Primjena egzaktne aritmetike na ocjenu točnosti numeričkog proračuna

Jaguljnjak Lazarević, A.[1], Lazarević D.[2] i Dvornik J.[3]

Sažetak
U radu je na sustavnom nizu primjera riješenih egzaktnom aritmetikom prikazan utjecaj pogreške zaokruživanja na točnost numeričkog proračuna. Analitička rješenja u obliku cijelih brojeva ili razlomaka uspoređena su s numeričkim rješenjima, dobivenim direktnim ili iteracijskim postupcima. S napretkom računala i algoritama za proračune cjelobrojnom aritmetikom ovim će pristupom biti moguće realnije ocjeniti točnost numeričkog proračuna inženjerskih problema. Ipak, uvijek će biti proračuna za koje nema pouzdana odgovora.

Ključne riječi: egzaktna aritmetika, pogreška zaokruživanja, broj uvjetovanosti

[1] **Doc. dr. sc. Antonia Jaguljnjak Lazarević, dipl. ing. građ.**, Sveučilište u Zagrebu, Rudarsko-geološko-naftni fakultet, Zavod za rudarstvo i geotehniku, Pierottijeva 6, 10000 Zagreb, e-mail: ajagu@rgn.hr
[2] **Prof. dr. sc. Damir Lazarević, dipl. ing. građ.**, Sveučilište u Zagrebu, Građevinski fakultet, Zavod za tehničku mehaniku, Fra Andrije Kačića-Miošića 26, 10000 Zagreb, e-mail: damir@grad.hr
[3] **Prof. emer. dr. sc. Josip Dvornik, dipl. ing. građ.**, Sveučilište u Zagrebu, Građevinski fakultet, Zavod za tehničku mehaniku, Fra Andrije Kačića-Miošića 26, 10000 Zagreb, e-mail: dvornik@grad.hr

1 Uvod

Svaki proračun konstrukcije temelji se na skupu nužnih aproksimacija u odnosu na realnu (izvedenu) konstrukciju i stvarna djelovanja. Zbog pretpostavaka proračuna neizbježno unosimo pogreške koje za dobro modeliranu građevinu procjenjujemo na 10%, a za loše modeliranu na čak 30%, pri čemu su iz procjene isključene pogreške zbog aproksimacija dinamičkih utjecaja poput vjetra i potresa [1].

Izvori neizbježnih pogrešaka postoje u svakom od četiri koraka realizacije konstrukcije: pri aproksimaciji konstrukcije matematičkim modelom, zatim pri tvorbi numeričkog modela, prilikom numeričkog proračuna i pri izvedbi konstrukcije [2]. Razlike između realnih i proračunskih vrijednosti pretežno nastaju zbog idealizacija koje koristimo prilikom tvorbe matematičkog i numeričkog modela te odstupanje izvedene konstrukcije od projekta. Zato se smatra da proračunska točnost mora biti veća od uobičajene točnosti rezultata barem za jedan red veličine, odnosno grešku u računskim operacijama treba ograničiti na $0{,}1 - 1$ %. U ovom radu prikazan je način izdvajanja pogreške zaokruživanja (koja je sastavni dio svakog numeričkog proračuna) i procjena utjecaja loše odabranog numeričkog modela na gubitak značajnih znamenaka.

2 Broj uvjetovanosti kao ocjena pogreške

Poznato je iz numeričke analize [3,4], gornja se granica ukupne pogreške zaokruživanja prilikom rješavanja linearnog sustava algebarskih jednadžbi $\mathbf{Ku} = \mathbf{f}$ može prikazati u obliku dokazane nejednakosti:

$$\frac{\|\delta\mathbf{u}\|}{\|\mathbf{u}\|} \leq \frac{\kappa(\mathbf{K})}{1 - \kappa(\mathbf{K})\dfrac{\|\delta\mathbf{K}\|}{\|\mathbf{K}\|}}\left(\frac{\|\delta\mathbf{K}\|}{\|\mathbf{K}\|} + \frac{\|\delta\mathbf{f}\|}{\|\mathbf{f}\|}\right), \tag{1}$$

gdje je $\mathbf{u}$ rješenje, matrica $\mathbf{K}$ i vektor $\mathbf{f}$ su ulazni podaci, $\kappa(\mathbf{K}) = \|\mathbf{K}^{-1}\|\|\mathbf{K}\|$ broj uvjetovanosti sustava, a $\|\cdot\|$ neka vektorska ili matrična norma. Ako relativna pogreška ulaznih podataka ne prelazi vrijednost strojne preciznosti, tada prema [5] izraz (1) možemo zapisati kao

$$\frac{\|\delta\mathbf{u}\|}{\|\mathbf{u}\|} = 10^{-s} \leq \kappa(\mathbf{K}) \cdot 10^{-p}, \tag{2}$$

gdje je p broj značajnih znamenaka strojnoga zapisa i s broj očekivanih točnih znamenaka rezultata. Logaritmiranjem izraza (2) dolazimo do „jednostavnoga“ izraza za procjenu gubitaka značajnih znamenaka $s \geq p - \log\kappa(\mathbf{K})$. Jednostavnost izraza kvari netrivijalan način određivanja broja uvjetovanosti, jer treba invertirati matricu, što je za velike sustave spor i memorijski zahtjevan postupak.

