



Posudek oponenta závěrečné práce

Student: Lukáš Jurásek
Oponent práce: Ing. Daniel Langr, Ph.D.
Název práce: Efektivní LU rozklad pro řídké matice
Obor: Teoretická informatika

Datum vytvoření: 31. 5. 2019

Hodnotící kritérium:	Způsob hodnocení – následující škálou 1 až 4:
1. Splnění zadání	<u>1=zadání splněno,</u> 2=zadání splněno s menšími výhradami, 3=zadání splněno s většími výhradami, 4=zadání nesplněno
<i>Popis kritéria:</i> Posuďte, zda předložená ZP dostatečně a v souladu se zadáním obsahově vymezuje cíle, správně je formuluje a v dostatečné kvalitě naplňuje. V komentáři uveďte body zadání, které nebyly splněny, posuďte závažnost, dopady a případně i příčiny jednotlivých nedostatků. Pokud zadání svou náročností vybočuje ze standardů pro daný typ práce nebo student případně vypracoval ZP nad rámec zadání, popište, jak se to projevilo na požadované kvalitě splnění zadání a jakým způsobem toto ovlivnilo výsledné hodnocení.	
<i>Komentář:</i> Práce dostatečně splňuje zadání, které se zabývá možnostmi paralelizace a efektivní implementace Doolittlova algoritmu pro LU rozkladu řídkých matic a příslušnými heuristikami pro minimalizaci vzniku nenulových maticových prvků.	
Hodnotící kritérium:	Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):
2. Písemná část práce	69 (D)
<i>Popis kritéria:</i> Zhodnoťte přiměřenost rozsahu předložené ZP vzhledem k obsahu, tj. zda všechny části ZP jsou informačně bohaté a ZP neobsahuje zbytečné části. Dále posuďte, zda předložená ZP je po věcné stránce v pořádku, případně vyskytují-li se v práci věcné chyby nebo nepřesnosti. Zhodnoťte dále logickou strukturu ZP, návaznosti jednotlivých kapitol a pochopitelnost textu pro čtenáře. Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 26/2017, článek 3. Posuďte, zda student využil a správně citoval relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami. Zhodnoťte, zda převzatý software a jiná autorská díla, byly v ZP použity v souladu s licenčními podmínkami.	

Komentář:

K věcné a formální stránce práce mám následující připomínky:

- str. 1 - "1.část" -> "1. část" (chybí mezera), podobně u 2. a 3. části.
- Úvod je velmi stručný. Minimálně mi zde chybí představení LU rozkladu, tj. co dělá a k čemu se používá. To samé platí o zmiňovaném Doolittlově algoritmu. Rovněž např. autor zmiňuje minimalizaci tzv. "fill-in", aniž by vysvětlil, co "fill-in" znamená. Úvod by měl tvořit ucelený a pochopitelný text, který čtenáře "vtáhne" do problému, což se v tomto případě neděje.
- str. 4 - Samostatný řádek s tečkou, která by měla být součástí vzorce.
- str. 5 - "V ideálním případě by tyto pole..." -> "tato pole".
- str. 5 - "...byla seřazena primárně podle čísla řádku a sekundárně podle čísla sloupce..." - stačí napsat "seřazena lexikograficky". Netuším, co autor myslí tím "ideálním případem", ale obecně toto tvrzení určitě neplatí. Vhodnost řazení silně závisí na aplikaci. Např. pro násobení matic je vhodnější mít "levou" matici seřazenou podle uvedeného klíče řádek-sloupec, ale "pravou" matici podle klíče "sloupec-řádek". Pro "stromové" zpracování matice je vhodné tzv. Z-Mortonovo řazení. Pro blokové zpracování zase řazení prvků s ohledem na příslušnost ke konkrétním blokům.
- str. 8 - Seznamy jsou buď součástí jedné věty nebo jejich položky tvoří samostatné věty. V prvním případě jsou položky odděleny čárkou a začínají obecně malým písmenem. V opačném případě tečkou a začínají velkým písmenem. Na konci každé položky musí být každopádně čárka (případně středník) nebo tečka.
- str. 11 - "Doolittlův algoritmus[5]" -> "Doolittlův algoritmus [5]" - chybí mezera. Navíc, obecně, je vhodnější citace dávat až na konec věty, pokud to není matoucí.
- Bylo by vhodné doplnit, jak se LU rozklad následně použije pro řešení soustav rovnic s více pravými stranami.

- Poměrně silně mi chybí přestavení verze Doolittlova algoritmu pro řídké formáty matic. Uvedený algoritmus (Algorithm 1) je zapsaný pro husté matice, tj. používá přímou indexaci maticových prvků, která má u hustých matic konstantní složitost. Použití této verze algoritmu pro např. formáty CSR či CSC by asi nebylo moc efektivní. Autor tyto skutečnosti sice diskutuje, ale již neukazuje algoritmy LU rozkladu pro (např. zmiňované) řídké formáty. Poté rovnou přechází k paralelizaci, která rovněž vychází z algoritmu vhodného pro husté matice. Zde by byla vhodná i řešerše a představení řešení pro LU rozklady řídkých matic. Ty jsou sice diskutovány v kapitole 4, ale jen velmi stručně bez uvedení konkrétních použitých přístupů, metod a algoritmů.

- str. 14 - "...avšak je nastane zpomalení..." - pravděpodobně jen "avšak nastane".
- str. 14 - "\npořadí" -> "v" by nemělo docházet k zalomení řádku.
- str. 25 - přetečení šířky řádku.

- sekce 5.3 - Implementace a paralelizace algoritmů, které tvoří jádro celé práce, jsou představeny velmi stručně. Zde by měl být mnohem detailnější popis a paralelní verze algoritmů by měly být uvedeny ve formě pseudokódu pro matice uložené v konkrétních řídkých formátech.

