Metoda nelinearnih najmanjših kvadratov

Poročilo 1. vaje

Jernej Sabadin

Univerza *v Ljubljani*
Fakulteta *za elektrotehniko*

Mentorja: Prof. Dr. Igor Škrjanc, Univ. Dipl. Inž. El., doc. dr. Dejan Dovžan

`Predmet: Inteligentni sistemi za podporo odločanju`
`Datum: 10. Marec 2023`

Contents

Ključne besede: Nelinearni najmanjši kvadrati, Levenberg-Marquad optimizacija

---

1 Naloge

- Opis problema
- Teoretičnno ozadje
- Implementacija algoritmov v okolju Matlab
- Analiza rezultatov

2 Opis problema

V nalogi želimo poiskati lokacijo izgubljenega mobilnega telefona. Mobitel se nahaja nekje v koordinatnem sistemu (slika 1). Mobitel so zaznale tri bazne postaje. Bazne postaje nam podajo oceno razdalje do telefona. Do teh razdalj dostopamo z matlabovimi funkcijami ($ping\_stolp\_1, ping\_stolp\_2, ping\_stolp\_3$), ki vrnejo ob vsakem klicu razdaljo do izgubljenega telefona. Naloga je čim bolj natančno določiti koordinate mobilnega telefona.

Problem lahko predstavimo na spodnji sliki. Krožnice imajo radije enake razdaljam do izgubljenega telefona, ki jih pridobimo s klicem zgoraj omenjenih funkcij. Na spodnji sliki smo funkcije klicali večkrat, zato je na sliki tudi več krožnic. Center vsake krožnice pripada koodrinatam bazne postaje.

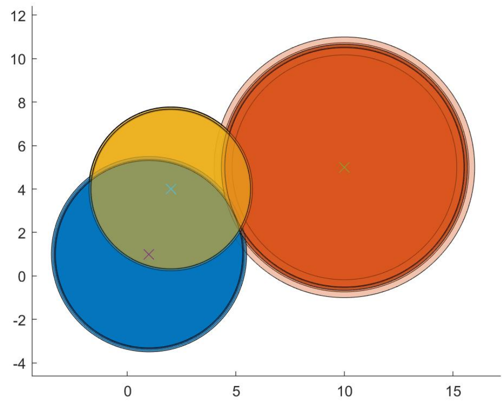

Fig. 1: Izris lokacij baznih postaj z x-i in njihovih domnevnih razdalj do izgubljenega telefona, predstavljenih z krožnicami

3 Teoretični uvod

V nalogi bomo lokacijo telefona poiskali z izbrano metodo nelinearne lokalne optimizacije. Nelinearna lokalna optimizacija se uporablja za iskanje optimalne rešitve kriterijske fukcije $I(\underline{\theta})$, katere gradient je nelinearen v parameterih $\underline{\theta}$. Vključuje iskanje optimuma kriterijske funkcije v majhnem območju prostora parametrov okoli začetne točke iskanja.

V nalogi bomo kriterijsko funkcijo napake zapisali z enačbo

$$I(\underline{\theta}) = \sum_{i=1}^{N} f^2(i, \underline{\theta}) \tag{1}$$

oziroma v vektroski obliki

$$I(\underline{\theta}) = \underline{f}^T \underline{f} \quad \text{pri} \quad \underline{\mathrm{f}} = [f(1,\underline{\theta}), f(2,\underline{\theta}), ..., f(N,\underline{\theta}),]^T \tag{2}$$

kjer je

$$\underline{f} = \underline{y} - \underline{\hat{y}}(\underline{\theta}) \tag{3}$$

razlika med merjenim izhodom $\underline{y}$ in izhodom modela $\underline{\hat{y}}(\underline{\theta})$, ki je nelinearen v parameterih.

