

---

# Statistička optimizacija nenegativne matrične faktorizacije

Lana Čaldarević  
Lana Reščec  
Petar Rudan

13. lipnja 2016.

---

---

## Sadržaj

<b>1</b>	<b>Uvod</b>	<b>3</b>
<b>2</b>	<b>Nenegativna matrična faktorizacija</b>	<b>4</b>
2.1	Metoda alternirajućih nenegativnih najmanjih kvadrata . . .	6
2.2	Metoda nenegativnih najmanjih kvadrata pomoću BPP-a . . .	7
<b>3</b>	<b>Statistička optimizacija - testiranje hipoteza</b>	<b>9</b>
<b>4</b>	<b>Rezultati</b>	<b>14</b>

---

## 1 Uvod

Nenegativna matrična faktORIZACIJA (NMF) je metoda dimenzionalne redukcije gdje je svaki element nenegativna vrijednost i koja kao rezultat daje aproksimaciju malog ranga s također nenegativnim elementima. Zahvaljujući upravo toj nenegativnosti, rezultati se mogu prirodno opisati: svaki podatak se može prikazati kao linearna kombinacija komponenti koje imaju fizičku interpretaciju. NMF metoda se pokazala vrlo korisnom za raznorazne probleme u strojnom učenju i data miningu i ima primjenu primjerice u područjima text mininga i bioinformatike. Jedan od najbržih algoritama NMF-a je Block Principal Pivoting metoda (BPP) kojom ćemo se mi i baviti. Optimizacija u svakoj iteraciji traži računanje velikog broja skupog problema najmanjih kvadrata. Naša metoda, koja je poboljšana verzija BPP metode, kreće s malim podskupom redaka i stupaca početne matrice i koristi hipotetičke testove na temelju kojih postupno povećava veličinu problema.

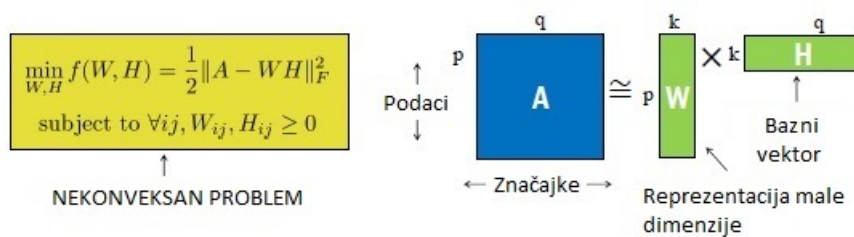
---

## 2 Nenegativna matrična faktorizacija

NMF matematički prikazujemo na sljedeći način. Neka je  $A \in \mathbb{R}^{m \times n}$  u kojoj je svaki element nenegativan i neka je  $k < \min\{m, n\}$ . NMF želi pronaći faktore  $W \in \mathbb{R}^{m \times k}$  i  $H \in \mathbb{R}^{k \times n}$  s nenegativnim elementima takve da je  $A \approx WH$ . Aproksimacija ove jednadžbe najčešće se mjeri Frobeniusovom normom, pa se postupak svodi na rješavanje sljedećeg nekonveksnog optimizacijskog problema:

$$\min_{W, H} f(W, H) = \frac{1}{2} \|A - WH\|_F^2$$

$$\forall i, j, W_{ij}, H_{ij} \geq 0$$



Slika 1: Grafički prikaz NMF-a

Rješavanju ovog problema pristupalo se na mnoge načine i razvijeni su brojni algoritmi. Neki od njih su ANLS (metoda alternirajućih nenegativnih najmanjih kvadrata), NNLS (metoda nenegativnih najmanjih kvadrata) i HALS (metoda hijerarhijskih alternirajućih najmanjih kvadrata). Mi se bavimo pristupom rješavanju NMF pomoću NNLS metode, odnosno BPP metode koja je jedan od najučinkovitijih NNLS algoritama.

