Gyökkeresés, nemlineáris egyenletrendszerek megoldása

Kormányos Andor

Komplex Rendszerek Fizikája Tanszék

Áttekintés

- Motiváció, néhány általánosság a megoldási módszerekről
- “Rosszul viselkedő” függvények
- Egyváltozós módszerek
  - függvény derivált felhasználását nem igénylő módszerek
  - függvény derivált felhasználásával: Newton-Ralphson módszer
- Polinomok gyökeinek keresése
- Többváltozós, nemlineáris egyenletrendszerek
- Newton-Ralphson módszer többdimenzióban

Motiváció: nemlineáris függvényillesztés

A χ^2 költségfüggvényt szeretnénk minimalizálni:

$$\chi^2 = \sum_i \frac{[y_i - y(x_i|\mathbf{a})]^2}{\sigma_i^2}$$

Általános esetben ez a következőre vezet:

$$\frac{\partial \chi^2}{\partial a_j} = 2 \cdot \sum_{i=1}^{N} \left[\frac{f(x_i|\mathbf{a}) - y_i}{\sigma_i^2} \cdot \left. \frac{\partial f(x|a)}{\partial a_j} \right|_{x=x_i} \right] = 0$$

- Ha f az $\mathbf{a}$ paraméterektől nem függő bázisfüggvények lineárkombinációja $\Rightarrow$ probléma lineáris
- egyébként egy nemlineáris egyenletrendszer

Nemlineáris egyenlet(rendszerek) megoldása

Egydimenziós eset:

$$f(x) = 0$$

egy nemlineáris függvény gyökeit keressük.

Többdimenziós eset:

$$\mathbf{F}(\mathbf{x}) = 0$$

Expliciten:

$$f_i(x_1, x_2, ..., x_M) = 0 \qquad i = 1...N$$

- az f_i függvények az x_j változókban nemlineárisak
- több függvény gyökét keressük szimultán módon

A megoldás léte

A megoldás léte

- $N > M$ esetben általában nincsen megoldás
- $N \leq M$ esetben sem garantált
- ha van megoldás, akkor nem garantált, hogy egyértelmű
- könnyen előfordulhat, hogy a gyökök nem diszkrétek, egész intervallum gyök (pl két egymást metsző gömb)

Inverzfüggvény-tétel

- ha egy f függvény egy x pontban folytonos, és a deriváltja nem nulla, akkor az x pont egy környezetében invertálható
- ez általánosítható többdimenziós esetre is
- $N \leq M$ esetben tudunk mondani valamit a megoldás létére

Iteratív módszerek

A gyökkeresés módszerei mindig iteratívak

- kiindulunk egy valamilyen értékekből vagy egy intervallumból, amin belül a gyököt sejtjük
- a gyököt lépésről lépésre közelítjük meg
- ezért a gyököt általában csak közelítőleg kapjuk meg

Az iteráció nem mindig konvergál:

- el kell tudni dönteni, hogy konvergál-e?
- ha igen, akkor jó helyre konvergált-e?
- ha nem, akkor meg kell próbálni másik helyről kiindulni

Mitől jobb egyik vagy másik módszer?

Függvénykiértékelések száma

- minél kevesebb függvénykiértékelés egy adott iterációs lépésben
- minél kisebb iteratív lépésszám, vagyis a kovergencia sebessége
- minél pontosabban meg akarjuk határozni a gyököt

Mitől jobb egyik vagy másik módszer?

Függvénykiértékelések száma

- minél kevesebb függvénykiértékelés egy adott iterációs lépésben
- minél kisebb iteratív lépésszám, vagyis a kovergencia sebessége
- minél pontosabban meg akarjuk határozni a gyököt

Stabilitás

- általában tudnunk kell, hogy nagyjából hol van a gyök és a közeléből indítani keresést
- garantálja-e valami, hogy a keresés ott is marad a közelben
- nem divergál-e el a végtelenbe

Gyök bekeretezése egy dimenzióban

Példa gyök bekeretezhetőségére

- ha a függvény folytonos $a < x < b$ intervallumon
- és $f(a) < 0$ és $f(b) > 0 \quad \Rightarrow \quad$ létezik gyök az intervallumon

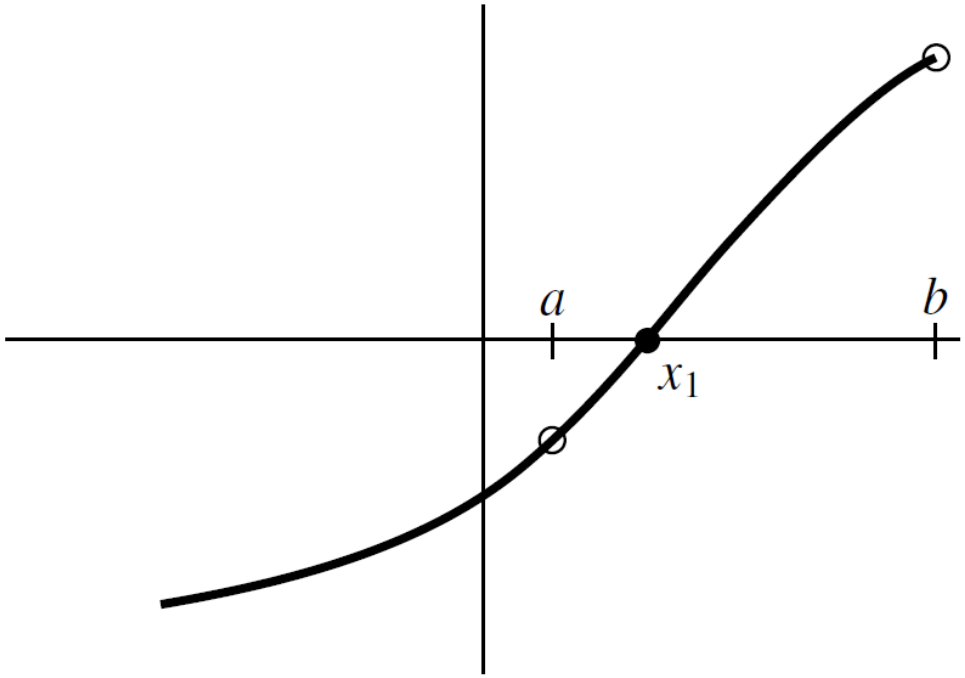

Figure: Ha a függvény előjelet vált a és b között, akkor ott van egy gyöke.
©Numerical Recipes

