Problem najmanjih kvadrata

4. dio kolegija Numeričke metode financijske matematike

Nela Bosner

Problem najmanjih kvadrata

Primjer

- *U modelu portfelja kad je broj vrijednosnica jako velik računanje sustava linearnih jednadžbi s matricama velikih dimenzija postaje vrlo zahtjevno.*
- *Capital Asset Pricing Model (CAMP) s druge strane nudi efikasnije računanje, bez korištenja* $\mathbf{C}^{-1}$.
- *U tom modelu pretpostavlja se da svi investitori koriste iste očekivane povrate, iste standardne devijacije i korelacije za sve vrijednosnice.*
- *Nadalje, pretpostavit ćemo da na raspolaganju imamo:*
    - *neriskantnu vrijednosnicu s povratom* μ_n *i standardnom devijacijom* $\sigma_n = 0$
    - *određeni portfelj riskantnih vrijednosnica tzv. tržišni portfelj s očekivanim povratom* $\mu_T = E(R_T)$ *i standardnom devijacijom* σ_T

Primjer (nastavak)

- *određeni broj riskantnih vrijednosnica označenih indeksom j s očekivanim povratima $\mu_j = E(R_j)$ i standardnim devijacijama σ_j*

- *Mi želimo predvidjeti povrat j-te vrijednosnice na temelju poznavanja povrata tržišnog portfelja.*
- *Koristit ćemo linearno predviđanje povrata R_j u ovisnosti o R_T, koje je predstavljeno funkcijom*

$$f(\beta_0, \beta_1) = \beta_0 + \beta_1 R_T,$$

gdje su β_0 i β_1 parametri koje na neki način moramo odabrati.

Primjer (nastavak)

- *Najbolje linearno predviđanje znači da moramo naći β_0 i β_1 takve da je očekivana kvadrirana greška predviđanja dana sa*

$$E((R_j - (\beta_0 + \beta_1 R_T))^2)$$

minimalna, što znači da predviđanje u prosjeku bude što je bliže moguće vrijednosti R_j.

- *Očekivanu kvadriranu grešku predviđanja dalje možemo raspisati kao*

$$\begin{aligned} E((R_j - (\beta_0 + \beta_1 R_T))^2) =& E(R_j^2) - 2\beta_0 E(R_j) - 2\beta_1 E(R_T R_j) + \\ & + \beta_0^2 + 2\beta_0\beta_1 E(R_T) + \beta_1^2 E(R_T^2) \end{aligned}$$

Primjer (nastavak)

- *Minimum ćemo naći tako da parcijalne derivacije po* β_0 *i* β_1 *gornjeg izraza izjednačimo s 0. Dobit ćemo*

$$
\begin{aligned}
0 &= -E(R_j) + \beta_0 + \beta_1 E(R_T) \\
0 &= -E(R_T R_j) + \beta_0 E(R_T) + \beta_1 E(R_T^2)
\end{aligned}
$$

- *Rješenje ovog sustava linearnih jednadžbi je*

$$
\begin{aligned}
\hat{\beta}_1 &= \frac{E(R_T R_j) - E(R_T)E(R_j)}{E(R_T^2) - E(R_T)^2} = \frac{\sigma_{jT}}{\sigma_T^2} \\
\hat{\beta}_0 &= E(R_j) - \hat{\beta}_1 E(R_T) = E(R_j) - \frac{\sigma_{jT}}{\sigma_T^2} E(R_T)
\end{aligned}
$$

Primjer (nastavak)

- *Hessian od* $E((R_j - (\beta_0 + \beta_1 R_T))^2)$ *glasi*

$$\begin{bmatrix} 1 & E(R_T) \\ E(R_T) & E(R_T^2) \end{bmatrix}$$

pa je pozitivno definitna matrica, i zaista se radi o minimumu.

- *Dakle, najbolje linearno predviđanje za* R_j *glasi*

$$\hat{R}_j = \hat{\beta}_0 + \hat{\beta}_1 R_T = E(R_j) + \frac{\sigma_{jT}}{\sigma_T^2}(R_T - E(R_T)).$$

- *U praksi gornji izraz za* $\hat{R}_j$ *ne može se direktno koristiti jer obično ne znamo* $E(R_T)$, $E(R_j)$, σ_{jT} *i* σ_T^2.

Primjer (nastavak)

- *Zato se koristi **linearna regresija** u kojoj se ti nepoznati parametri zamijenjuju aproksimativnim vrijednostima.*
- *Da bismo mogli primijeniti linearnu regresiju, pretpostavimo da imamo bivariatni vremenski niz opažanja $(R_{j,t}, R_{T,t})_{t=1}^{n}$ povrata j-te vrijednosnice i tržišnog portfelja.*
- *Tada nam model linearne regresije predviđa da je*

$$R_{j,t} = \beta_0 + \beta_1 R_{T,t} + \epsilon_t,$$

gdje su nam β_0 i β_1 ponovo nepoznanice a ϵ_t slučajni šum.
- *Koeficijenti regresije β_0 i β_1 mogu se odrediti **metodom najmanjih kvadrata**.*

Primjer (nastavak)

- *Rezultat problema najmanjih kvadrata su vrijednosti $\tilde{\beta}_0$ i $\tilde{\beta}_1$ koje minimiziraju sumu kvadrata greški*

$$\sum_{t=1}^{n}(R_{j,t}-(\beta_0+\beta_1 R_{T,t}))^2.$$

- *U ovom poglavlju upoznat ćemo se sa tehnikama za rješavanje ovog problema, a pomoću njih dobiju se koeficijenti*

$$\begin{aligned}
\tilde{\beta}_1 &= \frac{\sum_{t=1}^{n} R_{j,t}(R_{T,t}-\bar{R}_T)}{\sum_{t=1}^{n}(R_{T,t}-\bar{R}_T)^2} = \frac{\sum_{t=1}^{n}(R_{j,t}-\bar{R}_j)(R_{T,t}-\bar{R}_T)}{\sum_{t=1}^{n}(R_{T,t}-\bar{R}_T)^2} \\
&= \frac{s_{jT}}{s_T^2} \\
\tilde{\beta}_0 &= \bar{R}_j - \tilde{\beta}_1 \bar{R}_T,
\end{aligned}$$

Primjer (nastavak)

gdje su

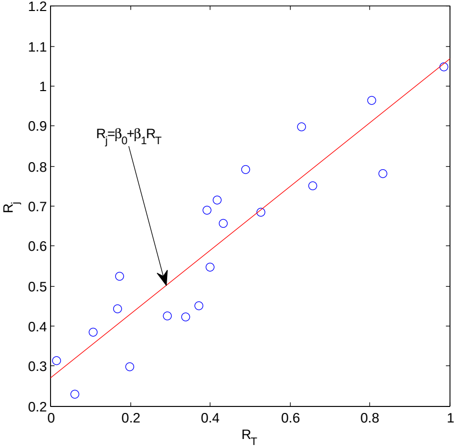
$$\bar{R}_j = \frac{1}{n}\sum_{t=1}^{n} R_{j,t}, \qquad \bar{R}_T = \frac{1}{n}\sum_{t=1}^{n} R_{T,t}.$$

- *Procjena dobivena metodom najmanjih kvadrata je*

$$\tilde{R}_j = \tilde{\beta}_0 + \tilde{\beta}_1 R_T = \bar{R}_j + \frac{s_{jT}}{s_T^2}(R_T - \bar{R}_T),$$

  što je diskretna verzija najboljeg linearnog predviđanja za R_j.
- *Dakle,* $\tilde{\beta}_1$ *je aproksimacija* $\hat{\beta}_1$.
- *Na kraju, CAMP nam daje rezultat*

$$\mu_j - \mu_n = \hat{\beta}_1(\mu_T - \mu_n).$$



Slika: Pravac $\tilde{R}_j = \tilde{\beta}_0 + \tilde{\beta}_1 R_T$ dobiven metodom najmanjih kvadrata

Regresija

- Regresija je jedna od najviše korištenih statističkih metoda.
- Od dostupnih podataka imamo
  - jednu promatranu zavisnu varijablu Y koja predstavlja reakciju promatranog modela na neki ulaz
  - n nezavisnih poznatih varijabli X_j na temelju kojih se vrši predviđanje

  sve varijable su izmjerene u m različitih mjerenja, gdje je često $m > n$.
- Označimo sa
  - Y_i vrijednost varijable Y u i-tom mjerenju
  - $X_{i,1},\ldots,X_{i,n}$ vrijednosti varijabli $X_1,\ldots,X_n$ u i-tom mjerenju

  gdje je $i = 1, \ldots, m$.
- Zadatak modela regresije je naći način na koji je Y povezan sa $X_1,\ldots,X_n$, zatim procjenu uvjetnog očekivanja od Y i predviđanje budućih vrijednosti od Y.

- Model višestruke **linearne regresije** koja povezuje Y i $X_1, \ldots, X_n$ je

$$Y_i = \beta_1 X_{i,1} + \cdots + \beta_n X_{i,n} + \epsilon_i,$$

  gdje je ϵ_i slučajni šum.
- $\beta_1, \ldots, \beta_n$ su nepoznati koeficijenti koje želimo odrediti.
- ϵ_i se često nazivaju greškama mjerenja, ali to nije uvijek slučaj.
- Pretpostavljamo da su ϵ_i bijeli šum, tj.
  - $\epsilon_1, \ldots, \epsilon_m$ su nezavisne i imaju iste distribucije,
  - sa očekivanjem 0 i konstantnom varijancom σ_ϵ^2.
- U tom slučaju je

$$\mathrm{Cov}(\epsilon_1, \ldots, \epsilon_m) = \sigma_\epsilon^2 I_m.$$

- Ako definiramo matricu **X** i vektore **b** i **y** sa

$$\mathbf{X} = \begin{bmatrix} X_{1,1} & \cdots & X_{1,n} \\ \vdots & & \vdots \\ X_{m,1} & \cdots & X_{m,n} \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_m \end{bmatrix},$$

tada sumu kvadrata grešaka možemo napisati kao

$$S(\mathbf{b}) = \sum_{i=1}^{m} \left(Y_i - \sum_{j=1}^{n} \beta_j X_{i,j} \right)^2 = (\mathbf{y} - \mathbf{X}\mathbf{b})^T (\mathbf{y} - \mathbf{X}\mathbf{b}).$$

- Rezultat metode najmanjih kvadrata je $\tilde{\mathbf{b}}$ sa svojstvom

$$S(\tilde{\mathbf{b}}) = \min S(\mathbf{b}).$$

- Da bismo bili konzistentni sa oznakama u numerici, od sada pa na dalje označavat ćemo:

$$A = \mathbf{X}, \quad x = \mathbf{b}, \quad b = \mathbf{y}.$$

Napomena

Ulogu vektora b je promijenjena, što je rezultat nesretne podudarnosti u statističkim i numeričkim oznakama za različite vektore.

Opis linearnog problema najmanjih kvadrata

Matrična formulacija

- Za $A \in \mathbb{R}^{m \times n}$, uz pretpostavku, $m \geq n$, i b rješavamo problem
$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2$$
tj. određujemo x tako da minimizira rezidual $r = Ax - b$
$$\min_x \|r\|_2$$
- Ako je $\text{rang}(A) < n$, onda rješenje x ovog problema očito **nije** jedinstveno, jer mu možemo dodati bilo koji vektor iz nul-potprostora od A, a da se rezidual ne promijeni.
- Među svim rješenjima x problema najmanjih kvadrata uvijek postoji jedinstveno rješenje x najmanje norme, tj. koje još minimizira i $\|x\|_2$.