BPP algoritam uključuje računanje  $W^T W, W^T A, H H^T, H A^T$ , u rješavanju problema najmanjih kvadrata, što može biti vrlo skupa operacija kada su  $p$  i/ili  $q$  veliki. Zbog toga ćemo reinterpretirati problem najmanjih kvadrata kao problem statističke procjene uz činjenicu da je ulazna matrica  $A$  generirana kroz stohastički proces.

Kako su  $W$  i  $H$  u ranim iteracijama daleko od ciljane vrijednosti odnosno dobre aproksimacije, nema previše smisla koristiti sve podatke da bismo ažurirali podatke u ANLS-u i tako baratati velikim problemom. Umjesto toga

primjenjujemo poduzorkovanje redaka  $W$  i  $A$  te stupaca  $H$  i  $A$ . Kod takvog poduzorkovanja, prema Centralnom graničnom teoremu znamo da je procjenjivanje najmanjih kvadrata asimptotički normalno, što nam omogućuje da oblikujemo statistički hipotetički test kojim možemo odrediti jesmo li dostigli zadovoljavajuće rješenje. Prednost ovog unaprjeđenja je prirodni kriterij zaustavljanja, upravo nivo značajnosti hipotetičkog testa.

---

## 2.1 Metoda alternirajućih nenegativnih najmanjih kvadrata

Predstaviti ćemo najprije algoritam ANLS-a za rješavanje NMF-a kako bi bio jasniji pristup u BPP metodi. U ovoj metodi najprije su varijable podijeljene u dvije grupe koje se zatim ažuriraju po redu. Algoritam je dakle:

1. Inicijaliziramo  $H \in \mathbb{R}^{k \times n}$  s nenegativnim elementima.
2. Ponavljati do kriterija zaustavljanja:

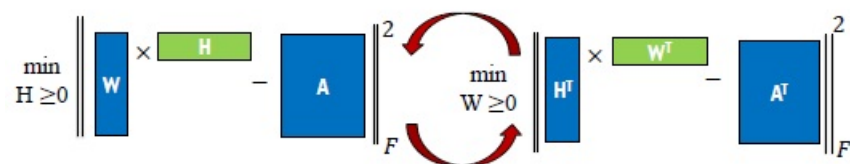
$$\min_{W \geq 0} \|H^T W^T - A^T\|_F^2$$

gdje je  $H$  fiksiran i zatim

$$\min_{H \geq 0} \|WH - A\|_F^2$$

gdje je  $W$  fiksiran.

3. Stupci matrice  $W$  su normirani u  $L_2$  normi, a retci matrice  $H$  skalirani prema tome.



Slika 2: Grafički prikaz ANLS-a

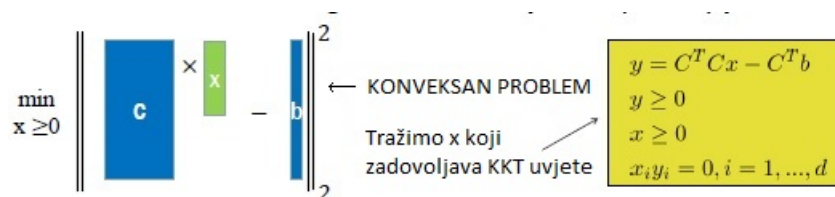
## 2.2 Metoda nenegativnih najmanjih kvadrata pomoću BPP-a

Primijetimo da su potproblemi u ANLS-u zapravo primjeri NNLS-a. Iako je početni problem nekonveksan, potproblemi su konveksni pa se istima mogu naći optimalna rješenja. Klasični algoritam za NNLS problem je *active set* metoda koja traži optimalne aktivne i pasivne skupove mijenjajući varijable među skupom kojim se barata u trenutnoj iteraciji. NNLS metoda se svodi na rješavanje konveksnog problema:

$$\min_{x \geq 0} \|Cx - b\|_2^2$$

gdje su  $C \in \mathbb{R}^{n \times d}$ ,  $x \in \mathbb{R}^{d \times 1}$  i  $b \in \mathbb{R}^{n \times 1}$  uz Karush-Kuhn-Tucker uvjete:

$$\begin{aligned} y &= C^T C x - C^T b \\ y &\geq 0 \\ x &\geq 0 \\ x_i y_i &= 0, i = 1, \dots, d \end{aligned}$$



