Posudek oponenta na bakalářskou práci

Kanonický rozklad tenzorů násobení polynomů a násobení matic na GF(2)

Vojtěch Obhlídal

1 Obecné shrnutí

Práce si klade za cíl rozklad tenzorů, které odpovídají násobení matic a násobení polynomů, na součet co nejmenšího počtu tenzorů ranku 1, tedy nalezení horního odhadu ranku těchto tenzorů. Motivací pro zkoumání uvedených tenzorů je, že jejich rank je roven minimálnímu počtu skalárních násobení při výpočtu součinu matic nebo součinu polynomů v algoritmech, kde se výsledek získá jako součet součinů lineárních kombinací vstupů nad daným tělesem.

Rank diskutovaných tenzorů je důležitý především pro těleso reálných čísel a jsou významné především rozklady, jejichž koeficienty jsou celá čísla. Hledání takových rozkladů lze rozdělit na dva kroky. V prvním kroku je nalezen rozklad nad dvouprvkovým tělesem GF(2) a ve druhém kroku se tento rozklad využije pro hledání rozkladu v celých číslech. Rozklad nad GF(2) lze ve druhém kroku využít různými způsoby. Jeden z možných postupů hledá celočíselné koeficienty, jejichž paritu určuje řešení nad GF(2). Při dalším možném postupu se hledají racionální koeficienty a řešení nad GF(2) je využito jako informace, které z nich jsou nulové a které nenulové. Pokud uvažujeme rozklady s koeficienty $\{-1, 0, 1\}$, pak jsou tyto postupy ekvivalentní. Pokud koeficienty mohou obsahovat nenulová sudá celá čísla, pak ekvivalentní nejsou. V práci je první postup využit nepřímo pro důkaz existence rozkladu nad GF(2), pokud existuje rozklad v celých číslech. Pro vlastní výpočty rozkladů v celých číslech byl testován druhý postup.

Vzhledem k tomu, že se práce zabývá rankem tenzoru nad konečným tělesem, je vhodné doplnit, že test existence rozkladu daného tenzoru na daný počet komponent ranku 1 je pro libovolné konečné těleso NP-úplná úloha. Polynomiální algoritmus pro tuto úlohu tedy pravděpodobně neexistuje a je potřeba hledat co nejefektivnější postupy, i když budou mít asymptoticky exponenciální složitost.

Úlohu nalezení rozkladu tenzoru na daný počet komponent ranku 1 lze vyjádřit jako řešení soustavy polynomiálních rovnic nad daným tělesem. V prvním kroku řešení se uvažuje těleso GF(2), jehož prvky jsou 0 a 1 a je tedy relativně snadné převést řešení této soustavy na problém splnitelnosti výrokových formulí v CNF (conjunctive normal form). Problém splnitelnosti formulí ve tvaru CNF je znám pod označením SAT problém a pro jeho řešení existuje dostupný software, tedy SAT řešiče. Těchto řešičů existuje větší počet a pro úlohy řešené v práci se osvědčuje SAT řešič Minisat 1.14.

Práce ze zabývá především rozklady tenzoru násobení polynomů, které mají nejvýše 7 koeficientů. Hlavním výsledkem práce je nalezení rozkladů těchto tenzorů nad GF(2), které splňují různé varianty požadavku na symetrii. Práce ukazuje, že tato metoda umožňuje nalézt rozklady, které dokazují stejné horní odhady ranku nad GF(2) pro tenzor násobení polynomů s počty koeficientů $n = 5, 6, 7$ jako v [1], kde jsou prezentovány také odpovídající rozklady v celých

číslech. Nalezení odpovídajících rozkladů v celých číslech se pro tyto hodnoty n v práci nepodařilo.

2 Připomínky k textu

2.1 Sekce 2.3, Převod soustavy rovnic na SAT problém a jeho řešení

Vysvětlení převodu soustavy polynomiálních rovnic nad GF(2) na formuli v CNF považuji za nedostatečné. Práce sice obsahuje zavedení formulí v CNF, ale sestrojení formule pro uvažovanou soustavu popisuje pouze tím, že je použit program Bosphorus. Tento program provádí určité optimalizace vstupní soustavy, ale pak použije standardní přepis rovnic nad GF(2) do SAT problému. Protože způsob přepisu do SAT je pro efektivitu řešení podstatný, je důležité vysvětlit, jak je proveden a případně, jaké optimalizace jsou možné. Toto v práci chybí a místo toho je na straně 28 uveden příklad výstupu programu Bosphorus pro příklad soustavy rovnic. Výstup je uveden v technickém formátu DIMACS, který se používá jako vstup pro SAT řešič, ale není vhodný pro účely popisu formule v textu. Uvedenou formuli lze poměrně snadno popsat s využitím běžného značení. Pro každý monom v soustavě na konci předchozí strany 27 obsahuje pomocnou proměnnou a klauzule odpovídající operacím násobení a sčítání. Například, první rovnice soustavy je

$$x_7 \cdot x_1 + x_{10} \cdot x_3 + x_{13} \cdot x_5 + 1 = 0 \; .$$

Po přidání pomocných proměnných y_1, y_2, y_3 se tato rovnice změní na soustavu rovnic

$$\begin{aligned} x_7 \cdot x_1 &= y_1 \\ x_{10} \cdot x_3 &= y_2 \\ x_{13} \cdot x_5 &= y_3 \\ y_1 + y_2 + y_3 &= 1 \end{aligned}$$

a každá z těchto rovnic je přepsána na minimální formuli v CNF, která ji vyjadřuje. Například první rovnici vyjadřuje formule

$$(x_7 \vee \neg y_1) \wedge (x_1 \vee \neg y_1) \wedge (\neg x_7 \vee \neg x_1 \vee y_1) \; ,$$

další dvě rovnice jsou analogické a poslední rovnici vyjadřuje formule

$$(y_1 \vee y_2 \vee y_3) \wedge (y_1 \vee \neg y_2 \vee \neg y_3) \wedge (\neg y_1 \vee y_2 \vee \neg y_3) \wedge (\neg y_1 \vee \neg y_2 \vee y_3) \; .$$

Použitím tohoto postupu pro všechny rovnice v soustavě se získá formule, která je až na číslování proměnných shodná s formulí na straně 28.

2.2 Sekce 2.4, Převod rozkladu z GF(2) na $\mathbb{Z}$

V této sekci je uvedena konstrukce soustavy rovnic s proměnnými v reálných číslech, která je odvozena ze vstupního rozkladu nad GF(2) tak, že v maticích A a B se nuly zachovávají a jedničky se nahrazují proměnnou. V matici C se všechny prvky nahrazují proměnnou. Jako vysvětlení tohoto rozdílu je uvedeno,

že ve známých rozkladech jsou v maticích A a B koeficienty z množiny $\{-1, 0, 1\}$, zatímco v matici C jsou koeficienty v $\mathbb{Z}$. Vzhledem k tomu, že $\{-1, 0, 1\}$ je podmnožina $\mathbb{Z}$, má zřejmě autor na mysli, že v matici C se vyskytují celá čísla s absolutní hodnotou větší než 1. Z toho, že všechny prvky matice C se nahrazují proměnnými, vyplývá, že uvedená metoda nevyužívá informaci o matici C z nalezeného rozkladu nad GF(2). Soustava je pak řešena dosazovací metodou bez dalšího upřesnění. V rámci obhajoby je potřeba použitou metodu řešení vysvětlit podrobněji. V rámci tohoto vysvětlení by bylo dobré zodpovědět následující otázky.

- Jakým způsobem se vybírají proměnné a rovnice pro jejich vyjádření, které se využije k dosazení do zbývajících rovnic?
- Využívá se předpoklad, že nenulové prvky matic A a B jsou z množiny $\{-1, 1\}$ a splňují tedy rovnici $x^2 = 1$?
- K jaké soustavě vede uvedený postup v případě, kdy tato metoda nedojde k řešení? Lze v těchto případech dojít k závěru, že soustava nemá řešení?

Pro násobení polynomů, které mají 5 až 7 koeficientů, jsou v práci uvedeny rozklady nad GF(2), které splňují různé varianty podmínek na symetrii. Pro žádný z nich se výše uvedenou metodou nepodařilo nalézt odpovídající rozklad nad $\mathbb{Z}$. Pomocí formulí v CNF lze vyjádřit sčítání celých čísel omezené velikosti. Bylo by vhodné vyzkoušet také hledání rozšíření na celočíselné řešení s využitím této metody, kdy vstupní rozklad nad GF(2) určuje paritu celočíselných koeficientů, tedy nejnižší bit v jejich binárním zápisu.