Gyök bekeretezés módszer korlátai

A függvény szinguláris viselkedése

- a függvény nem folytonos
- valahol divergál
- valamelyik deriváltja divergál

A gyök léte nem féltetlenül jelent előjelváltást

- lehet, hogy a gyök minimumhelyen van
- de a minimumhely nem feltétlen gyök
- jól el kell találni a bekeretezés pontjait

Egy intervallumon belül több gyök is lehet

Szakaszonként folytonos, divergens függvény

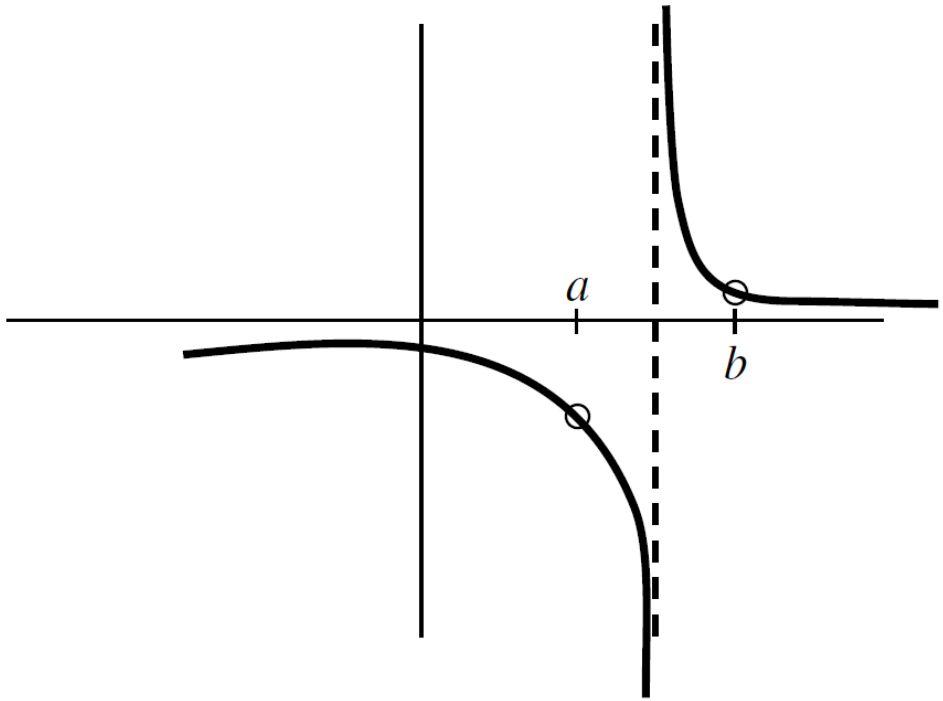

Figure: Problémás eset: függvénynek szakadása van. ©Numerical Recipes

Függvény rosszul viselkedő deriválttal

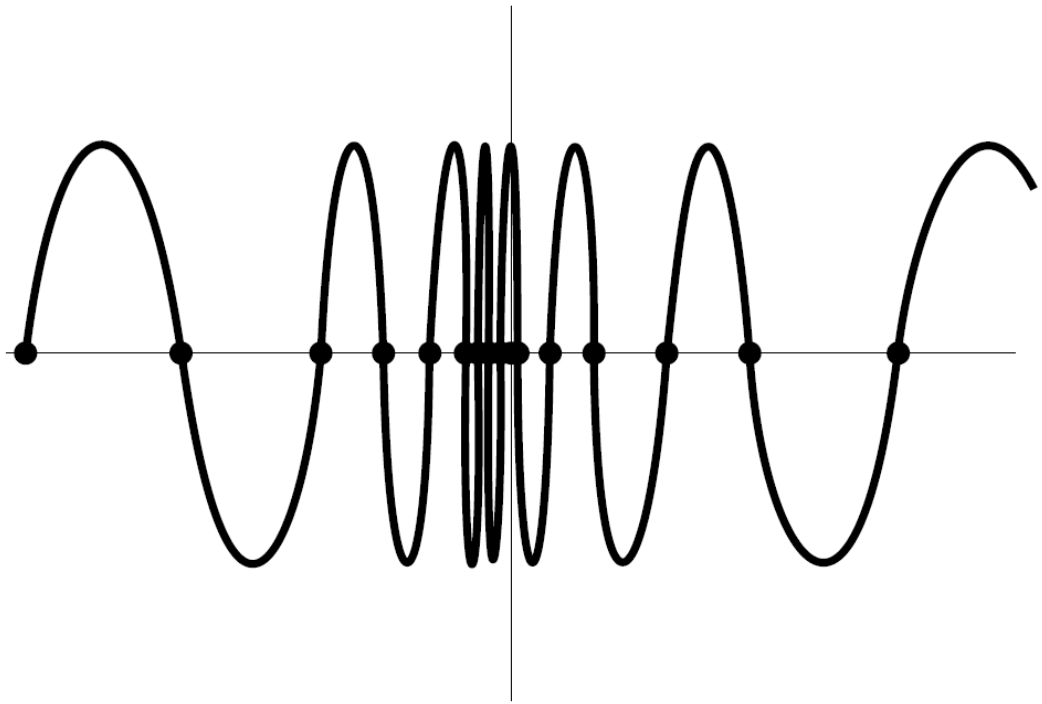

Figure: Problémás eset: a derivált divergál, a gyökök besűrűsödnek
©Numerical Recipes