- Iz geometrijske interpretacije problema najmanjih kvadrata odmah vidimo da je za rješenje x, Ax ortogonalna projekcija vektora b na $\mathrm{Im}(A)$.
- To se lako može provjeriti ako definiramo diferencijabilnu funkciju

$$\phi(x) = \frac{1}{2}\|Ax - b\|_2^2,$$

i izjednačimo $\nabla\phi(x) = 0$.
- Tada možemo raspisati $\phi(x)$ kao

$$\begin{aligned}
\phi(x) =& \frac{1}{2}(Ax - b)^T(Ax - b) = \frac{1}{2}x^T A^T Ax - x^T A^T b + \frac{1}{2}b^T b = \\
=& \frac{1}{2}\sum_{i,j=1}^{n} x_i (A^T A)_{ij} x_j - \sum_{i=1}^{n} x_i (A^T b)_i + \frac{1}{2}b^T b = \\
=& \frac{1}{2}\sum_{i}(A^T A)_{ii} x_i^2 + \frac{1}{2}\sum_{i\neq j}(A^T A)_{ij} x_i x_j - \sum_{i=1}^{n}(A^T b)_i x_i + \frac{1}{2}b^T b.
\end{aligned}$$

- Izračunajmo sada k-tu parcijalnu derivaciju od $\phi(x)$ i izjednačimo ju sa nulom.

$$\frac{\partial}{\partial x_k}\phi(x) = (A^TA)_{kk}x_k + \frac{1}{2}\sum_{j\neq k}(A^TA)_{kj}x_j + \frac{1}{2}\sum_{i\neq k}(A^TA)_{ik}x_i - (A^Tb)_k =$$

$$\underset{\hookrightarrow}{=}^{\{(A^TA)_{ji}=(A^TA)_{ij}\}} \sum_{i=1}^{n}(A^TA)_{ki}x_i - (A^Tb)_k = (A^TAx - A^Tb)_k.$$

- Dakle,

$$\nabla\phi(x) = A^TAx - A^Tb,$$

a iz $\nabla\phi(x) = 0$ slijedi

$$A^T(Ax - b) = A^Tr = 0,$$

ili da rješenje problema najmanjih kvadrata zadovoljava sustav **normalnih jednadžbi**

$$A^TAx = A^Tb.$$

Ovime smo dokazali sljedeći teorem.

Teorem

Skup svih rješenja problema najmanjih kvadrata označimo s

$$\mathcal{S} = \{x \in \mathbb{R}^n \mid \|Ax - b\|_2 = \min\}.$$

Tada je $x \in \mathcal{S}$ ako i samo ako vrijedi sljedeća relacija ortogonalnosti

$$A^T(b - Ax) = 0.$$

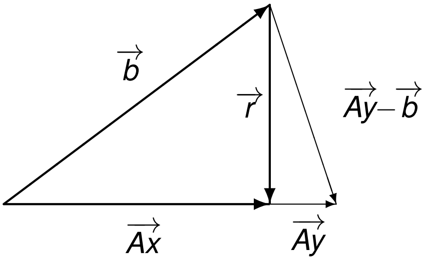
- Ako sa $\overrightarrow{b}$, $\overrightarrow{Ax}$ i $\overrightarrow{r}$ označimo vektore u vektorskom prostoru $\mathbb{R}^m$, pri čemu je x je rješenje problema najmanjih kvadrata, tada imamo da je

$$\overrightarrow{b} = \overrightarrow{Ax} - \overrightarrow{r},$$

a zbog $A^T r = 0$ je $(Ay)^T r = 0$ za svaki $y \in \mathbb{R}^n$, odnosno

$$\overrightarrow{r} \perp \mathrm{Im}(A).$$

- Na kraju možemo zaključiti da je $\overrightarrow{Ax}$ dobiven iz $\overrightarrow{b}$, tako što mu se oduzela komponenta okomita na $\mathrm{Im}(A)$, pa je $\overrightarrow{Ax}$ zaista ortogonalna projekcija od $\overrightarrow{b}$ na $\mathrm{Im}(A)$.



Slika: Okomitost reziduala rješenja x problema $\|Ax - b\|_2 \to \min$ na $\mathrm{Im}(A)$.

- Vratimo se ponovo sustavu normalnih jednadžbi.
- Matrica $A^T A$ je simetrična i pozitivno semidefinitna, a sustav normalnih jednadžbi je uvijek konzistentan, jer je

$$A^T b \in \mathrm{Im}(A^T) = \mathrm{Im}(A^T A).$$

- Kada smo računali rješenje sustava $\nabla\phi(x) = 0$, odnosno sustava normalnih jedadžbi, da bi ono bilo zaista minimum funkcije ϕ moramo provjeriti Hessian.
- Vrijedi

$$\frac{\partial^2}{\partial x_l \partial x_k}\phi(x) = (A^T A)_{kl},$$

što znači da je

$$H\phi = A^T A.$$

- $H\phi$ je
  1. pozitivno definitna u slučaju da je matrica A punog stupčanog ranga, pa tada postoji jedinstveni minimum, i on je rješenje sustava
  $$A^T Ax = A^T b$$
  2. pozitivno semidefinitna u slučaju da matrica A nema puni stupčani rang, pa se tada minimum postiže na čitavom afinom potprostoru.
     - To možemo provjeriti na sljedeći način.
     - Neka je x rješenje problema najmanjih kvadrata, i neka je i $x + z$ također rješenje istog problema.
     - Tada x i $x + z$ moraju zadovoljavati $A^T r = 0$, pa imamo
     $$0 = A^T[A(x + z) - b] = A^T(Ax - b) + A^T Az = A^T Az.$$
     - Ako gornju jednakost skalarno pomnožimo sa z, dobit ćemo da je $\|Az\|_2 = 0$, odakle slijedi da je $Az = 0$ odnosno $z \in \mathrm{Ker}(A)$ (z je u jezgri od A).

- Dakle skup rješenja u ovom slučaju čini skup
$$\mathcal{S} = x + \mathrm{Ker}(A).$$
- Ako je $x \perp \mathrm{Ker}(A)$, onda je
$$\|x + z\|_2^2 = \|x\|_2^2 + \|z\|_2^2,$$
pa je x jedinstveno rješenje problema najmanjih kvadrata koje ima minimalnu 2-normu.

Primjer

- *Provjerimo sada rezultat linearne regresija iz primjera sa početka poglavlja.*
- *Matrični oblik tog problema glasi:*

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2$$

gdje su

$$A = \begin{bmatrix} 1 & R_{T,1} \\ \vdots & \vdots \\ 1 & R_{T,n} \end{bmatrix}, \quad x = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad b = \begin{bmatrix} R_{j,1} \\ \vdots \\ R_{j,n} \end{bmatrix}.$$

- *Sustav normalnih jednadžbi za ovaj problem onda glasi*

Primjer (nastavak)

$$A^T Ax = A^T b$$

$$\begin{bmatrix} n & \sum_{i=1}^n R_{T,i} \\ \sum_{i=1}^n R_{T,i} & \sum_{i=1}^n R_{T,i}^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n R_{j,i} \\ \sum_{i=1}^n R_{T,i} R_{j,i} \end{bmatrix}.$$

- *Rješenje ovog sustava je*

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^n R_{T,i} R_{j,i} - \frac{\left(\sum_{k=1}^n R_{T,k}\right)\left(\sum_{\ell=1}^n R_{j,\ell}\right)}{n}}{\sum_{i=1}^n R_{T,i}^2 - \frac{\left(\sum_{k=1}^n R_{T,k}\right)^2}{n}}$$

$$\tilde{\beta}_0 = \frac{\sum_{i=1}^n R_{j,i}}{n} - \tilde{\beta}_1 \frac{\sum_{i=1}^n R_{T,i}}{n}$$

- *Malim manipulacijama suma dobije se rezultat naveden u početnom primjeru.*

Numeričko rješavanje problema najmanjih kvadrata

Postoji nekoliko načina rješavanja problema najmanjih kvadrata u praksi. Obično se koristi jedna od sljedećih metoda:

1. rješavanje sustava normalnih jednadžbi,
2. transformacija u linearni sustav većih dimenzija i njegovo rješavanje,
3. rješavanje problema najmanjih kvadrata pomoću **QR faktorizacije.**
4. rješavanje problema najmanjih kvadrata pomoću **dekompozicije singularnih vrijednosti (SVD).**

Rješavanje sustava normalnih jednadžbi

- Ova metoda je najbrža, ali je najmanje točna.
- Koristi se kad je A^TA pozitivno definitna (A punog ranga) i kad je njena uvjetovanost mala:

$$\kappa(A^TA) = \|A^TA\|_2 \|(A^TA)^{-1}\|_2 = \|A\|_2^2 \|A^{-1}\|_2^2 = \kappa(A)^2.$$

- Matrica A^TA rastavi se faktorizacijom Choleskog, a zatim se riješi linearni sustav

$$A^TAx = A^Tb.$$

- Ukupan broj aritmetičkih operacija za računanje A^TA, A^Tb, te zatim faktorizaciju Choleskog je $mn^2 + \frac{1}{3}n^3 + O(n^2)$.
- Budući da je obično $m \geq n$, onda je prvi član dominantan u ovom izrazu, a potječe od formiranja A^TA.

Transformacija u linearni sustav većih dimenzija

- Ako matrica A ima puni rang po stupcima, onda problem najmanjih kvadrata možemo transformirati i na linearni sustav različit od sustava normalnih jednadžbi.
- Simetrični linearni sustav
$$\begin{bmatrix} I & A \\ A^T & 0 \end{bmatrix} \begin{bmatrix} r \\ x \end{bmatrix} = \begin{bmatrix} b \\ 0 \end{bmatrix},$$
ekvivalentan je sustavu normalnih jednadžbi.
- Ako napišemo prvu i drugu blok-komponentu
$$r + Ax = b, \quad A^T r = 0,$$
onda uvrštavanjem r-a iz prve blok-jednadžbe u drugu dobivamo sustav
$$A^T(b - Ax) = 0.$$
- Ovaj sustav ima bitno manji raspon elemenata i bolju uvjetovanost od sustava normalnih jednadžbi.

Rješavanje problema najmanjih kvadrata pomoću QR faktorizacije

Teorem (QR dekompozicija)

- *Neka je $A \in \mathbb{C}^{m\times n}$, uz $m \geq n$. Tada postoji unitarna matrica $Q \in \mathbb{C}^{m\times m}$ takva da je*

$$Q^* A = R = \begin{bmatrix} R_1 \\ 0 \end{bmatrix},$$

gdje je $R \in \mathbb{C}^{m\times n}$, a $R_1 \in \mathbb{C}^{n\times n}$ gornje trokutasta matrica s nenegativnim dijagonalnim elementima.

- *Neka je $A \in \mathbb{R}^{m\times n}$, uz $m \geq n$. Tada postoji ortogonalna matrica $Q \in \mathbb{R}^{m\times m}$ takva da je*

$$Q^T A = R = \begin{bmatrix} R_1 \\ 0 \end{bmatrix},$$

gdje je $R \in \mathbb{R}^{m\times n}$, a $R_1 \in \mathbb{R}^{n\times n}$ gornje trokutasta matrica s nenegativnim dijagonalnim elementima.

Napomena

Ako izvršimo particiju matrica

$$Q = \begin{array}{cc} [\,Q_1 & Q_2\,] \\ n & m-n \end{array}$$

onda iz prethodnog teorema slijedi

$$A = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_1 \\ 0 \end{bmatrix} = Q_1 R_1 .$$

Dakle, QR dekompoziciju možemo napisati u skraćenom obliku

$$A = Q_1 R_1 ,$$

pri čemu je $Q_1 \in \mathbb{C}^{m\times n}$ ortonormirana matrica, a $R_1 \in \mathbb{C}^{n\times n}$ gornjetrokutasta matrica s nenegativnim dijagonalnim elementima.