Slika 3: Grafički prikaz NNLS-a

Ograničenja kod ovih pristupa je taj da se varijable mijenjaju uz uvjet da zadovoljavaju nenegativnost vektora rješenja pazeći na to da se objektna funkcija smanjuje u svakoj iteraciji. To rezultira time da se najčešće samo jedna varijabla mijenja po iteraciji, što čini algoritam prilično sporim kada je broj nepoznanica velik. Upravo BPP metoda rješava ovaj problem, što ju čini znatno bržom.

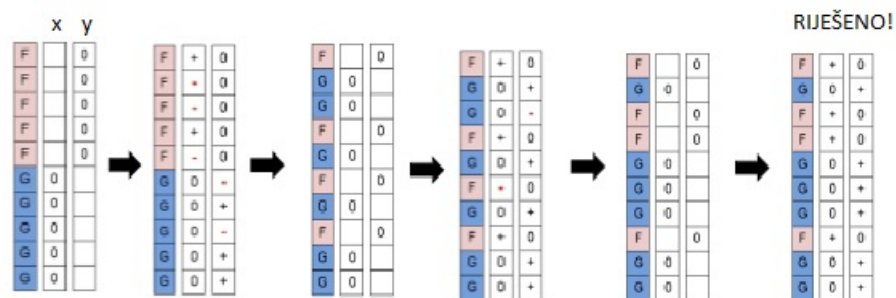
Kako bismo riješili NNLS problem BPP pristupom, prvi korak je podijeliti skup indeksa  $1, \dots, d$  na dva skupa  $F$  i  $G$  tako da u uniji čine cijeli indeksni skup, a da su pritom disjunktne. Označimo s  $x_F, x_G, y_F$  i  $y_G$  podskupove varijabli s odgovarajućim indeksima, te s  $C_F$  i  $C_G$  podmatrice od  $C$

s odgovarajućim indeksima stupaca. Vrijednosti  $x_G$  i  $y_F$  inicijaliziraju se na nulu, pa  $x = (x_F, x_G)$  i  $y = (y_F, y_G)$  zadovoljavaju zadnji od KKT uvjeta. Sada možemo riješiti jednadžbe:

$$x_F = (C_F^T C_F)^{-1} C_F^T b$$

$$y_G = C_G^T C_F x_F - C_G^T b$$

Ako su  $x_F \geq 0$  i  $y_G \geq 0$ , tada je rješenje  $x = (x_F, 0)$ . U suprotnom, skupovi  $F$  i  $G$  se ažuriraju mijenjanjem varijabli za koje taj uvjet ne vrijedi i proces se ponavlja dok se ne nađe rješenje.



Slika 4: Grafički prikaz BPP-a



### 3 Statistička optimizacija - testiranje hipoteza

Računamo  $x = (C^T C)^{-1} C^T b$ , gdje je  $C \in \mathbb{R}^{N \times P}$ ,  $b \in \mathbb{R}^{N \times 1}$ . Ako označimo  $i$ -ti redak matrice  $C$  sa  $c_i^T$ ,  $i$ -tu komponentu vektora  $b$  sa  $b_i$ , gornju jednakost možemo zapisati kao:

$$x_N = \left( \frac{1}{N} \sum_{i=1}^N c_i c_i^T \right)^{-1} \left( \frac{1}{N} \sum_{i=1}^N c_i b_i \right).$$

Vidimo da u svakoj iteraciji ANLS algoritma, ažuriramo vrijednost od  $x$  rješavanjem problema najmanjih kvadrata. Kako broj redaka matrice  $C$  i broj komponenti vektora  $b$ ,  $N$ , može biti jako velik, računanje  $C^T C$  i  $C^T b$  može biti vrlo skupo.