Gyökök és minimumok konfigurációi

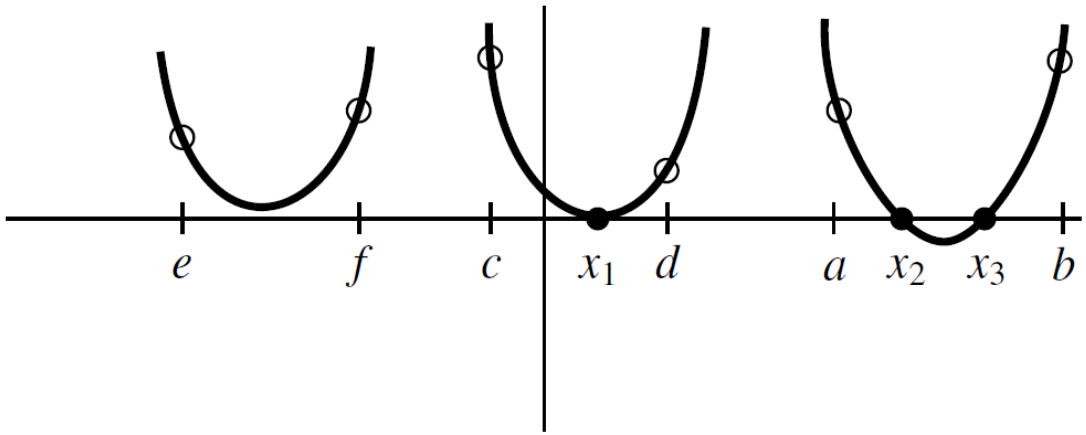

Figure: Fekete pöttyök jelölik a gyököket, karikák a keretezés pontjait. A keretezés nem biztos, hogy jelzi a gyök jelenlétét. ©Numerical Recipes

Egydimenziós módszerek

A gyökkeresés a függvény analízisével kezdődik!

A következő módszereket tekintjük át röviden

- felezés
- szekáns
- regula falsi
- Brent
- Newton-Ralphson

Felezős módszer

Kiindulás

- valamilyen módon sikerült bekeretezni a gyököt : $[a, b]$ intervallum
- $f(a) \cdot f(b) < 0$, azaz a függvény előjelet vált

Iteratív lépés

- megfelezzük az intervallumot és kiértékeljük a függvényt
- a felezőpontban vett függvényérték előjele megegyezik valamelyik végpont előjelével
- a felezőpont felváltja azt az intervallum végpontot, amelyben vett függényérték előjellel megegyezik

A felezős módszer leállási feltétele

A konvergenciához szükséges lépések száma

- a kiindulási intervallum hossza: ϵ_0
- a gyök értékére előírt pontosság: ϵ
- a módszer az intervallumot felezi, így a lépések száma

$$n = \log_2 \frac{\epsilon_0}{\epsilon}$$

Leállási feltétel

- ha az intervallum hossza eléri ϵ-t, megállunk
- gyöknek az intervallum felezőpontját tekintjük
- figyelem: a számítógép számábrázolása nem egyformán pontos pl a 0 és 10^8 körül (lásd Progalap, "Adattípusok" előadás)
- ezért az ϵ előírt pontosság függhet attól, hogy milyen tartományban keressük a gyököt

A felezős módszer tulajdonságai

Fő előny:

- a módszer mindig konvergál

Konvergencia gyorsasága: lineáris

- ha a gyök n lépés után egy ϵ_n intervallumra van lokalizáva, akkor a következő lépésben az intervallum nagysága feleződik

$$\epsilon_{n+1} = \frac{1}{2}\epsilon_n \qquad \text{vagyis} \qquad \epsilon_{n+1} = c\epsilon_n$$

A felezős módszer tulajdonságai

Fő előny:

- a módszer mindig konvergál

Konvergencia gyorsasága: lineáris

- ha a gyök n lépés után egy ϵ_n intervallumra van lokalizáva, akkor a következő lépésben az intervallum nagysága feleződik

$$\epsilon_{n+1} = \frac{1}{2}\epsilon_n \qquad \text{vagyis} \qquad \epsilon_{n+1} = c\epsilon_n$$

Fő gondok:

- csak előjelváltó esetben működik
- csak egy gyököt talál meg
- ha a függvénynek szakadása van, és ezért vált előjelet, akkor a módszer nem a gyököt, hanem a szakadási pontot találja meg

A szekáns módszer

Kiindulás

- valamilyen módon megközelítettük a gyököt
- nem is kell feltétlen bekeretezni
- egy x_1 és x_2 pontból indulunk

Iteratív lépés

- tekintjük az $f(x_{i-1})$ és $f(x_i)$ értékeket
- x_{i-1} és x_i között lineárisan interpoláljuk a függvény
- az x_{i+1} értéke a lineáris interpoláció gyöke lesz
- $f(x_i)$-t és $f(x_{i+1})$-t használjuk a következő lépésben