Korolar

Ako je $A \in \mathbb{C}^{n\times n}$ regularna matrica, tada je matrica Q jedinstvena.

- Kao što znamo, QR faktorizaciju možemo izračunati na više načina.
- Najčešća su dva načina računanja kod kojih se ortogonalna matrica Q dobije uzastopnim množenjem elementarnih ortogonalnih matrica, kao što su: reflektori ili rotacije.

QR faktorizacija pomoću Householderovih reflektora – podsjetnik

- Householderove reflektore primijenjujemo direktno na stupce matrice, i to od dijagonale na dolje.
- Neka je $A = [\ a_1^{(1)} \quad \cdots \quad a_n^{(1)}\ ] \in \mathbb{R}^{m\times n}$ za $m \geq n$.
- Ako je $a_1^{(1)} \neq 0$, stavimo li

$$e^{(1)} = e_1 \in \mathbb{R}^m,$$

znamo naći Householderov reflektor H_1 takav da je

$$H_1 a_1^{(1)} = \alpha_1 e^{(1)}.$$

- Tada je

$$A^{(2)} = H_1 A^{(1)} = [\ H_1 a_1^{(1)} \quad \cdots \quad H_1 a_n^{(1)}\ ] =$$

$$= \begin{bmatrix} \alpha_1 & * & * & \cdots & * \\ 0 & & & & \\ \vdots & a_2^{(2)} & a_3^{(2)} & \cdots & a_n^{(2)} \\ 0 & & & & \end{bmatrix}.$$

- Ako je $a_1^{(1)} = 0$, stavimo $H_1 = I$.
- Ako je $a_2^{(2)} \neq 0 \in \mathbb{R}^{m-1}$, postoji Householderova matrica $\bar{H}_2 \in \mathbb{R}^{(m-1)\times(m-1)}$ takva da je

$$\bar{H}_2 a_2^{(2)} = \alpha_2 e_1,$$

uz $e_1 \in \mathbb{R}^{m-1}$.
- Za

$$H_2 = \begin{bmatrix} 1 & 0 \\ 0 & \bar{H}_2 \end{bmatrix}$$

je

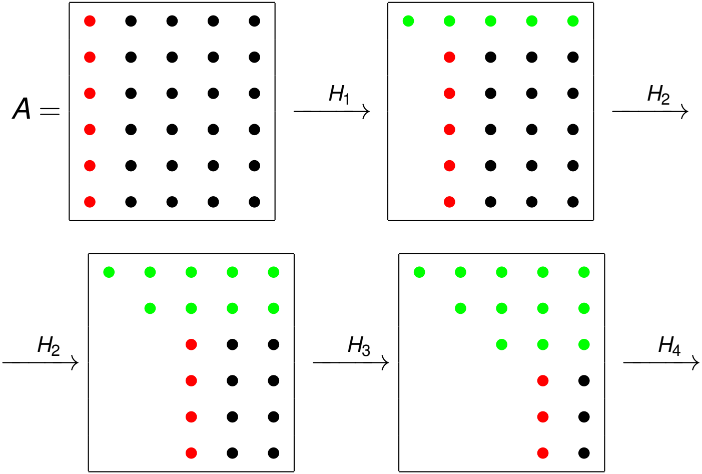
$$
\begin{aligned}
H_2 A^{(2)} &= \begin{bmatrix} 1 & 0 \\ 0 & \bar{H}_2 \end{bmatrix} \begin{bmatrix} \alpha_1 & * & * & \cdots & * \\ 0 & & & & \\ \vdots & a_2^{(2)} & a_3^{(2)} & \cdots & a_n^{(2)} \\ 0 & & & & \end{bmatrix} \\
&= \begin{bmatrix} \alpha_1 & * & * & \cdots & * \\ 0 & \alpha_2 & & & \\ \vdots & \vdots & \bar{H}_2 a_3^{(2)} & \cdots & \bar{H}_2 a_n^{(2)} \\ 0 & 0 & & & \end{bmatrix} \\
&= \begin{bmatrix} \alpha_1 & * & * & \cdots & * \\ 0 & \alpha_2 & * & \cdots & * \\ 0 & 0 & & & \\ \vdots & \vdots & a_3^{(3)} & \cdots & a_n^{(3)} \\ 0 & 0 & & & \end{bmatrix}
\end{aligned}
$$

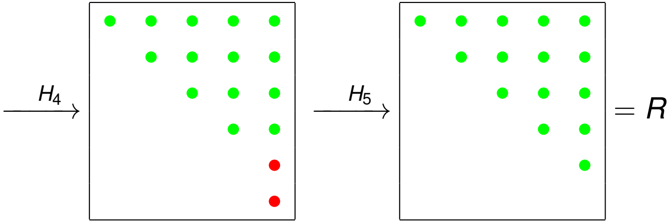
- Nastavljamo tako dalje, svaki puta smanjujući dimenziju problema i radeći sa

$$A^{(k)}(k:m,k:n) \quad \text{i} \quad \bar{H}_k \in \mathbb{R}^{(m-k+1)\times(m-k+1)},$$

a $H_k \in \mathbb{R}^{m\times m}$ definiramo sa

$$H_k = \begin{bmatrix} I_{k-1} & 0 \\ 0 & \bar{H}_k \end{bmatrix}.$$


$A =$
H_1
H_2
H_2
H_3
H_4


H_4
H_5
$= R$

- Na kraju imamo

$$H_n H_{n-1} \cdots H_1 A = \begin{bmatrix} \alpha_1 & * & * & \cdots & * \\ 0 & \alpha_2 & * & \cdots & * \\ & & \ddots & & \\ & & & \ddots & \\ 0 & 0 & & & \alpha_n \\ 0 & 0 & & & 0 \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & & & 0 \end{bmatrix} = R,$$

tj. $A = QR$, gdje je $Q = H_1 \cdots H_n$.

- Želimo li u R nenegativnu dijagonalu, prethodnu jednakost slijeva još pomnožimo matricom

$$H_{n+1} = \mathrm{diag}(\mathrm{sign}(\alpha_1), \ldots, \mathrm{sign}(\alpha_n), 1, \ldots, 1),$$

koja je ortogonalna.

Napomena

- *I kod QR faktorizacije se može* **pivotirati** *i to tako da se u k-tom koraku u matrici $A^{(k)}$ stupac najveće norme (od dijagonale na dolje $a_k^{(k)}, \ldots, a_n^{(k)}$) dovede na pivotno mjesto i ponište njegovi elementi ispod dijagonale.*
- *To se koristi kad želimo naći rang matrice, jer su dijagonalni elementi matrice R sortirani padajuće po apsolutnim vrijednostima.*
- *Imamo*

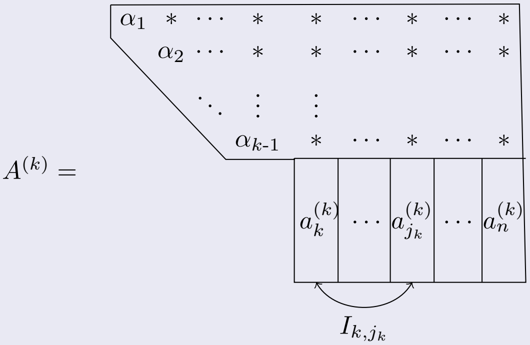
$$H_n H_{n-1}(\cdots H_2((H_1(AI_{1,j_1}))I_{2,j_2})\cdots I_{n-1,j_{n-1}}) = R,$$

$$Q = H_1 \cdots H_n, \quad P = I_{1,j_1} \cdots I_{n-1,j_{n-1}},$$

tj.

$$Q^T AP = R, \quad \Longrightarrow \quad AP = QR.$$

Napomena (nastavak)



- $\|a_{j_k}^{(k)}\|_2 = \max_{j=k,\ldots,n} \|a_j^{(k)}\|_2$
- *k-ti i j_k-ti stupac zamijene mjesta pomoću permutacije $I_{k,j_k} \longrightarrow A^{(k)} \cdot I_{k,j_k}$, i tada se računa H_k.*

Napomena (nastavak)

- *Kad se nakon pivotiranja u tekućem koraku i poništavanja ispoddijagonalnih elemenata u tekućem stupcu, na dijagonali nađe 0, tada znamo da je donji desni $(m-r) \times (n-r)$ blok matrice R jednak nulmatrici.*
- *U tom slučaju je matrica R, a onda i matrica A, ranga r.*

$$R = \left[\begin{array}{ccccc} * & * & * & * & * \\ & * & * & * & * \\ & & * & * & * \\ & & & & \\ & & & & \\ & & & & \end{array}\right] \begin{array}{l} \left.\begin{array}{l} \\ \\ \\ \end{array}\right\} r \\ \\ \\ \\ \end{array}$$

QR faktorizacija pomoću Givensovih rotacija – podsjetnik

- Givensove rotacije poništavaju element po element matrice A.
- Za dobivanje QR faktorizacije potrebno je poništiti sve elemente donjeg trokuta matrice A, i to tako da se jednom poništeni element (jednak nuli) više ne mijenja.
- Način na koji biramo kojim redom ćemo ih poništavati zove se **pivotna strategija**.
- Najčešća pivotna strategija je poništavanje po stupcima:

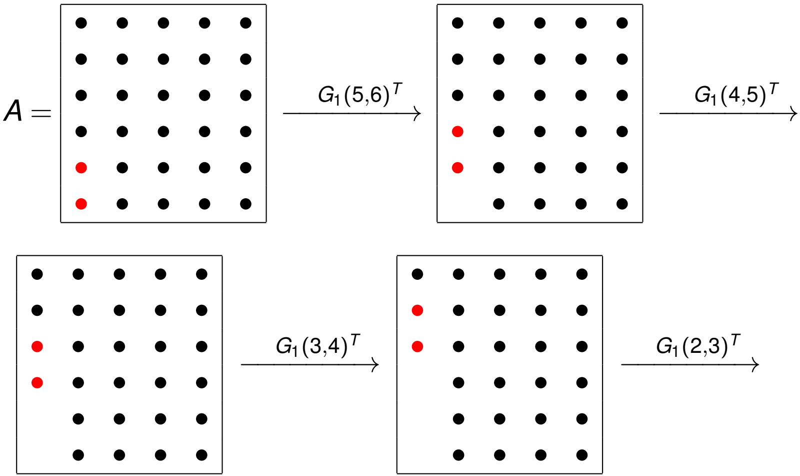
$$\begin{bmatrix} * & * & * & * & * \\ 5 & * & * & * & * \\ 4 & 9 & * & * & * \\ 3 & 8 & 12 & * & * \\ 2 & 7 & 11 & 14 & * \\ 1 & 6 & 10 & 13 & 15 \end{bmatrix}.$$

