****

**Shakime Richards Sparks**

**2018170667**

****



OpenMP es una interfaz de programación que permite programar en paralelo en entornos de memoria compartida, como las CPU multinúcleo. Proporciona un conjunto de directivas de compilador y una API para programas C, C++ y FORTRAN, que permite a los desarrolladores especificar regiones de código que pueden ejecutarse en paralelo. Al utilizar OpenMP, los desarrolladores pueden aprovechar la potencia de procesamiento disponible en los sistemas multinúcleo, lo que se traduce en una ejecución más rápida y eficiente de sus programas. Las directivas se utilizan para indicar a la biblioteca de ejecución OpenMP cómo paralelizar el código dentro de la región paralela designada [1].

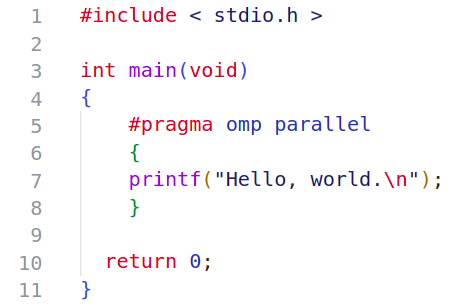


Según [1], cuando OpenMP encuentra una directiva de región paralela, crea tantos hilos como núcleos de procesamiento haya en el sistema. Por ejemplo, si el sistema tiene dos núcleos de procesamiento, OpenMP crea dos subprocesos para ejecutar la región paralela. Del mismo modo, si el sistema tiene cuatro núcleos de procesamiento, OpenMP crea cuatro subprocesos para ejecutar la región paralela.

Este funcionamiento se denomina***número de hilos por defecto*** y viene determinado por el número de núcleos de procesamiento disponibles en el sistema. Ahora bien, los programadores también pueden especificar manualmente el número de subprocesos que se crearán utilizando la API OpenMP. Pueden hacerlo estableciendo la cláusula ***num\_threads*** en la directiva de región paralela.



Cuando OpenMP encuentra la directiva ***omp parallel*** la identifica como una región paralela.



El hilo original se denominará hilo maestro con thread-id 0. Cuando el hilo maestro llega a esta línea, bifurca hilos adicionales para realizar el trabajo incluido en el bloque siguiendo el comportamiento del #pragma construct. El bloque es ejecutado por todos los hilos en paralelo [2].

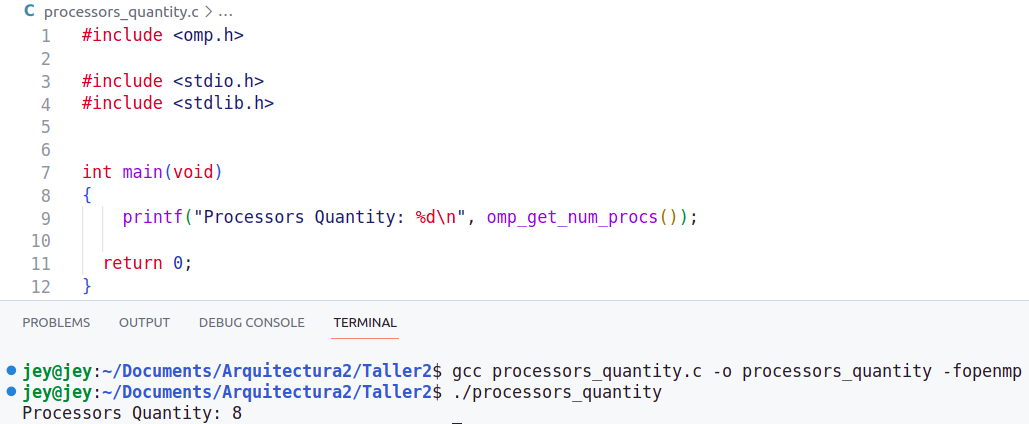


Las funciones OpenMP se incluyen en un header file llamado ***omp.h***. Sistemas linux como dover y foxcroft tienen gcc/g++ instalado con soporte OpenMP. Todo lo que de debe hacer es utilizar la bandera ***-fopenmp*** en la línea de comandos [2]:





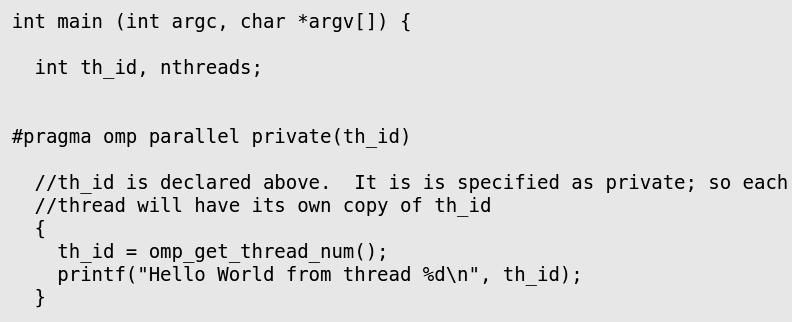
Según [2], utiliza la función omp\_get\_num\_procs().





Estas variables son privadas para cada subproceso, lo que significa que cada subproceso tendrá su propia copia local, no se inicializan y su valor no se mantiene para su uso fuera de la región paralela. Los contadores de iteración de bucle en las construcciones de bucle OpenMP son privados por defecto [2].

Se definen con **#pragma omp parallel private(variable):**



Las variables privadas son importantes en OpenMP porque aseguran el comportamiento correcto en un programa paralelo al permitir que cada subproceso tenga su propia copia independiente de una variable, y también brindan un mejor rendimiento al reducir la necesidad de que los subprocesos accedan a las variables compartidas.



Una variable compartida es visible y accesible por todos los subprocesos simultáneamente. Por defecto, todas las variables de la región de trabajo compartido son compartidas excepto el contador de iteración del bucle. Pero se pueden declarar explícitamente como **#pragma omp parallel shared(variable)**

Las variables compartidas deben utilizarse con cuidado, ya que pueden provocar condiciones de carrera. Por lo que, OpenMP proporciona varias formas de actualizar variables compartidas de manera segura, dentro de estas están: reducción, bloqueo, atomic, entre otras.



La directiva **#pragma omp flush** se utiliza para garantizar la consistencia entre la memoria temporal de un hilo y la memoria principal del sistema, o entre la vista de memoria de varios hilos [5]. Cuando se utilizan múltiples hilos en un programa OpenMP, cada hilo tiene su propia memoria caché que puede contener valores de las variables compartidas que son diferentes a los de otras cachés. Esto puede causar inconsistencias en el valor de las variables compartidas y errores en el programa. Para resolver este problema, se emplea esta directiva para asegurar que los valores de las variables compartidas sean actualizados y sincronizados entre los hilos antes de continuar con la ejecución.

Dependiendo del tipo de flush, la operación puede ser más o menos fuerte en la sincronización de la memoria. Una "strong flush" garantiza la consistencia entre la memoria temporal de un hilo y la memoria principal del sistema, mientras que una "weak flush" solo garantiza la consistencia entre la vista de memoria de los hilos.

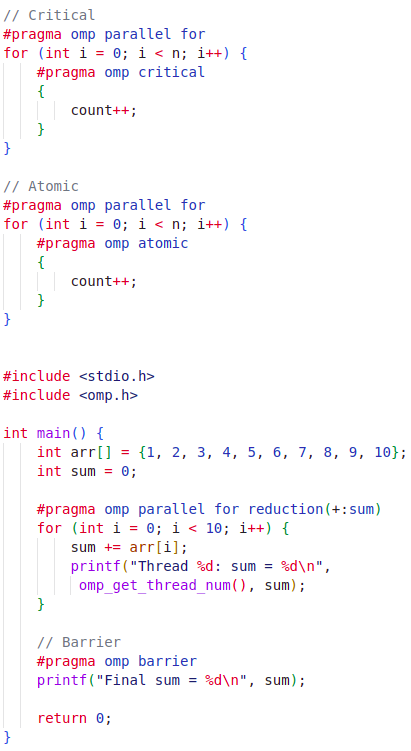
**NOTA:** *Considerar que la directiva #pragma omp flush no garantiza que todas las operaciones anteriores a la directiva estén completas antes de continuar con la ejecución. Por lo tanto, es recomendable utilizar otras directivas de sincronización, como #pragma omp barrier o #pragma omp critical, en combinación con #pragma omp flush para garantizar una sincronización completa y efectiva entre los hilos.*



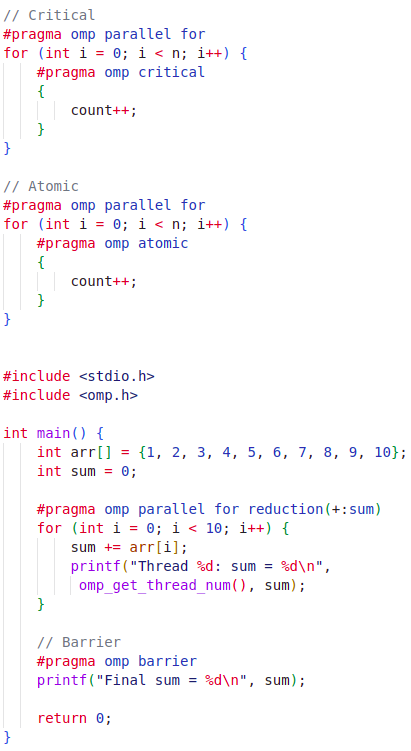
La directiva #**pragma omp single** es utilizada para indicar que un bloque de código debe ser ejecutado por un solo hilo dentro del equipo paralelo [2]. El propósito de esta directiva es garantizar que el bloque de código sea ejecutado por un solo hilo, lo que puede ser útil en situaciones donde solo una tarea debe ser realizada una vez, como por ejemplo inicializar una variable, imprimir un mensaje, leer un archivo.



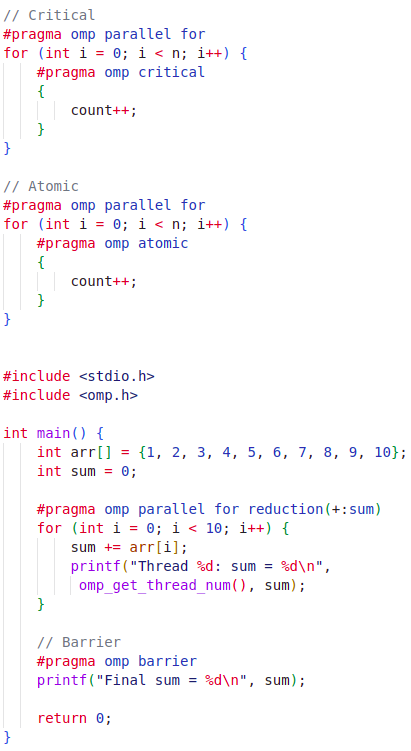
**Bloqueo (Locking):** OpenMP proporciona una cláusula ***critical*** que permite que una variable compartida sea actualizada de forma segura mediante el bloqueo de acceso a la variable mientras un hilo la actualiza. Esto asegura que ningún otro hilo pueda acceder a la variable al mismo tiempo. Suele utilizarse para proteger datos compartidos de condiciones de carrera [2]. Por ejemplo, para actualizar una variable compartida **count** de forma segura en un bucle utilizando un bloqueo, se debe hacer lo siguiente:



**Atomic:** OpenMP proporciona una cláusula ***atomic*** que permite que una variable compartida sea actualizada de forma segura mediante una operación atómica (como incremento, decremento o intercambio) realizada por un solo hilo a la vez. No hace que toda la sentencia sea atómica; sólo la actualización de memoria es atómica. Un compilador puede utilizar instrucciones de hardware especiales para obtener un mejor rendimiento que cuando se utiliza critical [2]. Por ejemplo, el mismo ejemplo anterior utilizando atomic, se debe hacer lo siguiente:



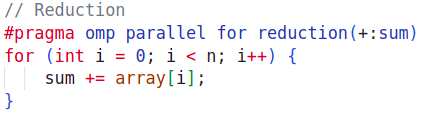
**Barrier:** OpenMP proporciona una cláusula ***barrier*** que se utiliza para sincronizar los hilos en un equipo, lo que significa que todos los hilos deben alcanzar el mismo punto en el código antes de que se pueda continuar con la ejecución. Esto puede ser útil en situaciones en las que se necesita asegurar que ciertas tareas en paralelo se completen antes de continuar con el resto del programa. De hecho, toda estructura de trabajo compartido tiene una barrera de sincronización implícita al final [2]. Por ejemplo, para calcular la suma de los elementos en un arreglo en paralelo:

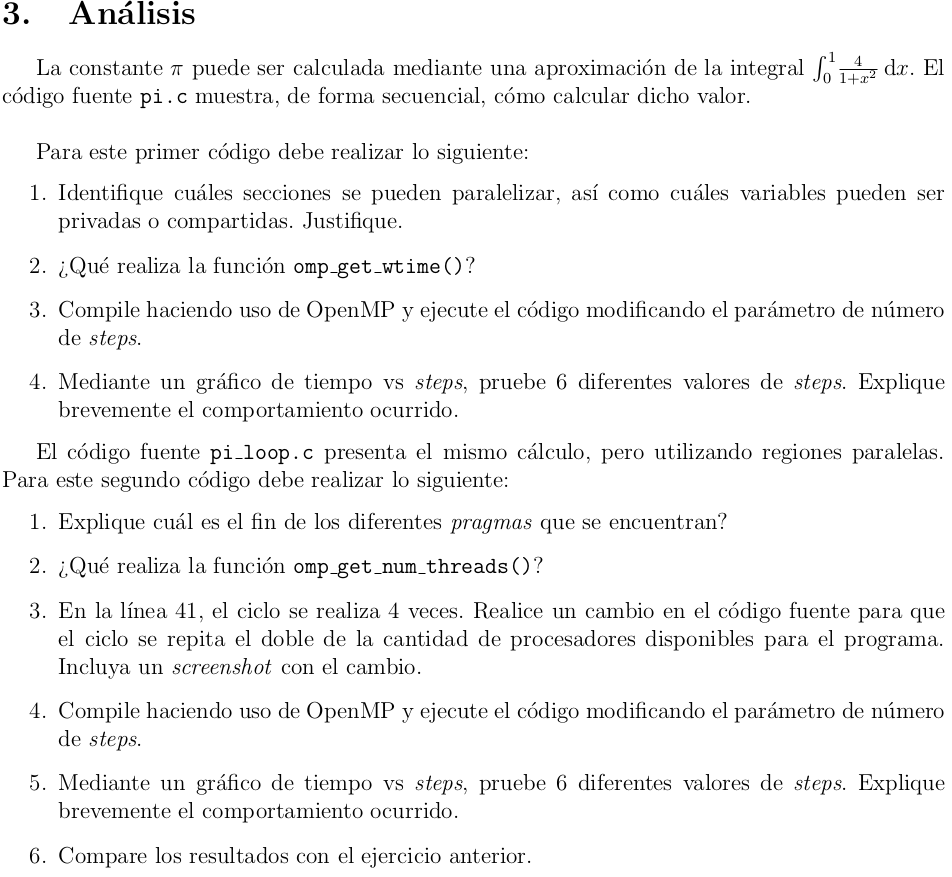


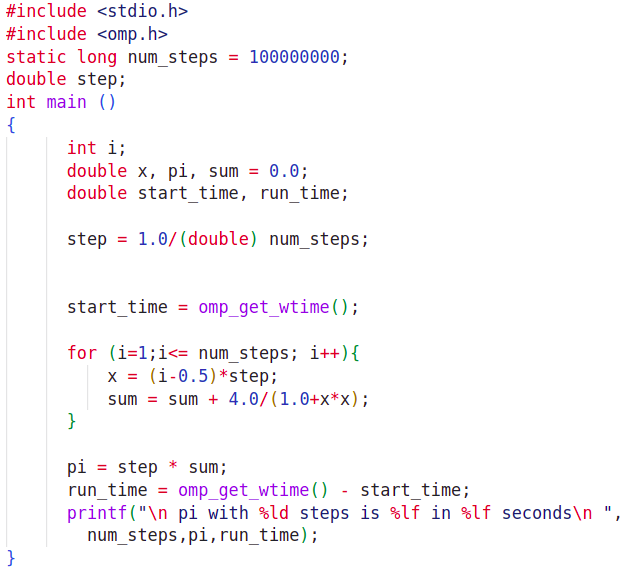