A szekáns módszer

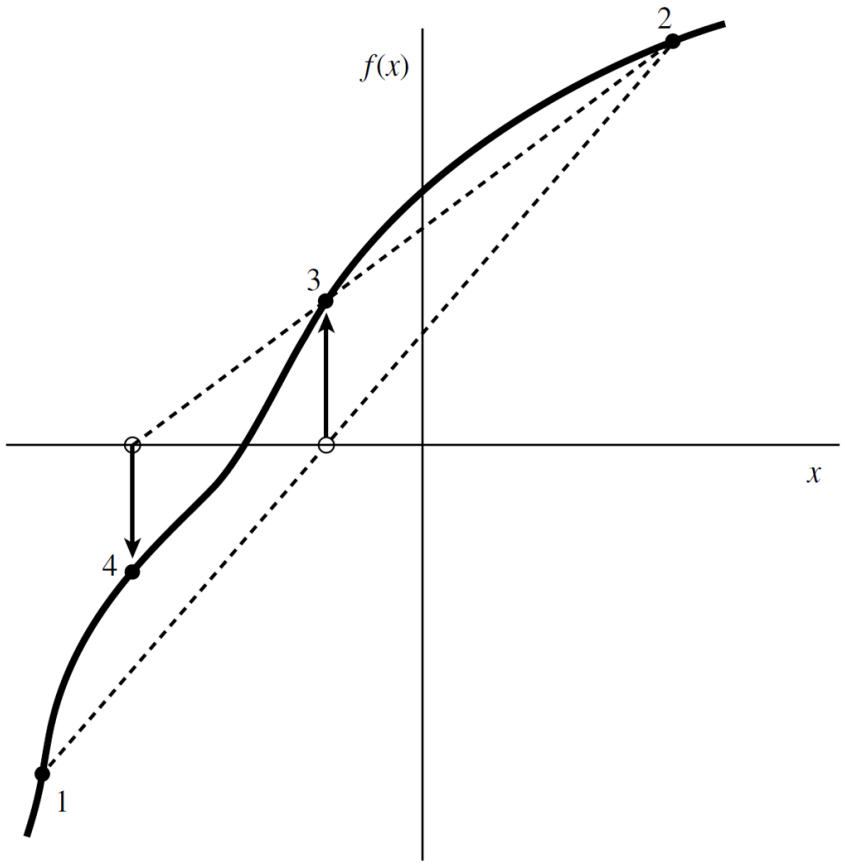

Figure: Szaggatott vonal mutatja a lineáris interpolációt az első két interációs lépésben. ©Numerical Recipes

A szekáns módszer tulajdonságai

Gyorsabban konvergál, mint a felezős módszer

- ott az intervallum mindig a felére csökkent
- itt a konvergencia sebessége ϵ gyenge hatványával megy

$$\lim_{i \to \infty} |\epsilon_{i+1}| = c \cdot |\epsilon_i|^{\alpha}, \qquad \alpha \simeq 1.618$$

A módszer nem tartja bekeretezve a gyököt

- nem csak ott működik, ahol a függvény előjelet vált
- de ha a gyök lokális minimum környékén van, akkor eltéved
- akár a végtelenbe is elmehet

Akkor működik jól, ha a függvény jól közelíthető egyenessel.

A regula falsi módszer algoritmusa

Kiindulás

- valamilyen módon sikerült bekeretezni a gyököt
- adott tehát egy $[x_1, x_2]$ intervallum
- $f(x_1)f(x_2) < 0$, azaz a függvény előjelet vált

Iteratív lépés

- $[x_1, x_2]$-n lineárisan interpolálunk
- megkeressük az egyenes és az x-tengely metszéspontját: x_3
- az x_{i-1}, x_i és x_{i+1} közül azt a kettőt tartjuk meg, amelyeknél a függvény előjele különböző

Némileg gyorsabb, mint a felezős módszer

Regula falsi módszer

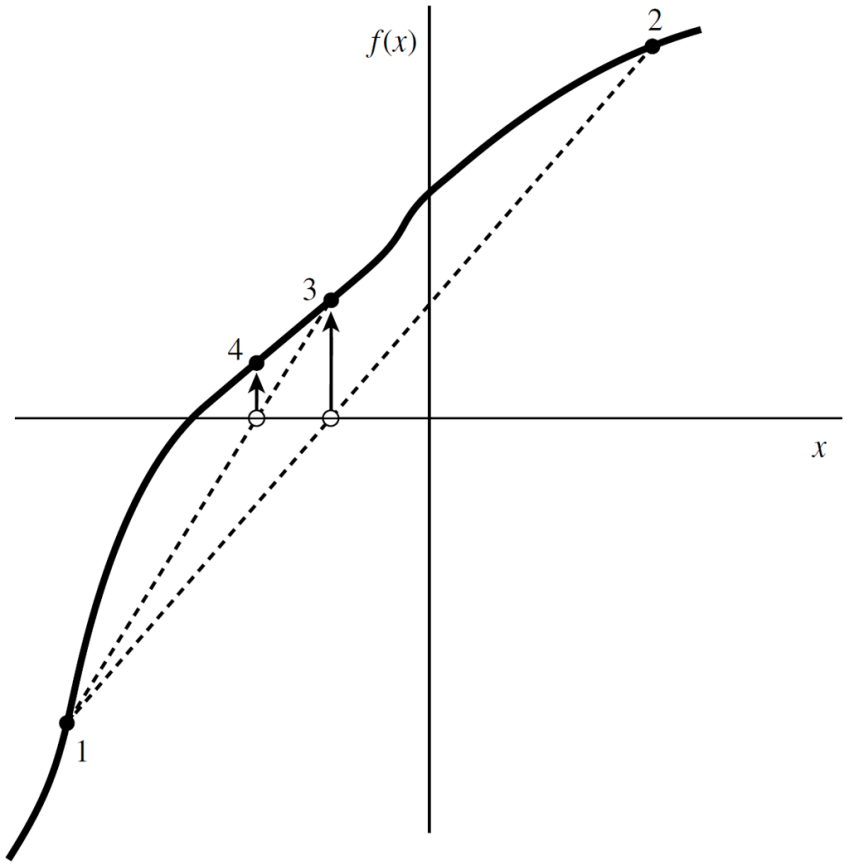

Figure: Szaggatott vonal mutatja a lineáris interpolációt az első két interációs lépésben. ©Numerical Recipes