Enačbo (1) imenujemo problem nelinearnih najmanjših kvadratov. Metoda nelinearnih najmanjših kvadratov služi za prilagajanje matematičnega modela nizu podatkovnih točk. Model je funkcija ene ali več spremenljivk, cilj metode pa je najti vrednosti teh spremenljivk, ki minimizirajo razliko med modelom in dejanskimi vrednostmi podatkov. Optimizacijskega problema ni mogoče rešiti z uporabo preprostih algebraičnih tehnik, kot je to mogoče v primeru linearnih najmanjših kvadratov. Namesto tega se uporablja iterativni pristop, pri katerem se vrednosti parametrov posodabljajo zaporedno, da se minimizira kriterijska funkcija.

V nadaljevanju opišemo postopek iskanja optimalnih parametrov v smislu iskanja optimuma kriterijske funkcije (1). Postopek je natačno opisan v knjijigi I. Škrjanca [1].

Z razvojem gradienta po Taylorju dobimo naslednji zapis

$$\underline{\mathrm{g}}_k \approx \underline{\mathrm{g}}_{k-1} + \underline{\mathrm{H}}_{k-1}\Delta\underline{\theta}_{k-1} \tag{4}$$

Kjer je H Hessova matrika drugih delnih odvodov kriterijske funkcije. Opisuje ukrivljenost kriterijske funkcije okoli trenutne točke v prostoru parametrov.

V nadaljevanu upoštevamo, da je gradient $I(\underline{\theta})$ glede na parameter θ_j naslednji:

$$g_j = \frac{\partial I(\underline{\theta})}{\partial \theta_j} = 2\sum_{i=1}^{N} f(i)\frac{\partial f(i)}{\partial \theta_j} \tag{5}$$

in da je gradient v vektorski obliki

$$\underline{\mathrm{g}} = 2\underline{J}^T \underline{f} \tag{6}$$

kjer je J Jacobijeva matrika, ki vsebuje odvode vektorske funkcije $\underline{f}$ po vseh parametrih $\underline{\theta}$.

$$\underline{J} = \begin{bmatrix} \frac{\partial f(1)}{\partial \theta_1} & \cdots & \frac{\partial f(1)}{\partial \theta_n} \\ \vdots & & \vdots \\ \frac{\partial f(N)}{\partial \theta_1} & \cdots & \frac{\partial f(N)}{\partial \theta_n} \end{bmatrix} \tag{7}$$

S pomočjo enačbe (4) lahko sedaj zapišemo naslednje:

$$2\underline{J}_k^T \underline{f}_k = 2\underline{J}_{k-1}^T \underline{f}_{k-1} + 2\underline{J}_{k-1}^T \underline{J}_{k-1}\Delta\underline{\theta}_{k-1} \tag{8}$$

kjer $2\underline{J}_{k-1}^T\underline{J}_{k-1}$ predstavlja aproksimacijo Hessove matrike v enačbi (4).

Ker iterativno iščemo ničlo gradienta, lahko levo stran enačbe postavimo na nič. Tako dobimo

$$\Delta\underline{\theta}_{k-1} = -(\underline{J}_{k-1}^T\underline{J}_{k-1})^{-1}\underline{J}_{k-1}^T \underline{f}_{k-1} \tag{9}$$

$$\underline{\theta}_k = \underline{\theta}_{k-1} + \Delta\underline{\theta}_{k-1} \tag{10}$$

V naši nalogi bomo uporabili Levenberg-Marquardtovo metodo:

$$\Delta\underline{\theta}_{k-1} = -(\underline{J}_{k-1}^T\underline{J}_{k-1} + \alpha_{k-1}\underline{I})^{-1}\underline{J}_{k-1}^T \underline{f}_{k-1} \tag{11}$$

kjer je člen $\alpha_{k-1}\underline{I}$ regularizacijski parameter in rešuje problem slabe pogojenosti matrike $\underline{J}_{k-1}^T\underline{J}_{k-1}$.

Inverz matrike v enačbi (11) se pridobi z reševanjem

$$(\underline{J}_{k-1}^T\underline{J}_{k-1} + \alpha_{k-1}\underline{I})\underline{\mathrm{p}}_{k-1} = (\underline{J}_{k-1}^T\underline{f}_{k-1}) \tag{12}$$

Za majhne vrednosti α_{k-1}, se Levenberg-Marquardt algoritem približuje Gauss-Newtnovem algoritmu opisan v enačbi (9) in (10). Za večje vrednosti α_{k-1} pa se približuje metodi najhitrejšega spusta

$$\underline{\theta}_k = \underline{\theta}_{k-1} - \eta_{k-1}\underline{p}_{k-1} \quad \text{kjer je} \quad \underline{\mathrm{p}}_{k-1} = \underline{I}\,\underline{\mathrm{g}}_{k-1} \tag{13}$$

Blizu točke optimuma nam da Gauss-Newtnova metoda dobre rezultate, zato vzamemo majhen α_{k-1}, daleč stran od optimuma pa lahko Gauss-Newtnova metoda divergira, zato vzamemo velik α_{k-1}. Tako je za dovolj velike α_{k-1} matrika $\underline{J}_{k-1}^T\underline{J}_{k-1} + \alpha_{k-1}\underline{I}$ pozitivno definitna in gibanje v smeri optimuma je zagotovljeno.

Parameter α_{k-1} nastavljamo tako, da je na začetku neko majhno število. Potem pa se parameter α_{k-1} zmanjšuje, če se bližamo optimumu oziroma povečuje, če se vrednost kriterijske funkcije povečuje.

4 Implementacija algoritma v okolju Matlab

Najprej definiramo vektorsko funkcijo f.

$$f(x_1,x_2) = \begin{bmatrix} w_1(\sqrt{(x_1-p_{1,1})^2+(x_2-p_{1,2})^2} - d_{1,1}) \\ w_2(\sqrt{(x_1-p_{2,1})^2+(x_2-p_{2,2})^2} - d_{2,1}) \\ w_3(\sqrt{(x_1-p_{3,1})^2+(x_2-p_{3,2})^2} - d_{3,1}) \\ w_1(\sqrt{(x_1-p_{1,1})^2+(x_2-p_{1,2})^2} - d_{1,2}) \\ w_2(\sqrt{(x_1-p_{2,1})^2+(x_2-p_{2,2})^2} - d_{2,2}) \\ w_3(\sqrt{(x_1-p_{3,1})^2+(x_2-p_{3,2})^2} - d_{3,2}) \\ \vdots \\ w_1(\sqrt{(x_1-p_{1,1})^2+(x_2-p_{1,2})^2} - d_{1,N}) \\ w_2(\sqrt{(x_1-p_{2,1})^2+(x_2-p_{2,2})^2} - d_{2,N}) \\ w_3(\sqrt{(x_1-p_{3,1})^2+(x_2-p_{3,2})^2} - d_{3,N}) \end{bmatrix}$$

Na spodnji sliki prikažemo obliko kriterijske funkcije, izračunane po enačbi (1).

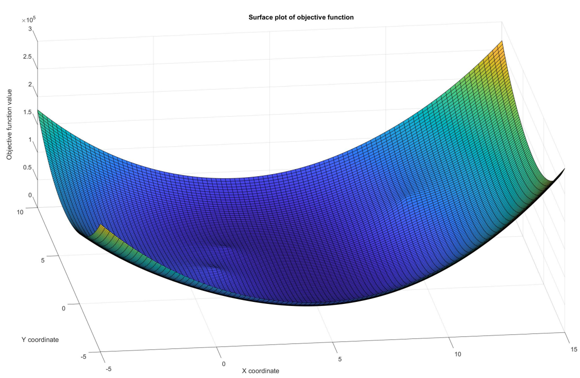

Fig. 2: Izris kriterijske funkcije

Prikažemo tudi konturni diagram.

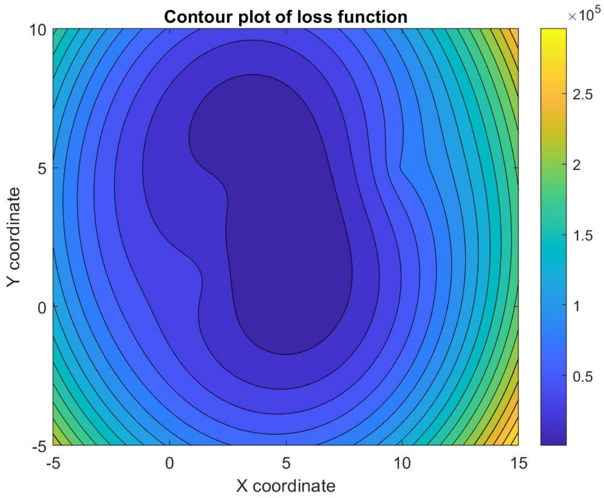

Fig. 3: Konturni diagram kriterijske funckcije

Optimizacije se lotimo z Levenberg-Marquardtovo metodo. Za začetno točko algoritma izberemo $(x_1, x_2) = (5, 5)$. Algoritem se zaključi, ko je norma razlike med dvema zaporednjima vektorjema parametrov dovolj majhna, oziroma ko je norma vektorja f dovolj majhna. V našem primeru vzamemo za toleranco 10^{-6}. V Levenberg-Marquardtovemu algoritmu omejimo tudi parameter α, tako da ne naraste na zelo velika števila.

Najprej se lotimo optimizacije kjer ne upoštevamo uteži ($w = [w_1, w_2, w_3] = [1, 1, 1]$).

Na spodnjih slikah prikažemo kako se kordinati x in y oziroma parametra x_1 in x_2 spreminjata z večanjem števila meritev. Na sliki izrišemo tudi vrednosti parametrov $\pm$ vrednosti varianc parametrov.

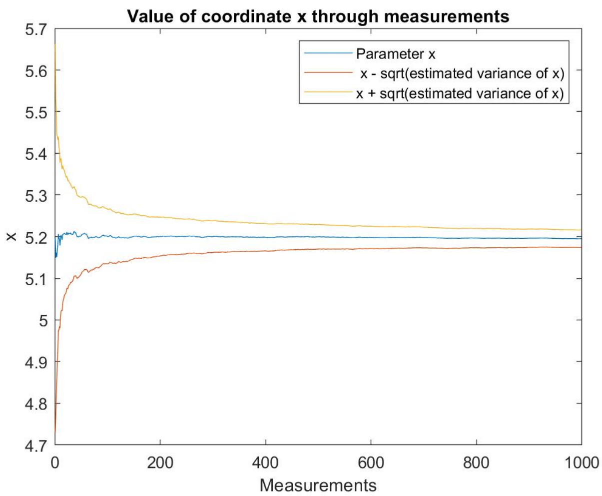

Fig. 4: Spreminjanje parametra x in njegove variance skozi meritve

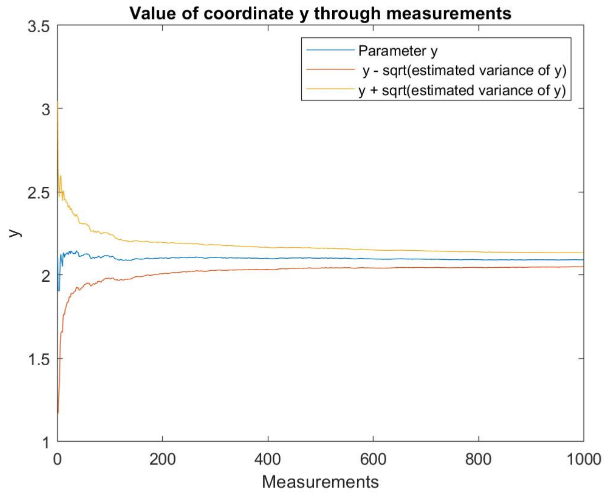

Fig. 5: Spreminjanje parametra y in njegove variance skozi meritve

Variance smo ocenili s pomočjo SVD (Singular Value Decomposition) razcepa Jacobijeve matrike. SVD razcep in ocena kovariančne matrike parametrov:

$$J = USV^T$$

$$\text{cov}(\hat{\theta}) = VS^{-2}V^T$$

kjer so U dimenzij $N \times N$ in V dimenzij $n \times n$ ortogonalni matriki, S pa je diagonalna matrika dimenzij $m \times n$, katere diagonalni elementi so pozitivne singularni vrednosti matrike J. Metoda SVD je numerično stabilna in primerna za probleme, ki vključujejo tanke matrike.

