The role of reordering in parallel algorithms for factorization of sparse matrix

Lect.dr. Bogdan OANCEA
Universitatea Artifex, Bucureşti

Sparse matrix computations and especially the solution of linear systems are at the heart of many scientific computing applications. Much could be gained if we were able to accelerate such computations by efficiently using parallel computers. However, solving a linear system with a sparse matrix is a difficult problem and the algorithms used for dense matrices cannot be used anymore. This paper presents a preliminary step for solving a linear system with a sparse matrix by the direct method - the reordering of the matrix. To avoid the fill-in during the matrix factorization, before doing the actual factorization the matrix is reordered using one of the algorithms described in this paper.
***Keywords**: sparse matrices, reordering algorithms.*

Introducere
Rezolvarea sistemelor liniare de ecuații stă la baza multor aplicații numerice din cele mai diverse domenii: dinamica fluidelor, modele economice etc. Problema rezolvării acestor sisteme devine dificilă atunci când matricea sistemului este de mari dimensiuni şi trebuie utilizați algoritmi paraleli.
Una dintre metodele des utilizate în rezolvarea acestor sisteme este metoda directă bazată pe factorizarea matricei sistemului. Metoda directă de rezolvare a sistemelor liniare de tipul $Ax=b$ constă în două etape:
1. se factorizează matricea sistemului $A=LU$ unde L şi U sunt matrice triunghiulare
2. se rezolvă două sisteme cu matrice triunghiulare $Ly=b$ şi $Ux=y$;

În cazul în care matricea sistemului este rară, rezolvarea acestuia este o problemă dificilă pentru că algoritmii utilizați în cazul matricelor dense nu sunt eficienți în această situație. De aceea trebuie proiectați noi algoritmi care să țină seama de caracteristicile speciale ale matricelor rare. În continuare vom prezenta câteva dintre cele mai importante aspecte legate de factorizarea matricelor rare prin algoritmi paraleli.
Există o strânsă legătură între teoria grafurilor şi rezolvarea sistemelor liniare cu matrice rare. Astfel, se poate defini graful de adiacență al unei matrice rare ca fiind un graf $G=(V,E)$ ale cărui n noduri reprezintă cele n necunoscute. Muchiile grafului reprezintă relații binare ce se stabilesc între ecuațiile sistemului astfel: între nodurile i şi j există o muchie atunci când $a_{ij} \neq 0$, adică ecuația i implică necunoscuta j. Acest graf este un graf orientat cu excepția cazului în care matricea este simetrică. În figura 1 sunt reprezentate două matrice împreună cu grafurile asociate lor.

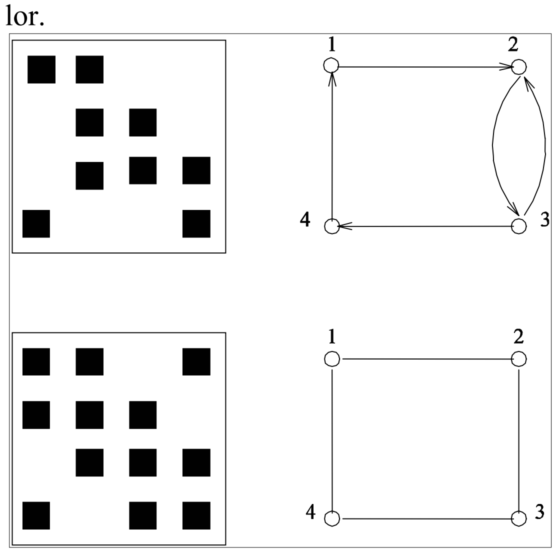

Fig.1. Două matrice rare şi grafurile de adiacență

Cu ajutorul modelului grafurilor putem de exemplu extrage paralelismul din eliminarea gaussiană determinând necunoscutele care sunt independente la un anumit pas al algoritmului. Aceste necunoscute sunt cele care nu depind unele de altele, aşa cum arată relațiile binare stabilite de graful de adiacență al

matricei sistemului. Liniile ce corespund acestor necunoscute pot fi utilizate simultan în procesul de pivotare.
Cele două situaţii extreme sunt reprezentate de cazul în care matricea este diagonală, caz în care toate necunoscutele sunt independente, respectiv cazul matricelor dense când fiecare necunoscută depinde de toate celelalte necunoscute. Dacă matricea este rară, ne situăm undeva între cele două extreme şi putem extrage suficient paralelism.