Presudni korak kojim bi se smanjila količina računanja je interpretirati retke matrice  $C$  kao uzorke iz stohastičkog procesa pa možemo uvesti stohastički proces  $b = c^T x + \epsilon$ , gdje je  $\epsilon$  greška. Tada  $x_N$  postaje maksimalni procjenitelj vjerojatnosti uz pretpostavku da su greške  $\epsilon$  zavisne i jednako distriburiane sa očekivanjem 0 i konstantom varijancom.

Ono što ćemo mi napraviti je da ćemo u svakoj iteraciji aproksimirati  $x_N$  sa  $N_t < N$  uzoraka od  $c$  i  $b$ .

$$x_{N_t} = \left( \frac{1}{N_t} \sum_{i=1}^{N_t} c_i c_i^T \right)^{-1} \left( \frac{1}{N_t} \sum_{i=1}^{N_t} c_i b_i \right).$$

Ako pretpostavimo da vrijedi sljedeće:

$$\mathbb{E}[cc^T] \text{ pozitivno definitno,}$$

$$\mathbb{E}[\epsilon|c] = 0,$$

$$\text{Var}[\epsilon|c] = \sigma^2$$

tada je procjenitelj  $x_N$  za  $x$  ( $x \sim N(\mu, \sigma^2)$ ) asimptotički normalan (slijedi iz teksta autora M. Verbeeka: A GUIDE TO MODERN ECONOMETRICS, 2000.):

$$x_{N_t} \sim N(\mu, \Sigma), \quad \Sigma = \frac{\mathbb{E}[cc^T]^{-1} \sigma^2}{N_t}.$$

Varijanca  $\Sigma$  je inverzno proporcionalna broju uzoraka  $N_t$ , te možemo primjeniti kako je  $x_{N_t}$  precizniji što je broj uzoraka  $N_t$  veći.

Tu distribuciju možemo aproksimirati kao

$$x_{N_t} \sim N(\mu_t, \Sigma_t),$$

gdje je:

$$\begin{aligned}\mu_t &= \left( \frac{1}{N_t} \sum_{i=1}^{N_t} c_i c_i^T \right)^{-1} \left( \frac{1}{N_t} \sum_{i=1}^{N_t} c_i b_i \right), \\ Q_{cc,t} &= \frac{1}{N_t - 1} \sum_{i=1}^{N_t} c_i c_i^T, \\ \sigma_t^2 &= \frac{1}{N_t - 1} \sum_{i=1}^{N_t} \epsilon_i^2, \\ \Sigma_t &= \frac{Q_{cc,t}^{-1} \sigma_t^2}{N_t}.\end{aligned}$$

Procjenjivanje ove distribucije za predloženo ažuriranje aproksimacije  $x_{N_t}$  omogućuje nam korištenje statističkih hipoteza za određivanje pouzdanosti te aproksimacije. Računamo  $\rho_t$ , vjerojatnost da predložen smjer ažuriranja nije unutar  $90^\circ$  od prave vrijednosti aproksimacije  $\mu - x_{N_{t-1}} \approx \mu_t - x_{N_{t-1}}$ , gdje je  $x_{N_{t-1}}$  aproksimacija iz prethodne iteracije. Sa  $p_{N_t}$  označimo funkciju gustoće distribucije  $N(\mu_t, \Sigma_t)$  predložene aproksimacije,  $\rho_t$  računamo kao:

$$\rho_t = \int_{\Omega} p_{N_t}(x) dx, \quad \Omega = \{x : \langle x - x_{N_{t-1}}, \mu_t - x_{N_{t-1}} \rangle < 0\}.$$

