Posudek oponenta diplomové práce

Řešení problému nejmenších čtverců s maticemi o proměnlivé hustotě nenulových prvků

Bc. Ilona Riegrová

Shrnutí obsahu práce

Práce se zaměřuje na numerické metody řešení problému nejmenších čtverců

$$\min_x \|Ax - b\|_2 ,$$

přičemž matice A má speciální nenulovou strukturu takovou, že ji lze rozdělit na řídkou a hustou část. Této struktuře je pak možné přizpůsobit i numerické metody.

Autorka práce prezentuje přehled numerických metod využívajících řídko-husté struktury matice A k efektivnějšímu řešení dané úlohy. Po teoretickém popisu metod v prvních třech kapitolách se v numerických experimentech věnuje porovnání uvažovaných přístupů.

Hodnocení práce

- **Téma** práce je aktuální, práce těží ze současných publikací vycházejících v renomovaných časopisech. Podle mého názoru je však téma pojato velmi široce a kladlo tak na autorku příliš vysoké nároky ohledně orientace v příslušné literatuře. To se projevilo i na textu práce.
- **Práce** je logicky členěna a **matematický text** je většinou dobře zformulován. Nicméně, v práci se najdou místa která nejsou příliš srozumitelná nebo využívají zbytečně silných předpokladů (například, že A je čtvercová a regulární).
- **Zdroje** jsou správně a v dostatečné míře citovány, práce obsahuje malé množství tiskových chyb.
- **Vlastním příspěvkem** autorky je především shrnutí aktuálních poznatků o uvažovaných metodách a provedení numerických experimentů, ve kterých srovnává vybrané metody.
- Práci celkově hodnotím jako **průměrnou**.

Podrobnější komentáře

- Práce popisuje velké množství přístupů a algoritmů, avšak pouze povrchním způsobem. V kapitole 1 je tento přístup pochopitelný, protože se jedná o popis dobře známé problematiky. Kapitoly 2 a 3 však tvoří jádro celé práce, jedná se o málo známé přístupy a proto měly být důkladněji zpracovány, vysvětleny a prezentovány ve formě algoritmů. Tyto dvě kapitoly nejsou dobře čitelné bez toho, aniž by čtenář zároveň nahlížel do citovaných zdrojů.

- Numerické experimenty zaslouží na jednu stranu pochvalu, protože zformulovat a naprogramovat uvažované metody není jednoduché a vyžaduje velkou dávku trpělivost a pečlivosti. Srovnání metod z pohledu počtu iterací přineslo jistě zajímavé výsledky. Na druhou stranu, srovnání metod z pohledu časové náročnosti je zavádějící ze dvou důvodů:
  - K testování byl použit Matlab, ve kterém velmi záleží na kódu a na využití vestavěných funkcí. Časové srovnání tak bez dalších technik (viz otázka 4) nelze považovat za věrohodné. Lepším (i když také ne vždy spravedlivým) způsobem srovnání časové náročnosti metod by bylo odhadnout počet aritmetických operací na jednu iteraci.
  - Místo řešení soustav s trojúhelníkovými maticemi autorka počítá skutečné inverze trojúhelníkových matic. Inverzi jedné matice často počítá v kódu několikrát. Pokud by šlo pouze o faktory příslušné k husté části matice, které jsou malé, tak by použití inverze nebylo tolik významné. Autorka však počítá i inverze velkých řídkých faktorů, které jsou pak plné. Tím uměle navyšuje výpočetní i paměťovou náročnost algoritmů a jde tak přímo proti filosofii uvažovaných metod, které se snaží využít řídkosti.

Závěr

Předložená práce dosáhla cílů zadání, svým rozsahem a úrovní zpracování splňuje požadavky kladené na diplomovou práci. Doporučuji ji uznat jako diplomovou práci.

V Praze, 11. června 2020

doc. RNDr. Petr Tichý, Ph.D.
KNM, MFF UK

Otázky k diskusi

1. Po prezentaci algoritmu PCGLS na straně 18 autorka tvrdí, že existují další dvě modifikace této metody, CGNR a CGNE. Jaký je rozdíl mezi CGLS a CGNR?
2. Co znamená, že LSQR je numericky spolehlivější než CGLS? (strana 19)
3. Je možné srovnat tři uvažované metody (aktualizace, kombinace částečných řešení a Schurova doplňku) z pohledu počtu operací, případně počtu operací na jednu iteraci a z hlediska paměťové náročnosti?
4. V Matlabu je možné použít příkaz `profile` ke zjištění časové náročnosti jednotlivých částí kódu (například `profile on` před výpočtem, `profile off` po výpočtu, a `profile report` k vygenerování přehledu). Zkusila jste použít tento způsob testování časové náročnosti matlabovského kódu?