Problémás függvény

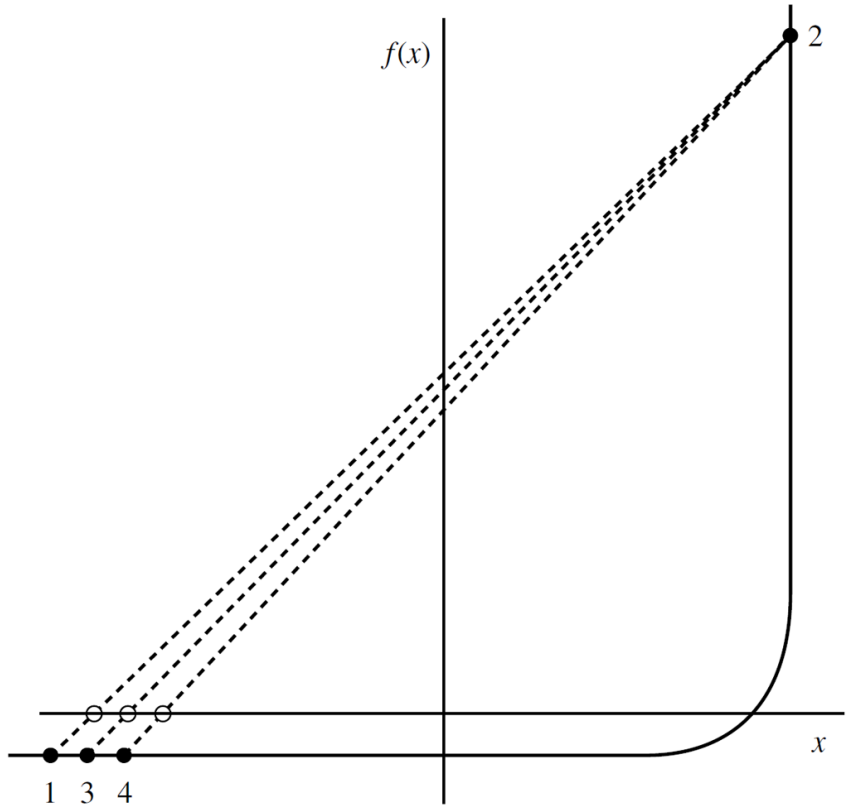

Figure: Egy eset, amikor a szekáns és a regula farsi módszer is lassan konvergál.
©Numerical Recipes

A belassulást elkerülő módszerek

Vannak függvények, ahol az elvileg gyorsabb módszerek lemaradnak

- pl. a regula falsi lehet lassabb a felezős módszernél is
- ilyenkor érdemes a módszereket kombinálni
- figyeljük a konvergencia sebességét
- ha lassú, akkor más módszerre váltunk

A belassulást elkerülő módszerek

Vannak függvények, ahol az elvileg gyorsabb módszerek lemaradnak

- pl. a regula falsi lehet lassabb a felezős módszernél is
- ilyenkor érdemes a módszereket kombinálni
- figyeljük a konvergencia sebességét
- ha lassú, akkor más módszerre váltunk

Wijngaarden–Dekker–Brent-módszer (röviden Brent)

- alapesetben az ún. inverz kvadratikus módszert használja
- vagyis nem egyenessel, hanem kvadratikus függvénnyel interpolálunk
- ehhez nem elég az intervallum két végpontja, hanem három ponttal dolgozunk

A módszerben vannak ellenőrzési pontok

- ha az új pont a kereten kívül esik
- ha konvergencia túl lassú

$\Rightarrow$ akkor inkább felezi az intervallumot

A függvény deriváltjának felhasználása

Az eddigit módszerek nem használták a függvény deriváltját

- viszont lehet, hogy az is expliciten adott
- ekkor gyorsabb módszer lehetséges

A Newton–Ralphson-módszer

- egyetlen x_1 pontból indulunk
- meghúzzuk a függvény érintőjét, ehhez kiszámoljuk az $f'(x_1)$ deriváltat
- az x tengellyel vett metszet lesz az új pont

Iterációs lépés:

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}$$

A Newton–Ralphson-módszer

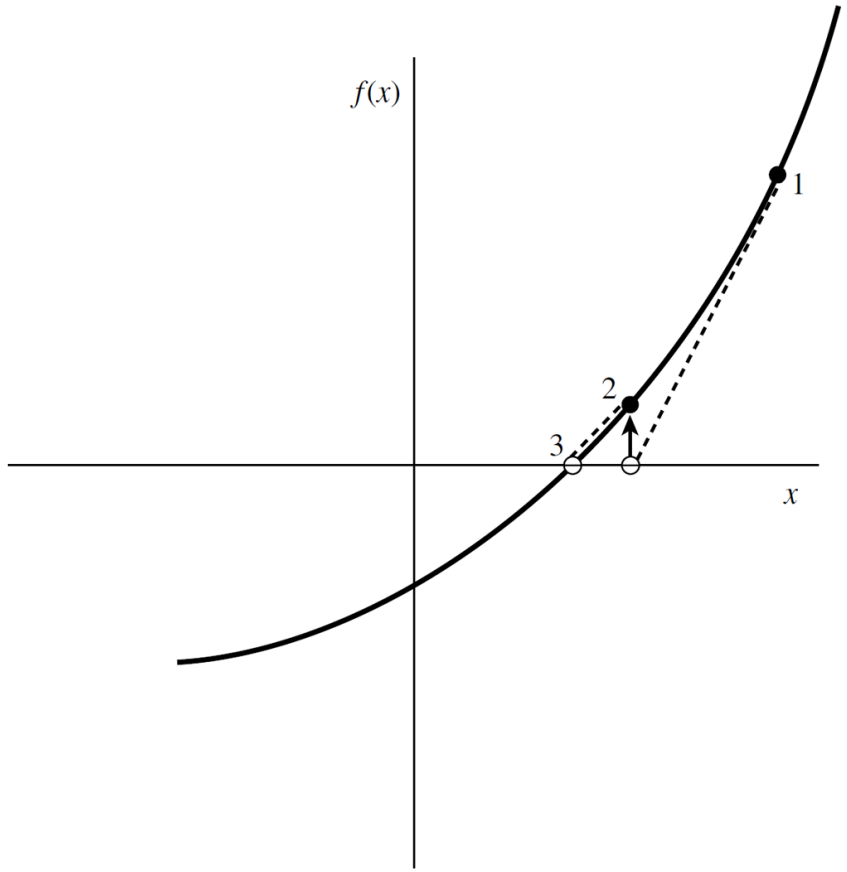

Figure: Szaggatott vonal mutatja a függvény érintőjét az első két interációs lépésben. ©Numerical Recipes

A Newton–Ralphson-módszer tulajdonságai

Kellően sima függvény esetén jól működik, ugyanis

- a gyök körüli Taylor-sorból elég csak az első tagot megtartani

$$f(x+\delta) \approx f(x) + f'(x)\delta + \frac{1}{2}f''(x)\delta^2 + \ldots$$

- ezzel a gyöktől való eltérés egy jó közelítése

$$\delta \approx -\frac{f(x)}{f'(x)}$$

Belátható, hogy a konvergencia sebessége négyzetes:

$$\epsilon_{i+1} = c \cdot \epsilon_i^2$$

Problámás helyzetek

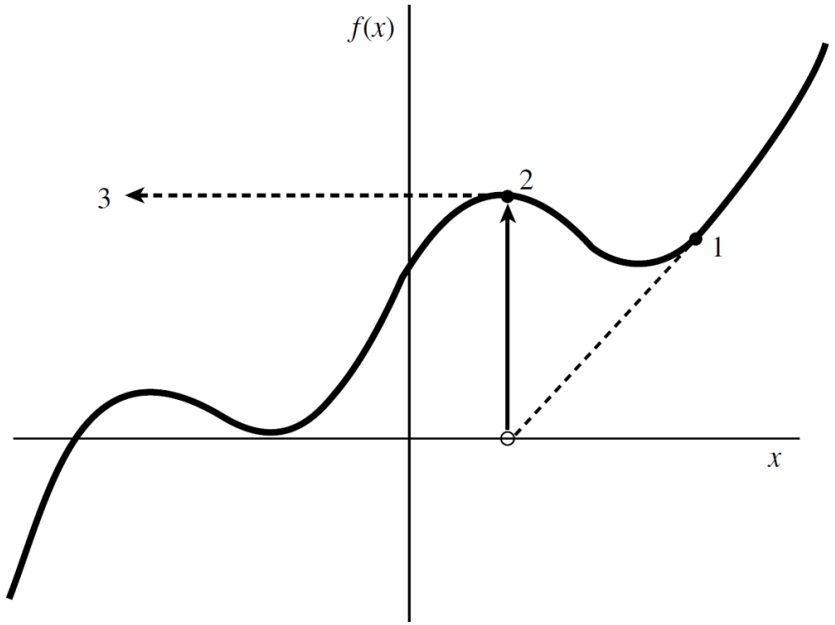

Figure: Mivel $f'(x_2) \approx 0$, ezért $x_3 = x_2 - \frac{f(x_2)}{f'(x_2)} \to -\infty$. ©Numerical Recipes

Problámás helyzetek

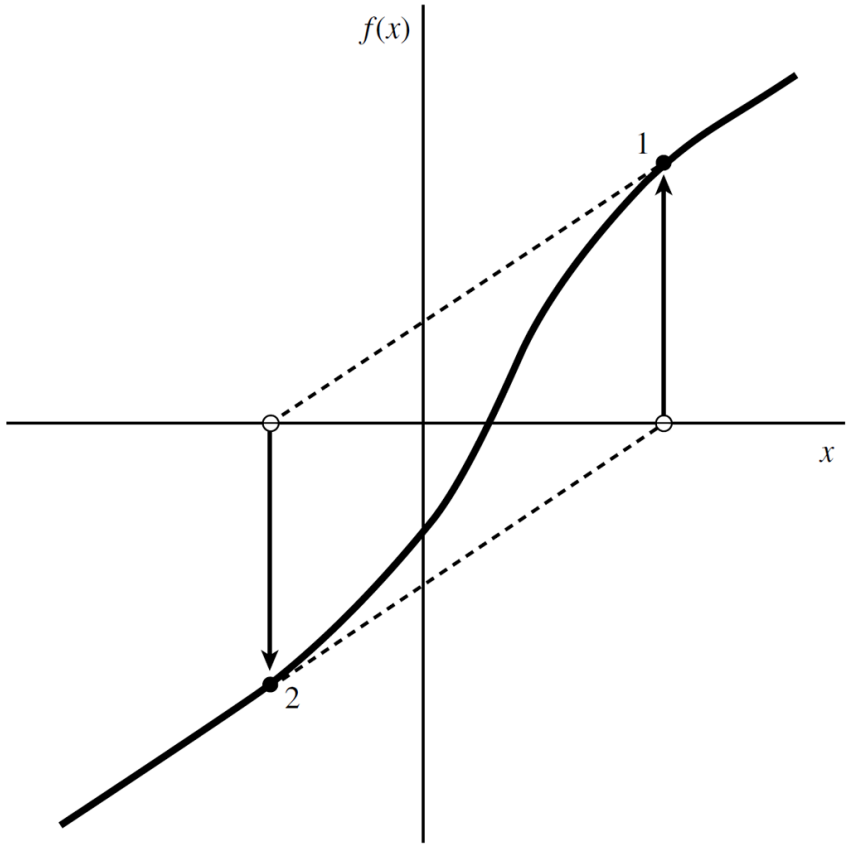

Figure: (közel) Végtelen ciklus és lassú konvergencia alakulhat ki. ©Numerical Recipes