OpenMP proporciona una cláusula de reducción permite que una variable compartida sea actualizada de forma segura mediante una operación de reducción específica (como suma, multiplicación, mínimo o máximo) realizada por todos los hilos en paralelo. El resultado final se almacena en una variable privada [4].

Se define con **#pragma omp reduction (operador : sum)**.Por ejemplo, la siguiente directiva de OpenMP actualiza una variable compartida **sum** de forma segura mediante una operación de reducción de suma.







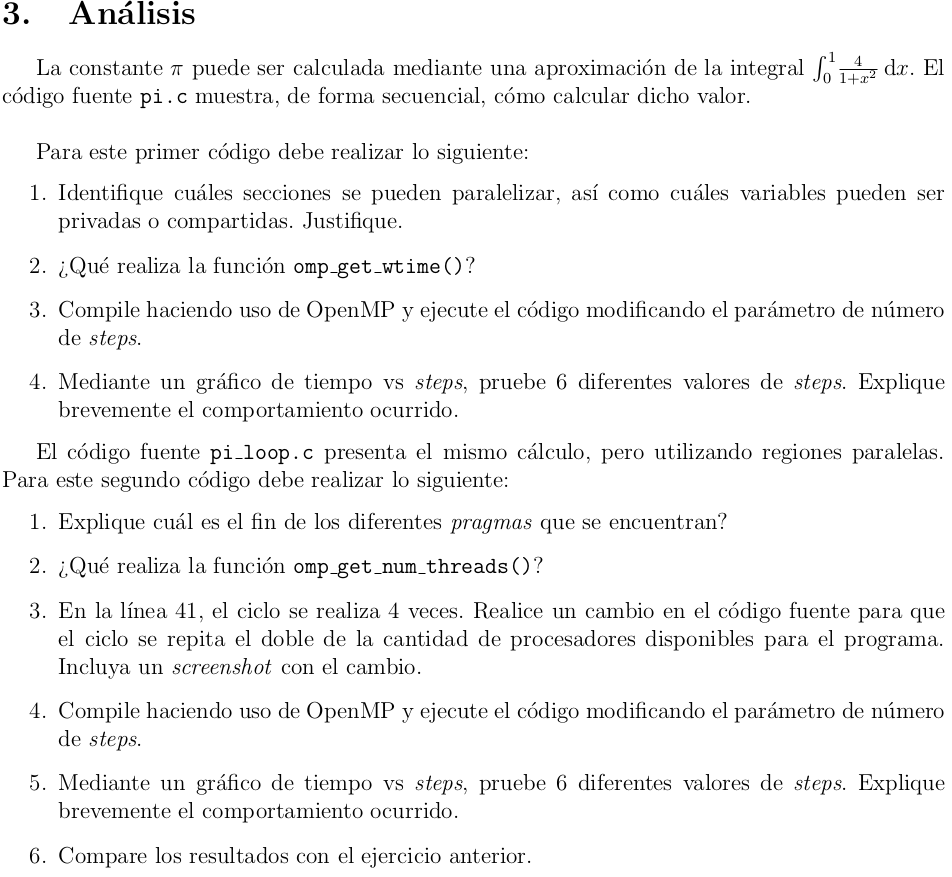
En el código dado, la única sección que se puede paralelizar es el bucle for, dado que las otras secciones del código, utilizan solamente el resultado final del bucle.