Reordonarea matricelor rare

Permutările de linii şi/sau coloane reprezintă o operaţie de bază utilizată în implementările paralele ale algoritmilor de factorizare a matricelor. Noţiunea de permutare se referă la tehnica interschimbării de linii ale matricei *A* pentru a evita pivoţii cu valori mici. În mod formal, aceste interschimbări se pot descrie prin *matrice de interschimbare*, adică matrice obţinute din matricea identitate cu liniile rearanjate. Produsul unor astfel de matrice de interschimbare poartă numele de *matrice de permutare*.
O matrice de permutare se memorează de obicei sub forma unui vector *p* cu *n* componente: *p(k)* va memora indexul coloanei ce conţine valoarea 1 în matricea de permutare, pe linia *k*. De exemplu, dacă matricea de permutare este:

$$P = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

atunci ea va fi memorată sub forma p = [4 1 3 2]. Înmulţirea la dreapta cu o matrice de permutare este echivalentă cu o permutare a coloanelor în timp ce înmulţirea la stânga înseamnă o permutare a liniilor.
Este uşor de arătat că atunci când liniile matricei sunt permutate ordinea în care sunt scrise ecuaţiile este schimbată iar atunci când coloanele sunt permutate se realizează o reordonare a necunoscutelor sistemului. În cazul matricelor rare se utilizează de regulă permutările simetrice care au proprietatea importantă că menţin elementele diagonale tot pe diagonală dar în altă ordine. Astfel de permutări simetrice se pot scrie: $A = P^T AP$.
Interpretarea unei permutări simetrice este simplă: matricea rezultată corespunde reordonării necunoscutelor apoi reordonării ecuaţiilor în acelaşi mod. Din punct de vedere al teoriei grafurilor interpretarea unei permutări simetrice este aceea a redenumirii nodurilor grafului asociat matricei fără a modifica muchiile grafului.
Pentru a înţelege importanţa reordonării matricelor rare vom considera un exemplu concret. În figura 2 avem reprezentată o matrice rară precum şi graful de adiacenţă al acesteia. Dacă se aplică procedura de eliminare gaussiană asupra acestei matrice, încă de la primul pas al algoritmului vom obţine un fenomen de *umplere* (fill-in) foarte pronunţat. Practic, factorii *L* şi *U* vor deveni matrice dense. Atunci când linia *k* este pivot, elementul *A(i,j)* care era zero înaintea factorizării devine diferit de zero după factorizare pentru toţi *i,j>k* astfel încât *A(i,k) ≠0* şi *A(k,j)≠ 0*.

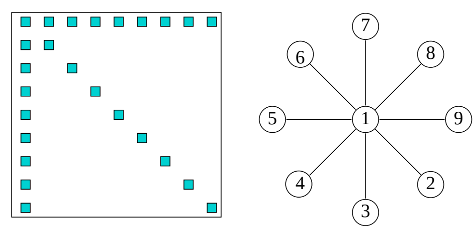

Fig. 2. O matrice rară 9x9 şi graful de adiacenţă

Dacă matricea este reordonată ca în figura 3 se observă că forma grafului de adiacenţă nu s-a schimbat, numai nodurile acestuia au fost redenumite. Avantajul acestei reordonări a matricei este acela că după factorizare vom obţine tot matrice rare (practic factorii *L* şi *U* vor avea aceeaşi structură ca şi *A*), evitându-se astfel fenomenul de umplere. Pe lângă faptul că reordonarea evită fenomenul de umplere, alt avantaj este acela că o reordonare bună poate conduce la creşterea gradului de paralelism - pot exista mai multe linii care vor fi folosite ca pivot simultan. Acest tip de paralelism este disponibil doar datorită faptului că matricea *A* este rară.

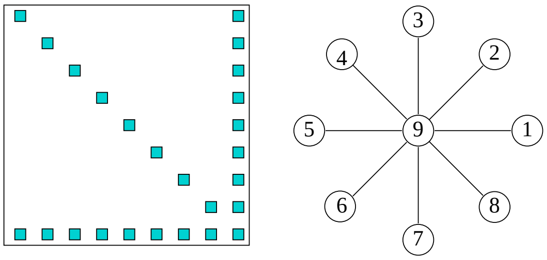

Fig. 3. Matricea rară după reordonare

Primul algoritm de reordonare pe care îl prezentăm este algoritmul Cuthill-McKee. În termeni de teoria grafurilor, algoritmul Cuthill-McKee este numit Breadth First Search. Totuşi există şi o diferenţă: în algoritmul BFS, nodurile unui nivel sunt traversate întotdeauna în ordinea în care ele sunt listate pe când în algoritmul Cuthill-McKee nodurile unui nivel sunt traversate de la nodul cu gradul cel mai mic la cel cu gradul cel mai mare. Algoritmul este prezentat în continuare.

```
LSet = i1;
next = 2;
mark(i1) = 1;
iperm(1) = i1;
WHILE (next < n) DO
      NextLSet = Ø
      Traversează LSet în ordinea crescătoare a gradurilor nodurilor şi
      pentru fiecare nod traverat execută:
                  Pentru fiecare nod i vecin al nodului j astfel ca
                  mark(i) = 0 execută:
                  NextLSet = NextLSet  U {i};
                  mark(i) = 1;
                        iperm(next) = i;
                  next = next + 1;
      LSet = NextLSet
ENDWHILE
```

