Мр Златко Удовичић

ПРИБЛИЖНО РЈЕШАВАЊЕ ЈЕДНАЧИНА

У литератури у којој се обрађује проблем приближног рјешавања једначина итеративним методама акценат се обично ставља на генерисање низа који конвергира ка рјешењу задане једначине, док се готово уопште, пажња не посвећује проблему заустављања итеративног процеса. Неједнакости помоћу којих се процјењује разлика између тачног и приближног рјешења увијек су доказане, али њихова примјена приликом рјешавања конкретних задатака у правилу остаје недовољно истакнута. Неискусног читаоца такав приступ може навести да погрешно схвати појам приближног рјешења, што касније може, како је показано у тексту, имати веома непријатне посљедице. Циљ овог чланка, у којем је као илустративна метода анализирана тзв. метода просте итерације, је да се допринесе превазилажењу поменутих недостатака.

1. Увод

Дефиниција 1. Нека су A и B непразни скупови, нека $f\colon A \to B$ и нека је $b \in B$. Једначина је сваки предикат дужине један[1] дефинисан на скупу A са

$$P(x): \text{``}f(x) = b\text{''}. \tag{1}$$

Рјешење једначине (1) је свако $\xi \in A$ за које је $P(\xi) = \top$. Уколико је $\eta \in A$ такво да је $P(\eta) = \perp$, онда η није рјешење једначине (1).

Другим ријечима, $\xi \in A$ је рјешење једначине (1) ако и само ако је $f(\xi) = 0$, односно ако и само ако ξ задовољава једначину (1).

Примјер 1. Нека је $A = B = \mathbf{R}$ и нека је $f(x) = x^2 + 5x - 6$. Број 1 јесте рјешење једначине $x^2 + 5x - 6 = 0$ јер је $f(1) = 0$, док број 3 није рјешење те једначине јер је $f(3) = 18 \neq 0$.

Ријешити задану једначину значи одредити сва њена рјешења или доказати да задана једначина нема рјешења. На жалост, занемарљиво мали број једначина које се јављају приликом рјешавања разних проблема може се ријешити. Због тога се развијају методе приближног рјешавања једначина.

Дефиниција 2. Нека је у дефиницији 1 A нормирани простор. Приближно рјешење једначине (1), одређено са тачношћу $\varepsilon > 0$, је свако $\xi^* \in A$ за које важи $\|\xi - \xi^*\| < \varepsilon$, гдје је ξ (тачно) рјешење једначине (1).

[1] Предикат дужине n, дефинисан на скупу A, је свака функција $P: A^n \to \{\top, \perp\}$.

Дакле, приближно рјешење једначине (1) је свако $\xi^* \in A$ које се налази у ε околини рјешења ξ. Приближно ријешити једначину значи за свако њено рјешење одредити барем једно приближно рјешење. Из претходне дефиниције се види да приближно рјешавати има смисла само оне једначине које имају рјешења. Основна подјела метода за приближно рјешавање једначина је на директне и итеративне.

Директним методама се, након коначно много корака, долази до рјешења једначине (поступак за одређивање корijена полинома другог, трећег или четвртог степена, Гаусова метода рјешавања система линеарних једначина, ...). Теоретски, примјена ових метода доводи до (тачног) рјешења задане једначине. Међутим, због грешака у рачунању, које су готово неизбјежне, директним методама се у правилу добија само приближно рјешење задане једначине.

Итеративним методама се, након бесконачно много корака, долази до рјешења једначине (разне методе за приближно рјешавање алгебарских и трансцендентних једначина, итеративне методе за рјешавање система линеарних једначина, ...). Због немогућности њихове практичне реализације, итеративне методе се заустављају након коначно много корака и тиме се добија само приближно рјешење задане једначине. У даљем тексту пажња ће се посветити итеративним методама, а као модел метода се узима тзв. метода просте итерације (аналогно разматрање се може спровести и за било коју другу итеративну методу).