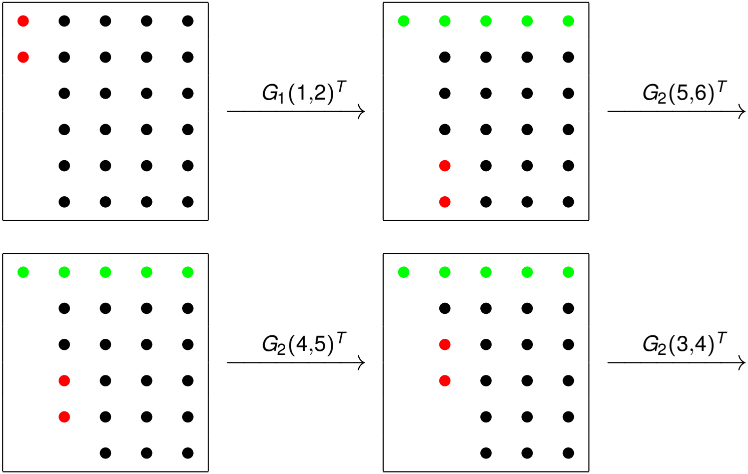
- Na poziciji (i, j) element se poništava Givensovom rotacijom $G_j(i-1, i)$, gdje su $i-1$ i i pivotni indeksi.
- Na kraju, za $A \in \mathbb{R}^{m \times n}$, $m \geq n$ dobivamo da je

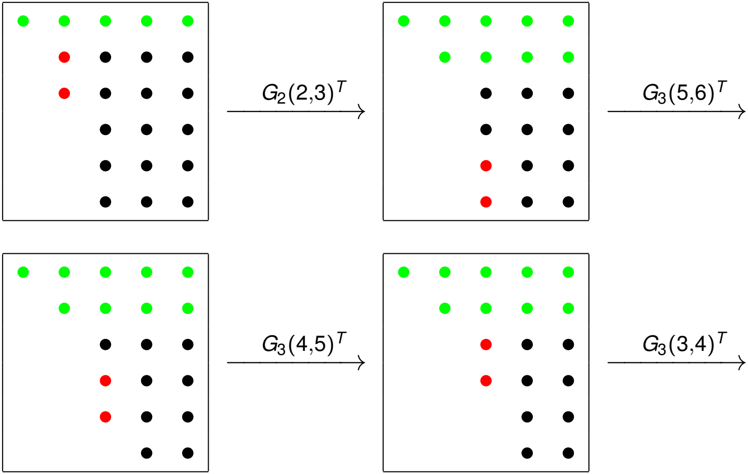
$$G_n(n, n+1)^T \cdots G_n(m-2, m-1)^T G_n(m-1, m)^T \cdots G_2(2,3)^T \cdots G_2(m-2, m-1)^T \cdot$$
$$\cdot G_2(m-1, m)^T \cdot G_1(1,2)^T \cdots G_1(m-2, m-1)^T G_1(m-1, m)^T A = R,$$

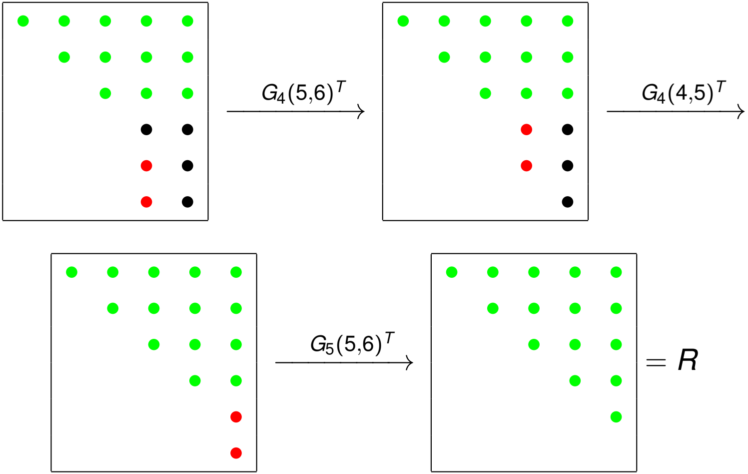
tj. $A = QR$, gdje se matrica Q tada dobiva kao produkt odgovarajućih Givensovih rotacija

$$Q = G_1(m-1, m) \cdots G_1(1,2) G_2(m-1, m) \cdots G_2(2,3) \cdots G_n(m-1, m) \cdots G_n(n, n+1).$$


$A =$
$G_1(5,6)^T$
$G_1(4,5)^T$
$G_1(3,4)^T$
$G_1(2,3)^T$


$G_1(1,2)^T$
$G_2(5,6)^T$
$G_2(4,5)^T$
$G_2(3,4)^T$


$G_2(2,3)^T$
$G_3(5,6)^T$
$G_3(4,5)^T$
$G_3(3,4)^T$


$G_4(5,6)^T$
$G_4(4,5)^T$
$G_5(5,6)^T$
$= R$

Rješenje problema najmanjih kvadrata

Postoje dvije različite situacije kod rješavanja problema najmanjih kvadrata $\min \|Ax - b\|_2$.

Matrica A je punog stupčanog ranga

- U tom slučaju rješenje problema je jednako rješenju sustava normalnih jedadžbi

$$x = (A^T A)^{-1} A^T b.$$

- Sada napišemo QR faktorizaciju matrice A

$$A = QR = Q_1 R_1,$$

  gdje je Q_1 ortonormalna $m \times n$ matrica, a R_1 $n \times n$ regularna trokutasta matrica i uvrstimo u rješenje.
- Dobivamo

$$\begin{aligned} x &= (A^T A)^{-1} A^T b = (R_1^T Q_1^T Q_1 R_1)^{-1} R_1^T Q_1^T b \\ &= (R_1^T R_1)^{-1} R_1^T Q_1^T b = R_1^{-1} R_1^{-T} R_1^T Q_1^T b = R_1^{-1} Q_1^T b. \end{aligned}$$

- Dakle, x se dobiva primjenom "invertirane" skraćene QR faktorizacije od A na b.
- Preciznije, da bismo našli x, rješavamo trokutasti linearni sustav
$$R_1 x = Q_1^T b.$$
- Na ovakav se način najčešće rješavaju problemi najmanjih kvadrata.
- Nije teško pokazati da je cijena računanja $2mn^2 - \frac{2}{3}n^3$.

Matrica A nema puni stupčani rang

- U ovom slučaju prvo trebamo odrediti rang matrice A i zbog toga se koristi **QR faktorizacija sa stupčanim pivotiranjem.**
- Ako matrica A ima rang $r < n$, onda njena QR faktorizacija sa pivotiranjem ima oblik

$$AP = QR = Q \begin{array}{cc} \begin{bmatrix} R_{11} & R_{12} \\ 0 & 0 \\ 0 & 0 \end{bmatrix} & \begin{array}{c} r \\ n-r \\ m-n \end{array} \\ \begin{array}{cc} r & n-r \end{array} & \end{array},$$

gdje je R_{11} regularna reda r, R_{12} neka $r \times (n-r)$ matrica, a matrica P je $n \times n$ matrica permutacija.

- Kod rješavanja problema najmanjih kvadrata tada imamo

$$\begin{aligned} \|b - Ax\|_2^2 &= \|Q^T(b - Ax)\|_2^2 = \|Q^T b - (Q^T AP)(P^T x)\|_2^2 \\ &= \left\| \begin{bmatrix} c \\ d \end{bmatrix} - \begin{bmatrix} R_{11} & R_{12} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix} \right\|_2^2 \\ &= \|(c - R_{12}z) - R_{11}y\|_2^2 + \|d\|_2^2, \end{aligned}$$

gdje je

$$P^T x = \begin{bmatrix} y \\ z \end{bmatrix} \begin{array}{c} r \\ n-r \end{array}, \quad Q^T b = \begin{bmatrix} c \\ d \end{bmatrix} \begin{array}{c} r \\ m-r \end{array} \Rightarrow c = Q(:, 1:r)^T b.$$

- Prema tome, ako tražimo x koji minimizira normu reziduala, tada on mora zadovoljavati

$$(c - R_{12}z) - R_{11}y = 0, \Longrightarrow x = P \begin{bmatrix} y \\ z \end{bmatrix} = P \begin{bmatrix} R_{11}^{-1}(c - R_{12}z) \\ z \end{bmatrix}$$

- Sljedeći korak je pronalaženje vektora z takvog da $\|x\|_2$ ima minimalnu normu, ali to nije jednostavno iz oblika rješenja
$$x = P \begin{bmatrix} R_{11}^{-1}(c - R_{12}z) \\ z \end{bmatrix},$$
smeta nam podmatrica R_{12}, jer kad bi ona bila jednaka nul-matrici tada bi se minimalno rješenje moglo postići za $z = 0$.
- Ako stavimo da je $z = 0$ tada dobivamo **osnovno rješenje**
$$x = P \begin{bmatrix} R_{11}^{-1}c \\ 0 \end{bmatrix},$$
koje inače ne mora imati minimalnu normu u skupu svih rješenja, ali ga je jednostavno izračunati i ima najviše r elemenata različitih od nule.

- Do rješenja problema najmanjih kvadrata sa minimalnom normom možemo, s druge strane, doći pomoću **potpune ortogonalne dekompozicije**.
- U jednakosti $AP = QR$ možemo izvesti još jednu QR faktorizaciju, i to na sljedeći način.
- Trebamo izračunati $n \times n$ ortogonalnu matricu Z takvu da je

$$\begin{bmatrix} R_{11}^T \\ R_{12}^T \end{bmatrix} = Z^T \begin{bmatrix} L_{11}^T \\ 0 \end{bmatrix}, \text{ tj. } \begin{bmatrix} R_{11} & R_{12} \end{bmatrix} = \begin{bmatrix} L_{11} & 0 \end{bmatrix} Z,$$

što nazivamo **LQ faktorizacijom**. Odnosno, vrijedi

$$Z \begin{bmatrix} R_{11}^T \\ R_{12}^T \end{bmatrix} = \begin{bmatrix} L_{11}^T \\ 0 \end{bmatrix} \begin{array}{c} r \\ n-r \end{array} \quad \text{tj.} \quad \begin{bmatrix} R_{11} & R_{12} \end{bmatrix} Z^T = \begin{bmatrix} L_{11} & 0 \end{bmatrix}$$

gdje je L_{11} $r \times r$ regularna donje trokutasta matrica.

- Tada slijedi

$$Q^T APZ^T = (Q^T AP)Z^T = \begin{bmatrix} R_{11} & R_{12} \\ 0 & 0 \\ 0 & 0 \end{bmatrix} Z^T$$

$$= \begin{bmatrix} L_{11} & 0 \\ 0 & 0 \end{bmatrix},$$

odnosno, možemo zaključiti da je

$$Q^T AS = L = \begin{array}{cc} \begin{bmatrix} L_{11} & 0 \\ 0 & 0 \end{bmatrix} & \begin{array}{c} r \\ m-r \end{array} \\ \begin{array}{cc} r & n-r \end{array} & \end{array},$$

gdje je $S = PZ^T$.

- Time smo eliminirali blok R_{12}.
- Primijetimo da je $\mathrm{Ker}(A) = \mathrm{Im}(S(1:n, r+1:n))$.

- Kod rješavanja problema najmanjih kvadrata tada imamo

$$\begin{aligned}\|Ax-b\|_2^2 &= \|Q^T(Ax-b)\|_2 = \|(Q^TAS)S^Tx - Q^Tb\|_2 \\ &= \left\| \begin{bmatrix} L_{11} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} w \\ v \end{bmatrix} - \begin{bmatrix} c \\ d \end{bmatrix} \right\|_2^2 \\ &= \|L_{11}w - c\|_2^2 + \|d\|_2^2,\end{aligned}$$

gdje je

$$S^Tx = \begin{bmatrix} w \\ v \end{bmatrix} \begin{array}{c} r \\ n-r \end{array} \qquad Q^Tb = \begin{bmatrix} c \\ d \end{bmatrix} \begin{array}{c} r \\ m-r \end{array}.$$

- Jasno je, da ako x treba minimizirati normu reziduala, tada moramo imati $L_{11}w - c = 0$, odnosno $w = L_{11}^{-1}c$, a da bi x imao minimalnu normu tada mora biti $v = 0$.