Iako postoje mišljenja da je gornja granica pogreške (2) previše pesimistična odnosno presigurna, pa čak i da pogreška zaokruživanja ne može ugroziti dobivene rezultate [6], čini se da treba biti oprezan. U nastavku je opisana primjena egzaktne aritmetike na prognozu pogreške temeljene na izrazu (2).

3 Realizacija postupka

Temeljna je zamisao odrediti sve članove osnovnog sustava **K**, **u** i **f** u obliku cijelih brojeva i razlomaka, čime isključujemo utjecaj pogreške zaokruživanja. Tako realizirani sustav i njegovo rješenje nazivamo etalonom. Da bismo proveli postupak moramo upotrijebiti programske pakete koji podržavaju simboličku algebru, a sustav možemo riješiti izravnim ili obratnim postupkom. U potonjem slučaju komponente vektora **u** također su racionalni brojevi. Poželjno je da etalon opisuje neki realni problem.

Nakon što smo **K** i **f** realizirali racionalnim brojevima koristimo ih kao ulazne podatke za numerički proračun i pohranjujemo ih u strojnom obliku dvostruke preciznosti. Budući da zapisivanje brojeva u memoriji računala sadrži pogrešku zaokruživanja nastaje perturbirani sustav koji će tijekom rješavanja primjenom nekoga numeričkog postupka više ili manje akumulirati pogrešku. Na kraju iz usporedbe cjelobrojnog i numeričkog rješenja možemo odrediti točan gubitak značajnih znamenaka odnosno iznos pogreške.

Gomilanje pogreške u računskim operacijama može biti brzo i značajno ako je riječ o loše uvjetovanom sustavu ili o nestabilnom algoritmu koji dodatno inducira pogreške. Brojne analize iz područja proračuna građevinskih konstrukcija potvrđuju vezu između loše uvjetovanosti sustava i lošeg projektiranja odnosno modeliranja, čak iako podrazumjevamo stabilan postupak proračuna [2].

Za realizaciju etalona odabran je Lagrangeov trikubični konačni element u obliku kvadra. Element je C^0 kompatibilan s koordinatnim funkcijama u obliku Lagrangeova polinoma trećeg stupnja. Sadrži 64 pravilno raspoređena čvora sa tri translacijska stupnja slobode što ukupnu daje 192 nepoznanice. Formulacija konačnog elementa pri kojemu su eliminirani svi utjecaji koji pogoršavaju broj uvjetovanosti, a članovi matrice krutosti određeni u obliku racionalnih brojeva dana je u [7]. Kao opterećenje odabrano je čisto savijanje kojemu pripada analitičko rješenje koje ne sadrži beskonačne vrijednosti, a desna je strana sustava određena obrnutim postupkom. Modeli su pridržani minimalno potrebnim brojem pravilno raspoređenih veza.

Uspostavljene su dvije grupe etalona: dobro uvjetovani modeli s Lagrangeovim elementima jednakih stranica i modeli kod kojih je omjer stranica izrazito nepovoljan (izduženi i pločasti elementi) čime je pogoršana uvjetovanost sustava.

Numerički proračuni provedeni su programskim paketom *Wolfram Mathenatica* na dva načina: izravnim multifrontalnim postupkom IMF (algoritam pod naredbom *LinearSolve* uz štedni zapis matrice) i klasičnom metodom konjugiranih gradijenata CG [8] (programiranom u *Mathematici*).

4 Rezultati proračuna

Dio rezultata proračuna - usporedbe etalonskih i numeričkih rješenja (dobivenog IMF postupkom) do stupnja diskretizacije 7×7×7 preuzet je iz [7, 9]. Proširenje na veće modele i druge numeričke postupke realizirano je radnom stanicom nabavljenom sredstvima istraživačkog projekta Hrvatske zaklade za znanost [10].

Osnovni podaci o proračunskim modelima s procjenom gubitka značajnih znamenaka prema izrazu $\log \kappa(\mathbf{K})$ priloženi su u tablici 1. Za proračun broja uvjetovanosti upotrijebljen je algoritam pod naredbom *MatrixConditionNumber* uz izbor norme dva. Spektralna norma ili norma dva odabrana je jer daje najmanju vrijednosti broja uvjerovanosti [5].

