

Hledání lokálního maxima funkce algoritmus »postup06«

Petr Byczanski

Ústav geoniky AV ČR

Abstrakt : Lokální maximum diferencovatelné funkce je hledáno postupnou změnou argumentu . V okolí každého postupového bodu je funkční průběh aproximován pouze několika členy Taylorova rozvoje . Změna argumentu určená na základě aproximativního průběhu je primárně omezena zachováním jistého stupně platnosti aproximace ! Toto je jádrem prezentovaného algoritmu . Velikost oblasti použitelnosti aproximace je určována jediným řídicím parametrem .

Velmi jednoduchý algoritmus »postup06« je určen pro nalezení argumentu lokálního maxima funkce F , když platí

- (A) $F: D \subset E_n \rightarrow E_1$ má spojité první parciální derivace
 - (B) je zadán výchozí bod X , o němž je předpokládáno, že se z něj dá spojitou změnou dosáhnout lokálního maxima .
-

Principiální myšlenkové řešení je : Začínáme v zadaném bodě . Provádíme myšlené infinitezimální změny úměrné gradientu funkce . Pohybujeme se tím po teoretické křivce směřující do lokálního maxima resp. do bodu vzrůstu funkčních hodnot nade všechny meze . Tento cílový bod nemusíme dosáhnout, když se zastavíme ve stacionárním bodě nebo narazíme na hranici definičního oboru .

Předkládaný algoritmus je diskrétní analogií uvedeného myšlenkového postupu . V okolí každého postupového bodu nahradíme přesný průběh funkce F pouze prvními dvěma členy Taylorova rozvoje . Lineární aproximační funkce určuje jistou postupovou polopřímku . Po ní se pohybujeme nejvýše do okamžiku, kdy bychom vyjeli ..

- .. z oblasti alespoň přibližné platnosti aproximace
- .. z definičního oboru

Díky první podmínce je podstatná část změny zachycena v aktuální aproximaci . V následujícím postupovém bodě proto určitě dojde ke zvětšení funkční hodnoty a určitě se příliš neodchýlíme od teoretické postupové křivky vedoucí z předchozího postupového bodu do cílového bodu .

Posloupnost postupových kroků ukončíme při dosažení téměř nulové velikosti gradientu funkce resp. při směřování mimo definiční obor .

Míra platnosti aproximace je zadána jediným řídicím parametrem .

Použité označování : $\partial_i F$ označuje první parciální derivaci F podle i -té složky argumentu . Po vynechání indexu označuje ∂F vektor tvořený prvními parciálními derivacemi .

Přesný myšlenkový postup

Pro enumeraci souvislé změny si zavedeme číselný parametr t s nulovou hodnotou odpovídající začátku . Teoretická postupová křivka $x(t)$ splňuje

$$x(0) = X \quad (1a)$$

$$\frac{d}{dt} x(t) = \partial F(x(t)) \quad (1b)$$

Rovnice (1b) zajišťuje "co nejrychlejší" postup k cíli . Fakticky nás nezajímá celé řešení , ale pouze koncový bod !

Skutečný diskretní postup

Soustavu diferenciálních rovnic (1) nebudeme řešit přesně .

Z výchozího bodu X se začneme pohybovat podél tečné přímky . Při jistém odklonu od teoretické křivky (1) první postupový krok ukončíme . Dostáváme se tím do jistého postupového bodu , který bude obecně odlišný od cílového bodu .

V tomto novém postupovém bodě celou úvahu zopakujeme . Vede z něj trochu jiná teoretická křivka k požadovanému cíli . Opět je popsána soustavou (1) , v rovnici (1a) bude na pravé straně aktuální postupový bod . Podél tečné přímky vedené počátkem opět postoupíme blíže k cíli .

Tento proces opakujeme až do jeho vhodného ukončení .

Výchozí daný bod a všechny další body , do nichž dojdeme , budeme nazývat postupovými body . Když vynecháme jejich očíslování , opět si je označíme pomocí X .

Postupový krok

Teoretickým základem každého kroku postupu je rozvoj

$$F(X + \Delta x) = F(X) + \partial F(X)^T \times \Delta x + ..?.. \quad (2)$$

Závislost na změně argumentu Δx budeme aproximovat pouze prvními dvěma členy rozvoje

$$f + v^T \times \Delta x \quad (3)$$

kde

$$f \equiv F(X) \quad (4a)$$

$$v \equiv \partial F(X) \quad (4b)$$

Odpovídající přibližná postupová křivka je dána rovnicemi

$$\Delta x(0) = \text{nulový vektor} \quad (5a)$$

$$\frac{d}{dt} \Delta x(t) = v \quad (5b)$$

Řešením je

$$\Delta x(t) = t \cdot v \quad (6)$$

Tato aktuální postupová polopřímka se přimyká k teoretické postupové křivce (1) odpovídající aktuálnímu výchozímu postupovému bodu tím více, čím menší je Δx . Proto alespoň část postupu podle (1) můžeme nahradit částí postupu podle (6).

Zásadní otázkou je: Jak daleko smíme podle (6) postoupit?

