Dr. hab. inż. Sergiy Fialko Kraków, dnia 24 marca 2021 r.

Politechnika Krakowska im. Tadeusza Kościuszki

RECENZJA

rozprawy doktorskiej

mgra. inż. Konrada Sewiłłó-Jopek

**„Parallel multi-thread multi-frontal solver using element partition trees”**

1. PODSTAWA OPRACOWANIA RECENZJI.

Podstawę opracowania recenzji stanowi list Przewodniczącego Rady Dyscypliny Informatyka Techniczna i Telekomunikacja dr hab. inż. Marka Kisiel-Dorochnickiego, prof. n. AGH, z dnia 16.02.2021 r. zlecający wykonanie recenzji. Przedmiotem oceny jest praca doktorska, opracowana przez mgr inż. Konrada Sewiłłó-Jopek „Parallel multi-thread multi-frontal solver using element partition trees”. Promotorami przewodu są prof. dr hab. Maciej Paszyński oraz dr hab. Anna Paszyńska. Rozprawa doktorska liczy 112 stron, składa się z 4 rozdziałów i listy źródeł literaturowych mieszczącej 75 pozycji.

1. OCENA PRZEDMIOTU ROZPRAWY

W prace doktorskiej jest przedstawiona jedną z możliwych implementacji solwera wielofrontalnego dla rozwiązywania układów równań liniowych algebraicznych z macierzami rzadkimi wynikającymi z zastosowania metody elementów skończonych. Ten solwer jest przeznaczony dla rozwiązania problemów adaptacyjnych hp na komputerach wielordzeniowych z pamięcią wspólną architektury Symmetric Multiprocessing (SMP). Zrównoleglenie pozostało wykonane na podstawie wielowątkowości. Aktualność pracy polega na następującym:

* Tworzeniu specyficznego solwera bezpośredniego dla macierzy rzadkich, zorientowanego na problemy *hp* adaptacji, ponieważ solwery uniwersalne czasami nie są najlepsze przy rozwiązaniu takich zadań.
* Liczba rdzeni i przepustowość pamięci RAM komputerów wielordzeniowych szerokiego użycia stale rosną, co wraz ze wzrostem objętości tej pamięci umożliwia rozwiązanie wielu zagadnień MES bez odwołania do dysku, co z kolei otwiera drogę do tworzenia oprogramowania charakteryzującego się znacznym przyspieszeniem wraz ze wzrostem liczby wątków. Jeszcze kilka lat temu rozwiązanie takich zadań na komputerach SMP było nie do pomyślenia bez wykorzystania pamięci dyskowej.
1. OCENA OSIĄGNIĘTYCH WYNIKÓW

Rozdział 1 zawiera wprowadzenie w metody wielofrontalne. Odnotowano, że początkową informacją dla dobrze znanych algebraicznych wielofrontalnych solwerów jest źródłowa rzadka macierz przedstawiona w jednym z formatów skompresowanych. Jednocześnie autor stosuje inne podejście, w którym początkową informacją dla solwera jest frontalne drzewo użyte w tej pracy jako *element partition tree*. Drzewo frontalne pozwala zarządzać całym procesem faktoryzacji numerycznej, która odbywa się w macierzach gęstych stosunkowo nie wielkiego rozmiaru – macierzach frontalnych. W takiej czy innej formie takie podejście było już spotykane w literaturze. Przedstawiono także główne etapy klasycznego solwera wielofrontalnego – tworzenie grafu sąsiedztwa macierzy źródłowej, uporządkowanie w celu zredukowania zapełnieni w macierzy sfaktoryzowanej, tworzenie drzewa eliminacji (drzewa frontalnego), jego uporządkowanie, oraz poszukiwanie super wierzchołków w celu zwiększenia wydajności procedury obliczania dopełnienia Schura opartej na algorytmie mnożenia macierzy gęstych. Podano listę osiągnięć naukowych.

Rozdział 2 przedstawia metody i algorytmy zastosowane w tej pracy. Niewątpliwym zainteresowaniem i naukową nowością jest bisection weighted by element size and order algorithm, który ma na celu skonstruowanie uporządkowanego *element partition trees* dla problemów *hp* adaptacji metody elementów skończonych. Elementarne przykłady pokazują działanie metody polegającej na agregowaniu krok-po-kroku i faktoryzacji w pełni zebranej części macierzy frontalnej, a także poprawieniu niekompletnie zebranej (pozostałej) jej części, która w jednym z następnych kroków metody weźmie udział w agregowaniu kolejnej macierzy frontalnej. W przypadku klasycznych zagadnień elementów skończonych podobne metody zastosowano wcześniej, jednak dla problemów adaptacyjnych, biorąc pod uwagę ich specyfikę, o ile recenzentowi wiadomo, po raz pierwszy. Dla proponowanej metody wielofrontalnej podano oszacowanie operacji zmiennoprzecinkowych na etapie faktoryzacji oraz na etapie podstawień bezpośrednich i wstecznych, a także oszacowanie transferów danych w pamięci RAM. Ponadto dokonuje się oszacowania użycia energii elektrycznej.