Privadas:

* run\_time y start\_time: Son variables privadas al estar declaradas dentro del ámbito de la función main(). Cuando se ejecuta el programa, cada hilo tendrá su propia instancia de estas variables y las actualizará de manera independiente.
* i: La variable i es una variable de control del bucle y su valor se utiliza para determinar el índice del array, debe almacenarse como una variable privada.
* pi: La variable pi es una variable de local y no se actualiza en el bucle, por lo que puede almacenarse como una variable privada.

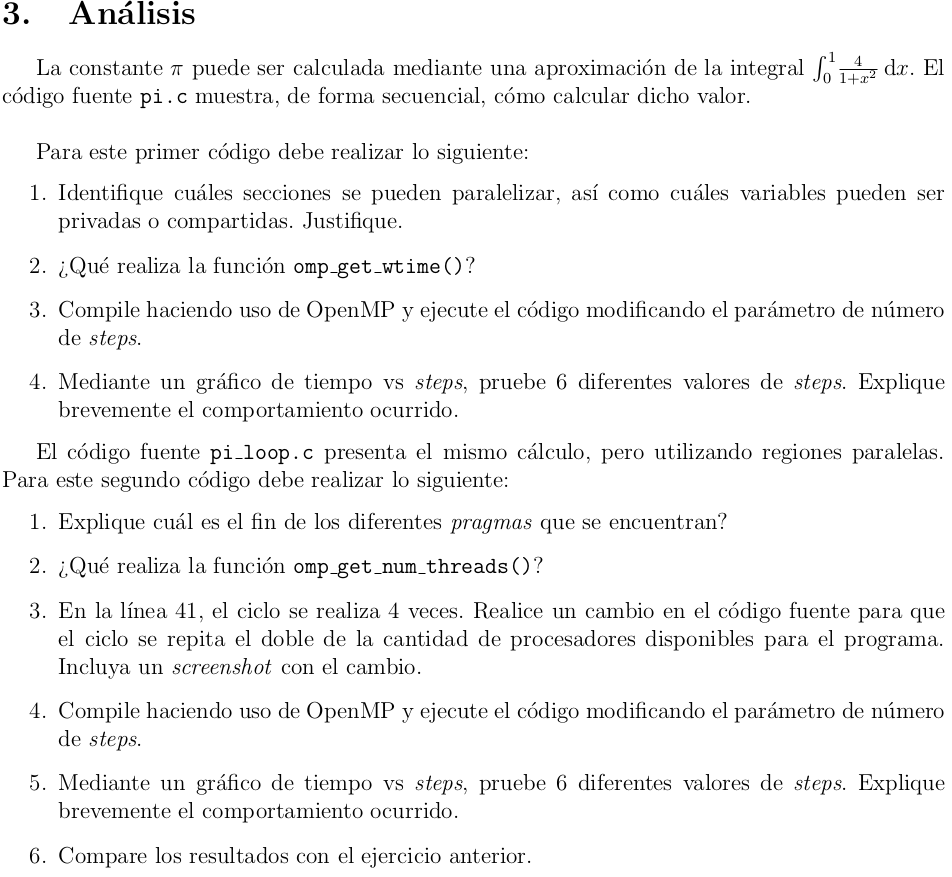
Compartidas:

* sum: La variable sum es una variable utilizada dentro del bucle para acumular la suma de los términos. Como el valor de sum se actualiza en cada iteración y se utiliza en la siguiente iteración, no se puede declarar como privada, ya que cada hilo tendría su propia copia y los resultados finales no serían correctos. En cambio, se declara como compartida, lo que permite que todos los hilos accedan a la misma variable y actualicen su valor de manera segura. Se puede usar la cláusula **reduction** para realizar una operación de reducción en la variable compartida al final de cada iteración del bucle.
* step: La variable step es una constante global y no se actualiza en el bucle, pero si se utiliza en el cálculo de pi, por lo que puede ser una variable compartida.