Délka kroku postupu

Podstatou předkládaného algoritmu je zachování jistého stupně platnosti aproximace. Intuitivně je jasné, že pokud 2. člen v (2) podstatně převýší neznámý 3. člen \dots , budeme jej moci zanedbat. Zvýšení aproximační hodnoty pak bude automaticky znamenat také zvýšení skutečné funkční hodnoty. Tim se určitě přiblížíme k cíli.

Stěžejní relaci zajišťující jistý stupeň platnosti aproximace můžeme napsat ve tvaru

$$|F(X + t \cdot v) - f - t \cdot v^T \times v| \leq \frac{t \cdot v^T \times v}{\mu} \quad (7)$$

Nalevo je absolutní hodnota neznámého členu \dots , v čitateli napravo je známý 2. člen rozvoje. Konstanta μ ve jmenovateli je ostře větší než 1. Reguluje míru platnosti aproximace. Udává totiž dílek aproximační změny funkční hodnoty, o který se může lišit skutečná změna funkční hodnoty.

Musíme nalézt co největší hodnotu parametru t splňujícího relaci (7). Vzhledem ke kontextu použití postačují pouze hodnoty z množiny

$$\{ \dots, 1/8, 1/4, 1/2, 1, 2, 4, 8, \dots \}$$

Největší vyhovující prvek nalezneme binárním půlením. Začneme buď nějakou "rozumnou" hodnotou, např. $t = 1$, anebo výslednou hodnotou z předchozího kroku. Pro tuto počáteční hodnotu ověříme splnění relace (7). V kladném případě budeme opakovaně zkoušet $t \leftarrow t \cdot 2$. Při prvním neúspěchu se vrátíme k předchozí hodnotě. V záporném případě budeme opakovaně zkoušet $t \leftarrow t/2$. Při prvním úspěchu jsme hotovi.

Ověřování relace (7) vyžaduje výpočet funkční hodnoty v bodě $X + t \cdot v$. V tomto "průzkumném" bodě se (na rozdíl od postupového bodu) nevyčísľují derivace!

Nalezené t určí další postupový bod

$$X \leftarrow X + t \cdot v \quad (8)$$

Ukončení postupových kroků

Teoreticky je postup ukončen, když je $v \equiv \partial F(X)$ nulové anebo když jsme se velmi přiblížili hranici definičního oboru a směřujeme mimo něj.

Numericky nulovost gradientu funkce většinou nenastane! Proto je žádoucí přidat nějakou uživatelskou podmínku ukončení přirozeně odpovídající konkrétní

řešené úloze (např. dolní mez velikosti v atp.). Nutné podmínky "strojového" ukončení jsou :

- .. nedojde ke změně polohy postupového bodu
($X + t \cdot v$ je numericky shodné s X)
- .. nedojde ke vzrůstu funkční hodnoty
($t \cdot v^T \times v$ je numericky nula)

Volba řídicího parametru

Jediným řídicím parametrem předkládaného algoritmu je konstanta μ figurující v (7) .

Kdybychom extrémně použili hodnotu 1 , mohlo by se stát , že v jednom postupovém kroku vůbec nedojde ke zvětšení funkční hodnoty . Proces hledání maxima by vůbec nemusel skončit .

Při použití hodnoty větší než 1 určitě dojde v každém kroku ke zvětšení funkční hodnoty .

Jedním z hledisek pro volbu μ může být , jak velké odchylení od teoretické křivky (1) připustíme . Menší odchylce odpovídá větší μ .

Jiné mnohem praktičtější hledisko je , snažit se o co nejmenší počet postupových kroků . Příliš malá hodnota blízká 1 může vést ke vzrůstu počtu kroků díky velkému odklonu od křivky vedoucí k cíli . Příliš velké hodnoty vedou ke vzrůstu počtu kroků díky častějšímu přerušování postupu . Optimální bude něco "uprostřed" . Praktické pokusy ukázaly , že vhodnou je hodnota v rozmezí 1.5 až 2 .

Poznámky

Stěžejní myšlenka - dodržení zachování jistého stupně platnosti aproximace - byla již úspěšně použita také v [1] a [2] .

Použití kvadratické aproximace vede ke dvěma kvalitativně odlišným situacím : Aproximační funkce je buď neomezená (jako zde) nebo omezená . V druhém případě je vhodným kandidátem na další postupový bod poloha maxima aproximační funkce , samozřejmě pokud leží v oblasti přibližné platnosti aproximace .

Pokus s použitím aproximace pomocí tří členů Taylorovy řady byl učiněn za účelem prozkoumání , zdali nebude lokální maximum nalezeno rychleji . Srovnání bylo provedeno na jedné konkrétní úloze . Na ní byl "složitější" algoritmus asi 30x pomalejší oproti zde prezentovanému algoritmu . Hlavní příčinou zpomalení je asi (opakované) provádění diagonalizace symetrické matice koeficientů kvadratického členu .

-
- [1] Byczanski,P. : Řešení soustavy nelineárních rovnic , algoritmus »postup01«
Sborník semináře Moderní matematické metody v inženýrství
Ostrava VŠB-TUO 2000 s.22-25
- [2] Byczanski,P. : Řešení soustavy nelineárních rovnic , algoritmus »postup08«
Sborník konference MATLAB 2000
Praha Humusoft 2000 s.61-66