Tablica 1. Osnovni podaci o diskretizaciji uz teorijsku procjenu broja izgubljenih značajnih znamenaka

Stupanj diskretizacije	Broj elemenata	Broj nepoznanica	Broj izgubljenih znamenaka $\log \kappa(\mathbf{K})$		
			Kocka	Ploča	Štap
1×1×1	1	186	4,5	9,5	8,7
2×2×2	8	1 023	5,2	9,7	8,8
3×3×3	27	2 994	5,6	10,0	9,1
4×4×4	64	6 585	6,0	10,2	9,3
5×5×5	125	12 282	6,2	10,5	9,5
6×6×6	216	20 571	6,5	10,7	9,7
7×7×7	343	31 938	6,7	10,9	9,9
8×8×8	512	46 869	6,8	11,0	10,0
9×9×9	729	65 850	16,7	17,8	16,9
10×10×10	1 000	89 367	7,3	10,8	10,6
11×11×11	1 331	117 096	7,2	11,4	10,4
12×12×12	1 728	151 953	7,3	11,5	10,5

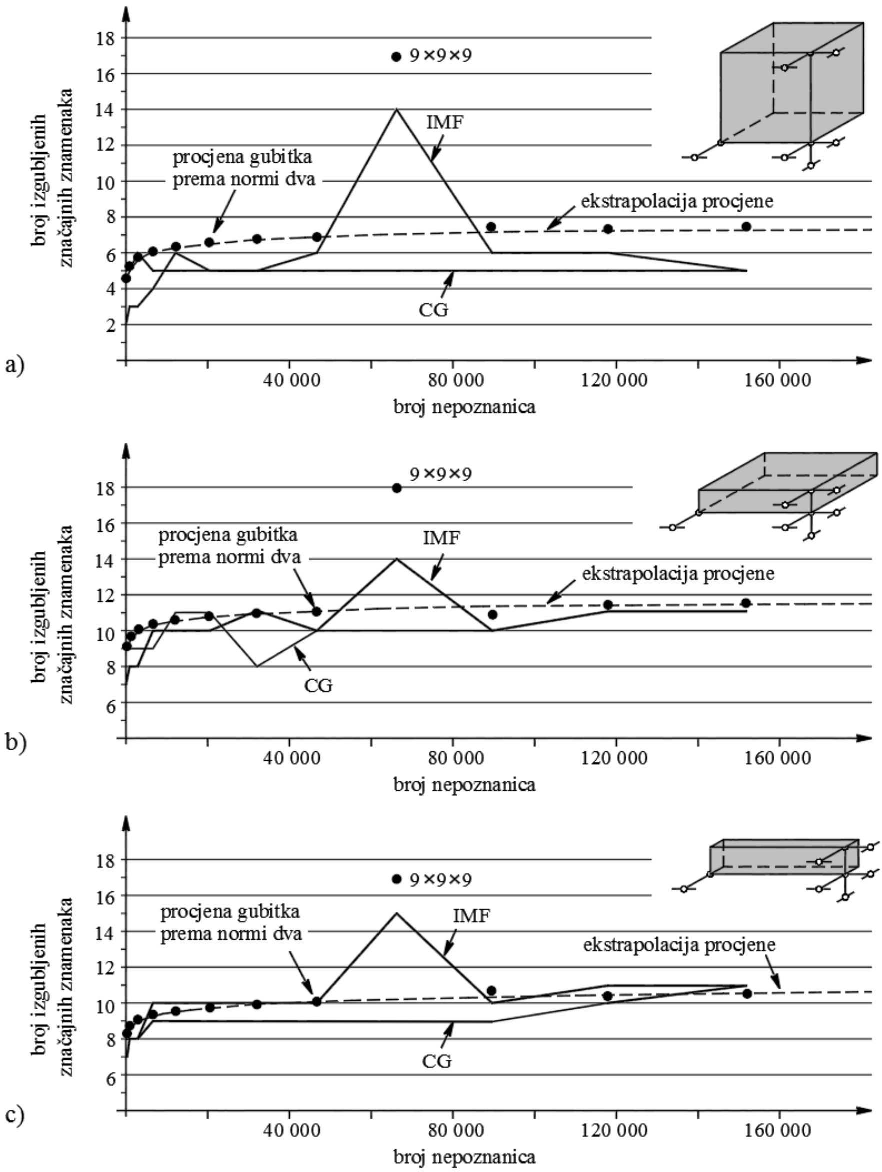
Rezultati su očekivani u dva smjera: s povećanjem stupnja diskretizacije raste broj računskih operacija i gomilanje pogreške zaokruživanja, a sustav postaje lošije uvjetovan uz isti stupanj diskretizacije ali loš oblik elementa.