- str. 29 - "standart" -> "standard".
- str. 30 - "...byly také vygenerovány náhodné matice..." - zcela chybí popis metody generování náhodných matic.
- sekce 6.4 - Chybí návaznost textu na konkrétní výsledky uvedené v grafech - obrázky nejsou z textu vůbec referované.

- Ve výsledcích postrádám naprosto základní měření pro paralelní algoritmus, a to je měření škálovatelnosti (čas v závislosti na počtu vláken).

- Dále by bylo vhodné porovnat implementované metody LU rozkladu s existujícími řešeními, které poskytují např. knihovny představené v kapitole 4.

- Závěr je velmi stručný a vůbec neobsahuje dosažené výsledky práce. Zde by měly být stručně shrnuty naměřené experimenty a závěry udělané na základě jejich analýzy.

Při práci s referencemi jsem nenarazil na žádný problém.

Hodnotící kritérium:

Způsob hodnocení – bodové hodnocení 0 až 100 bodů
(známka A až F):

3. Nepísemná část, přílohy

70 (C)

Popis kritéria:

Dle charakteru práce se případně vyjádřete k nepísemné části ZP. Například: SW dílo – kvalita vytvořeného programu a vhodnost a přiměřenost technologií, které byly využité od vývoje až po nasazení. HW – funkční vzorek – použité technologie a nástroje, Výzkumná a experimentální práce – opakovatelnost experimentů

Komentář:

Zdrojové kódy byly napsané v C++. Defakto celá implementace se nachází v jediné třídě, což mi po pravdě připadá jako ne moc šťastné řešení. Vhodnější by bylo např. vytvořit sadu tříd pro matice (různého typu, různého formátu, jako šablony pro podporu prvků různých typů, apod.) a zvlášť implementovat algoritmy, které s maticemi pracují. Přístup "jedna třída, která umí vše" není v praxi příliš doporučován a v budoucnu vede na problémy s rozšiřováním a udržitelností kódu.

Další připomínky:

- "using namespace std;" - Nikdy nepoužívat v hlavičkových souborech!
- Hlavičkový soubor Matrix.h obsahuje direktivy #include i pro hlavičkové soubory knihovny C++, které v Matrix.h vůbec nejsou potřeba. Co jsem našel, tak se jedná o iostream, queue, map, algorithm a limits. Tyto hlavičkové soubory by měly být vloženy ve zdrojovém souboru Matrix.cpp. (Obecně tento problém způsobuje větší závislosti, prodlužuje dobu předkladu, apod.).
- Standard C++ nedefinuje typ "uint" - jeho použití proto vede na nepřenositelnost kódu.

Hodnotící kritérium:

Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):

4. Hodnocení výsledků, jejich využitelnost

80 (B)

Popis kritéria:

Dle charakteru práce zhodnoťte možnosti nasazení výsledků práce v praxi nebo uveďte, zda výsledky ZP rozšiřují již publikované známé výsledky nebo přinášející zcela nové poznatky.

Komentář:

Dosažené výsledky dávají jistý vhled do paralelizace Doolittlova algoritmu a heuristik pro minimalizace vzniku nenulových prvků během LU rozkladu. Víceméně z práce vyplývá, že vlastní LU rozklad je dobře paralelizovatelný, zatímco u heuristik je to podstatně horší, což je důležité zjištění. Pro větší relevanci výsledků by zde bylo vhodné porovnání autorovy implementace s existujícími implementacemi, minimálně u vlastního LU rozkladu.

Hodnotící kritérium:

Způsob hodnocení – nehodnotí se

5. Otázky k obhajobě

Popis kritéria:

Uveďte případné dotazy, které by měl student zodpovědět při obhajobě ZP před komisí (body oddělte odrážkami).

Otázky:

- 1) Str. 15 - "Výstupem algoritmu je permutační vektor..." - Není vůbec uvedeno, co se s tímto permutačním vektorem dělá. Slouží k proházení maticových řádků? Sloupců? Jaký má takové proházení vliv na řešenou soustavu rovnic?
- 2) Sekce 3.1 - U heuristik MMDO a RCMK jsou diskutovány možnosti paralelizace (sekce 3.1.2.1 a 3.1.4.1). Pro heuristiky MDO a CMK ale taková diskuze chybí. Proč?
- 3) Algoritmus 4 (Markowitzova strategie) obsahuje operace prohození dvou řádků a sloupců. Tato operace lze snadno provést na hustých maticích. Jak ji ale provádět na maticích uložených v řídkých formátech? Existují nějaké varianty tohoto algoritmu specializované pro konkrétní řídké formáty?

Hodnotící kritérium:

Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):

6. Celkové hodnocení

70 (C)

Popis kritéria:

Shrňte stránky ZP, které nejvíce ovlivnily Vaše celkové hodnocení. Celkové hodnocení nemusí být aritmetickým průměrem či jinou hodnotou vypočtenou z hodnocení v předchozích jednotlivých kritériích. Obecně platí, že bezvadně splněné zadání je hodnoceno klasifikačním stupněm A.

Text hodnocení:

Celkově práce zadání splňuje a přináší zajímavé poznatky z oblasti paralelizace LU rozkladu řídkých matic a heuristik na minimalizaci vzniku nenulových prvků. Většina zmíněných nedostatků je spíše drobného charakteru, ale jako zásadní považují absenci detailního popisu autorem navržené paralelizace algoritmů pro konkrétní řídké formáty a jejich prezentaci ve formě pseudokódu (autor pouze odkazuje na uvedené algoritmy, které ale jsou určeny primárně pro husté matice; navíc jsou pouze sekvencí).

Podpis oponenta práce: