Calcolo Parallelo e Distribuito

Prodotto Matrice-Vettore
Strategia 3

caso: non esatta divisibilità
isoefficienza

Docente: Prof. L. Marcellino
Tutor: Prof. P. De Luca

PROBLEMA: Prodotto Matrice-Vettore

Progettazione
di un algoritmo parallelo
per architettura MIMD

per il calcolo del prodotto
di una matrice A pr un vettore b:

matrice A: N righe, M colonne
Vettore b: M elementi

III STRATEGIA

Decomposizione 1: BLOCCHI di RIGHE

+

Decomposizione 2: BLOCCHI di COLONNE

=

Decomposizione 3:

BLOCCHI Righe&Colonne

Abbiamo calcolato…

Calcolo **di speedup ed efficienza**

(**def classica**)

… nel caso dell’esatta divisibilità!

Speed-up/efficienza (**def classica**)

matrice A: N righe, M colonne
Vettore b: M elementi

GRIGLIA qxp

Attenzione:

- è possibile riconoscere la I strategia dalla III quando $p=1$
- è possibile riconoscere la II strategia dalla III quando $q=1$
- cosa succede se **mod(N,q) ≠0** e/o **mod(M,p)≠0** ?

III Strategia: speed-up/efficienza (**def classica**)

matrice A: N righe, M colonne
Vettore b: M elementi

GRIGLIA qxp

- **Attenzione:**

  cosa succede se **mod(N,q) ≠0** e/o **mod(M,p)≠0** ?

Alcuni core si occuperanno di blocchi di matrice (e se serve del vettore) di dimensione maggiore

Es: <u>RIGHE</u> **mod(N,q)≠0**

Il numero di righe che avanza
(cioè il resto della divisione)
viene ridistribuito a tutti i core riga che hanno prima coordinata strettamente minore del resto

III Strategia: speed-up/efficienza (**def classica**)

matrice A: N righe, M colonne
Vettore b: M elementi

GRIGLIA qxp

- **Attenzione:**

  cosa succede se **mod(N,q) ≠0** e/o **mod(M,p)≠0** ?

Alcuni core si occuperanno di blocchi di matrice (e se serve del vettore) di dimensione maggiore

Es: RIGHE **mod(N,q)≠0**

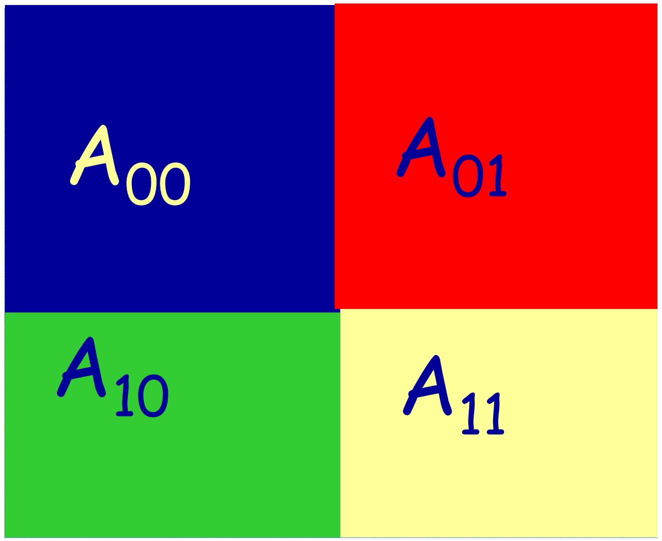
I core che hanno coordinata riga strettamente minore del resto
hanno una riga in più della matrice.
Nessuna variazione invece per il blocco relativo al vettore!

III Strategia: speed-up/efficienza (**def classica**)

matrice A: N righe, M colonne
Vettore b: M elementi

Es: se **mod(N,q) ≠0**

A



GRIGLIA
qxp=4=2x2

$\dim[b_{loc}]= M/p$

$\dim[A_{10}]=(N/q\ )x(M/p)$

$\dim[A_{11}]=(N/q)x(M/p)$

$\dim[A_{00}]=(N/q + 1)x(M/p)$

$\dim[A_{01}]=(N/q + 1)x(M/p)$

III Strategia: speed-up/efficienza (**def classica**)

