

Programmation Parallèle

TP - Introduction à OpenMP

Oguz Kaya
oguz.kaya@universite-paris-saclay.fr

Pour compiler un code `programme.cpp` avec OpenMP et générer l'exécutable `programme`, saisir

```
g++ -O2 -std=c++11 -fopenmp programme.cpp -o programme
```

Part 1

Hello World

Ex. 1

- Écrire un programme qui ouvre une région parallèle dans laquelle chaque thread imprime son identifiant.
- En suite, dans la même region parallèle, afficher "Hello World" par un seul thread.
- Essayer de faire varier le nombre de threads par les trois manières que l'on a rencontré dans le cours (la variable d'environnement `OMP_NUM_THREADS`, la fonction `omp_set_num_threads()` et la clause `num_threads()`). Quel est l'ordre de préséance entr'eux?

Part 2

La somme d'un tableau d'entiers

Ex. 2

- Écrire un programme C/C++ qui initialise un tableau `A[N]` de flottants tel que `A[i] = i` pour $0 \leq i < N$.
- Faire une deuxième boucle qui calcul la somme des éléments de `A[N]`.
- Paralléliser les deux boucles avec `#pragma omp for`.
- Maintenant, paralléliser la deuxième boucle avec `#pragma omp sections` ayant 4 sections tel que chaque section parcourt $N/4$ itérations de la boucle. Veillez à ce qu'il n'y ait pas de concurrence parmi les threads en effectuant la réduction des sommes partielles.
- Comparer le temps d'exécution séquentielle et le temps d'exécution parallèle du programme.

Part 3

Mergesort en parallèle avec OpenMP sections

Le but de cet exercice est de trier un tableau d'entiers en parallèle à l'aide de l'OpenMP sections. Le fichier `mergesort.cpp` fournit déjà un code squelette qui alloue et initialise le tableau `A[N]` à trier ainsi que `temp[N]` que l'on utilise comme "buffer" dans l'algorithme de mergesort.

Ex. 3

- Créer une région parallèle avec 4 sections (ou 8, s'il y a au moins 8 cœurs dans la machine, à vérifier avec la commande `lscpu`) qui trie chacune $N/4$ éléments consécutifs du tableau `A[N]` avec `std::sort`.
- Dans la même région parallèle, après avoir terminé ses 4 sections, créer 2 sections chacune fusionnant deux tableaux triés de taille $N/4$ en un seul tableau trié de taille $N/2$, sur le buffer `temp[N]`. Pour ce faire, profiter de la fonction `merge` fournit dans le squelette.
- Finalement, en dehors de la région parallèle, fusionner les deux tableaux triés de taille $N/2$ en un seul tableau de taille N , soit le tableau `A[N]` trié.

- d) Comparer le temps d'exécution séquentielle (avec 1 seul thread) et le temps d'exécution parallèle du programme. Quelle est l'accélération obtenue? Quelle est l'efficacité?

Part 4

Le calcul de π

Le nombre π peut être défini comme l'intégrale de 0 à 1 de $f(x) = \frac{4}{1+x^2}$. Une manière simple d'approximer une intégrale est de discrétiser l'ensemble d'étude de la fonction en utilisant N points.

On considère l'approximation suivante avec $s = \frac{1}{N}$:

$$\pi \approx \int_0^1 \frac{4}{1+x^2} dx \approx \sum_{i=0}^{N-1} s \times \frac{f(i \times s) + f((i+1) \times s)}{2}$$

On écrira un programme qui parallélise une approximation de la valeur de π en utilisant OpenMP. On effectuera deux façons différentes de parallélisation. Un code squelette `calcul-pi.cpp` est fourni pour travailler dessus.

Ex. 4

- D'abord, écrire le code séquentiel qui calcule correctement la valeur π .
- On peut alors répartir le calcul de π parmi P threads. Au premier, on parallélisera tout simplement la boucle principale à l'aide de `#pragma omp for`. Tester la performance en utilisant de différents nombres de threads.
- Pour la deuxième stratégie "faite à la main", chaque thread parcourra N/P indices consécutifs dans la région parallèle pour calculer son `piLocal`. C'est à dire, on ne va pas utiliser `#pragma omp for`; on va plutôt répartir les itérations parmi les threads nous-même. Une fois que c'est calculé, les threads additionneront les uns après les autres leur `piLocal` dans `pi` qui soit une variable partagée par tous les threads (il faudrait utiliser une région `critical` ou `atomic` afin d'éviter le problème d'écriture concurrente en cet étape). Tester la performance en utilisant de différents nombres de threads.

Part 5

La conjecture de Goldbach

Le but de cet exercice est de démontrer l'intérêt de l'utilisation de la clause `schedule` de l'OpenMP afin d'améliorer les performances d'un code parallélisé. Un code squelette `goldbach.cpp` est fourni pour travailler dessus.

Dans le monde de la théorie des nombres, d'après la conjecture de Goldbach: chaque nombre pair supérieur à deux est la somme de deux nombres premiers.

Dans cet exercice, vous êtes fournis un code séquentiel qui teste cette conjecture. Le code permet de trouver le nombre de paires de Goldbach pour un nombre pair donné i (c'est à dire le nombre de paires de nombres premiers P_1, P_2 tels que $P_1 + P_2 = i$) pour $i = 4, \dots, 8192$. Le coût de calcul est proportionnel à i^2 dans l'itération i , donc pour obtenir une performance optimale avec plusieurs threads la charge de travail doit être distribuée en utilisant intelligemment la clause `schedule` de l'OpenMP.