Blagi porast procjene gubitaka značajnih znamenaka kod većih modela poremećen je skokom broja uvjetovanosti kod diskretizacije 9×9×9 koji procjenjuje gubitak **svih** značajnih znamenaka strojnog zapisa dvostruke preciznosti (čak i za model kocke). Pretpostavljamo da dijagonala matrice krutosti sadrži član male vrijednosti u odnosu na ostale pa raste raspon vlastitih vrijednosti. Omjer maksimalne i minimalne vlastite vrijednosti direktno utječe na vrijednost broja uvjetovanosti [2,11].

Usporedba procijenjenog i stvarnog gubitaka značajnih znamenaka za pojedine modele prikazana je na slici 1. Ekstrapolacija procjene gubitka određena je na temelju prvih sedam stupnjeva diskretizacije [7].

Iz grafičkoga prikaza vidljivo je da su točni gubici znamenaka vrlo blizu teorijske procjene. Kod lošije uvjetovanih modela (pločasti i štapni elementi) s veliki brojem nepoznanica (> 90 000) točan broj i procjena se podudaraju. Nešto veće odstupanje primjećujemo kod modela ploče i štapa s manjim brojem nepoznanica, te kod modela kocke, bez obzira na stupanj diskretizacije, što potvrđuje tvrdnje o većoj „otpornosti“ rijeđih i/ili pravilnih mreža na pogrešku zaokruživanja.

Skok točnog broja izgubljenih znamenaka za diskretizaciju 9×9×9 svih modela zabilježen je samo kod rješavanja IMF postupkom, dok za CG taj stupanj diskretizacije nije problematičan i nema gomilanja pogreške zaokruživanja. Ovaj je rezultat očekivan jer je poznat utjecaj dekompozicije matrice na gubitak znamenaka [2, 5].



Slika 1. Ovisnost teorijske procjene i točnog gubitka znamenaka o broju nepoznanica za model: a) kocke, b) ploče i c) štapa

5 Zaključak

Prema rezultatima u ovom radu neizbježna pogreška zaokruživanja sigurno smanjuje točnost naših numeričkih proračuna. Na jednostavnom primjeru čistog savijanja, diskretiziranog s nekoliko stotina do 1,5 10^5 nepoznanica, riješenog primjenom dvostruke preciznosti, točan gubitak značajnih znamenaka iznosi 5 za dobro te 11 za loše uvjetovane modele, pri čemu teorijska procjena prema normi dva dobro prognozira vrijednost gubitka. Jasno, pomak opterećen pogreškom zaokruživanja ulazni je podatak za proračune reakcija, deformacija, naprezanja i unutarnjih sila.

Zahvale

Ovaj je rad financirala Hrvatska zaklada za znanost projektom IP – 2014 – 09 – 2899.

Literatura

[1] Dvornik J., Lazarević D.; Manjkavosti proračunskih modela inženjerskih konstrukcija; Građevinar, 2005., vol. 57, br. 4, str. 227 - 326.
[2] Dvornik J., Lazarević D., Bićanić N.; O načelima i postupcima proračuna građevinskih konstrukcija, Sveučilište u Zagrebu Građevinski fakultet, 2019., Zagreb.
[3] Varga, R. S.: Matrix Iterative Analysis, Springer–Verlag, Berlin Heidelberg, 2000.
[4] Stoer, J.; Bulirsch, R.: Introduction to Numerical Analysis, Springer–Verlag, New York, 1992.
[5] Bathe, K. J.: Finite Element Procedures, Prentice Hall, New Jersey, 1996.
[6] Bischoff, M.; Wall, W. A.; Bletzinger, K.–U.; Ramm, E.: Models and Finite Elements for Thin–walled Structures, Encyclopedia of Computational Mechanics, Volume 2, Solids and Structures, Chapter 3, John Wiley & Sons, Ltd., West Sussex, 2004.
[7] Jaguljnjak Lazarević A., Dvornik J., Frgić L.: Utjecaj pogreške zaokruživanja na točnosti proračuna konstrukcije, Građevinar, 2011.,vol. 63, br. 11, str. 911 – 921.
[8] Hestens M. R., Stiefel E.; Methods of Conjugate Gradients for Solving Linear System, Journal of Research of the National Bureau of Standards, 1952, Vol. 49, no.6, pp. 409 – 436.
[9] Jaguljnjak Lazarević A., Dvornik J., Frgić L.; Influence of Round-Off Errors on Numerical Computation Accuracy, 6th International Congress of Croatian Society of Mechanics, Proceedings Croatian Society of Mechanics, 2009. str. 48-48.
[10] https://www.grad.unizg.hr/yoda
[11] Demell, J. W.: Applied Numerical Linear Algebra, SIAM, Philadelphia, 1997.