matrice A: N righe, M colonne
Vettore b: M elementi

Es: se **mod(N,q) ≠0**

GRIGLIA
qxp=4=2x2

$\dim[A_{00}]=(N/q + 1)\times(M/p)$

$\dim[A_{01}]=(N/q + 1)\times(M/p)$

$\dim[b_{loc}]= M/p$

$\dim[A_{10}]=(N/q)\times(M/p)$

$\dim[A_{11}]=(N/q)\times(M/p)$

P_{00} P_{01}

P_{10} P_{11}

P_{00}: $\dim[r_{00}]=(N/q + 1)$, P_{01}: $\dim[s_{01}]=(N/q + 1)$

P_{10}: $\dim[r_{10}]=(N/q)$, P_{11}: $\dim[s_{11}]=(N/q)$

III Strategia: speed-up/efficienza (**def classica**)

matrice A: N righe, M colonne
Vettore b: M elementi

GRIGLIA qxp

I strategia per collezione vettori

$$S_{qxp}(NxM) = T_1(NxM)/T_{qxp}(NxM) =$$
$$= N[2M-1] \,/\, (\boxed{[N/q+1]}\,[\,2M/p-1\,] + \boxed{(N/q+1)}\,(p-1))$$

II strategia per collezione vettori

$$S_{qxp}(NxM) = T_1(NxM)/T_{qxp}(NxM) =$$

II strategia per collezione vettori

$$= N[2M-1] \,/\, (\boxed{[N/q+1]}\,[\,2M/p-1\,] + \boxed{(N/q+1)}\,\log_2(p))$$

III Strategia: speed-up/efficienza (**def classica**)

matrice A: N righe, M colonne
Vettore b: M elementi

GRIGLIA qxp

- **Attenzione:**
  cosa succede se **mod(N,q) ≠0** e/o **mod(M,p)≠0** ?

Alcuni core si occuperanno di blocchi di matrice (e se serve del vettore) di dimensione maggiore

Es: <u>COLONNE</u> **mod(M,p)≠0**

Il numero di colonne che avanza
(cioè il resto della divisione)
viene ridistribuito a tutti i core colonna che hanno seconda coordinata strettamente minore del resto

III Strategia: speed-up/efficienza (**def classica**)

matrice A: N righe, M colonne
Vettore b: M elementi

GRIGLIA qxp

- **Attenzione:**

  cosa succede se **mod(N,q) ≠0** e/o **mod(M,p)≠0** ?

Alcuni core si occuparanno di blocchi di matrice (e se serve del vettore) di dimensione maggiore