További kérdések a N–R-módszerrel kapcsolatban

1). Ha $f'(x)$ nem adott expliciten, vagy nem számolható könnyen, elvileg használhatnánk a numerikus deriált közelítését is:

$$f'(x_i) \approx \frac{f(x_i + dx) - f(x_i)}{dx}$$

Egy dimenzióban nem érdemes numerikus deriváltat használni

- nem elég pontos, így lassítja a konvergenciát
- a szekáns módszer gyorsabb

További kérdések a N–R-módszerrel kapcsolatban

1). Ha $f'(x)$ nem adott expliciten, vagy nem számolható könnyen, elvileg használhatnánk a numerikus deriált közelítését is:

$$f'(x_i) \approx \frac{f(x_i + dx) - f(x_i)}{dx}$$

Egy dimenzióban nem érdemes numerikus deriváltat használni

- nem elég pontos, így lassítja a konvergenciát
- a szekáns módszer gyorsabb

2.) Ha a Newton–Ralphson-módszer nem konvergál

- ilyenkor érdemes hibrid módszert választani, a Brent-módszerhez hasonlóan
- pl más módszerrel bekeretezzük a gyököt
- ha már közel járunk, akkor N–R-módszerrel pontosítjuk

Polinomok gyökei

Az n-ed rendű polinomnak n gyöke van, nem feltétlenül mind különböző

- ha a polinom együtthatói együtthatók valósak, akkor
  - a gyökök valósak, vagy
  - komplex konjugált párok
- ha a polinom együtthatói komplexek, akkor a komplex gyökök között általában nincs kapcsolat

Ha egy gyök többszörös, akkor ott a nem csak a függvény, hanem a deriváltja is eltűnik, ezért a N-R módszer sem biztos, hogy működik

Polinomok gyökeinek megtalálása

Általános stratégia: a gyököket egyesével keressük

- valamilyen módszerrel egy gyököt megtalálunk: r
- leosztjuk a polinomot $(x - r)$-rel

$$Q(x) = P(x)/(x - r)$$

- $Q(x)$ meghatározására létezik egyszerű algoritmus
- az eggyel alacsonyabb fokú polinom egy gyökét ismét valami elemi algoritmussal keressük
- ha komplex aritmetikával dolgozunk, akkor a leosztás a komplex gyökök esetén is működik

Polinomok gyökeinek megtalálása

Egy másik stratégia:

- egy olyan mátrixot felírni, amelynek a sajátértékei megegyeznek a polinom gyökeivel
- mátrix sajátérékeinek számolására is vannak kifinomult eljárások
- általában lassabb, mint a leosztásos eljárás, de egyes esetekben pontosabb lehet

Polinomok gyökeinek megtalálása

Egy másik stratégia:

- egy olyan mátrixot felírni, amelynek a sajátértékei megegyeznek a polinom gyökeivel
- mátrix sajátérékeinek számolására is vannak kifinomult eljárások
- általában lassabb, mint a leosztásos eljárás, de egyes esetekben pontosabb lehet

Háttér: tudjuk, hogy a $\mathbf{A}$ mátrix sajátértékei az ún *karakterisztukus polinom* gyökei: $P(x) = \det[\mathbf{A} - x\mathbf{I}]$.

Ezt megfordítva: ha $P(x) = \sum_{i=0}^{m} a_i x^i$ akkor

$$A = \begin{pmatrix} -\frac{a_{m-1}}{a_m} & -\frac{a_{m-2}}{a_m} & \dots & -\frac{a_0}{a_m} \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

sajátértékei megyeznek $P(x)$ gyökeivel.

Polinomok gyökeinek megtalálása

Problémák:

- a gyökkeresés sok lépést igényel, felhalmozódnak a numerikus hibák
- a megtalált gyök, amivel osztunk, sem teljesen pontos

Javítási lehetőség

- megkeressük a gyököket
- egyenként pontosítjuk őket a Newton–Ralphson-módszerrel