Дефиниција 3. Нека су X и Y нормирани простори, нека је $S \subseteq X$ и нека $\varphi\colon X \to Y$. Функција φ је контракција на скупу S, са коефицијентом контракције $q \in [0, 1)$ ако

$$(\forall x, y \in S)\ \|\varphi(x) - \varphi(y)\| \le q\|x - y\|.$$

Теорема 1. (Банахова теорема о непокретној тачки) *Нека је X Банахов простор, нека је скуп $S \subseteq X$ затворен у простору X и нека $\varphi\colon S \to S$. Ако је функција φ контракција на скупу S, онда низ генерисан рекурентном формулом*

$$x_n = \varphi(x_{n-1}), \qquad n \in \mathbf{N}$$

конвергира ка јединственом рјешеу $\xi \in S$ једначине $x = \varphi(x)$ независно од избора почетне апроксимације $x_0 \in S$.

Доказ ове теореме може се наћи нпр. у [2].

Разлика између n-те апроксимације x_n и тачног рјешења ξ може се процијенити на један од сљедећа два начина:

$$\|x_n - \xi\| \le \frac{q}{1-q}\|x_n - x_{n-1}\| \tag{2}$$

$$\|x_n - \xi\| \le \frac{q^n}{1-q}\|x_1 - x_0\| \tag{3}$$

Конкретизацијом простора X и Y, односно функције φ, као посљедице Банахове теореме о непокретној тачки, добијају се тврђења која су релативно једноставна за практичну употребу.

Теорема 2. *Ако функција $\varphi\colon [a,b] \to [a,b]$, диференцијабилна на интервалу (a,b), има особину да постоји позитивна константа $q < 1$ таква да је*

$$(\forall x \in (a,b))\ |\varphi'(x)| \le q, \tag{4}$$

онда низ генерисан рекурентном формулом $x_n = \varphi(x_{n-1})$, $n \in \mathbf{N}$ конвергира ка јединственом рјешењу $\xi \in [a,b]$ једначине $x = \varphi(x)$, независно од избора почетне апроксимације $x_0 \in [a,b]$.

Није тешко доказати (примјеном Лагранжеве теореме о средњој вриједности) да је услов (4) довољан да функција φ, у теореми 2, буде контракција на интервалу $[a,b]$.

Теорема 3. *Нека је A реална квадратна матрица реда n и нека постоји позитивна константа $q < 1$ таква да је $\|I - A\| \le q$ (у било којој матричној норми). Тада низ генерисан рекурентном формулом*

$$x_{n+1} = (I - A)x_n + b, \qquad n \in \mathbf{N}$$

конвергира ка јединственом рјешењу ξ система линеарних једначина $Ax = b$ независно од избора почетне апроксимације $x_0 \in \mathbf{R}^n$.

Поступак приближног рјешавања задане једначине реализује се у три фазе:

а) локализација рјешења задане једначине,

б) генерисање низа апроксимација и

в) заустављање итеративног процеса.

У првој фази се одређује околина траженог рјешења задане једначине у којој су задовољене претпоставке теореме на којој се базира итеративна метода. Обично су ове теореме формулисане тако да задовољење њихових претпоставки у некој околини рјешења, осим конвергенције низа апроксимација ка том рјешњу, обезбјеђује и егзистенцију и јединственост одговарајућег рјешења у тој околини. На жалост, није ријеткост да се приликом приближног рјешавања једначине ова фаза изоставља, без обзира на њен есенцијални значај за коректно рјешење постављеног проблема.