- U tom slučaju je

$$x = S\begin{bmatrix} w \\ 0 \end{bmatrix} = \begin{bmatrix} S(:,1:r) & S(:,r+1:n) \end{bmatrix}\begin{bmatrix} w \\ 0 \end{bmatrix} = S(:,1:r)w.$$

- Vrijedi da je

$$x \in \mathrm{Im}(S(:,1:r)) \perp \mathrm{Im}(S(:,r+1:n)) = \mathrm{Ker}(A),$$

  što smo pokazali da mora vrijediti za jedinstveno rješenje problema najmanjih kvadrata sa minimalnom normom.
- Dakle, konačno rješenje problema najmanjih kvadrata sa minimalnom normom, za $w = L_{11}^{-1}c = L_{11}Q(:,1:r)^T b$ glasi

$$x = S\begin{bmatrix} w \\ 0 \end{bmatrix} = S\begin{bmatrix} L_{11}^{-1}Q(:,1:r)^T b \\ 0 \end{bmatrix}.$$

Zadaci

Zadatak

Riješite problem najmanjih kvadrata iz primjera sa početka poglavlja. Opažanja povrata j-te vrijednosnice i tržišnog portfelja u vremenskim instancama dana su u datoteci

`primjer_regresija_vrijednosnice.mat`

na adresi

`http://www.math.hr/~nela/nmfm.html`

Vaš zadatak je:

1. *generirati matricu A i vektor b u problemu najmanjih kvadrata,*
2. *izračunati QR faktorizaciju matrice A s pivotiranjem pomoću MATLAB-ove funkcije* `qr()`*,*

Zadatak (nastavak)

3. *izračunati rješenje problema najmanjih kvadrata* $[\ \tilde{\beta}_0 \quad \tilde{\beta}_1\ ]^T$,
4. *nacrtati graf sa prikazanim točkama opažanja* $(R_{T,t}, R_{j,t})$ *i pravcem* $R_{j,t} = \tilde{\beta}_0 + \tilde{\beta}_1 R_{T,t}$, *sa pravilno označenim osima.*

Uputa: *QR faktorizacija s pivotiranjem u MATLAB-u se računa tako da se kao izlazne vrijednosti stave tri varijable:* `[Q,R,P]=qr(A)`.

Rješavanje problema najmanjih kvadrata pomoću dekompozicije singularnih vrijednosti

Teorem (Dekompozicija singularnih vrijednosti (SVD))

Neka je $A \in \mathbb{R}^{m\times n}$ matrica ranga r. Tada postoje ortogonalne matrice $U \in \mathbb{R}^{m\times m}$ i $V \in \mathbb{R}^{n\times n}$ takve da je na jedinstven način određena dijagonalna matrica

$$U^T AV = \Sigma = \begin{array}{cc} \begin{bmatrix} \Sigma_+ & 0 \\ 0 & 0 \end{bmatrix} & \begin{array}{l} r \\ m-r \end{array} \\ \begin{array}{cc} r & n-r \end{array} & \end{array}$$

gdje je $\Sigma_+ = diag(\sigma_1, \ldots, \sigma_r)$, uz $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$. Kažemo da je $A = U\Sigma V^T$ ***dekompozicija singularnih vrijednosti (SVD)*** *matrice A.*

Napomena

Ako izvršimo particiju matrica

$$U = \begin{array}{cc} [\,U_1 & U_2\,] \\ r & m-r \end{array} \qquad V = \begin{array}{cc} [\,V_1 & V_2\,] \\ r & n-r \end{array}$$

onda iz prethodnog teorema slijedi:

$$\begin{bmatrix} \Sigma_+ & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix} A [\, V_1 \quad V_2 \,],$$

odnosno

$$\Sigma_+ = U_1^T A V_1 \quad i \quad A = U_1 \Sigma_+ V_1^T.$$

Matrica A je punog stupčanog ranga

- U tome slučaju rješenje problema je jednako rješenju sustava normalnih jedadžbi
$$x = (A^T A)^{-1} A^T b.$$
- Sada napišemo SVD matrice A
$$A = U\Sigma V^T = U_1 \Sigma_+ V^T,$$
gdje je U_1 ortonormalna $m \times n$ matrica, V ortogonalna $n \times n$ matrica, a Σ_+ $n \times n$ dijagonalna matrica, i uvrstimo u rješenje.
- Dobivamo
$$\begin{aligned} x =& (A^T A)^{-1} A^T b = (V\Sigma_+ U_1^T U_1 \Sigma_+ V^T)^{-1} V\Sigma_+ U_1^T b \\ =& (V\Sigma_+^2 V^T)^{-1} V\Sigma_+ U_1^T b = V\Sigma_+^{-2} V^T V\Sigma_+ U_1^T b \\ =& V\Sigma_+^{-2}\Sigma_+ U_1^T b = V\Sigma_+^{-1} U_1^T b. \end{aligned}$$
- Dakle, x se dobiva primjenom “invertiranog” skraćenog SVD-a od A na b.

Matrica A nije punog stupčanog ranga

- Uobičajeno se SVD primjenjuje u metodi najmanjih kvadrata i kad matrica A nema puni stupčani rang, a za razliku od QR faktorizacije ne moraju se raditi dodatne faktorizacije.
- Rješenja su istog oblika, samo što moramo znati izračunati "inverz" matrice Σ kad ona nije regularna, tj. kad ima neke nule na dijagonali.
- Takav inverz zove se generalizirani inverz i označava sa Σ^+ ili $\Sigma^\dagger$.
- U slučaju da je

$$\Sigma = \begin{bmatrix} \Sigma_+ & 0 \\ 0 & 0 \end{bmatrix},$$

pri čemu je Σ_+ regularna, onda je

$$\Sigma^\dagger = \begin{bmatrix} \Sigma_+^{-1} & 0 \\ 0 & 0 \end{bmatrix}.$$

- Neka matrica A ima rang $r < n$.
- Rješenje x koje minimizira $\|Ax - b\|_2$ može se karakterizirati na sljedeći način.
- Neka je $A = U\Sigma V^T$ SVD od A i neka je

$$A = U\Sigma V^T = [U_1, U_2, U_3] \begin{bmatrix} \Sigma_+ & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} [V_1, V_2]^T = U_1 \Sigma_+ V_1^T,$$

  gdje je Σ_+ regularna matrica reda r, matrice U_1 i V_1 imaju r stupaca, matrice U_2 i V_2 imaju $n - r$ stupaca, a matrica U_3 ima $m - n$ stupaca.
- Tada se sva rješenja problema najmanjih kvadrata mogu napisati u formi

$$x = V_1 \Sigma_+^{-1} U_1^T b + V_2 z,$$

  gdje je z proizvoljni vektor.

- Rješenje x koje ima minimalnu 2-normu je ono za koje je $z = 0$, tj.
$$x = V_1 \Sigma_+^{-1} U_1^T b.$$
- Prethodne tvrdnje ćemo sada provjeriti.
- Korištenjem unitarne invarijantnosti 2-norme, dobivamo

$$\begin{aligned}
\|Ax - b\|_2^2 &= \|U^T(Ax - b)\|_2^2 = \left\| \begin{bmatrix} U_1^T \\ U_2^T \\ U_3^T \end{bmatrix} (U_1 \Sigma_+ V_1^T x - b) \right\|_2^2 \\
&= \left\| \begin{bmatrix} \Sigma_+ V_1^T x - U_1^T b \\ -U_2^T b \\ -U_3^T b \end{bmatrix} \right\|_2^2 \\
&= \|\Sigma_+ V_1^T x - U_1^T b\|_2^2 + \|U_2^T b\|_2^2 + \|U_3^T b\|_2^2.
\end{aligned}$$

- Očito, prethodni izraz je minimiziran kad je prva od tri norme u posljednjem redu jednaka 0, tj. ako je

$$\Sigma_+ V_1^T x = U_1^T b,$$

ili

$$x = V_1 \Sigma_+^{-1} U_1^T b.$$

- Stupci matrica V_1 i V_2 su međusobno ortogonalni, pa je $V_1^T V_2 z = 0$ za sve vektore z.
- Odavde vidimo da x ostaje rješenje problema najmanjih kvadrata i kad mu dodamo $V_2 z$, za bilo koji z, tj. ako je

$$x = V_1 \Sigma_+^{-1} U_1^T b + V_2 z.$$

- To su ujedno i sva rješenja, jer stupci matrice V_2 razapinju nul-potprostor $\mathrm{Ker}(A)$.
- Osim toga, zbog spomenute ortogonalnosti vrijedi i

$$\|x\|_2^2 = \|V_1 \Sigma_+^{-1} U_1^T b\|_2^2 + \|V_2 z\|_2^2,$$

a to je minimalno za $z = 0$.

- Rješenje problema najmanjih kvadrata korištenjem SVD-a je najstabilnije, a može se pokazati da je, za $m \gg n$, njegovo trajanje približno jednako kao i trajanje rješenja korištenjem QR-a.
- Za manje m, trajanje je približno $4mn^2 - \frac{4}{3}n^3 + O(n^2)$.

Generalizirani inverz

- Ako želimo proširiti pojam inverza (X) i na matrice koje nisu regularne ili čak nisu kvadratne, onda zahtijevamo da on mora zadovoljavati malo oslabljene uvjete nego standardni inverz:

$$AX = XA = I.$$

- Najpoznatiji generalizirani inverz je tzv. **Moore–Penroseov inverz**, koji je određen sa sljedeća četiri uvjeta.
  Moore–Penroseovi uvjeti:
  1. $AXA = A$
  2. $XAX = X$
  3. $(AX)^* = AX$
  4. $(XA)^* = XA$

Za generalizirani inverz vrijede sljedeća svojstva

- Neka je $A \in \mathbb{C}^{m\times n}$. Tada postoji jedinstvena matrica $X \in \mathbb{C}^{n\times m}$, koja zadovoljava Penroseove uvjete. Ta matrica ima oblik
$$A^\dagger = V \begin{bmatrix} \Sigma_+^{-1} & 0 \\ 0 & 0 \end{bmatrix} U^*, \quad \text{pri čemu je } A = U \begin{bmatrix} \Sigma_+ & 0 \\ 0 & 0 \end{bmatrix} V^*$$
singularna dekompozicija matrice A.
- Za proizvoljnu matricu $A \in \mathbb{C}^{m\times n}$ vrijedi:
  1. $\left(A^\dagger\right)^\dagger = A$
  2. $\left(\bar{A}\right)^\dagger = \overline{\left(A^\dagger\right)}$
  3. $\left(A^T\right)^\dagger = \left(A^\dagger\right)^T$
  4. $\mathrm{rang}(A) = \mathrm{rang}(A^\dagger) = \mathrm{rang}(AA^\dagger) = \mathrm{rang}(A^\dagger A)$
  5. Ako matrica $A \in \mathbb{C}^{m\times n}$ ima rang n, tada je
$$A^\dagger = (A^*A)^{-1}A^* \quad \text{i} \quad A^\dagger A = I_n.$$

6. Ako matrica $A \in \mathbb{C}^{m \times n}$ ima rang m, tada je
$$A^\dagger = A^*(AA^*)^{-1} \quad \mathrm{i} \quad AA^\dagger = I_m.$$
7. Ako je $A = FG$ i $\mathrm{rang}(A) = \mathrm{rang}(F) = \mathrm{rang}(G)$, tada je
$$A^\dagger = G^\dagger F^\dagger.$$
8. Ako su U i V unitarne matrice, tada je
$$(UAV)^\dagger = V^* A^\dagger U^*.$$

- Dakle, kod problema najmanjih kvadrata za $m \geq n$, $r = \mathrm{rang}(A) \leq n$, i za

$$V = [\ V_1 \quad V_2\ ], \quad \Sigma = \begin{bmatrix} \Sigma_+ & 0 \\ 0 & 0 \end{bmatrix}, \quad U = [\ U_1 \quad U_2\ ],$$

pri čemu su $U \in \mathbb{R}^{m\times m}$ i $U_1 \in \mathbb{R}^{m\times r}$, $\Sigma \in \mathbb{R}^{m\times n}$ i $\Sigma_+ \in \mathbb{R}^{r\times r}$, $V \in \mathbb{R}^{n\times n}$ i $V_1 \in \mathbb{R}^{n\times r}$, bez obzira da li je matrica punog ranga ili nije, rješenje glasi

$$\begin{aligned} x = & V_1\Sigma_+^{-1}U_1^T b = [\ V_1 \quad V_2\ ] \begin{bmatrix} \Sigma_+^{-1} & 0 \\ 0 & 0 \end{bmatrix} [\ U_1 \quad U_2\ ]^T b \\ = & V \begin{bmatrix} \Sigma_+^{-1} & 0 \\ 0 & 0 \end{bmatrix} U^T b = A^\dagger b \end{aligned}$$

- U slučaju kada je matrica kvadratna i regularna tada je $A^\dagger = A^{-1}$, pa je rješenje problema najmanjih kvadrata $x = A^{-1}b$ ujedno i rješenje sustava linearnih jednadžbi $Ax = b$.