Módszerek komplex gyökök megtalálására

- ld. Numerical Recipes: Laguerre-módszer

Gyökök keresése több dimenzióban

Több dimenzióban nincsen általánosan jó gyökkereső módszer.
PI két dimenzióban

$$\begin{aligned} f(x,y) &= 0 \\ g(x,y) &= 0 \end{aligned}$$

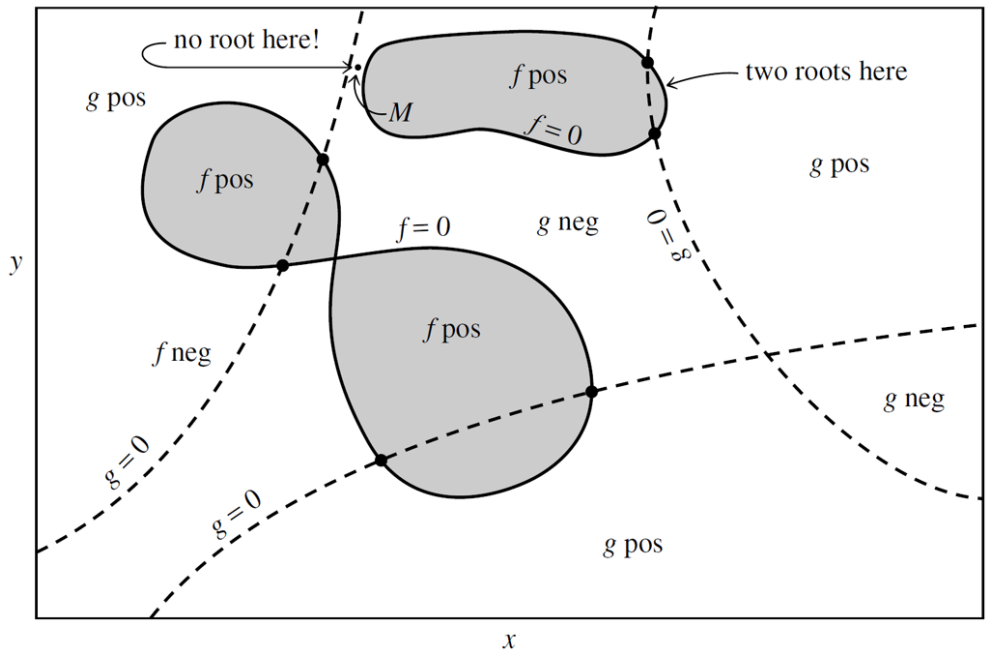

Figure: Az egyes függvények zéró kontúrvonalainak a metszéspontjait keressük. A gyökök száma, *a priori*, ismeretlen. ©Numerical Recipes

Gyökök keresése több dimenzióban

Mindenképpen tudnunk kell részleteket is a problémáról

- hány gyököt várunk
- ezeket kb. hol várjuk

Jobb híján

- olyan módszert keresünk, ami egy gyököt meg tud találni
- abban az esetben, ha a gyökhöz elég közelről indulunk

Newton–Ralphson-módszer több dimenzióban

Tekintsük a következő egyenletrendszert

$$f_i(x_1, x_2, ..., x_N) = f_i(\mathbf{x}) = 0 \qquad i = 1...N$$

Taylor sorfejtés

$$f_i(\mathbf{x} + \delta\mathbf{x}) = f_i(\mathbf{x}) + \sum_{j=1}^{N} \frac{\partial f_i}{\partial x_j} \delta x_j + \dots$$

A deriváltat most a teljes Jacobi-mátrix helyettesíti:

$$J_{ij} = \frac{\partial f_i}{\partial x_j}$$

Newton–Ralphson-módszer több dimenzióban

Tekintsük a következő egyenletrendszert

$$f_i(x_1, x_2, ..., x_N) = f_i(\mathbf{x}) = 0 \qquad i = 1...N$$

Taylor sorfejtés

$$f_i(\mathbf{x} + \delta\mathbf{x}) = f_i(\mathbf{x}) + \sum_{j=1}^{N} \frac{\partial f_i}{\partial x_j} \delta x_j + \dots$$

A deriváltat most a teljes Jacobi-mátrix helyettesíti:

$$J_{ij} = \frac{\partial f_i}{\partial x_j}$$

Ezzel az $\mathbf{F} = (f_1(\mathbf{x}), f_2(\mathbf{x}), \dots f_N(\mathbf{x}))^T$ függvény Taylor-sora

$$\mathbf{F}(\mathbf{x} + \delta\mathbf{x}) = \mathbf{F}(\mathbf{x}) + \mathbf{J}\delta\mathbf{x} + ...$$

Newton–Ralphson-módszer több dimenzióban

F függvény Taylor-sora

$$\mathbf{F}(\mathbf{x} + \delta\mathbf{x}) = \mathbf{F}(\mathbf{x}) + \mathbf{J}\delta\mathbf{x} + \ldots$$

- keressük azt a $\delta\mathbf{x}$ vektort, ami a gyök irányába lépteti az aktuális legjobb becslést

Ha $\mathbf{F}(\mathbf{x} + \delta\mathbf{x}) = 0$ akkor a $\mathbf{J}\delta\mathbf{x} = -\mathbf{F}$ lineáris egyenletrendszert kell megoldani

Iterációs lépés:

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \delta\mathbf{x}$$

Ha **J**-t nehéz analitikusan számolni, akkor, jobb híján, a véges differenciák segítségével numerikusan is számolhatóak a parciális deriváltak

Newton–Ralphson-módszer több dimenzióban

Az egydimenziós esetben láttuk, hogy a N-R módszer bizonyos esetben eldivergál.

Hogyan tudjuk eldönteni, hogy a többdimenziós esetben számolt $\delta\mathbf{x}$ lépés elfogadható-e?

Newton–Ralphson-módszer több dimenzióban

Az egydimenziós esetben láttuk, hogy a N-R módszer bizonyos esetben eldivergál.

Hogyan tudjuk eldönteni, hogy a többdimenziós esetben számolt $\delta\mathbf{x}$ lépés elfogadható-e?

- tekintsük az $f = |\mathbf{F}|^2 = \mathbf{F}^T \cdot \mathbf{F} \geq 0$ függvényt
- ha $\mathbf{F} = 0$ akkor f nyilván minimális
- akkor fogadjuk el a $\delta\mathbf{x}$ lépést, ha az adott iterációban az f értéke csökken

Newton–Ralphson-módszer több dimenzióban

Az egydimenziós esetben láttuk, hogy a N-R módszer bizonyos esetben eldivergál.

Hogyan tudjuk eldönteni, hogy a többdimenziós esetben számolt $\delta\mathbf{x}$ lépés elfogadható-e?

- tekintsük az $f = |\mathbf{F}|^2 = \mathbf{F}^T \cdot \mathbf{F} \geq 0$ függvényt
- ha $\mathbf{F} = 0$ akkor f nyilván minimális
- akkor fogadjuk el a $\delta\mathbf{x}$ lépést, ha az adott iterációban az f értéke csökken

Mi van, ha $\delta\mathbf{x}$ lépés nem elfogadható?

- visszalépés (backtracking)

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \lambda\delta\mathbf{x} \qquad 0 < \lambda < 1$$

- λ választására vannak kifinomult módszerek, lásd Numerical Recipies, Sect. IX

Köszönöm a figyelmet!