

# Metodologia PCAM - AB1

Bruno Baldissera Carlotto	10724351
Bruno Gazoni	7585037
Gabriel Eluan Calado	10734453
João Villaça	10724239
Matheus Steigenberg Populim	10734710

## Particionamento

O problema de soma do produto de duas matrizes em uma terceira com a abordagem a seguir tem complexidade cúbica, visto que precisa efetuar duas multiplicações de matriz independentes (complexidade cúbica) e somar as matrizes resultantes (complexidade quadrada), tendo uma complexidade final  $O(2n^3 + n^2)$ . Considerando que, para a multiplicação de matrizes, cada par composto por uma linha  $A_i$  da matriz A e uma coluna  $B_j$  da matriz B resulta em uma posição  $R_{ij}$  na matriz resultado, e cada posição desta é computacionalmente independente, as tarefas devem particionar estas computações para maior eficiência, implicando paralelizar multiplicações de cada linha  $A_i$  por coluna  $B_j$  correspondente dos pares de matrizes de entrada e escrita de seu resultado na posição adequada  $R_{ij}$  de cada matriz de resultado. Além disso as computações  $R_1 = AB$  e  $R_2 = CD$  são independentes entre si também e serão paralelizadas. Por fim, tendo  $R_1$  e  $R_2$  somamos estas, sendo essa soma dependente das multiplicações anteriores e seus dados resultantes. O particionamento dessa soma é efetuado a nível de posições, sendo independente cada posição  $R_{ij}$  da matriz resultante  $R_3$  (sendo gerada por  $R_{1ij} + R_{2ij}$ ). As matrizes de entrada e matriz de saída não terão suas operações de E/S paralelizadas. Como os dados de dimensão das matrizes serão informados via entrada padrão, serão utilizados como critérios de parada para os laços que iteram pelas matrizes, sendo assim prevista de antemão sua quantidade.

## Comunicação

Deve se estabelecer comunicação entre as 4 matrizes de entrada, que são 2 pares de matrizes (A, B e C, D) para calcular os produtos. Os resultados  $R_1 = AB$  e  $R_2 = CD$  destes pares serão somadas para resultar na matriz final. A dependência de dados nas multiplicações ocorre a nível de pares linha  $i$  e coluna  $j$  das matrizes de entrada e sua posição correspondente  $ij$  da matriz resultante, sendo assim necessária sincronização dos resultados apenas associados aos índices  $ij$  e não entre todos os resultados. Em um escopo maior, a computação da soma  $R_3 = R_1 + R_2$  aguarda a computação das multiplicações descritas anteriormente.

## Aglomeração

Tomando como base o número de processos disponíveis como  $T$ , os cálculos são feitos em duas instâncias paralelas para multiplicar os dois pares de matrizes, sendo que, dentro dessas instâncias, cada processo será responsável por operar multiplicando com respeito a uma linha e uma coluna específicas, paralelizando os dados que não possuem dependências entre si, como discutido na seção de comunicação. Após todo esse processo,

serão geradas duas matrizes com os produtos, que serão somadas, e assim teremos a matriz final. Através da diretiva de paralelização de loops, haverá  $T/2$  processos disponíveis para cada instância, que atenderão ao maior número possível de linhas e colunas paralelamente, de acordo com a carga de trabalho representada pela dimensão das matrizes de entrada. A complexidade resultante, tendo como base  $O(n)$  para a versão sequencial, será assim  $O(n/T + c)$ , sendo  $c$  a constante relativa a operações de E/S não paralelizadas. Para otimizar o uso de memória, o resultado da soma das matrizes  $R1$  e  $R2$  será guardado na própria matriz  $R1$  ao invés de ser instanciada uma nova matriz  $R3$ .

## **Mapeamento**

A versão paralela deve considerar processadores equipotentes, visto que as operações desenvolvidas por cada processador pelo programa são as mesmas sobre um conjunto de dados paralelizados. A aplicação das regiões paralelas será tal que um compilador que não consiga associar as diretivas necessárias para paralelização resulte ainda assim em uma versão sequencial e plenamente funcional do código. Não será considerada, entretanto, portabilidade a nível de código para processadores heterogêneos, que poderão ser empregados para operações de E/S que não foram paralelizadas mas que necessitarão de alterações específicas a nível de código para tanto.