2.3 Další připomínky

- Str 12, řádek 10. Vpravo mají být sloupcové vektory, tedy $\mathrm{vec}(E^T), \mathrm{vec}(F^T)$ místo $\mathrm{vec}(E^T)^T, \mathrm{vec}(F^T)^T$. Podobně, na straně 14, řádek 14, jsou ve vyjádření násobení polynomů pomocí tenzorů použity řádkové matice $\overline{p}^T$ a $\overline{q}^T$, ale mají být sloupcové, tedy bez transpozice.
- Na straně 18 na konci sekce 1.3.2.1 dochází ke změně významu symbolu n. V předchozích částech textu označuje n stupeň polynomu a od tohoto místa dále označuje počet koeficientů, který je o 1 větší. To komplikuje porovnání vzorců v různých částech textu a je to pravděpodobně důvod pro některé překlepy. Například na str 14, řádek 10 jsou chybně uvedeny exponenty n, m, p, které mají být $n+1, m+1, p+1$. Na řádku 12 na stejné straně je chybně uveden exponent $(n + 1)(m + 1)(n + m - 1)$, který má být $(n + 1)(m + 1)(n + m + 1)$.
- Autor práce zaměňuje rank tenzoru a počet komponent rozkladu. Rank tenzoru je minimální počet komponent rozkladu, ale toto minimum často není známo. Například, na straně 19, řádek 15 je rovnost $\mathrm{rank}(\mathcal{P}_{n+1,m+1}) = R$, kde R je počet komponent rozkladu. Tento počet se nemusí rovnat ranku tenzoru, rovnost tedy obecně neplatí a algoritmus platnost rovnosti nevyžaduje. Podobně, na straně 29 je rank tenzoru součástí vstupu algoritmu, ale tento rank algoritmus nevyužívá a vstupem je počet komponent rozkladu.

- Na straně 24 je uvedena tabulka počtů proměnných a počtů rovnic v soustavě, která popisuje rozklad tenzoru, ale není řečeno, o jaký tenzor se jedná. Není tedy jasné, co tabulka znamená. Podle počtu rovnic lze usoudit, že se jedná o tenzor násobení polynomů s využitím symetrie, ale toto by mělo být v textu explicitně řečeno.

- V Poznámce 2.6 není zřejmé, o jakou disjunkci $\vee$ se jedná. Pro konjunkci $\wedge$ je uveden běžný způsob zkrácení formulí a je zdůvodněn tím, že soustavu booleovských formulí můžeme považovat za soustavu algebraických rovnic. Není zřejmé, jak je toto míněno, protože uvedený způsob zkrácení se používá běžně i v případech, kdy se jedná výlučně o jeden typ formulí. Navíc, Poznámka 1.11 ukazuje převod ANF na logickou formuli s operací XOR, ale převod mezi formulí v CNF a formulí s operací XOR formuli podstatně mění a nejde tedy pouze o změnu značení.

- Na straně 31 v sekci 3.1.1 se vychází z předpokladu, že algoritmus násobení polynomů musí provést součin $a_0 b_0$, tedy součin absolutních členů polynomů. To sice pravděpodobně pro dobré algoritmy platí, ale tvrzení na řádku 13, že v citaci [1] lze nalézt odvození tohoto faktu, není pravda. V [1] je toto formulováno jako vhodný předpoklad, nikoli jako tvrzení.

3 Hodnocení

Souvislost počtu skalárních násobení v algoritmech pro výpočet násobení matic a polynomů s rozklady odpovídajících tenzorů je v práci vysvětlena srozumitelně. Převod soustavy rovnic nad tělesem GF(2) na úlohu pro SAT řešič je vysvětlen pouze odkazem na software Bosphorus a není tedy vysvětlen ani jeho základní princip, který je poměrně jenoduchý. Pokud jde o převod řešení na celá čísla, je popsán způsob, jak úlohu vyjádřit jako soustavu rovnic v reálných číslech s poznámkou, že soustava je řešena dosazovací metodou. Tato metoda není obecně efektivní a není vysvětleno za jakých podmínek může být efektivní ve zkoumaném kontextu. S pomocí dosazovací metody nebyly nalezeny rozklady pro tenzory násobení polynomů pro 5 nebo více koeficientů, což jsou hlavní případy, kterými se práce zabývá. Navrhuji klasifikaci C (dobře) ze stupnice A až F.

V Praze 18. srpna 2020

RNDr. Petr Savický, CSc.