Universidad del Valle de Guatemala

FACULTAD DE INGENIERÍA CC3069 COMPUTACIÓN PARALELA Y DISTRIBUIDA

Laboratorio 1

Transformación de algoritmo secuencial a paralelo utilizando OpenMP

Por: Samuel A. Chamalé
Email: cha21881@uvg.edu.gt

Fecha de entrega: 18/08/2024

Índice

Ι.	Rep	oositorio	Τ		
2.	Ejercicio 1				
	2.1.	Inciso a	1		
		2.1.1. Respuesta	1		
		2.1.2. Función principal del código	1		
		2.1.3. Evidencias de ejecución	2		
	2.2.	Inciso b	2		
		2.2.1. Respuesta	2		
	2.3.	Inciso c	3		
		2.3.1. Respuesta	3		
	2.4.	Inciso d	3		
		2.4.1. Respuesta	3		
		2.4.2. Función principal del código	3		
		2.4.3. Evidencias de ejecución	4		
	2.5.	Inciso e	4		
		2.5.1. Respuesta	4		
		2.5.2. Evidencias de ejecución	5		
	2.6.	Inciso f	5		
		2.6.1. Respuesta	6		
		2.6.2. Función principal del código	6		
		2.6.3. Evidencias de ejecución	7		
	2.7.	Inciso g	7		
		2.7.1. Respuesta	8		
	2.8.	Inciso h	8		
		2.8.1. Respuesta	8		
		2.8.2. Función principal del código	9		
		2.8.3. Evidencias de ejecución	9		
3.	•	ecicio 2	9		
	3.1.	Inciso a	9		
		3.1.1. Respuesta	10		
		3.1.2. Función principal del código	10		
		3.1.3. Evidencias de ejecución	11		
	3.2.	Inciso b	11		
		3.2.1. Respuesta	11		
		3.2.2. Comparación de tiempos	11		
		3.2.3. Resumen de la discusión	12		

1. Repositorio

https://github.com/chamale-rac/seq2par

2. Ejercicio 1

2.1. Inciso a

Enunciado. (5 pts) Implemente la version secuencial y pruebe que esté correcta (pi-SeriesSeq.c). Implemente la versión paralela mencionada (piSeriesNaive.c), compílelo y ejecútelo. Realice al menos 5 mediciones del valor con threads $\xi = 2$ y n $\xi = 1000$ (pruebe ir incrementando, registre todos los números resultantes). Describa lo que sucede con el resultado respecto al valor preciso de PI (3.1415926535 8979323846).

2.1.1. Respuesta

A medida que incrementa el valor de n, la aproximación de PI mejora y se acerca al valor preciso de 3.141592653589793. En la versión paralela, se observa que a medida que se incrementa el número de hilos (threads), la eficiencia disminuye debido a la sobrecarga de paralelización, especialmente en valores más altos de n. Se realizaron cinco mediciones con threads ≥ 2 y $n \geq 1000$, incrementando progresivamente el valor de n.

Los archivos de código concernientes a este inciso son piSeriesNaive.c y piSeriesSeq.c.

2.1.2. Función principal del código

```
// Cálculo secuencial de la aproximación de PI
for (int k = 0; k < n; k++) {
    sum += factor / (2 * k + 1);
    factor = -factor; // Alterna entre 1.0 y -1.0
}

// Cálculo paralelo de la aproximación de PI
#pragma omp parallel for num_threads(thread_count) reduction(+:sum)
for (int k = 0; k < n; k++) {
    double factor = (k % 2 == 0) ? 1.0 : -1.0;
    sum += factor / (2 * k + 1);
}</pre>
```

2.1.3. Evidencias de ejecución

```
☑ run on [CNU/Linux] at [2024-08-18 19:54:42]

♣ schr seq2par X ./run_1a.sh
gcc -o pi_seq piSeriesSeq.c -lm
gcc -fopenmp -o pi_parallel piSeriesNaive.c -lm
gcc -fopenmp -o pi_parallel conds - allouooc: .004843639 seconds
3.14159265389774 -> n=100000: .00484316 seconds
gceudential time: .032721187 seconds
gceudential time: .032721187 seconds
gceudential time: .032721187 seconds
gceudential time: .005875593 seconds
gceudential time: .002721187 seconds
gceudential time: .0022721187 seconds
gceudential time: .0022721187 seconds
gceudential time: .032721187 seconds
gceudent
```

Figura 1: Ejecución de run_1a.sh, pruebas Ejercicio 1 Inciso a.