- Za matricu A punog stupčanog ranga, za rješenje najmanjih kvadrata x i za aproksimaciju rješenja y vrijedi sljedeća ocjena:

$$\begin{aligned}\frac{\|y-x\|_2}{\|x\|_2} &= \frac{\|A^\dagger A(y-x)\|_2}{\|x\|_2}\cdot\frac{\|A\|_2}{\|A\|_2} \leq \|A\|_2\|A^\dagger\|_2\frac{\|Ay-Ax\|_2}{\|A\|_2\|x\|_2} \\ &\leq \kappa(A)\frac{\|Ay-Ax\|_2}{\|Ax\|_2} = \kappa(A)\frac{\|Ay-b+b-Ax\|_2}{\|Ax\|_2} \\ &= \kappa(A)\frac{\|r_y-r_x\|_2}{\|b+r_x\|_2}\end{aligned}$$

  pri čemu je $\kappa(A)=\|A\|_2\|A^\dagger\|_2$ broj uvjetovanosti, a $r_y=Ay-b$ i $r_x=Ax-b$ su reziduali kod kojih r_x ima minimalnu normu.
- To znači da relativna norma greška rješenja ovisi o broju uvjetovanosti matrice A i o razlici između reziduala.

- Numeričke metode za rješavanje problema najmanjih kvadrata primijenjene na loše uvjetovane matrice mogu dati vrlo netočne aproksimacije rješenja.
- Specijalno za sustave linearnih jednadžbi je $r_x = 0$, pa imamo
$$\frac{\|y - x\|_2}{\|x\|_2} \leq \kappa(A)\frac{\|r_y\|_2}{\|b\|_2},$$
gdje na desnoj strani imamo broj uvjetovanosti matrice A i relativnu normu reziduala.
- To znači da kada zaustavimo iterativnu metodu za rješavanje sustava linearnih jednadžbi u iteraciji u kojoj je postignuta mala ralativna norma reziduala, ako je matrica loše uvjetovana relativna norma greške ne mora biti mala i možemo imati netočnu aproksimaciju rješenja.

Zadaci

Zadatak

Riješite problem najmanjih kvadrata iz primjera sa početka poglavlja. Opažanja povrata j-te vrijednosnice i tržišnog portfelja u vremenskim instancama dana su u datoteci

`primjer_regresija_vrijednosnice.mat`

na adresi

`http://www.math.hr/~nela/nmfm.html`

Vaš zadatak je:

1. *generirati matricu A i vektor b u problemu najmanjih kvadrata,*
2. *izračunati SVD faktorizaciju matrice A pomoću MATLAB-ove funkcije* `svd()`*,*

Zadatak (nastavak)

3. *izračunati rješenje problema najmanjih kvadrata* $\begin{bmatrix} \tilde{\beta}_0 & \tilde{\beta}_1 \end{bmatrix}^T$*,*
4. *nacrtati graf sa prikazanim točkama opažanja* $(R_{T,t}, R_{j,t})$ *i pravcem* $R_{j,t} = \tilde{\beta}_0 + \tilde{\beta}_1 R_{T,t}$*, sa pravilno označenim osima.*

Problem određivanja numeričkog ranga

- Vidjeli smo da je kod rješavanja problema najmanjih kvadrata važno da li matrica A ima puni stupčani rang ili nema.
- U egzaktnoj aritmetici to je lako odrediti:
  - kod QR faktorizacije ako se pojavljuje 0 na dijagonali matrice R ona nije regularna, pa nije regularna niti matrica A,
  - kod SVD-a ako postoje singularne vrijednosti jednake 0 matrica A nije regularna.
- U aritmetici konačne preciznosti taj problem nije tako jednostavan: što znači da je neki broj jednak 0?
- Ako 0 mora biti rezultat neke računske operacije, onda ćemo umjesto nje vrlo često dobiti neki vrlo mali broj koji je rezultat grešaka zaokruživanja i grešaka računskih operacija.

- U toj situaciji teško je odrediti da li je taj vrlo mali broj rezultat operacije koja je u egzaktnoj aritmetici zaista i trebala dati vrlo mali broj, ili operacije koja je trebala dati 0.

Primjer

$$
\begin{aligned}
fl(1.000000000000001 - 1) &= 1.1102e - 015 \\
fl(1.0000000000000001 - 1 - 1e - 16) &= -1.0000e - 016
\end{aligned}
$$

- Vrlo često je kod numeričkog rješavanja nekog problema i svejedno na koji način smo dobili taj mali broj:
  - on će stvarati problema kao da je zaista jednak 0.

- Konkretno, u slučaju određivanja numeričkog ranga, ukoliko izračunamo singularnu vrijednost matrice A koja je npr. reda 10^{-16}, iako je matrica čak i trebala biti punog ranga, točnije rješenje ćemo dobiti ako problem rješavamo kao da matrica nije punog ranga.
- Sličan problem je i kod QR faktorizacije: pivotiranje se uvodi da bi se lakše numerički mogli odrediti oni elementi na dijagonali matrice R koje možemo poistovjetiti sa 0 (zbog padajućih apsolutnih vrijednosti dijagonalnih elemenata).
  - Zbog toga, kada na dijagonali dobijemo dovoljno mali broj, bez obzira trebao li on biti u egzaktnoj aritmetici jednak 0 ili ne, cijeli donji dijagonalni blok matrice R poistovijećujemo s 0.

Primjer

- *Zbog grešaka zaokruživanja, umjesto pravog R, izračunamo*

$$R' = \begin{bmatrix} R_{11} & R_{12} \\ 0 & R_{22} \\ 0 & 0 \end{bmatrix}.$$

- *Naravno, željeli bismo da je $\|R_{22}\|_2$ vrlo mala, reda veličine $\varepsilon\|A\|_2$, pa da je možemo "zaboraviti", tj. staviti $R_{22} = 0$ i tako odrediti numerički rang od A.*
- *Nažalost, to nije uvijek tako. Na primjer, bidijagonalna matrica*

$$A = \begin{bmatrix} \frac{1}{2} & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \frac{1}{2} \end{bmatrix}$$

Primjer

je skoro singularna ($\det(A) = 2^{-n}$)*, njena QR faktorizacija je* $Q = I$, $R = A$*, i nema niti jednog* R_{22} *koji bi bio po normi malen.*

- *Zbog toga koristimo pivotiranje, koje* R_{11} *pokušava držati što bolje uvjetovanim, a* R_{22} *po normi što manjim.*

Zadatak

U MATLAB-u generirajte bidijagonalnu matricu iz prethodnog primjera reda 100.

- *Izračunajte njenu QR faktorizaciju s pivotiranjem, i provjerite dijagonalne elemente matrice R.*
- *Izračunajte njen SVD, i provjerite njene singularne vrijednosti.*

Primjer

U datoteci

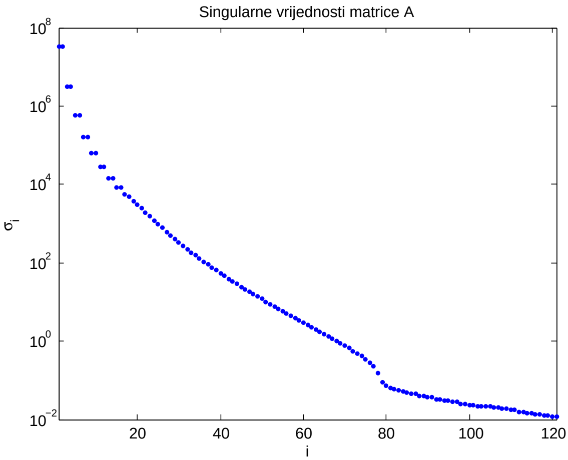
`primjer_osjetljivosti_pnk_Ax.mat`

na adresi

http://www.math.hr/~nela/nmfm.html

spremljeni su matrica A i vektor $\bar{x}$.

- *Izračunajmo najprije SVD matrice A i provjerimo da li je ona punog ranga.*
- *Vidimo da je matrica A punog ranga i da je njena uvjetovanost velika: $\kappa(A) = 2.924 \cdot 10^9$.*


Singularne vrijednosti matrice A
10^8
10^6
10^4
10^2
10^0
10^{-2}
σ_i
20
40
60
80
100
120
i

Primjer (nastavak)

- *Izračunajte* $\bar{b} = A \cdot \bar{x}$, *pri čemu treba ispasti da je* $\bar{b} = [\ 10000 \quad \cdots \quad 10000\ ]^T$.
- *U ovom slučaju* $\bar{x}$ *je egzaktno rješenje problema najmanjih kvadrata* $\min \|Ax - \bar{b}\|_2$.
- *Sada ćemo malo pokvariti vektor* $\bar{b}$ *i vidjeti kako to utječe na rješenje problema najmanjih kvadrata.*
- *Izračunajmo*
$$b = \bar{b} + \eta,$$
gdje su elementi od η *slučajni brojevi iz normalne distribucije.*
- *Ovime se elementi od* $\bar{b}$ *i* b *poklapaju u prve 4 vodeće znamenke, a u 5. znamenci se u prosjeku pojavljuje greška.*

Primjer (nastavak)

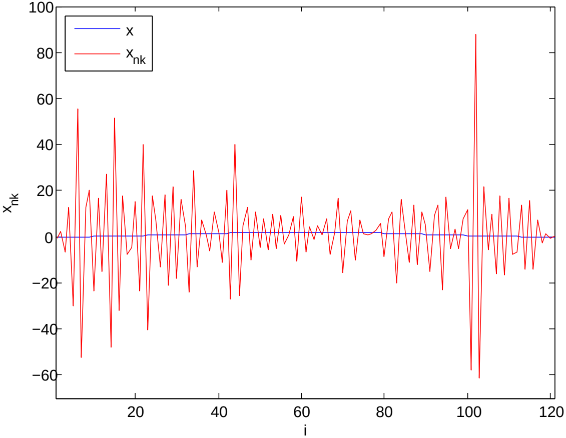
- *Riješimo sada problem najmanjih kvadrata $\min \|Ax - b\|_2$ pomoću SVD-a, i usporedimo ga sa $\bar{x}$.*
- *Rješenje ćemo označiti sa x_{nk}.*
- *Najprije ćemo provjeriti norme reziduala:*

$$\begin{aligned} \|A\bar{x} - \bar{b}\|_2 &= 0 \\ \|Ax_{nk} - b\|_2 &\approx 4.7955 \end{aligned}$$

- *Dalje, provjerimo normu razlike $x_{nk} - \bar{x}$:*

$$\|x_{nk} - \bar{x}\|_2 \approx 224.1275.$$

- *Dakle, možemo zaključiti da smo dobili minimalnu normu reziduala, ali greška je ogromna.*



Slika: Egzaktno rješenje i rješenje najmanjih kvadrata.