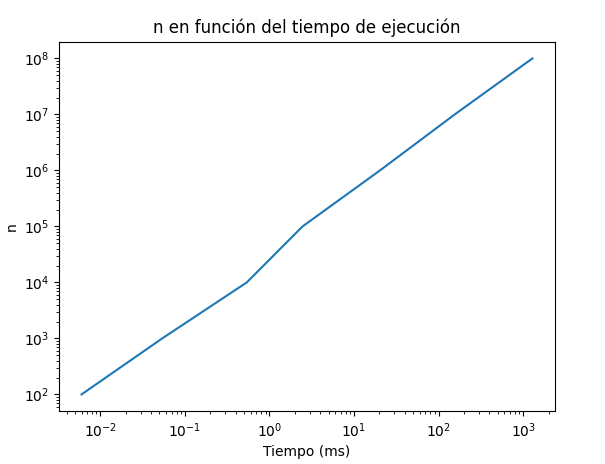
Según [7], la función omp\_get\_wtime() devuelve el tiempo actual del sistema en segundos, y se utiliza para medir el tiempo de ejecución de una sección específica del código. En el código dado, la función se utiliza para medir el tiempo total que tarda el bucle for en completarse y calcular el valor de pi.

Al guardar el tiempo inicial en la variable start\_time y restarle al tiempo final obtenido después del bucle, se calcula el tiempo total de ejecución del bucle y se muestra al finalizar la ejecución. Esto es útil para comparar el rendimiento de diferentes implementaciones o configuraciones de paralelización en OpenMP.