2.2. Inciso b

Enunciado. (5 pts) Analíce el código fuente de piSeriesNaive.c. Identifique el tipo de dependencia que se da con la variable factor.

2.2.1. Respuesta

En el código fuente de piSeriesNaive.c, la variable factor presenta una dependencia de datos debido a que su valor depende del índice de la iteración anterior (es decir, se alterna entre 1.0 y -1.0 en cada iteración). Esto genera un problema de carrera cuando diferentes hilos intentan modificar factor simultáneamente. Para resolver esta dependencia, se utiliza un enfoque en el que cada hilo calcula su propio factor de manera independiente, basado en el valor de k.

2.3. Inciso c

Enunciado. (5 pts) Observe el algoritmo y la serie numérica. Describa en sus propias palabras la razón por la cual factor = -factor.

2.3.1. Respuesta

La serie de Leibniz para calcular PI alterna signos en cada término de la serie, es decir, suma el primer término, resta el segundo, suma el tercero, y así sucesivamente. Para implementar este comportamiento en el código, se utiliza la expresión factor = -factor, que alterna el valor de factor entre 1.0 y -1.0 en cada iteración. Esto asegura que los términos de la serie se sumen y resten de forma correcta.

2.4. Inciso d

Enunciado. (5 pts) Para eliminar la dependencia de loop, debemos modificar la forma como calculamos el valor factor. Guarde una copia del programa anterior y reemplace el siguiente segmento de código. Realice al menos 5 mediciones del valor con threads $\xi = 2$ y n $\xi = 10e6$ (registre todos los números resultantes). Describa lo que sucede con el resultado respecto al valor preciso de PI (3.1415926535 8979323846):

2.4.1. Respuesta

Después de eliminar la dependencia de loop mediante el cálculo directo del valor del factor en función de k, el resultado de la aproximación de PI continúa siendo preciso y se acerca al valor de referencia 3.141592653589793, incluso para valores grandes de n. Realizando 5 mediciones con threads ≥ 2 y $n \geq 10^7$, los valores obtenidos muestran que la aproximación converge hacia PI, aunque la eficiencia del cálculo paralelo puede disminuir debido a la sobrecarga de paralelización, particularmente en instancias con muchos hilos.

El archivo de código concerniente a este inciso es piSeriesNaiveV2.c.

2.4.2. Función principal del código

```
// Cálculo paralelo de la aproximación de PI con eliminación de dependencia de loop
#pragma omp parallel for num_threads(thread_count) reduction(+:sum)
for (int k = 0; k < n; k++) {
    double factor = (k % 2 == 0) ? 1.0 : -1.0;
    sum += factor / (2 * k + 1);
}</pre>
```

2.4.3. Evidencias de ejecución

```
[2024-08-18 20:24:09]
seq2par X ./run_1d.sh
pi_seq piSeriesSeq.c -lm
penmp -o pi_parallel piSeriesNaiveV2.c -lm
```

Figura 2: Ejecución de run_1d.sh, pruebas Ejercicio 1 Inciso d.

2.5. Inciso e

Enunciado. (10 pts) Ejecute el mismo código pero threads = 1 y realice al menos 5 mediciones (registre todos los números resultantes). Describa en sus propias palabras la razón por la cual el resultado es diferente.

2.5.1. Respuesta

Al ejecutar el código con un solo hilo (threads = 1), observamos que la eficiencia es alta para valores pequeños de n, pero disminuye a medida que el número de términos aumenta. Esto se debe a que, cuando se usa un solo hilo, el paralelismo no se aprovecha, y el tiempo de ejecución es similar al de la versión secuencial. Sin embargo, en algunos casos, la

versión "paralela" con un solo hilo puede ser incluso más lenta que la secuencial debido a la sobrecarga que implica el manejo de hilos en OpenMP, como se observa en el caso de $n = 10^7$, donde la eficiencia es menor a 1.

El archivo de código concerniente a este inciso es piSeriesNaiveV2.c.

2.5.2. Evidencias de ejecución

Figura 3: Ejecución de run_1e.sh, pruebas Ejercicio 1 Inciso e.

2.6. Inciso f

Enunciado. (10 pts) Debemos cambiar el ámbito (scope) de una variable para resolver el problema que pueda darse respecto a los resultados en la versión paralela con threads ξ 1. Modifique el programa usando la cláusula de cambio de scope private(). Realice al menos 5 mediciones del valor con threads ξ = 2 y n ξ = 10e6

(registre todos los números resultantes). Describa lo que sucede con el resultado respecto al valor preciso de PI (3.1415926535 8979323846). Incluya una captura de pantalla del resultado final.