- U prethodnom primjeru imali smo loše uvjetovanu matricu, čija je najmanja svojstvena vrijednost

$$\sigma_{min} = 1.1610 \cdot 10^{-2} < 10^{-9} \cdot \|A\|_2$$

što je izgleda "dovoljno mala vrijednost" da utječe na numerički rang.
- Efekt toga je činjenica da smo malo pokvarili vektor b, a dobili smo totalno drugačije i oscilirajuće rješenje, daleko od očekivanog.
- Razmotrit ćemo sada dvije tehnike koje se koriste za stabiliziranje jako oscilirajućih rješenja.

Regularizacija

- Najčešće korištena metoda za stabiliziranje oscilirajućih rješenja problema najmanjih kvadrata je uvođenje uvjeta na rješenje x oblika
$$\|Q(x - x_0)\|_2^2 \leq \beta^2.$$
Ovdje su
  - x_0 opcionalna inicijalna aproksimacija od $\bar{x}$,
  - Q je matrična reprezentacija linearnog operatora uvjeta,
  - β^2 je konstanta koja određuje jačinu uvjeta.
- Aproksimacija x_λ dobiva se rješavanjem problema
$$\min_x \left(\|b - Ax\|_2^2 + \lambda \|Q(x - x_0)\|_2^2 \right),$$
gdje je parametar λ Lagrangeov multiplikator čija vrijednost ovisi o β^2.
- Rješenje je oblika
$$x_\lambda = (A^T A + \lambda^2 Q^T Q)^{-1} (A^T b + \lambda^2 Q^T Q x_0).$$

- Uspjeh regularizacije ovisi o izboru parametra λ, a za to postoji nekoliko načina.
- Najčešći izbor za Q je identiteta $I_n \in \mathbb{R}^{n\times n}$.
- U tom slučaju problem se može izraziti kao prošireni problem

$$\begin{bmatrix} b \\ \lambda x_0 \end{bmatrix} = \begin{bmatrix} A \\ \lambda I_n \end{bmatrix} \bar{x} + \begin{bmatrix} \eta \\ \lambda\gamma \end{bmatrix},$$

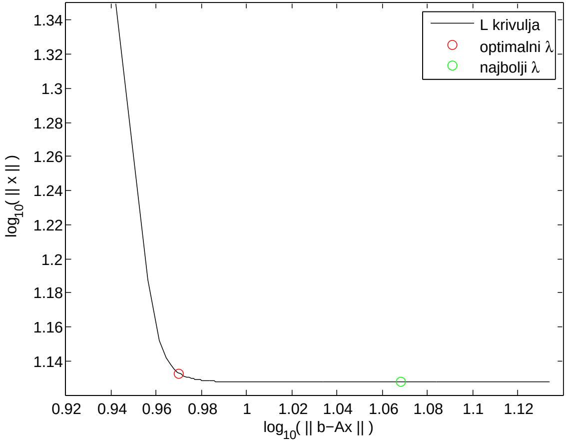
sa

$$\begin{bmatrix} \eta \\ \gamma \end{bmatrix} \sim N(0, I_{m+n}).$$

- Parametar λ postaje težinska konstanta koja bi trebala biti dovoljno velika da priguši oscilacije aproksimativnog rješenja x_λ tako da ga drži blizu x_0, a s druge strane dovoljno mala da ne prouzroči preveliki rast kvadrata norme $\|Ax_\lambda - b\|_2^2$.

Primjer

- *Sada ćemo primijeniti regularizaciju na naš primjer.*
- *Ako ne znamo kakvog nam je oblika rješenje najbolje je uzeti* $x_0 = 0$.
- *Najčešći način za odabir optimalnog parametra* λ *bazira se na L krivulji.*
- *Koordinate točaka na L krivulji predstavljaju* $\log_{10} \|x_\lambda\|_2$ *i* $\log_{10} \|Ax_\lambda - b\|_2$ *za rješenje problema* x_λ *pomoću regularizacije s parametrom* λ.
- *Odabire se ona vrijednost* λ *za koju je* $\|x_\lambda\|_2$ *ograničen na najbolji mogući način, dok istovremeno* $\|Ax_\lambda - b\|_2$ *nije prevelik.*
- *Takav* λ *odgovara točci u uglu L krivulje.*
- *Za naš primjer optimalni* λ *je* $\lambda_{opt} = 0.748$.



Slika: L krivulja za $0.1 \leq \lambda \leq 100$ sa točkama koje odgovaraju λ_{opt} i λ_{naj}.

Primjer (nastavak)

- *Dakle, rješavamo problem najmanjih kvadrata* $\min_x \|\mathbf{A}x - \mathbf{b}\|_2$*, gdje su*

$$\mathbf{A} = \begin{bmatrix} A \\ \lambda_{opt} I_n \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b \\ \lambda_{opt} x_0 \end{bmatrix},$$

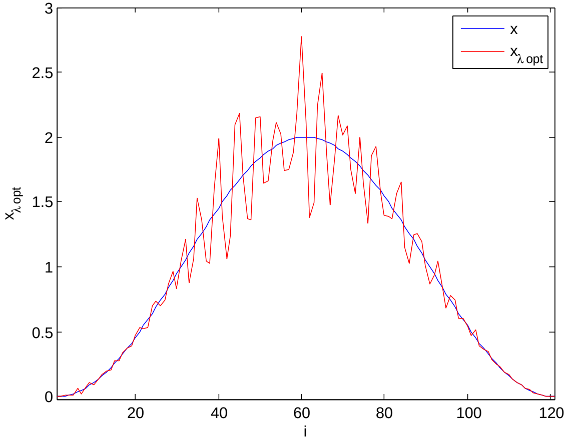
  koristeći SVD.
- *Za rješenje* $x_{\lambda opt}$ *ovog problema provjerit ćemo normu reziduala:*

$$\|Ax_{\lambda opt} - b\|_2 \approx 7.3050.$$

- *S druge strane je norma razlike* $x_{\lambda opt} - \bar{x}$*:*

$$\|x_{\lambda opt} - \bar{x}\|_2 \approx 2.2425.$$

- *Dakle, norma reziduala je malo narasla, ali greška je puno bolja nego kod rješenja najmanjih kvadrata.*



Slika: Egzaktno rješenje i rješenje regularizacije za λ_{opt}.

Primjer (nastavak)

- *Nekim statističkim metodama može se pokazati da aproksimaciju s najboljom greškom možemo dobiti za* $\lambda_{naj} = 77.5$.
- *Zato ćemo na kraju riješiti problem najmanjih kvadrata* $\min_x \|\mathbf{A}x - \mathbf{b}\|_2$, *za*

$$\mathbf{A} = \begin{bmatrix} A \\ \lambda_{naj} I_n \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b \\ \lambda_{naj} x_0 \end{bmatrix},$$

  koristeći SVD.
- *Za rješenje* $x_{\lambda naj}$ *ovog problema opet ćemo provjeriti normu reziduala:*

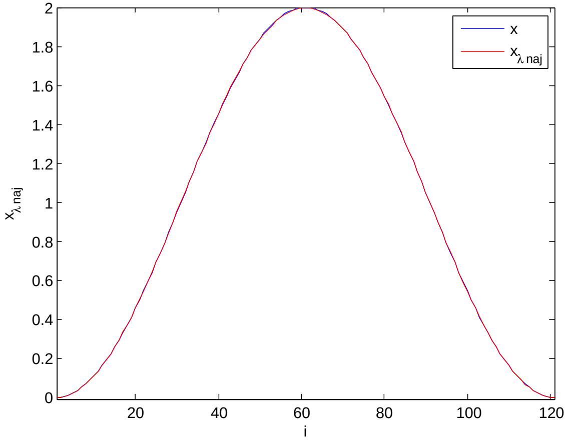
$$\|Ax_{\lambda naj} - b\|_2 \approx 20.1267.$$

Primjer (nastavak)

- *S druge strane je norma razlike* $x_{\lambda naj} - \bar{x}$*:*

$$\|x_{\lambda naj} - \bar{x}\|_2 \approx 0.0221.$$

- *U ovom slučaju norma reziduala je još malo narasla, ali greška je prihvatljivo mala.*



Slika: Egzaktno rješenje i rješenje regularizacije za λ_{naj}.

Krnja dekompozicija singularnih vrijednosti (TSVD)

- Za rješavanje loše uvjetovanih problema često se koristi krnja dekompozicija singularnih vrijednosti (TSVD), koja koristi aproksimaciju ranga $p < r = \text{rang}(A)$.
- Ako je $A = U\Sigma V^T$ SVD matrice A, tada je prema jednom teoremu za SVD

$$A_p = \sum_{i=1}^{p} \sigma_i u_i v_i^T,$$

najbolja aproksimacija ranga p matrice A.
- Za $m \geq n$, neka su matrice $U \in \mathbb{R}^{m\times m}$, $V \in \mathbb{R}^{n\times n}$ i $\Sigma \in \mathbb{R}^{m\times n}$ particionirane na sljedeći način

$$U = \begin{array}{c} \begin{bmatrix} U_1 & U_2 & U_3 \end{bmatrix} \\ \begin{array}{ccc} p & n-p & m-n \end{array} \end{array} \qquad V = \begin{array}{c} \begin{bmatrix} V_1 & V_2 \end{bmatrix} \\ \begin{array}{cc} p & n-p \end{array} \end{array} \qquad \Sigma = \begin{array}{cc} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \\ 0 & 0 \end{bmatrix} & \begin{array}{c} p \\ n-p \\ m-n \end{array} \\ \begin{array}{cc} p & n-p \end{array} & \end{array}$$

gdje je $\sigma_p > \zeta\sigma_1$ i $\sigma_{p+1} < \zeta\sigma_1$ za neku toleranciju ζ.

- Tada je
$$A_p = U_1 \Sigma_1 V_1^T = U(:, 1:p)\Sigma(1:p, 1:p)V(:, 1:p)^T.$$
- Rješavanje problema najmanjih kvadrata svodi se na minimizaciju $\|r_{svd}\|_2^2 = \|A\bar{x} - b\|_2^2$, gdje je
$$\|r_{svd}\|_2^2 = \|\Sigma_1 V_1^T \bar{x} - U_1^T b\|_2^2 + \|\Sigma_2 V_2^T \bar{x} - U_2^T b\|_2^2 + \|U_3^T b\|_2^2,$$
što je ekvivalentno minimizaciji prva dva izraza u gornjoj jednadžbi.
- TSVD postavlja $\sigma_i = 0$ za $i = p + 1, \ldots, n$ i minimizira samo prvi izraz.
- To je ekvivalentno rješavanju problema najmanjih kvadrata za matricu A_p
$$\min \|r_{tsvd}\|_2^2 = \min(\|\Sigma_1 V_1^T \tilde{x} - U_1^T b\|_2^2 + \|U_2^T b\|_2^2 + \|U_3^T b\|_2^2).$$
- Važno je odabrati pogodnu toleranciju ζ ili rang p, tako da norma reziduala i norma rješenja budu male.