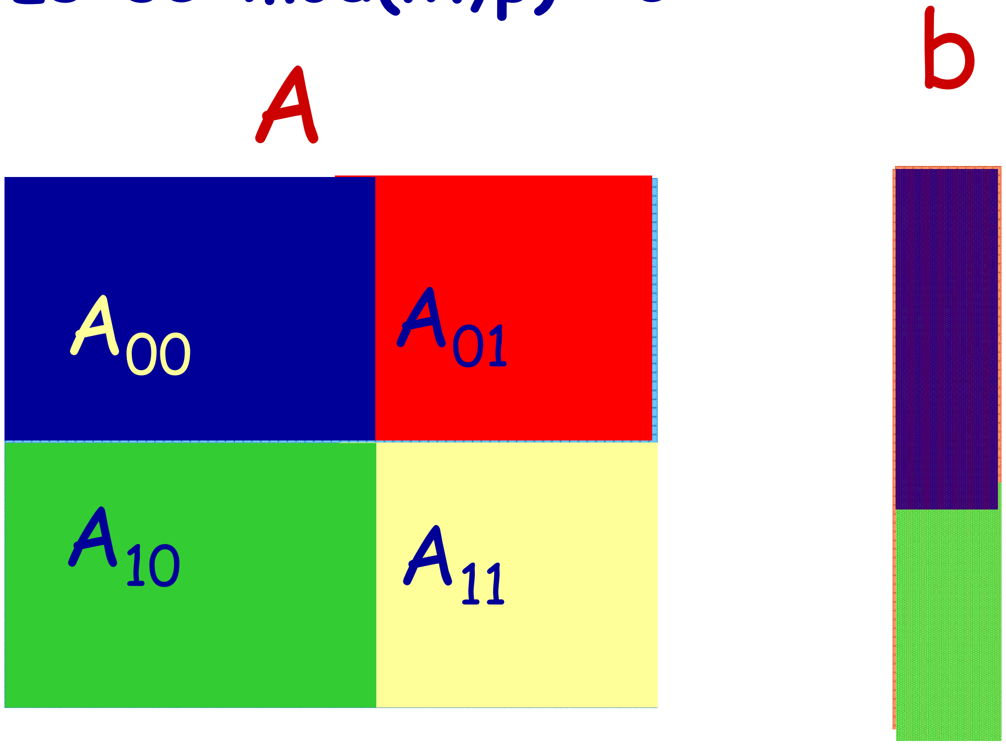
Es: COLONNE **mod(M,p)≠0**

I core che hanno coordinata colonna strettamente minore del resto hanno una colonna in più della matrice e anche un elemento in piu del vettore!

III Strategia: speed-up/efficienza (**def classica**)

matrice A: N righe, M colonne
Vettore b: M elementi

Es: se **mod(M,p) ≠0**



GRIGLIA
qxp=4=2x2

$dim[A_{01}]=(N/q)x(M/p)$
$dim[A_{11}]=(N/q)x(M/p)$
$dim[b_1]= M/p$

$dim[A_{00}]=(N/q\ )x(M/p+1)$

$dim[b_0]= M/p+1$

$dim[A_{10}]=(N/q\ )x(M/p+1)$

III Strategia: speed-up/efficienza (**def classica**)

matrice A: N righe, M colonne
Vettore b: M elementi

Es: se **mod(M,p) ≠0**

GRIGLIA
qxp=4=2x2

$dim[A_{00}]=(N/q)x(M/p+1)$

$dim[b_0]= M/p+1$

$dim[A_{10}]=(N/q)x(M/p+1)$

$dim[A_{01}]=(N/q)x(M/p)$
$dim[A_{11}]=(N/q)x(M/p)$
$dim[b_1]= M/p$

C_{00} C_{01}

C_{10} C_{11}

P_{00}: $dim[r_{00}]=N/q$, P_{01}: $dim[s_{01}]=N/q$
P_{10}: $dim[r_{10}]=N/q$, P_{11}: $dim[s_{11}]=N/q$

III Strategia: speed-up/efficienza (**def classica**)

matrice A: N righe, M colonne
Vettore b: M elementi

GRIGLIA qxp

I strategia per collezione vettori

$$S_{qxp}(NxM) = T_1(NxM)/T_{qxp}(NxM) =$$

$$= N[2M-1] \,/(N/q\, \boxed{[\,(2M/p+1)-1\,]} + N/q\,(p-1))$$

II strategia per collezione vettori

$$S_{qxp}(NxM) = T_1(NxM)/T_{qxp}(NxM) =$$

$$= N[2M-1] \,/(N/q\, \boxed{[\,(2M/p+1)-1\,]} + N/q\,\log_2(p))$$

III Strategia: speed-up/efficienza (**def classica**)

matrice A: N righe, M colonne
Vettore b: M elementi

GRIGLIA qxp

- **Attenzione:**

  cosa succede se **mod(N,q) ≠0** e/o **mod(M,p)≠0** ?

Alcuni core si occuperanno di blocchi di matrice (e se serve del vettore) di dimensione maggiore

Es: <u>RIGHE</u> **mod(N,q)≠0** ***e*** <u>COLONNE</u> **mod(M,p)≠0**

Il numero delle righe e delle colonne che avanzano
(cioè il resto della divisione)
verranno ridistribuiti a tutti i core riga e colonna che hanno prima
e **seconda coordinata strettamente minore del resto**

III Strategia: speed-up/efficienza (**def classica**)

matrice A: N righe, M colonne
Vettore b: M elementi

GRIGLIA qxp

- **Attenzione:**

  cosa succede se **mod(N,q) ≠0** e/o **mod(M,p)≠0** ?