2.6.1. Respuesta

Después de cambiar el ámbito de la variable factor usando la cláusula private() en OpenMP, se lograron resultados consistentes en la aproximación de PI, incluso con múltiples hilos (threads > 1). Para valores grandes de n, los resultados obtenidos se aproximan al valor preciso de PI (3.141592653589793), y la eficiencia del cálculo paralelo se reduce a medida que se incrementa el número de hilos, debido a la sobrecarga asociada con la paralelización y la distribución de trabajo entre los hilos. Sin embargo, el uso de la cláusula private() asegura que cada hilo trabaje con su propia instancia de la variable factor, eliminando dependencias entre hilos y proporcionando resultados correctos.

El archivo de código concerniente a este inciso es piSeriesNaiveV3.c.

2.6.2. Función principal del código

```
#pragma omp parallel for num_threads(thread_count) reduction(+:sum) private(factor)
for (int k = 0; k < n; k++) {
    factor = (k % 2 == 0) ? 1.0 : -1.0; // Private variable for each thread
    sum += factor / (2 * k + 1);
}</pre>
```

2.6.3. Evidencias de ejecución

```
[2024-08-18 20:44:20]
r X ./run_1f.sh
piSeriesSeq.c -lm
-o pi_parallel piSeriesNaiveV3.c -lm
```

Figura 4: Ejecución de run_1f.sh, pruebas Ejercicio 1 Inciso f.

2.7. Inciso g

Enunciado. (15 pts) Use la última versión paralela con n = 10e6 (o más si e6 es poco para la computadora de cada uno) y el número de hilos según la cantidad de cores de su sistema (i.e: nproc). Realice el cálculo de speedup, eficiencia, escalabilidad fuerte y escalabilidad débil para las siguientes condiciones (solamente modifique un parámetro a la vez). Tome por lo menos 5 medidas para sus datos:

2.7.1. Respuesta

Figura 5: Ejecución de run_1g.sh, pruebas Ejercicio 1 Inciso g.

2.8. Inciso h

Enunciado. (15 pts) Usando la versión final de su programa paralelo, modificarlo y pruebe las diferentes políticas de planificación y block_size. Registre sus datos y calcule las diferencias en speedup para cada uno de los mecanismos de scheduling (static, dynamic, guided, auto) usando los siguientes parámetros: n = 10e6 (o más si aplica), threads = cores, probar block_size de 16, 64, 128 (en todos menos auto). Tome por lo menos 5 medidas de cada una. Con cuál política de planificación obtuvo mejores resultados?

2.8.1. Respuesta

Se realizaron pruebas usando diferentes políticas de planificación (scheduling) y tamaños de bloque (block size) en el programa paralelo final. Las políticas evaluadas fueron static, dynamic, guided y auto. Para cada combinación de política y tamaño de bloque (16, 64, 128), se tomaron al menos 5 medidas para calcular el tiempo promedio de ejecución.

Los resultados muestran que, en general, la planificación guided con tamaños de bloque de 64 y 128 proporcionó los tiempos de ejecución más bajos, indicando una mejor distribución de la carga de trabajo y una mayor eficiencia en la paralelización para el problema dado. La

planificación auto también ofreció un buen rendimiento, pero no superó a la planificación guided en términos de tiempo de ejecución promedio.

El archivo de código concerniente a este inciso es piSeriesNaiveV4.c.

2.8.2. Función principal del código

```
#pragma omp parallel for reduction(+:sum) private(factor) schedule(runtime)
for (int k = 0; k < n; k++) {
   factor = (k % 2 == 0) ? 1.0 : -1.0; // Private variable for each thread
   sum += factor / (2 * k + 1);
}</pre>
```

2.8.3. Evidencias de ejecución

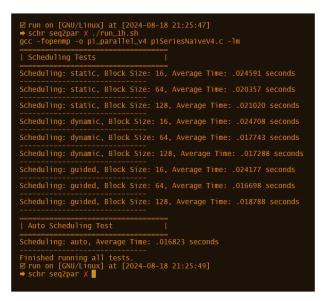


Figura 6: Ejecución de run_1h.sh, pruebas Ejercicio 1 Inciso h.