On vous demande par la suite de:

Ex. 5

- Paralléliser le code séquentiel en utilisant `omp for` sans préciser d'ordonnancement. Quelle est la stratégie d'ordonnancement de l'OpenMP par défaut (et la taille de bloc impliquée par ceci)?
- Réaliser la parallélisation avec l'ordonnancement `static` et la taille de bloc 256.
- Réaliser la parallélisation avec l'ordonnancement `dynamic`.
- Réaliser la parallélisation avec l'ordonnancement `dynamic` et la taille de bloc 256.
- Réaliser la parallélisation avec l'ordonnancement `guided`.
- Réaliser la parallélisation avec l'ordonnancement `guided` et la taille de bloc 256.
- Dans quelles stratégies avez-vous obtenu la meilleure performance? A-t-il du sens compte tenu de la complexité du calcul?

Part 6

Produit matrice-vecteur en OpenMP

Le but de cet exercice est d'écrire un programme qui calcule le produit d'une matrice A de taille $N \times N$ et d'un vecteur x de taille N :

$$b = Ax.$$

Utiliser le code squelette fourni dans le fichier `matvec.cpp` qui alloue les vecteurs x , b et la matrice A orientée par les lignes (c'est à dire, les premières N éléments correspondent à la première ligne $A(0, :)$, puis la deuxième ligne $A(1, :)$, etc.). La matrice A et le vecteur x sont initialisés par le code squelette.

Ex. 6

- Coder la version séquentielle dans le champs indiqué du code, puis mesurer le temps d'exécution.
- Implanter la version parallélisée avec `omp for` et mesurer le temps d'exécution.
- Réaliser une autre version basée sur OpenMP Tasks tel que chaque produit scalaire $b(i) = A(i, :)x(:)$ est effectué par une tâche.
- Afficher l'accélération et l'efficacité.
- Tester la performance pour toutes les tailles de dimension entre $2^0, \dots, 2^{12}$. A partir de quelle taille constatez-vous un gain de performance? Modifier la version parallèle tel qu'elle performe mieux pour les petites tailles de dimension. Pour ce faire, utiliser la clause d'OpenMP correspondante ou faire un branchement explicite (qui exécute la version séquentielle ou parallèle en fonction de la taille de dimension).

Part 7

Calcul de Fibonacci avec OpenMP Tasks

Dans cet exercice, nous allons paralléliser le calcul du nombre Fibonacci[N] à l'aide de l'OpenMP Tasks.

Ex. 7

- Réaliser une version récursive qui parallélise ce calcul avec `#pragma omp task` et `#pragma omp taskwait`
- Réaliser une version itérative qui calcule tous les Fibonacci[i] dans un tableau de taille N et parallélise ce calcul avec `#pragma omp task` et la clause `depend(in:...)` et `depend(out:...)`

Part 8

Mergesort en parallèle avec OpenMP Tasks

Paralléliser le code séquentiel donné pour le mergesort en utilisant OpenMP Tasks. Limiter la taille minimum de tâches créées à une constante K (e.g., 100000) tel que la récursion ne génère plus de tâches à partir de cette taille-là. Expérimenter avec ce paramètre pour trouver la meilleur accélération.

Part 9

Prefixe-somme en 2D avec OpenMP Tasks

Nous voulons calculer la prefixe-somme d'un tableau 2D $A[N][N]$ tel que $B[x][y] = \sum_{(i,j)=(0,0)}^{(x,y)} A[i][j]$. Une manière efficace de ce faire est à travers la programmation dynamique: pour chaque $B[x][y]$, on peut d'abord calculer $B[x][y-1]$, $B[x-1][y]$, et $B[x-1][y-1]$, ensuite calculer $B[x][y]$ comme la suite:

$$B[x][y] = B[x][y-1] + B[x-1][y] - B[x-1][y-1] + A[x][y]$$

pour tous x, y . En effet, cet équation implique certaines dépendences qu'il faudra respecter dans le calcul (e.g., $B[x][y-1]$, $B[x-1][y]$, $B[x-1][y-1]$ doivent être calculés avant $B[x][y]$).

Ex. 8

- a) Planter une version séquentielle qui tout d'abord initialise le tableau $A[x][y] = x + y$, puis calcule B en utilisant A avec deux boucles imbriquées sur x et y . Pourrait-on paralléliser ce code avec `omp for`? Si oui, combien de parallélisation (le nombre de threads actifs au maximum) pourrait-on avoir? Si no, pourquoi?
- b) Next, parallelize it using OpenMP Tasks this time, by creating N^2 tasks (one task for computing each $B[x][y]$), and adding necessary input/output dependencies. Measure the parallel execution time. Is it fast? Why/why not?
- c) Ensuite, paralléliser le calcul en utilisant OpenMP Tasks, en créant N^2 tâches (une tâche pour chaque calcul de $B[x][y]$) et en ajoutant les dépendances nécessaires entre les tâches. L'exécution, est-elle plus rapide? Pourquoi/pourquoi pas?
- d) Enfin, implanter une version avec tuilage, avec une taille bloc K . Dans ce cas, chaque tâche calculera un bloc de $K \times K$ éléments contigus de B (e.g., $B[0 \dots K-1][0 \dots K-1]$). On faudra $(N/K) \times (N/K)$ tâches et 4 boucles imbriquées dans cette version (les deux premières itèrent sur les indices de bloc, les deux dernières itèrent à l'intérieur d'un bloc, ce qui constitue une tâche). Pour préciser les dépendences, vous pouvez utiliser un tableau temporaire de taille $(N/K) \times (N/K)$.