Poznámky

Na tyto poznámky není nutné reagovat a slouží pouze jako doložení argumentů zmíněných v posudku. Slouží též jako zpětná vazba pro autorku práce. Samozřejmě se k nim může během obhajoby vyjádřit, pokud s nimi nesouhlasí či pokud cítí potřebu je uvést na pravou míru.

- Někdy je používána $\|\cdot\|$ a někdy $\|\cdot\|_2$, v obou případech by však mělo jít o stejnou normu.
- Autorka je často nekonsistentní v předpokladech týkajících se rozměrů matice (zda je matice čtvercová či obdélníková). Například LU rozklad je definován pouze pro čtvercové matice, pak je ale používán pro obdélníkové matice. Před použitím LU rozkladu pro obdélníkovou matici by bylo vhodné alespoň krátké vysvětlení, jak vzniká LU rozklad obdélníkové matice.
- Matice YX a XY v definice pseudoinverze nejsou ortogonální, ale symetrické. Můžeme o nich hovořit také jako o ortogonálních projektorech.
- Odvození na straně 6-7 předpokládá, že A je čtvercová a regulární. Pak je ovšem vše triviální a není co odvozovat. Mělo se předpokládat, že A je obdélníková s plnou sloupcovou hodností.
- Na straně 6 se na prvním řádku píše, že není potřeba řádková pivotace a na posledním řádku se píše, že je potřeba řádková pivotace.
- V (1.4) měl být použit symbol r pro reziduum, místo y, konzistentně s předchozím textem.
- Uprostřed strany 7 je uvedeno, že matice soustavy (1.4) je čtvercová, symetrická a regulární. To ovšem platí pouze za předpokladu, že A má plnou sloupcovou hodnost. Zároveň, má-li A plnou sloupcovou hodnost, pak je $A \neq 0$ a matice je indefinitní.
- Strana 8. V dolní části autorka hovoří o aplikaci reflexí a rotací a formálně je popisuje pomocí matic $P_1, \ldots, P_n$. Uvedený formální popis sedí na Householderovy reflexe, kterých je n a jsou symetrické. Formální popis však nesedí na Givensovy rotace, které jsou obecně nesymetrické a jejich počet je větší než n.
- Strana 10, 3. řádek. Choleského rozklad je uvažován pro čtvercové matice A. Co je tedy m na řádku 3? Pravděpodobně má autorka na mysli Choleského rozklad matice $A^T A$, kde $A \in \mathbb{R}^{m \times n}$. To by ale mělo být v textu explicitně řečeno, protože ve Větě 10, po které text bezprostředně následuje, je A čtvercová.
- Strana 10, poslední odstavec. Stabilní algoritmus pro SVD nevychází z Lanczosova procesu, ale z převedení matice A na horní či dolní bidiagonální tvar pomocí ortogonálních transformací. Následně je aplikován Demmel-Kahanův algoritmus, který spočte singulární rozklad bidiagonální matice.
- Strana 11, řádek -9. Pokud přeuspořádáváme sloupce, aplikujeme permutace zprava.

- Strana 12, odstavec 1.3.1.1. Odvození mohlo být lépe a podrobněji vysvětleno. Například mělo být zdůvodněno proč se x, které je řešením $U\Pi_2 x^T = c + z$, počítá ve dvou krocích. Z algoritmu 1 to pak čtenáři dojde, nicméně vysvětleno to mělo být již na straně 12.

- Proč jsou matice A v algoritmech 1, 2 a 3 čtvercové? Algoritmy přece potřebujeme použít pro obdélníkové matice.

- Strana 29. Příklad binárního řádkového stretchingu je velmi nesrozumitelně popsán. Není zřejmé, co je a_{m1} a a_{m2}. Soustava (2.8) je dle mého názoru špatně zformulována a není z ní jasné, jak se rozšíří původní matice systému. Počet neznámých v (2.8) nesedí s rozměry matice.

- Strana 29, (2.9). Pravděpodobně mělo být

$$z = \begin{pmatrix} x \\ s \end{pmatrix}.$$

- Strana 29. Tabulka 4.1 prezentuje rozměry m a n. Z tabulky plyne, že $\operatorname{rank}(A) = n$ i v případech, kdy je $m < n$ (což není možné). Není ani zřejmé, co se myslí tím, že obdélníková matice je SPD. Autorka tím pravděpodobně myslela $\operatorname{rank}(A) = \min(n, m)$ a matice $A^T A$ je SPD v případě $m > n$. To se ale může čtenář jen domýšlet.