Przedstawiono algorytm konstruowania uwarunkowania wstępnego polegający na powrocie do siatki zgrubnej po wykonaniu kilku etapów zagęszczenia siatki w procesie adaptacji. Po pierwsze, w lokalnych subdomenach otaczających punkty zagęszczenia siatki podczas procesu adaptacji, obliczając dopełnienie Schura, eliminują się równania związane z wewnętrznymi węzłami otoczenia punktów syngularności. Incomplete LU PCG solver jest następnie stosowany do pozostałych równań. Takie podejście może być skuteczne w przypadku zadań adaptacyjnych.

W rozdziale 3.1 wykonano porównanie wydajności przy użyciu uporządkowania element partition trees uzyskanych zarówno za pomocą procedur z biblioteki METIS, jak i przy użyciu procedur z bibliotek solwera MUMPS. Porównano 5 metod uporządkowania: bisections-weighted-by-element-size algorithm zastosowany do rekursywnego wywołania procedury METIS\_WPartGraphRecursive z biblioteki METIS, area & neighbors bottom-up algorithm, nested-dissections algorithm zastosowany z METIS\_NodeND, Approximate minimum degree (AMD) algorithm oraz optimal element partition tree found by the expensive dynamic programming procedure. Rozważano dwa zagadnienia. W pierwszym zagęszczanie siatki odbywało się w środku prostokątnego obszaru, a w drugim – na dwóch sąsiednich granicach. Dla każdego z zadań zastosowano elementy skończone o rządzie aproksymacji *p* od 2 do 5. Porównywano czas rozwiązywania problemów przy uporządkowaniu algorytmami area & neighbors bottom-up algorithm oraz bisections-weighted-by-element-size algorithm z algorytmem expensive dynamic programming procedure, przy tym rozważano siatki o różnej gęstości.

Dalej rozważano problem z kwadratową domeną pokrytą początkowo jednolitą kwadratową siatką. Następnie przeprowadzono lokalne zagęszczenie siatki w losowo wybranych elementach skończonych. Dokonano porównania liczby operacji zmiennoprzecinkowych dla proponowanego podejścia przy zastosowaniu procedur uporządkowania ND i AMD z biblioteki MUMPS z zaproponowanym w recenzowanej pracę bisections-weighted-by-element-size przy zagęszczaniu siatki w obszarze oraz przy granicę obszaru. Zgodnie z oczekiwaniami, ze względu na heurystyczny charakter algorytmów uporządkowania, w jednych przypadkach jedne algorytmy prowadziły do mniejszej liczby operacji, a w drugich – inne algorytmy demonstrowali swoją przewagę.

Dla dwóch zagadnień trójwymiarowych porównano liczbę operacji zmiennoprzecinkowych i czas rozwiązania uzyskany w proponowanym podejściu z wielofrontalnym solwerem MUMPS przy użyciu metody uporządkowania bisections weighted by element size, przy czym ukazana jest przekonująca przewaga proponowanego podejścia.

Szczególnie interesującym jest rozdział 3.3, poświęcony wielowątkowemu zrównolegleniu proponowanego solwera wielofrontalnego. Niestety, autor nie podaje żadnych informacji o tym, co dokładnie podlega zrównoleglaniu – drzewo frontalne, obliczenie dopełnienia Schura w macierzy frontalnej lub jedno i drugie (zrównoleglenie dwupoziomowe). Wiadomo tylko, że użyto biblioteki C ++ GALOIS. Podany *speedup* – przyspieszenie ze wzrostem liczby wątków, dla rozważanych problemów dwuwymiarowych mniej więcej odpowiada temu, które wykazuje solwer MUMPS – około 6,5 na 16 wątkach przy braku turbo boost.

Wyniki pokazane na rys. 61, słabo czytelne. Brakuje wyjaśnienia, ile wątków było utrzymane dla solwera MUMPS, brakuje również wyjaśnień, co oznacza w legendzie 1*th*, 2*th*,… .