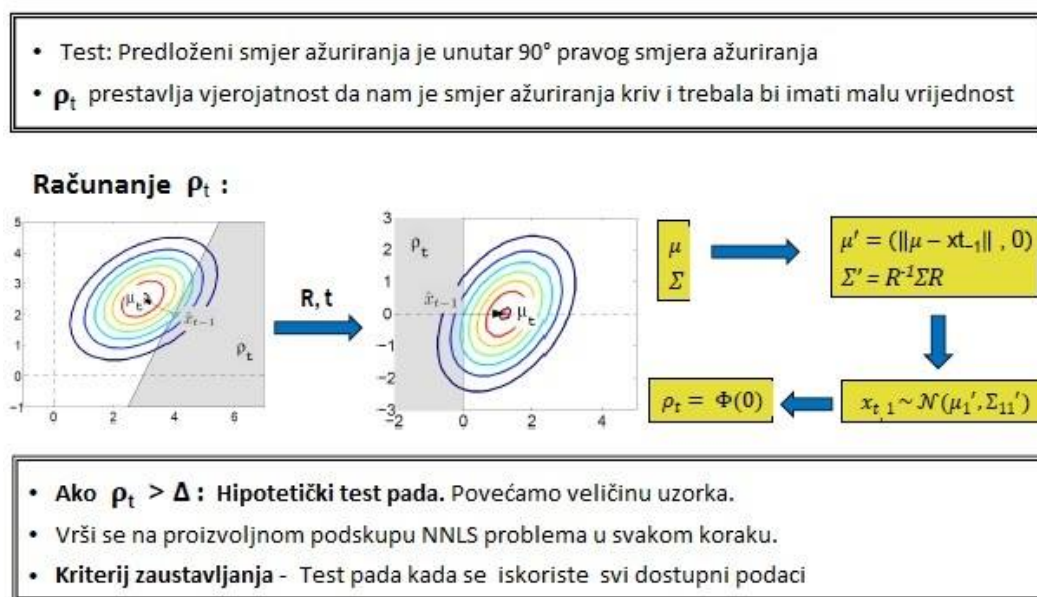
Kako bismo  $\rho_t$  izračunali efikasno, uvodimo transformaciju koordinatnog sustava tako da je  $x_{N_{t-1}}$  ishodište, a  $\mu_t - x_{N_{t-1}}$ , pravi smjer ažuriranja, je uz prvu koordinatnu os. Uz ovu transformaciju, distribucija predložene aproksimacije se mijenja, te imamo:

$$x_{N_t} \sim N(\mu'_t, \Sigma'_t), \quad \mu'_t = [\|\mu_t - x_{N_{t-1}}\|; 0_{P-1}], \quad \Sigma'_t = R^{-1} \Sigma_t R,$$

gdje je  $0_{P-1}$  vektor od  $P - 1$  nula, a  $R$  rotacijska matrica takva da je  $R(\mu_t - x_{N_{t-1}}) = \mu'_t$ .

Sada promatramo hiperravninu koja prolazi kroz ishodište i ortogonalna je na prvu koordinatnu os. To razdvaja  $\mathbb{R}^P$ , na dva poluprostora,  $\mathbb{R}_+^P$  i  $\mathbb{R}_-^P$ . Sada je  $\rho_t$  vjerojatnost "pada" u prostor  $\mathbb{R}_-^P$  i možemo ju lako izračunati pomoću marginalne distribucije od  $x_{N_t}(1)$ , prve komponente od  $x_{N_t}$ , koja je  $N(\mu'_t(1), \Sigma'(1, 1))$ , i  $\rho_t = \Phi(0)$ , gdje je  $\Phi(\cdot)$  kumulativna funkcija distribucije od  $x_{N_t}(1)$ .

Ako je  $\rho_t$  veće od praga  $\rho^*$ , postoji vjerojatnost da je predloženi smjer ažuriranja značajno drugačiji od pravog smjera i odbacujemo hipotezu. Tada moramo povećati veličinu uzoraka kako bi smanjili varijancu u ažuriranjima tako da ne odbacujemo daljne hipoteze. Ako je to nemoguće (npr. primjenili smo hipoteze na sve dostupne uzorke), prekidamo sa radom algoritma.