Критеријуми за заустављање итеративног процеса (у методи просте итерације) изводе се из неједнакости (2) (тзв. апостериорна оцјена грешке) и (3) (тзв. априорна оцјена грешке). Из неједнакости (2) слиједи да x_n апроксимира рјешење задане једначине са тачношћу $\varepsilon > 0$ ако је

$$\|x_n - x_{n-1}\| < \varepsilon \left(\frac{1}{q} - 1\right), \tag{5}$$

док из неједнакости (3) слиједи да је довољно генерисати

$$n = \left[\frac{\ln \frac{\varepsilon(1-q)}{\|x_1 - x_0\|}}{\ln q}\right] + 1 \tag{6}$$

чланова низа апроксимација да би се рјешење добило са тачношћу $\varepsilon > 0$. Обе ове релације се добијају тако што се захтијева да десна страна неједнакости (2), односно (3) буде мања од задане тачности $\varepsilon > 0$. Треба нагласити да се критеријум (6) добија након више мајорација, па је лошији од критеријума (5), у том смислу да може захтијевати одређивање већег броја апроксимација него што је то стварно потребно, да би се постигла захтијевана тачност. Са друге стране, критеријум (6) је бољи, у том смислу што није након сваког корака потребно упоређивати разлику између узастопних апроксимација са константом $\varepsilon(\frac{1}{q}-1)$.

2. Погрешно схватање појма приближног рјешења

Основа погрешног схватања појма приближног рјешења је у коментару који је дат након дефиниције 1. Наиме, $\xi^* \in A$ се сматра приближним рјешењем једначине (1) ако и само ако је $f(\xi^*)$ приближно једнако нули. Нешто коректнија, али и даље погрешна дефиниција приближног рјешења једначине (1) би гласила: $\xi^* \in A$ је приближно рјешење једначине (1), одређено са тачношћу $\varepsilon > 0$, ако и само ако је $\|f(\xi^*)\| < \varepsilon$ (наравно, ова „дефиниција“ подразумијева да је скуп B у дефиницији 1 нормирани простор). Узрок оваквог схватања је често присутна (погрешна) формулација задатка облика „Приближно ријешити једначину ... “, без задавања тачности са којом треба ријешити задану једначину. Поред тога, погрешно схватање појма приближног рјешења се узрокује и онда када се приближно рјешавање задане једначине сведе на само формално генерисање низа који (можда) конвергира ка рјешењу задане једначине (реализација друге фазе у поступку приближног рјешавања), без јасног критеријума о броју чланова низа које треба одредити (у оваквим ситуацијама је готово немогуће генерисати оптималан број чланова низа). Иако тривијалан, примјер који слиједи илиуструје какве погрешне закључке може изазвати погрешно схватање појма приближног рјешења.

Примјер 2. Нека је $f_1(x) = 10^{-5}x - 10^{-2}$ и нека је $f_2(x) = 10^5 x - 10^8$.

Једначина $f_1(x) = 0$ има јединствено рјешење $\xi = 10^3$, а приближно рјешење ове једначине, одређено са тачношћу $\varepsilon = 10^{-3}$, је сваки број из интервала $(1000-10^{-3}, 1000+10^{-3})$.

Са друге стране, $|f_1(x)| < \varepsilon \Leftrightarrow x \in (900, 1100)$. Дакле, према погрешној дефиницији појма приближног рјешења, сваки број из интервала $(900, 1100)$ би могао бити приближно рјешење једначине $f_1(x) = 0$ одређено са тачношћу $\varepsilon = 10^{-3}$, што знатно проширује интервал којем смије припадати приближно рјешење.

Једначина $f_2(x) = 0$ има јединствено рјешење $\xi = 10^3$, а приближно рјешење ове једначине, одређено са тачношћу $\varepsilon = 10^{-3}$, је такође сваки број из интервала $(1000 - 10^{-3}, 1000 + 10^{-3})$.

Будући да је $|f_2(x)| < \varepsilon \Leftrightarrow x \in (1000 - 10^{-8}, 1000 + 10^{-8})$, према погрешној дефиницији појма приближног рјешења, приближно рјешење једначине $f_2(x) = 0$, одређено са тачношћу $\varepsilon = 10^{-3}$, би могли бити само бројеви из интервала $(1000 - 10^{-8}, 1000 + 10^{-8})$, што знатно сужава интервал којем смије припадати приближно рјешење.