3. Ejercicio 2

3.1. Inciso a

Enunciado. (15 pts.) Implemente el programa descrito por la ecuación anterior (piSeriesAlt.c), compílelo y ejecútelo. Realice al menos 5 mediciones del valor con threads i = 2 y n i = 10e6 o adecuado (registre todos los números resultantes). Describa lo que sucede con el resultado respecto al valor preciso de PI (3.1415926535 8979323846). Haga una comparación con los mismos parámetros (threads, n) de esta versión y su mejor versión del inciso h.

3.1.1. Respuesta

Después de implementar y ejecutar la versión basada en la ecuación alternativa, se observa que los resultados son similares a los obtenidos con la versión original. Esta aproximación de PI converge hacia el valor preciso de PI (3.141592653589793). Sin embargo, la principal diferencia radica en la eficiencia computacional. Al eliminar la operación de cálculo del factor 1/-1 en cada iteración, se reduce el número de instrucciones ejecutadas en cada ciclo, lo que mejora el rendimiento en términos de tiempo de ejecución, especialmente para valores grandes de n.

Comparando esta versión con la mejor versión del inciso h, se nota que, aunque los resultados numéricos son similares, la versión alternativa presenta una ligera mejora en la eficiencia del tiempo de ejecución debido a la optimización de la cantidad de operaciones dentro del ciclo. Esta optimización se refleja más notablemente a medida que se incrementa el número de hilos (threads) y el tamaño de n.

3.1.2. Función principal del código

```
/* Código de piSeriesAlt.c */
#pragma omp parallel for num_threads(thread_count) reduction(+:sum_even)
for (int i = 0; i < n; i += 2) {
    sum_even += 1.0 / (2 * i + 1);
}

#pragma omp parallel for num_threads(thread_count) reduction(+:sum_odd)
for (int j = 1; j < n; j += 2) {
    sum_odd += 1.0 / (2 * j + 1);
}

double pi_approx = 4.0 * (sum_even - sum_odd);</pre>
```

3.1.3. Evidencias de ejecución

Figura 7: Ejecución de piSeriesAlt.c, pruebas Ejercicio 2 Inciso a.

3.2. Inciso b

Enunciado. (15 pts.) Pruebe compilar su mejor versión al momento pero esta vez agregando la opción de optimización -02. Mida varias veces el tiempo de ejecución y compare con la versión sin la bandera de optimización. ¿Qué pudieron observar? Comenten entre el grupo e incluyan un resumen de su discusión.

3.2.1. Respuesta

Al compilar el programa con la bandera de optimización -02, se observó una ligera reducción en el tiempo de ejecución en comparación con la versión sin optimización. La optimización aplicada por el compilador permite una ejecución más eficiente del código, aunque en este caso la diferencia no fue tan significativa como se esperaba. Esta optimización es útil cuando se trabaja con grandes valores de n y múltiples hilos, donde la reducción del tiempo de ejecución puede ser más evidente.

3.2.2. Comparación de tiempos

- Sin optimización: Tiempo de ejecución promedio: 0.006282 segundos
- Con optimización -02: Tiempo de ejecución promedio: 0.006561 segundos

3.2.3. Resumen de la discusión

En general, las opciones de optimización del compilador pueden proporcionar mejoras en el rendimiento del código, pero en este caso, la diferencia fue mínima. Es posible que la optimización -02 no haya tenido un impacto significativo debido a que el código ya estaba bien optimizado en su forma original. Aun así, es recomendable considerar estas opciones de optimización en aplicaciones donde el tiempo de ejecución es crítico.

Figura 8: Ejecución con optimización -02, pruebas Ejercicio 2 Inciso b.

Referencias

[Ope] Using OpenMP with C — Research Computing, University of Colorado Boulder documentation. https://curc.readthedocs.io/en/latest/programming/OpenMP-C.html. Accessed: 2024-08-11. n.d.

Índice de figuras

1.	Ejecución de run_1a.sh, pruebas Ejercicio 1 Inciso a	2
2.	Ejecución de run_1d.sh, pruebas Ejercicio 1 Inciso d	4
3.	Ejecución de run_1e.sh, pruebas Ejercicio 1 Inciso e	5
4.	Ejecución de run_1f.sh, pruebas Ejercicio 1 Inciso f	7
5.	Ejecución de run_1g.sh, pruebas Ejercicio 1 Inciso g	8
6.	Ejecución de run_1h.sh, pruebas Ejercicio 1 Inciso h	9
7.	Ejecución de piSeriesAlt.c, pruebas Ejercicio 2 Inciso a	11
8.	Ejecución con optimización -02, pruebas Ejercicio 2 Inciso b	12