Slika 5: Testiranje hipoteza

Mana ovog pristupa je ta da pretpostavka o normalnosti ne vrijedi nužno za jako raspršene matrice pa je tako ponašanje našeg algoritma u tim slučajevima nepredvidivo. Kako raste broj uzoraka,  $N$ , shodno opada nepouzdanost procjene od  $x$ . Upravo to je razlog zašto u početnim iteracijama nam je dovoljan manji uzorak kako bismo odredili smjer ažuriranja, a u daljnjim iteracijama prema konvergenciji, koristimo veći uzorak kako bismo smanjili

nepouzdanost rješenja. Operacije koje se trebaju izvesti sa svakom iteracijom tako su smanjene.

U konačnici naš algoritam izgleda ovako:

1. Inicijalizacija  $S_W$  i  $S_H$ , početnih uzoraka veličina  $W$  i  $H$ , na djelove  $p$  i  $q$ .
2. Inicijalizacija  $W \in \mathbb{R}^{m \times k}$  s nenegativnim elementima.
3. Proizvoljno permutirati retke i stupce matrice  $A$ , tako da odabir prvih  $n$  redaka ili stupaca od  $A$  predstavlja crtanje  $n$  proizvoljnih primjera.
4. Ponavljati rješavanje dva potproblema do konvergencije:

(a)  $\min_{H \geq 0} \|WH - A\|_F^2$  gdje je  $W$  fiksiran:

i. dok je  $S_W < p$ :

- A. Naći rješenje za podmatricu  $J$  proizvoljno odabranih stupaca iz  $H$  koristeći BPP metodu uz korištenje samo  $S_W$  redaka  $W$  i  $A$ .
- B. Podvrgnuti matricu  $J$  hipotetičkom testu kako bi se vidjelo je li rješenje odgovarajuće, tj. ažurira li se matrica u pravom smjeru.
- C. Ako test ne prođe, ažuriramo  $S_W = \min(p, 2S_W)$  i budući da smo povećali  $S_W$  s novim uzorcima matrice  $W$  koristeći  $H$  iz prethodne iteracije ponovno pokušamo izračunati matricu. S druge strane, kada test prođe izlazimo iz petlje.

ii. Rješavamo problem za prvih  $S_H$  stupaca matrice  $H$  koristeći  $S_W$  redaka od  $W$  i  $A$ .

(b)  $\min_{W \geq 0} \|H^T W^T - A^T\|_F^2$  uz  $H$  fiksiran analogno kao u (a).

---



## 4 Rezultati

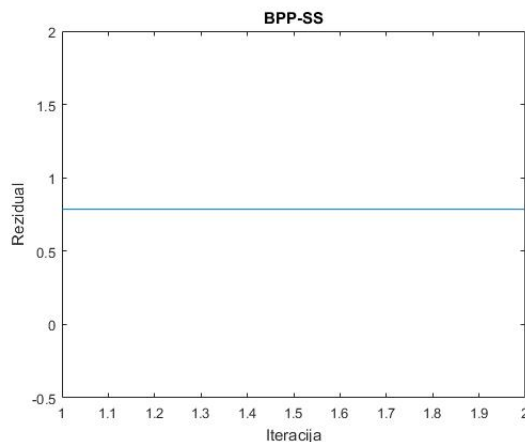
Slijedi prikaz rezultata naše metode. Algoritam smo implementirali u MATLAB-u te smo ga testirali na random podacima iz MATLAB-a.

Rezidual odnosno relativnu grešku ( $\|A - WH\|_F / \|A\|_F$ ) smo mjerili u svakom koraku algoritma. Također smo gledali potreban broj iteracija te usporedili sa brojem iteracija potrebnih za metodu alternirajućih nenegativnih najmanjih kvadrata. Metodu alternirajućih nenegativnih najmanjih kvadrata nismo kodirali, nego smo koristili gotovu MATLAB-ovu funkciju *nnmf*.

Prvo smo izračunali potreban broj iteracija za izvršenje našeg programa te smo sa dobivenim brojem iteracija pokrenuli MATLAB-ovu funkciju *nnmf* te usporedili dobivene rezultate.