3. Примјери

У четири примјера који слиједе рјешаване су једначине облика $x = \varphi(x)$. Свака од једначина је „приближно ријешена“ са заданом тачношћу $\varepsilon > 0$ само генерисањем низа $x_{n+1} = \varphi(x_n)$, а као критеријум заустављања итеративног поступка кориштена је неједнакост $\|x_{n+1} - x_n\| = \|\varphi(x_n) - x_n\| < \varepsilon$. Наравно, приближна рјешења постављених једначина одређена су и коректним поступком, при чему је указано на грешке које су изазване погрешним схватањем појма приближног рјешења.

Примјер 3. Одредити приближно рјешење једначине $x = \varphi(x)$, гдје је $\varphi(x) = \dfrac{x^2+1}{x}$, са тачношћу $\varepsilon = 0.5 \cdot 10^{-2}$.

Рјешење. Није тешко провјерити да задана једначина нема рјешења, па уопште нема смисла тражити њена приближна рјешења. Међутим, са̂мо генерисање низа $x_{n+1} = \varphi(x_n)$, са почетном вриједности $x_0 = 222.210$, уз критеријум заустављања $|\varphi(x_n) - x_n| < \varepsilon$, већ након четири итерације доводи до „приближног рјешења“ добијеног са заданом тачношћу. Резултати рачунања су дати у сљедећој табели.

n	x_n	$\varphi(x_n)$	$\lvert\varphi(x_n) - x_n\rvert < 0.5 \cdot 10^{-3}$
0	222.210	222.215	$\perp$
1	222.215	222.220	$\perp$
2	222.220	222.225	$\perp$
3	222.225	222.229	$\top$

Иако је пресликавање φ контракција на сваком затвореном подинтервалу интервала $(-\infty, -\frac{\sqrt{2}}{2}) \cup (\frac{\sqrt{2}}{2}, \infty)$ (није тешко провјерити да је $|\varphi'(x)| < 1 \Leftrightarrow x \in (-\infty, -\frac{\sqrt{2}}{2}) \cup (\frac{\sqrt{2}}{2}, \infty)$), није могуће примјенити теорему 2 јер не постоји интервал $[\alpha, \beta]$ такав да је $\varphi([\alpha,\beta]) \subseteq [\alpha,\beta]$. Заиста, ако би такав интервал постојао морали би бити испуњени услови

$$\varphi(\alpha) \geq \alpha \Leftrightarrow \alpha \geq 0, \qquad \varphi(\beta) \leq \beta \Leftrightarrow \beta \leq 0$$

који су очигледно у супротности са чињеницом да мора бити $\alpha < \beta$. На крају треба нагласити и то да је права $y = x$ коса асимптота функције $\varphi(x) = \dfrac{x^2+1}{x}$, па се разлика $|x_n - \varphi(x_n)|$, за довољно велико n, може учинити произвољно малом.

Примјер 4. Одредити приближно рјешење једначине $x = \varphi(x)$, гдје је $\varphi(x) = \ln(x + 0.01)$, које припада интервалу $I = [1.1, 1.4]$, са тачношћу $\varepsilon = 0.5 \cdot 10^{-3}$.

Рјешење. Будући да је $\varphi'(x) = \dfrac{100}{100x+1}$, функција φ је растућа на интервалу I, па из $\varphi(1.1) > 1.103$ и $\varphi(1.4) < 1.35$ слиједи да је $\varphi(I) \subseteq I$. Осим тога, $\max_{x\in[1.1,1.4]}|\varphi'(x)| = |\varphi'(1.1)| = \frac{100}{111} = q < 1$, па су на интервалу I задовољене претпоставке теореме 2, што значи да низ генерисан рекурентном формулом $x_{n+1} = \varphi(x_n)$, независно од избора почетне апроксимације $x_0 \in I$, конвергира ка јединственом (у интервалу I) рјешењу задане једначине. Према оцјени (5), x_n ће апроксимирати тражено рјешење за тачношћу ε ако је $|x_n - x_{n-1}| < \varepsilon(\frac{1}{q} - 1) = \frac{11}{2\cdot 10^5} = 0.55\cdot 10^{-4}$ (а не ако је $|x_n - x_{n-1}| = |\varphi(x_{n-1}) - x_{n-1}| < \varepsilon = 0.5\cdot 10^{-3}$). Прерано заустављење итеративног поступка изазвано примјеном погрешног критеријума илустровано је у сљедећој табели.