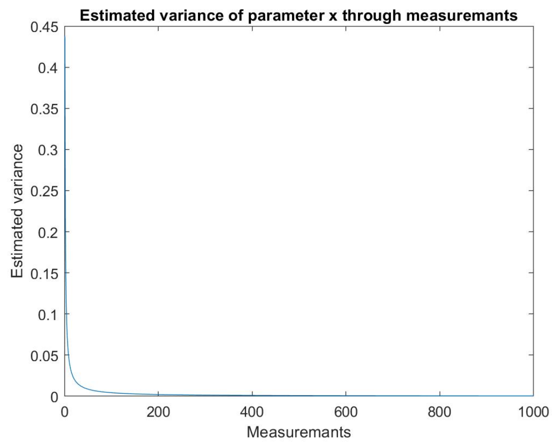

Fig. 6: Spreminjanje ocene variance parametra x skozi meritve

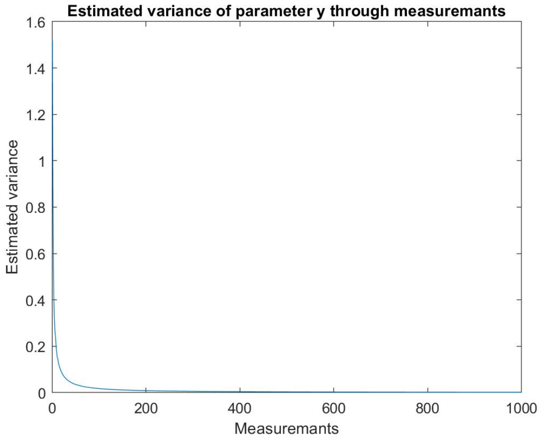

Fig. 7: Spreminjanje ocene variance parametra y skozi meritve

Opazimo, da pade varianca parametrov po približno 200 meritvah na vrednost 10^{-4} in tam ostane. To je vidno tudi na slikah (4) in (5).

Z upoštevanjem vseh 3000 meritev oziroma $N = 1000$ dobimo naslednje rezultate.

```
- Estitamed coordinates are: (5.195181e+00, 2.091816e+00).
- Mean of the error is -2.181051e-03.
- error variance is 3.549945e-02
- Estimated variance of the parameter x is 4.341564e-04 and variance of the parameter y is 1.756894e-03.
- Number of iterations: 6
```

Fig. 8: rezultati

Opazimo, da je varianca parametrov nekje v razredu 10^{-4}, povprečna vrednost vektorja $\underline{f} = \underline{y} - \underline{\hat{y}}(\underline{\theta})$ pa tudi nekje 10^{-4}.

Prikažemo tudi, kako se je povprečna vrednost $\underline{f}$ spreminjala z večanjem meritev.

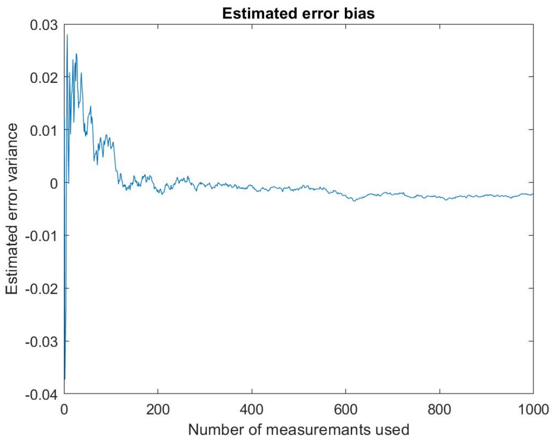

Fig. 9: Povprečna vrednost napake skozi meritve

prikažemo tudi kako se je varianca napake spreminjala z večanjem meritev.