Treba napomenuti kako metoda možda neće iskonvergirati k rješenju radi korištenja statističkih hipoteza. Na sljedećoj slici možemo vidjeti jedan takav primjer.

U testovima smo uzimali da je prag za odbacit hipotezu  $\rho^* = 0.4$ , te smo odabrali  $J = 10$ .

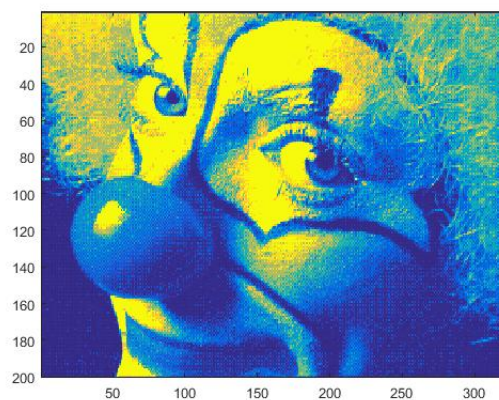


Slika 7: Primjer odbačenih hipoteza

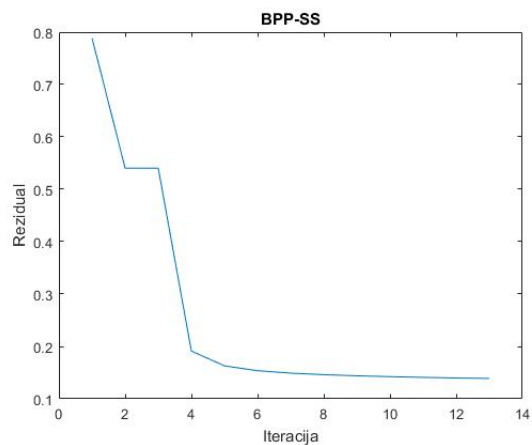
Uzeli smo matricu dimenzija  $p = 200$ ,  $q = 320$ , te smo za  $k$  uzeli 50. Primjetimo kako su hipoteze odbačene u svakom koraku algoritma, te iz tog razloga vrijednosti matrica  $W$  i  $H$ , koje su zadane kao random nenegativne

matrice iz uniformne distribucije, nisu promjenjene. Rezidual ostaje isti kao i na početku te nam daje do znanja da metoda nije iskonvergirala k rješenju.

U sljedećem primjeru možemo vidjeti uspješnu izvedbu naše metode uz iste podatke kao i maloprije.



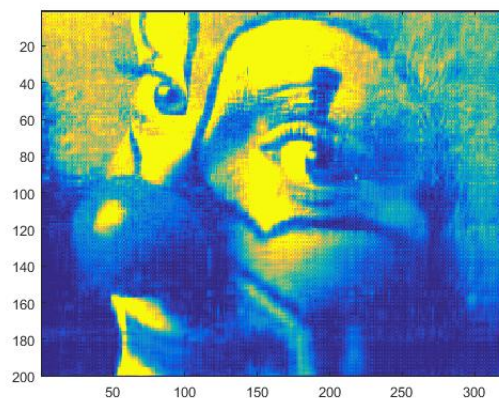
Slika 8: Matrica  $A$ , dimenzije  $200 \times 320$



Slika 9: Konvergencija metode k aproksimativnom rješenju

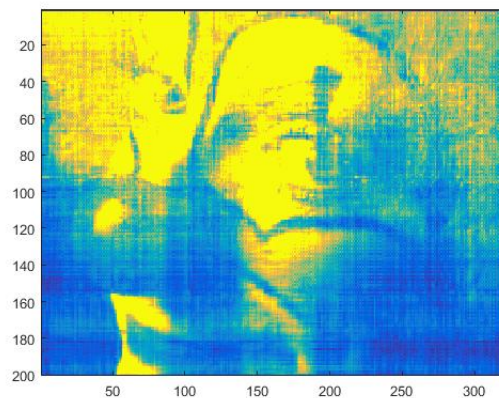
U samo 14 iteracija dobili smo aproksimativnu faktorizaciju matrice  $A$ , te smo dobili relativnu grešku **0.1389**. Pogledajmo aproksimativno rješe-

nje na slici dobiveno pomoću BPPSS kako bismo mogli bolje usporediti sa originalnom slikom.