Speedup pokazany na rys. 63 dla problemu trójwymiarowego z zagęszczaniem siatki w węzłach jest dość niewielki – około 4 na 16 wątkach. Warto by było jakoś to skomentować. Dla wyników pokazanych na ryc. 65 i 67 dla trójwymiarowego ciała z zagęszczaniem siatki odpowiednio do krawędzi i do powierzchni bocznych, przyspieszenie okazało się znacznie lepsze – do 9. Odpowiada to innym przykładam opublikowanym w różnych artykułach, oraz doświadczeniu recenzenta dla solwera MUMPS.

Wyniki pokazane na rys. 69-70, nie pozwalają wyjaśnić trwałość procesu iteracyjnego dla porównywanych solwerów iteracyjnych ILUPCG i zaproponowany w pracy local Schur complement + ILUPCG. Porównanie przez liczbę iteracji nie jest zbyt reprezentatywne, ponieważ w jednym solwerze iteracje mogą być długie, a w drugim – szybkie. Zależy to od ilości elementów niezerowych w operatorze uwarunkowania wstępnego. Oprócz tego warto by było przedstawić czas przygotowania uwarunkowania wstępnego.

4.0 UWAGI KRYTYCZNE

4.1 Od dawna znane są metody wykorzystujące w solwerach frontalnych informacje o topologii modelu obliczeniowego – łączeniu ze sobą elementów skończonych za pomocą węzłów. Jeżeli przypomnieć o pierwszym frontalnym solwerze opracowanym przez Bruce M. Irons – A frontal solution program for finite element analysis, Numerical Methods in Engineering, 1970, v. 2, pp. 5 – 32, to tam cały model obliczeniowy na pierwszym etapie jest przedstawiony w postaci oddzielnych elementów skończonych. Następnie rozpoczyna się proces agregowania. W określonej kolejności do ansamblu elementów skończonych jest dodawany kolejny element skończony, a jeśli na tym etapie są węzły całkowicie zebrane, to wszystkie równania z nimi związane eliminują się. Następnie dodaje się następny element skończony, i tak dalej. Istnieją również inne prace, w których proces faktoryzacji frontalnej opiera się na stworzenie uporządkowanego drzewa frontalnego. Dlatego idea stosowania *element partitioning trees* – szczególnego przypadku drzewa frontalnego, w takiej czy innej formie była używana w poprzednich pracach.

4.2 Porównanie skuteczności proponowanego podejścia z innymi metodami w pracy jest ograniczone jedynie przez porównanie z solwerem MUMPS. Nie ma porównań z bezpośrednimi solwerami supernodalnymi, a także z solwerami iteracyjnymi, które często okazują się znacznie szybsze niż bezpośrednie w przypadku dużych problemów trójwymiarowych.

5.0 PODSUMOWANIE

Główne osiągnięcia naukowe recenzowanej pracy to:

5.1 Opracowanie wielofrontalnego solwera dla rozwiązywania *hp* adaptacyjnych problemów MES, precyzyjnie uwzględniającego specyfikę tej klasy problemów i przeznaczonego do szeroko rozpowszechnionych wielordzeniowych komputerów o architekturze SMP.

5.2 Zrównoleglenie wielowątkowe oparte na bibliotece C ++ GALOIS, pozwoliło uzyskać stabilny speedup wraz ze wzrostem liczby wątków.

5.3 Zaproponowano oryginalne uwarunkowanie wstępne, oparte na częściową eliminacje równań w obszarach lokalnego zagęszczenia siatki, a następnie na zastosowaniu solwera ILUPCG.

Wymienione osiągnięcia świadczą o wysokim poziomie oryginalności oraz stanowią nowość naukową recenzowanej pracy doktorskiej.

Autor ma wystarczający stopień wiedzy teoretycznej z zakresu nauk technicznych w dyscyplinie informatyka techniczna i telekomunikacja, o czym świadczy prawidłowe sformułowanie problemu badawczego i wysoki poziom techniczny realizacji badań. Nie można było by wykonać tej pracy bez dogłębnej znajomości metod numerycznych, podstaw metody elementów skończonych, współczesnych technik programowania wielowątkowego.

Wszystkie badania zostały przeprowadzone na dobrym poziomie naukowym, a wiarygodność ich wyników uzasadniona jest na podstawie porównania solwera autora z wynikami otrzymanymi solwerem MUMPS, korzystającego wysokim zaufaniem w kołach naukowych.

Zgłoszone uwagi nie umniejszają meritum tej pracy. Uważam, że recenzowana praca „Parallel multi-thread multi-frontal solver using element partition trees” spełnia wymagania Ustawy z dnia 14 marca 2003 r. z późniejszymi zmianami "O stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki" i wnoszę o dopuszczenie jej Autora, mgr inż. Konrada Sewiłłó-Jopek do publicznej obrony wyżej wymienionej pracy.