Alcuni core si occuperanno di blocchi di matrice (e se serve del vettore) di dimensione maggiore

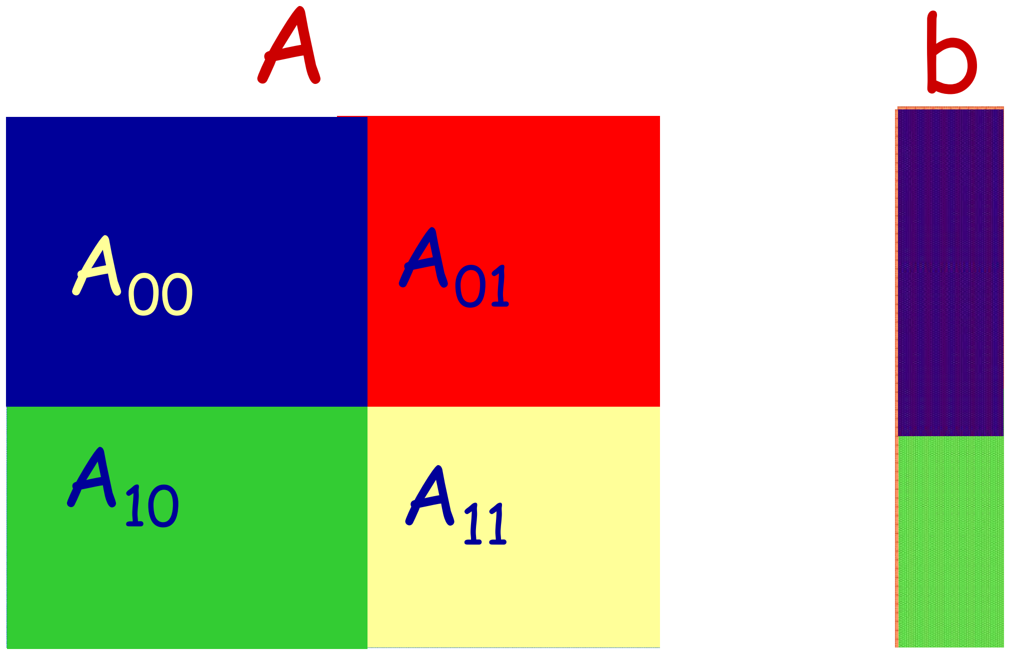
Es: RIGHE **mod(N,q)≠0** **e** COLONNE **mod(M,p)≠0**

I core che hanno coordinata riga <u>e</u> colonna strettamente minore del resto hanno <u>una</u> riga e <u>una</u> colonna in più della matrice e anche <u>un</u> elemento in piu del vettore!

III Strategia: speed-up/efficienza (**def classica**)

matrice A: N righe, M colonne
Vettore b: M elementi

Es: se **mod(N,q) ≠0, mod(M,p) ≠0**



GRIGLIA
qxp=4=2x2

$dim[A_{11}]=(N/q)x(M/p)$
$dim[b_1]= M/p$

$dim[A_{00}]=(N/q +1)x(M/p+1)$

$dim[b_0]= M/p+1$

$dim[A_{10}]=(N/q)x(M/p+1)$, $dim[A_{01}]=(N/q+1)x(M/p)$

III Strategia: speed-up/efficienza (**def classica**)

matrice A: N righe, M colonne
Vettore b: M elementi

GRIGLIA
$q \times p = 4 = 2 \times 2$

Es: se **mod(N,q) ≠0, mod(M,p) ≠0**

$dim[A_{11}]=(N/q)\times(M/p)$
$dim[b_1]= M/p$

$dim[A_{00}]=(N/q+1)\times(M/p+1)$

$dim[b_0]= M/p+1$

$dim[A_{10}]=(N/q)\times(M/p+1)$, $dim[A_{01}]=(N/q+1)\times(M/p)$

C_{00}	C_{01}
C_{10}	C_{11}

P_{00}: $dim[r_{00}]=N/q+1$, P_{01}: $dim[s_{01}]=N/q+1$

P_{10}: $dim[r_{10}]=N/q$, P_{11}: $dim[s_{11}]=N/q$

III Strategia: speed-up/efficienza (**def classica**)

matrice A: N righe, M colonne
Vettore b: M elementi

GRIGLIA qxp

I strategia per collezione vettori

$$S_{qxp}(N \times M) = T_1(N \times M)/T_{qxp}(N \times M) =$$
$$= N[2M-1] / (\boxed{[N/q+1]\ [\ 2M/p+1-1\ ]} + \boxed{(N/q+1)}\ (p-1))$$

II strategia per collezione vettori

$$S_{qxp}(N \times M) = T_1(N \times M)/T_{qxp}(N \times M) =$$
$$= N[2M-1] / (\boxed{[N/q+1]\ [\ 2M/p+1-1\ ]} + \boxed{(N/q+1)}\ \log_2(p))$$