- Rješenje pomoću TSVD je tada oblika

$$x_{tsvd} = \sum_{i=1}^{p} \frac{u_i^T b}{\sigma_i} v_i = V(:, 1:p)\Sigma(1:p, 1:p)^{-1} U(:, 1:p)^T b.$$

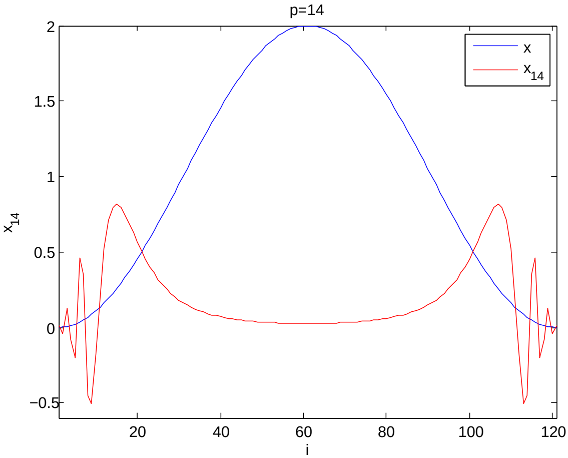
Primjer

- *Sada ćemo primijeniti TSVD ponovo na naš primjer.*
- *Prvo ponovo trebamo pogledati singularne vrijednosti matrice A, i uočiti indekse u kojima singularne vrijednosti padnu za jedan red veličine.*

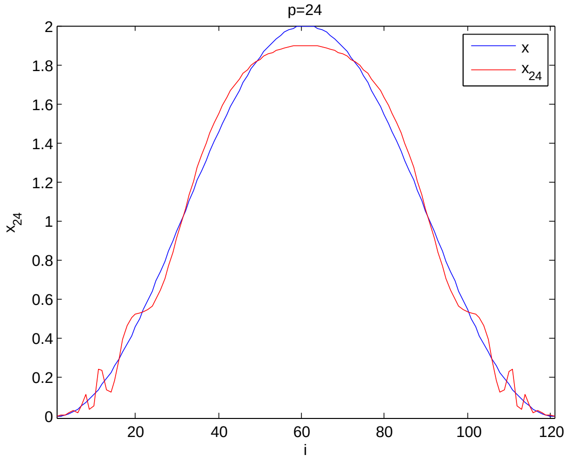
i	*1-2*	*3-4*	*5-8*	*9-14*	*15-24*
$\sigma_i \approx$	10^7	10^6	10^5	10^4	10^3

i	*25-36*	*37-51*	*52-68*	*69-78*	*79-121*
$\sigma_i \approx$	10^2	10^1	10^0	10^{-1}	10^{-2}

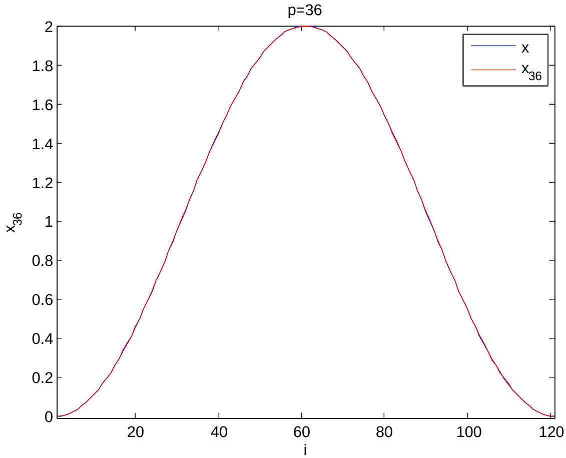
- *Dalje ćemo birati TSVD aproksimacije za* $p = 2, 4, 8, 14, 24, 36, 51, 68, 78, 121$, *i označiti ih sa* x_p.
- *Za aproksimacije* x_2, x_4, x_8 *i* x_{14}, *odgovarajući rang je ipak premali.*



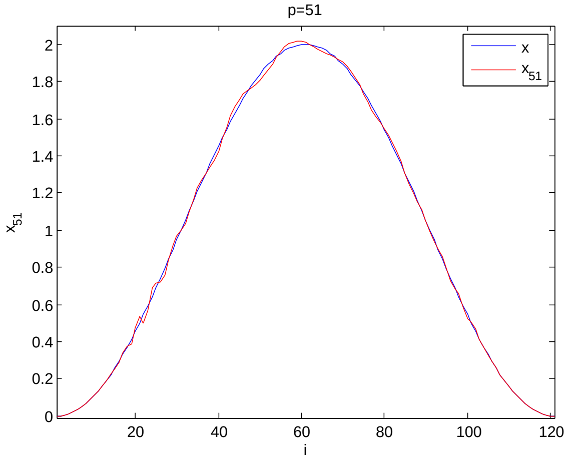
Slika: Egzaktno rješenje i TSVD rješenje za $p = 14$: $\|Ax_{14} - b\|_2 \approx 76515$, $\|x_{14} - \bar{x}\|_2 \approx 12.9085$.



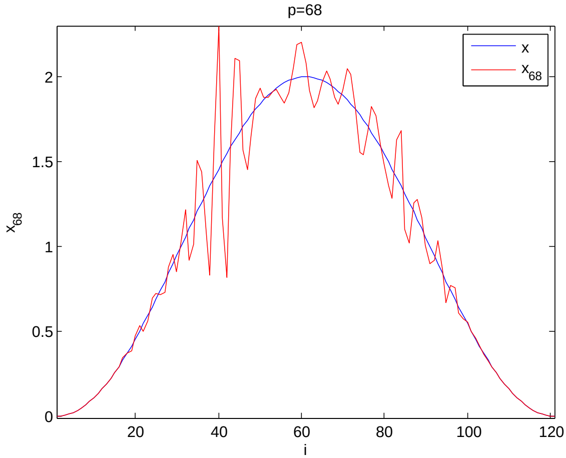
Slika: Egzaktno rješenje i TSVD rješenje za $p = 24$: $\|Ax_{24} - b\|_2 \approx 636.9091$, $\|x_{24} - \bar{x}\|_2 \approx 0.7172$.



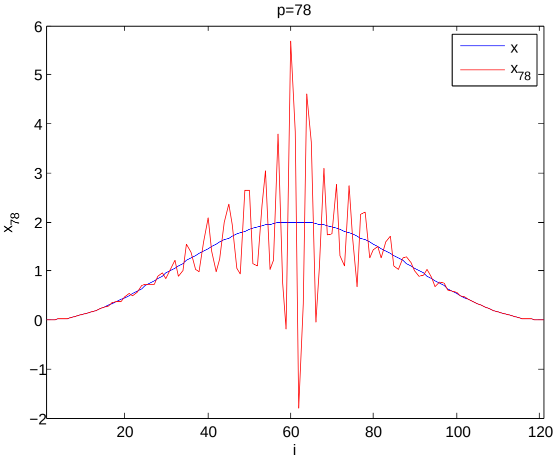
Slika: Egzaktno rješenje i TSVD rješenje za $p = 36$: $\|Ax_{36} - b\|_2 \approx 9.7512$, $\|x_{36} - \bar{x}\|_2 \approx 0.0144$.



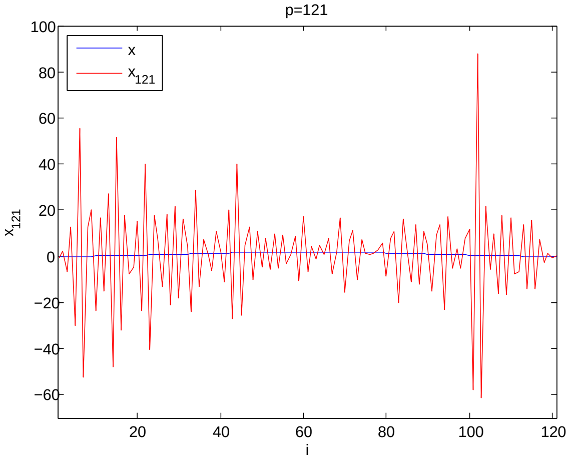
Slika: Egzaktno rješenje i TSVD rješenje za $p = 51$: $\|Ax_{51} - b\|_2 \approx 8.6826$, $\|x_{51} - \bar{x}\|_2 \approx 0.1774$.



Slika: Egzaktno rješenje i TSVD rješenje za $p = 68$: $\|Ax_{68} - b\|_2 \approx 7.5599$, $\|x_{68} - \bar{x}\|_2 \approx 1.8108$.



Slika: Egzaktno rješenje i TSVD rješenje za $p = 78$: $\|Ax_{78} - b\|_2 \approx 7.0118$, $\|x_{78} - \bar{x}\|_2 \approx 8.5269$.



Slika: Egzaktno rješenje i TSVD rješenje za $p = 121 = r$: $\|Ax_{121} - b\|_2 \approx 4.7955$, $\|x_{121} - \bar{x}\|_2 \approx 224.1275$.

Primjer (nastavak)

- *Možemo zaključiti da je najbolja aproksimacija postignuta za $p = 36$, i ona je čak i bolja od aproksimacije dobivene regularizacijom za λ_{naj}.*
- *Dakle, singularne vrijednosti matrice A manje od 10^2 možemo zanemariti i izjednačiti sa nulom:*

$$\sigma_i \leq 2.6365 \cdot 10^{-6}\sigma_1, \qquad i = 37, \dots, 121,$$

  i pri tome dobiti prilično zadovoljavajuću aproksimaciju rješenja.
- *To znači da singularni vektori vodećih singularnih vrijednosti koji formiraju matricu $A_{36} = U(:, 1:36)\Sigma(1:36, 1:36)V(:, 1:36)^T$ sadržavaju dovoljno informacija za rekonstrukciju matrice A,*

Primjer (nastavak)

$$\|A - A_{36}\|_2 = \sigma_{37} = 89.5076 \leq 2.6365 \cdot 10^{-6}\|A\|_2,$$

a pri tome je A_{36} bolje uvjetovana matrica od A

$$\kappa(A_{36}) = 3.1907 \cdot 10^5.$$

Generalizirani problem najmanjih kvadrata

- Kod linearne regresije koja povezuje Y i $X_1, \ldots, X_n$
$$Y_i = \beta_1 X_{i,1} + \cdots + \beta_n X_{i,n} + \epsilon_i,$$
pretpostavili smo da je
$$\mathrm{Cov}(\epsilon_1, \ldots, \epsilon_m) = \sigma_\epsilon^2 I_m.$$
- U slučaju kada $\mathrm{Cov}(\epsilon_1, \ldots, \epsilon_m)$ nije gornjeg oblika, a kada je ta matrica poznata do na skalirajući faktor
$$\mathrm{Cov}(\epsilon_1, \ldots, \epsilon_m) = \sigma^2 \mathbf{G},$$
tada rješavamo sustav normalnih jednadžbi **generaliziranog problema najmanjih kvadrata**
$$(\mathbf{X}^T \mathbf{G}^{-1} \mathbf{X})\mathbf{b} = \mathbf{X}^T \mathbf{G}^{-1} \mathbf{y}.$$
- Zbog toga se cijeli problem mora preformulirati u problem regresije sa matricom $\mathbf{G}^{-1/2}\mathbf{X}$ i vektorom $\mathbf{G}^{-1/2}\mathbf{y}$, gdje je $\mathbf{G}^{1/2}$ Choleski faktor matrice $\mathbf{G}$.