n	x_n	$\varphi(x_n)$	$\lvert\varphi(x_n) - x_n\rvert < 0.5\cdot 10^{-3}$	$\lvert\varphi(x_n) - x_n\rvert < 0.55\cdot 10^{-4}$
0	1.13400	*1.13453*	$\perp$	$\perp$
1	1.13453	1.13499	$\perp$	$\perp$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
16	1.13770	1.13776	$\top$	$\perp$
17	1.13776	**1.13781**	$\top$	$\top$

Дакле, приближно рјешење задане једначине, одређено са таћношћу $\varepsilon = 0.5\cdot 10^{-3}$, је $\xi^* = 1.138$, а не „приближно рјешење“ $\xi^* = 1.135$ које се добија примјеном погрешног критеријума заустављања.[2]

Примјер 5. Одредити приближно рјешење једначине $x = \varphi(x)$, гдје је $\varphi(x) = -\dfrac{1}{10}(\exp(5x) + 2)$, које припада интервалу $I = [-4, -0.1]$, са тачношћу $\varepsilon = 0.5\cdot 10^{-3}$.

n	x_n	$\varphi(x_n)$	$\lvert\varphi(x_n) - x_n\rvert < 0.5\cdot 10^{-3}$	$\lvert\varphi(x_n) - x_n\rvert < 10^{-3}$
0	-0.25000	-0.22865	$\perp$	$\perp$
1	-0.22865	-0.23188	$\perp$	$\perp$
2	-0.23188	$-\mathbf{0.23137}$	$\perp$	$\top$
3	-0.23137	$-\mathit{0.23145}$	$\top$	$\top$

Рјешење. Будући да је $\varphi'(x) = -\frac{1}{2}\exp(5x)$, функција φ је опадајућа на интервалу I, па из $\varphi(-4) < -0.1$ и $\varphi(-0.1) > -4$ слиједи да је $\varphi(I) \subseteq I$. Осим тога, $\max_{x\in[-4,-0.1]}|\varphi'(x)| = |\varphi'(-0.1)| = \frac{1}{2\sqrt{e}} = q < 1$, па су на интервалу I задовољене претпоставке теореме 2, што значи да низ генерисан рекурентном формулом $x_{n+1} = \varphi(x_n)$, независно од избора почетне апроксимације $x_0 \in I$, конвергира ка јединственом (у интервалу I) рјешењу задане једначине. Према оцјени (5), x_n ће апроксимирати тражено рјешење за тачношћу ε ако је $|x_n - x_{n-1}| < \varepsilon(\frac{1}{q} - 1) = \frac{1}{2}\cdot 10^{-3}(2\sqrt{e} - 1)$ (а не ако је $|x_n - x_{n-1}| = |\varphi(x_{n-1}) - x_{n-1}| < \varepsilon = 0.5\cdot 10^{-3}$). Будући да је $\frac{1}{2}\cdot 10^{-3}(2\sqrt{e} - 1) > 10^{-3}$, у претходној табели као критеријум заустављања је кориштена неједнакост $|x_n - x_{n-1}| < 10^{-3}$. Дакле, кориштењем погрешног критеријума заустављања, једна итерација је урађена непотребно, а

[2] Задана једначина има још једно рјешење које се налази у интервалу $(0.5, 1)$.

