaka,
- tj. podaci  $(x_k, f_k)$ , gdje je  $f_k := f(x_k)$ , za  $k = 0, \dots, n$ .

# Kriteriji aproksimacije — interpolacija

- Parametri  $a_0, \dots, a_n$  (mora ih biti **točno onoliko koliko i podataka**, tj.  $m = n$ ) određuju se iz uvjeta

$$\varphi(x_k; a_0, a_1, \dots, a_n) = f_k, \quad k = 0, \dots, n,$$

što je, općenito, **nelinearni** sustav jednažbi.

- **Linearnost** funkcije  $\varphi$  povlači da parametre  $a_k$  dobivamo iz sustava **linearnih jednažbi**
  - koji ima **točno  $n + 1$**  jednažbi za  $n + 1$  nepoznanica. Matrica tog sustava je **kvadratna**.

# Kriteriji aproksimacije — minimizacija pogreške

Minimizacija pogreške je drugi kriterij određivanja parametara aproksimacije. Funkcija  $\varphi$  bira se tako da se **minimizira** neka odabrana norma pogreške

$$e(x) = f(x) - \varphi(x)$$

u nekom odabranom prostoru funkcija  $\mathcal{F}$  za  $\varphi$ , na nekoj domeni  $X$ .

Ove aproksimacije, često zване i **najbolje aproksimacije po normi**, dijele se na

- **diskretne** — ako se norma pogreške  $e$  minimizira na **diskretnom** skupu podataka  $X$ ;
- **kontinuirane** — ako se norma pogreške  $e$  minimizira na **kontinuiranom** skupu podataka  $X$ .

# Kriteriji aproksimacije — minimizacija pogreške

Standardno se kao **norme pogreške** koriste

- 2-norma i
- $\infty$ -norma.

Za 2-normu

- pripadna se aproksimacija zove **srednjekvadratna**,
- a **metoda** za njeno nalaženje zove se **metoda najmanjih kvadrata**.

Funkcija  $\varphi$ , odnosno njeni **parametri**, se traže tako da bude

$$\min_{\varphi \in \mathcal{F}} \|e(x)\|_2.$$



# Kriteriji aproksimacije — minimizacija pogreške

- U diskretnom slučaju je  $X = \{x_0, \dots, x_n\}$ , pa raspisom prethodne relacije dobivamo

$$\sqrt{\sum_{k=0}^n (f(x_k) - \varphi(x_k))^2} \rightarrow \min,$$

- a u kontinuiranom slučaju

$$\sqrt{\int_a^b (f(x) - \varphi(x))^2 dx} \rightarrow \min.$$

Preciznije, **minimizira** se samo ono pod korijenom.

# Kriteriji aproksimacije — minimizacija pogreške

U slučaju  $\infty$ -norme pripadna se aproksimacija zove **minimaks aproksimacija**, a parametri se biraju tako da se nađe

$$\min_{\varphi \in \mathcal{F}} \|e(x)\|_{\infty}.$$

📍 U **diskretnom** slučaju traži se

$$\max_{k=0, \dots, n} |f(x_k) - \varphi(x_k)| \rightarrow \min,$$

📍 a u **kontinuiranom** slučaju

$$\max_{x \in [a, b]} |f(x) - \varphi(x)| \rightarrow \min.$$

# Kriteriji aproksimacije — minimizacija pogreške

Ovaj je tip aproksimacija **poželjniji** od srednjevadratnih,

- jer se traži da **maksimalna greška** bude **minimalna**,
- ali ih je općenito **mного teže izračunati** (na primjer, dobivamo problem minimizacije **nederivabilne** funkcije!).

**Za znatiželjne:** U praksi norme, pored funkcije mogu uključivati i **neke njene derivacije**. Primjer takve norme

$$\|f\| = \sqrt{\int_a^b [(f(x))^2 + (f'(x))^2] dx},$$

na prostoru  $C^1[a, b]$  svih funkcija koje imaju **neprekidnu prvu derivaciju** na  $[a, b]$ .

# Ključni problemi kod aproksimacije

Matematički problemi koje treba riješiti:

- egzistencija i jedinstvenost rješenja problema aproksimacije, što ovisi o tome
  - koje funkcije  $f$  aproksimiramo kojim funkcijama  $\varphi$  (dva prostora)
  - i kako mjerimo grešku  $e$  (norma);
- analiza kvalitete dobivene aproksimacije — vrijednost “najmanje” pogreške i ponašanje funkcije greške  $e$  (jer norma je ipak samo broj),
- konstrukcija algoritama za računanje najbolje aproksimacije.

# Veza aproksimacije i interpolacije — diskretno

U diskretnom slučaju,

- problem interpolacije na konačnom skupu točaka  $X$  (točke iz  $X$  su čvorovi interpolacije),

možemo smatrati specijalnim, ali posebno važnim slučajem

- aproksimacije po normi na skupu  $X$ , uz neku od
- standardnih normi na konačnodimenzionalnim prostorima (ovisi o tome odakle bирамо  $\varphi$ ).

Posebnost: uz minimizaciju norme pogreške  $\|e\| \rightarrow \min$ , dodatno tražimo da je

- minimum norme pogreške jednak nuli, tj.  $\min \|e\| = 0$ , što je onda ekvivalentno odgovarajućim uvjetima interpolacije.

## Veza aproksimacije i interpolacije — primjer

Primjer. Neka je  $X = \{x_0, \dots, x_n\}$  i tražimo aproksimacijsku funkciju  $\varphi$

• u prostoru  $\mathcal{P}_n$  svih polinoma stupnja najviše  $n$ .

Kao kriterij aproksimacije uzmimo neku  $p$ -normu ( $1 \leq p \leq \infty$ )

• vektora  $e$  grešaka funkcijskih vrijednosti na skupu  $X$ .

Za  $1 \leq p < \infty$ , zahtjev je

$$\|e\|_p = \|f - \varphi\|_p = \left( \sum_{k=0}^n |f(x_k) - \varphi(x_k)|^p \right)^{1/p} \rightarrow \min.$$

Za  $p = \infty$ , tražimo

$$\|e\|_\infty = \|f - \varphi\|_\infty = \max_{k=0, \dots, n} |f(x_k) - \varphi(x_k)| \rightarrow \min.$$

## Veza aproksimacije i interpolacije — primjer

Očito je  $\|e\|_p = 0$  ekvivalentno uvjetima interpolacije

$$f(x_k) = \varphi(x_k), \quad k = 0, \dots, n.$$

Međutim, nije jasno može li se to postići, tj.

- postoji li takva aproksimacijska funkcija  $\varphi \in \mathcal{P}_n$
  - za koju je minimum norme greške jednak nuli,
- tako da je  $\varphi$  i interpolacijska funkcija.

U nastavku, pokazat ćemo da je odgovor potvrđan za ovaj primjer. Razlog:

- Prostor  $\mathcal{P}_n$ , u kojem tražimo aproksimaciju, ima taman dovoljno veliku dimenziju.