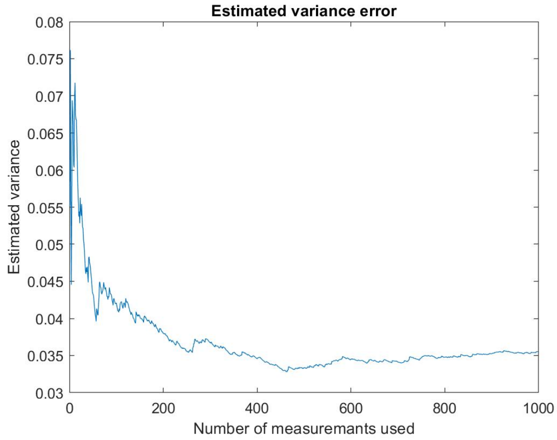

Fig. 10: Varianca napake skozi meritve

4.1 Uteženi nelinearni najmanjši kvadrati

V nadaljevanju smo upoštevali variance meritev v metodi optimizacije. Običajno se predpostavlja, da so meritve pošumljene z belim šumom. Tako smo uteži nastavili na vrednosti ($w = [w_1, w_2, w_3] = [\frac{1}{\sigma_1}, \frac{1}{\sigma_2}, \frac{1}{\sigma_3}] \cdot (\sum_{i=1}^{3} \frac{1}{\sigma_i})^{-1}$, kjer so $\frac{1}{\sigma_1}, \frac{1}{\sigma_2}, \frac{1}{\sigma_3}$ standarni odkloni pošumljenih meritev razdalj.

S tako izbiro uteži dobimo naslednje rezultate.

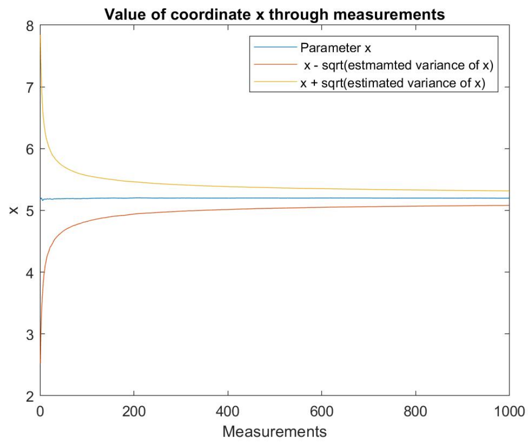

Fig. 11: Spreminjanje parametra x in njegove variance skozi meritve

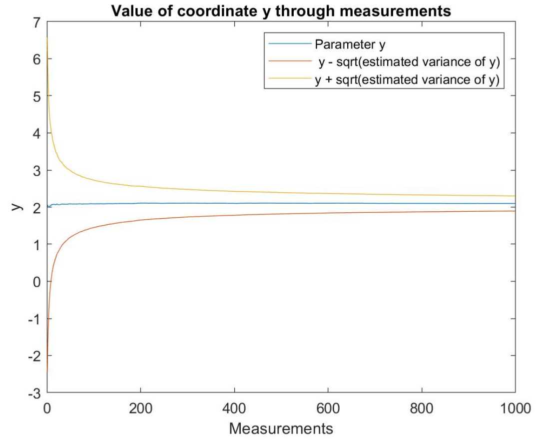

Fig. 12: Spreminjanje parametra y in njegove variance skozi meritve

Z upoštevanjem vseh 3000 meritev oziroma $N = 1000$ dobimo naslednje rezultate.

```
- Estitamed coordinates are: (5.197755e+00, 2.098919e+00).
- Mean of the error is -7.046590e-05.
- error variance is 6.299817e-04
- Estimated variance of the parameter x is 1.380311e-02 and variance of the parameter y is 4.141687e-02.
- Number of iterations: 5
```

Fig. 13: rezultati

Opazimo, da smo z vključitvijo uteži, zmanjšali standardno deviacijo napake in povprečno vrednost napake $\underline{f} = \underline{y} - \hat{\underline{y}}(\underline{\theta})$, vendar povečali varianco parametrov.