Slika 10: BPPSS

Sada pogledajmo metodu alternirajućih najmanjih kvadrata sa maksimalnim brojem iteracija 14, koliko je bilo potrebno BPPSS-u da iskonvergirao, radi usporedbe.



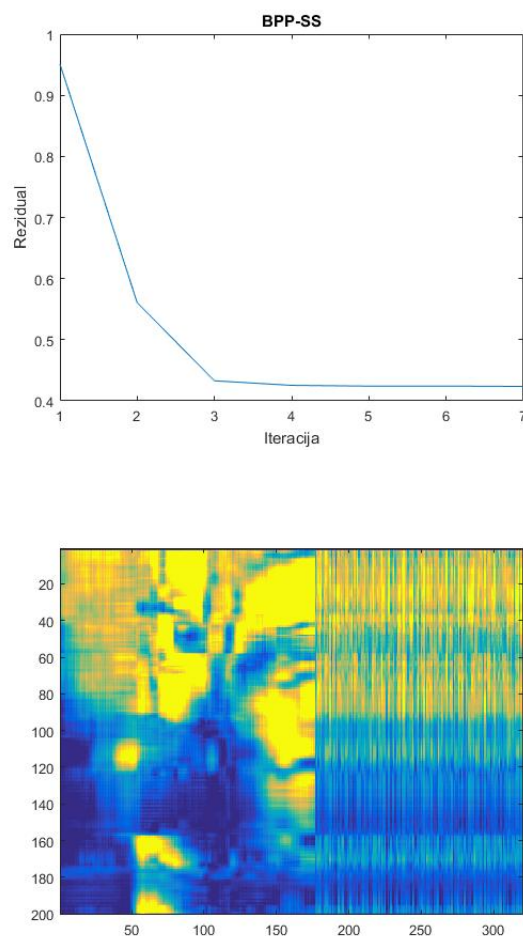
Slika 11: NMF

Dobiveni rezidual za NMF iznosi **0.4579**. Možemo zaključiti kako je u ovom slučaju BPPSS puno brže iskonvergirao aproksimativnom rješenju



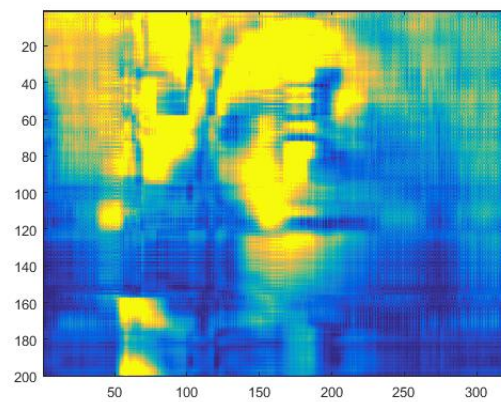
nego NNMF.

Sljedeći primjer pokazuje uspješno konvergiranje našeg algoritma k aproksimativnom rješenju, no nakon treće iteracije nema više promjena u ažuriranjima, što nas ne dovodi do dovoljno dobrog aproksimativnog rješenja što možemo vidjeti i na donjoj slici:



Slika 12: BPPSS

Dok je NNMF za 7 iteracija uspio naši poprilično dobro aproksimativno rješenje.



Slika 13: NMF

---

## Literatura

- [1] STATISTICAL OPTIMIZATION OF NON-NEGATIVE MATRIX FACTORIZATION: A. Korattikara, L. Boyels, M. Welling, J. Kim, H. Park
  - [2] FAST NONNEGATIVE MATRIX FACTORIZATION: AN ACTIVE-SET-LIKE METHOD AND COMPARISONS: J. Kim, H. Park
  - [3] MULTIVARIATE HYPOTHESIS TESTS FOR STATISTICAL OPTIMIZATION: A. Korattikara
-