приближно рјешење задане једначине, одређено са таћношћу $\varepsilon = 0.5 \cdot 10^{-3}$, је $\xi^* = -0.231$.

Примјер 6. Са тачношћу $\varepsilon = 0.5 \cdot 10^{-3}$ одредити приближно рјешење система линеарних једнача $Ax = b$, ако је $A = \begin{bmatrix} 0.5 & 0.3 \\ -0.3 & 0.65 \end{bmatrix}$ и $b = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. За процјену грешке користити униформну векторску (матричну) норму.

Рјешење. Будући да је $\|I-A\|_\infty = 0.95 = q < 1$, задовољене су претпоставке теореме 3, па низ вектора генерисан рекурентном формулом $x_{n+1} = (I-A)x_n + b$ конвергира (у униформној норми) ка јединственом рјешењу заданог система линеарних једначина независно од избора почетне апроксимације x_0. Као и у претходним примјерима, разлика између задане тачности $\varepsilon = 0.5 \cdot 10^{-3}$ и броја $\varepsilon(\frac{1}{q} - 1) = \frac{1}{38} \cdot 10^{-3} > 0.2 \cdot 10^{-4}$ је значајна, па примјена погрешног критеријума заустављања има за посљедицу прерано заустављење итеративног поступка што показује сљедећа табела.

n	x_n	$(I-A)x_n + b$	$\|b - Ax_n\|_\infty < 0.5 \cdot 10^{-3}$	$\|b - Ax_n\|_\infty < 0.2 \cdot 10^{-4}$
0	$\begin{pmatrix} 0.18900 \\ -3.02000 \end{pmatrix}$	$\begin{pmatrix} 0.18850 \\ -3.01970 \end{pmatrix}$	$\perp$	$\perp$
1	$\begin{pmatrix} 0.18850 \\ -3.01970 \end{pmatrix}$	$\begin{pmatrix} \mathit{0.18834} \\ \mathit{-3.01936} \end{pmatrix}$	$\perp$	$\perp$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
7	$\begin{pmatrix} 0.18867 \\ -3.01881 \end{pmatrix}$	$\begin{pmatrix} 0.18869 \\ -3.01883 \end{pmatrix}$	$\top$	$\perp$
8	$\begin{pmatrix} 0.18869 \\ -3.01883 \end{pmatrix}$	$\begin{pmatrix} \mathbf{0.18870} \\ \mathbf{-3.01885} \end{pmatrix}$	$\top$	$\top$

Дакле, приближно рјешење заданог система линеарних једначина, одређено са тачношћу $\varepsilon = 0.5 \cdot 10^{-3}$, је $(\xi^*, \eta^*) = (0.189, -3.019)$, док се примјеном погрешног критеријума заустављања добија „приближно рјешење" $(\xi^*, \eta^*) = (0.188, -3.019)$.

Литература

[1] Херцег, Д., Крејић, Н., ***Нумеричка анализа***, Природно-математички факултет и STYLOS, Нови Сад, 1997.

[2] Јовановић, Б., Радуновић, Д., ***Нумеричка анализа***, Математички факултет, Београд, 2003.

[3] Kazdan, J.L., *Solving equations an elegant legacy*, The teaching of mathematics, Vol. IV/2, 105–144, Beograd, 2001.

[4] Миловановић, В. Г., ***Нумеричка анализа I део***, Научна књига, Београд, 1991.

[5] Радуновић, Д., ***Нумеричке методе. Збирка задатака***, Академска мисао, Београд, 2000.

[6] Scitovski, R., *Numerička matematika*, Elektrotehnički fakultet Sveučilišta J. J. Strossmayera, Osijek, 1999.

[7] Золић, А., ***Рјешавање једне класе трансцедентних једначина и неједначина***, Настава математике, XLIV 3/4, 19–23, Београд, 1999.

[8] Zolić, A., *Zbirka zadataka iz numeričke matematike (riješeni zadaci s kolokvijuma i ispita)*, Univerzitet u Tuzli, 2003.