Speed-up/efficienza (**def classica**)

matrice A: N righe, M colonne
Vettore b: M elementi

GRIGLIA qxp

Attenzione:

- è possibile riconoscere la I strategia dalla III quando p=1
- è possibile riconoscere la II strategia dalla III quando q=1

Tutti i conti fatti per la III strategia nel caso in cui N non sia esattamente divisibile per q e/o M non sia esattamente divisibile per p, possono essere proiettati per la I strategia (p=1) e per la II strategia (q=1) provate a farvi i conti da soli!

III Strategia: **isoefficienza**

matrice A: N righe, M colonne
Vettore b: M elementi

Per il calcolo dell'isoefficienza è necessario separare i conti tra la I e la II strategia impiegata per la collezione dei risultati!!!

I strategia per collezione vettori

$$Oh = qxp\ (N/q\ [\ 2M/p-1\ ] + N/q\ (p-1)) - N[2M-1]=$$

$$= 2\ NM - pN + p^2\ N - p\ N - 2NM + N =$$

$$= p^2\ N - 2pN + N$$

È uguale a quello della II strategia (con I strategia per collezione vettori): overhead dipende dal numero delle righe e dal numero dei core lungo le colonne!

III Strategia: isoefficienza

matrice A: N righe, M colonne
Vettore b: M elementi

uguale a quello della II strategia (con I strategia per collezione vettori)

I strategia per collezione vettori

$I(q_0\, p_0,\ q_1\, p_1,\ N_0\, M_0) =$

$C = [N_1\,(-2p_1 + p_1^2 + 1)]/[N_0\,(-2p_0 + p_0^2 + 1)]$

$\cancel{N_1}\, M_1 = C\, N_0\, M_0 =$

$= \cancel{N_0}\, M_0\, [\cancel{N_1}\,(-2p_1 + p_1^2 + 1)]/[\cancel{N_0}\,(-2p_0 + p_0^2 + 1)]$

Nel calcolo delle nuove dimensioni N ed M, devo fissare il numero di righe e calcolare le colonne

III Strategia: **isoefficienza**

matrice A: N righe, M colonne
Vettore b: M elementi

Rifare i conti per la seconda strategia...

II strategia per collezione vettori

$$Oh = q x p\,(N/q\,[\,2M/p-1\,] + N/q\,\log_2(p)) - N[2M-1] =$$

$$= 2N\,M - pN + p\,N\,\log_2(p) - 2NM + N =$$

$$= p\,(N\,\log_2(p) - N + 1)$$

Anche in questo caso uguale alla II strategia (con II strategia per collezione vettori): Overhead dipende dal numero delle righe e dei core definiti lungo le righe della griglia virtuale!

II Strategia: isoefficienza

matrice A: N righe, M colonne
Vettore b: M elementi

uguale a quello della II strategia (con II strategia per collezione vettori)

II strategia per collezione vettori

$$I(q_0\, p_0,\ q_1\, p_1,\ N_0\, M_0) = C =$$

$$[N_1(-p_1+p_1 \log_2(p_1)+1)]/[N_0(-p_0+p_0\log_2(p_0)+1)]$$

$$\cancel{N_1}\, M_1 = C\, N_0\, M_0 =$$

$$= \cancel{N_0}\, M_0\, [\cancel{N_1}(-p_1+p_1 \log_2(p_1)+1)]/[\cancel{N_0}(-p_0+p_0\log_2(p_0)+1)]$$

Stesse osservazioni di prima, nel calcolo delle nuove dimensioni, posso fissare un qualunque numero di righe e calcolare il più opportuno numero di colonne

III Strategia: **isoefficienza**

I conti devono essere rifatti nel caso in cui
mod(N,q) ≠0 e/o **mod(M,p)≠0**

ma la scalabilità è compromessa

Domanda facile d'esame

Se ho una matrice A, di dimensione N=32, M=28; ed un vettore b di dimensione M=28 qual è la migliore strategia parallela, in termini di speedup (nella definizione classica), se ho 8 core?

Risposta di chi ha capito tutto?

La I... sempre (senza bisogno di fare conti)

Ed esclusa la I? Cioè tra la II e la III?

Conti, calcolatrice alla mano...