Fig. 4. Algoritmul de ordonare Cuthill McKee

Vectorul *iperm* memorează nodurile grafului în ordinea în care ele au fost traversate. Astfel, cu ajutorul informaţiilor conţinute în *iperm* se poate determina matricea de permutare. Experimental s-a observat că inversând numărătoarea nodurilor grafului dată de algoritmul Cuthill-McKee se obţine o reordonare foarte bună pentru eliminarea gaussiană.

Un alt algoritm de reordonare a matricelor rare este algoritmul ordonării după gradul minim - *Minimum Degree Ordering*. Să presupunem că am parcurs primii *k* paşi ai algoritmului de eliminare gaussiană. Vom nota cu c_i numărul de elemente diferite de zero în coloana *i* a părţii active a matricei $(n-k)\times(n-k)$ şi cu l_i numărul de elemente nenule de pe linia *i*. Se defineşte funcţia de cost: $C(i) = (ci - 1) * (li - 1)$. Algoritmul gradului minim alege următorul pivot la pasul *k* acela pentru care *C(i)* este minim şi *A(i,i)* îndeplineşte anumite condiţii privitoare la stabilitatea numerică.

O altă metodă de reordonare prezentată aici poartă numele de *bisecţie imbricată* (*nested bissection ordering*) şi se bazează pe partiţionarea unui graf. Vom exemplifica această metodă pentru graful de adiacenţă din figura 5.

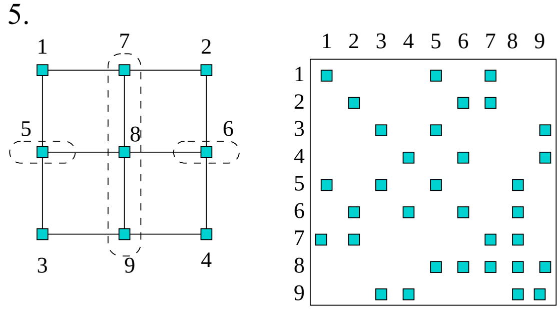

Fig. 5. Ordonarea unei matrice rare după metoda bisecţiei imbricate

Iniţial se va alege un set de noduri care după îndepărtare să împartă graful în două componente neconectate între ele de dimensiune (ca număr de noduri) aproximativ egală. Setul de noduri ales va fi numerotat după ce toate nodurile din cele două subdomenii au fost numerotate. Nodurile din aceste două subdomenii sunt numerotate recursiv folosind aceeaşi metodă. După ce toate nodurile grafului au fost numerotate matricea ordonată se obține punând în corespondență nodul *i* din graf cu ecuația *i*.

Concluzii

Reordonarea matricelor este o etapă obligatorie în rezolvarea sistemelor liniare de mari dimensiuni cu matrice rare prin metode directe pentru a evita fenomenul de umplere. În prezent există mai multe strategii de reordonare a matricelor rare în vederea factorizării. Trei dintre metodele de reordonare au fost prezentate în acest articol. Totuşi, nu există deocamdată nici un criteriu teoretic de a alege una dintre aceste metode. Experimental se constată că utilizarea unor algoritmi de reordonare conduce la reducerea gradului de umplere după factorizare dar se reduce în acelaşi timp şi gradul de paralelism posibil de exploatat în procesul de factorizare. În practică se va alege întotdeauna acel algoritm de reordonare care realizează un compromis între gradul de paralelism şi factorul de umplere.

Bibliografie

1. I.S. Duff, Henk A. van der Vorst. "Developments and trends in parallel solution of linear systems" *Technical Report, RAL-TR-1999-027,* RutherFord Appleton Laboratory, 1999.
2. V. Kumar, A. Grama, A. Gupta, and G. Karypis. "Introduction to Parallel Computing". *The Benjamin/Cummings Publishing Company,* 1994.
3. V. Kumar and A. Gupta. "Analyzing scalability of parallel algorithms and architectures". *Proceedings of the 1991 International Conference on Supercomputing*, June 1991.
4. J.W.H. Liu. "The role of elimination tree in sparse matrix factorization". *SIAM Journal on Matrix Analysis and Applications*, (11:134-172), 1990.
5. B. Oancea. "Metode numerice de calcul paralel pentru modelele economice din economie". *Teză de Doctorat,* Decembrie, 2002, ASE, Bucureşti.

J. A. Scott, "A new ordering strategy for frontal solvers". *Technical report RAL-TR-1998-056,* Rutherford Appleton Laboratory, 1998.