Spodaj prikažemo $\underline{f}$.

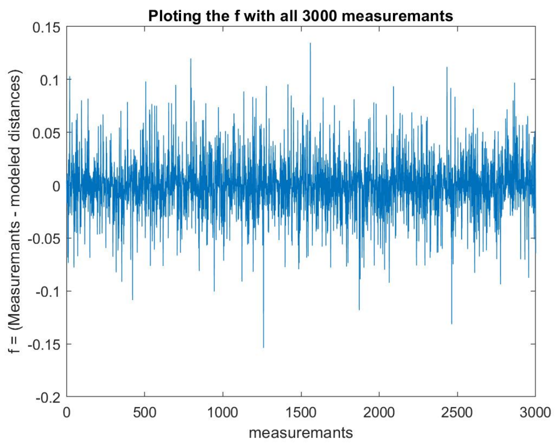

Fig. 14: razlika med meritvami in modelom

Prikažemo tudi, kako sta se povprečno vrednost napake in standardna deviacija napake spreminjali z večanjem meritev.

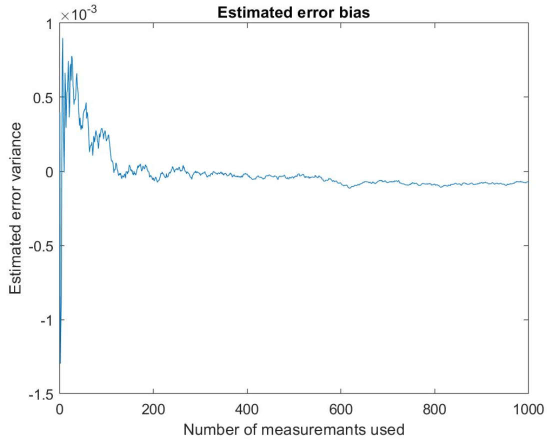

Fig. 15: Povprečna vrenost napake skozi meritve

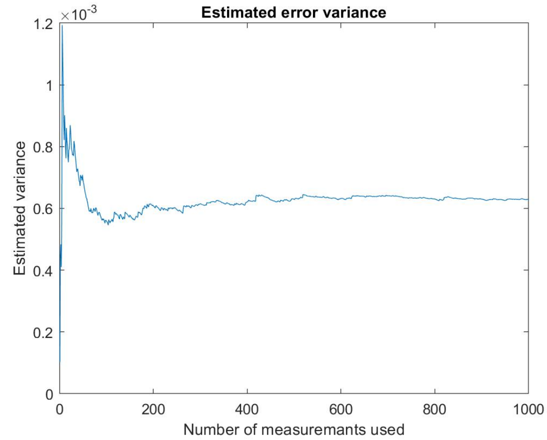

Fig. 16: Varianca napake skozi meritve

5 Zaključek

Pri nalogi smo se spoznali z metodo nelinarih najmanjših kvadratov. Lokacijo telefona smo poiskali s pomočjo Levenberg-Marquardtovega algoritma.

Opazili smo, da se z povečevanjem meritev varianca ocene lokacije telefona manjša. Opazimo tudi, da se z dodajanjem uteži, varianca napake in povprečna vrednost napake glede na optimizacijo brez uteži znižata, vendar pa se poveča varianca parametrov. Lahko rečemo, da smo z dodajanjem uteži izboljšali konsistentnost in pristranskost modela. Manjša kot je standardna deviacija napake bolj je model konsistenten. Bližje kot je povprečna vrednost napake nični, manj pristranski je model.

6 Reference

1 I. Škrjanc, Knjiga - Inteligentni sistemi za podporo odločanju, FE 2023

2 I. Škrjanc, Inteligentni sistemi za podporo odločanju, gradivo za predavanja, FE 2023