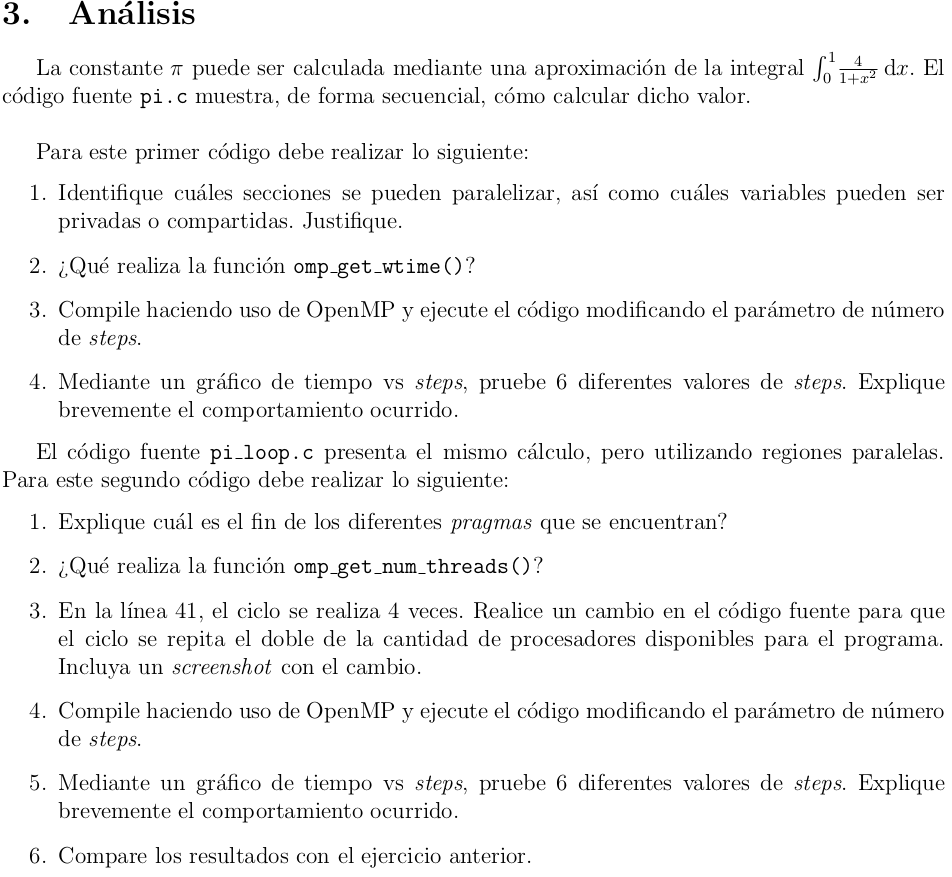


**Muestras:**

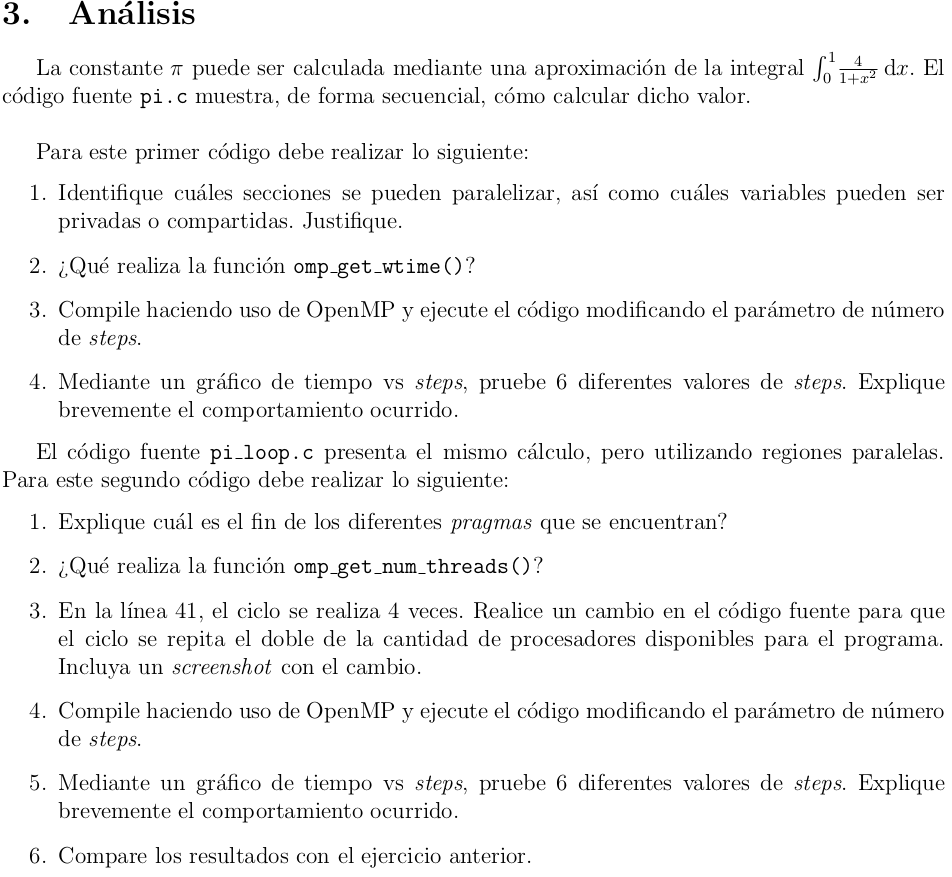
1. pi with 100 steps is 3.141601 in 0.000006 seconds
2. pi with 1000 steps is 3.141593 in 0.000054 seconds
3. pi with 10000 steps is 3.141593 in 0.000540 seconds
4. pi with 100000 steps is 3.141593 in 0.002466 seconds
5. pi with 1000000 steps is 3.141593 in 0.020307 seconds
6. pi with 10000000 steps is 3.141593 in 0.155542 seconds
7. pi with 100000000 steps is 3.141593 in 1.290154 seconds



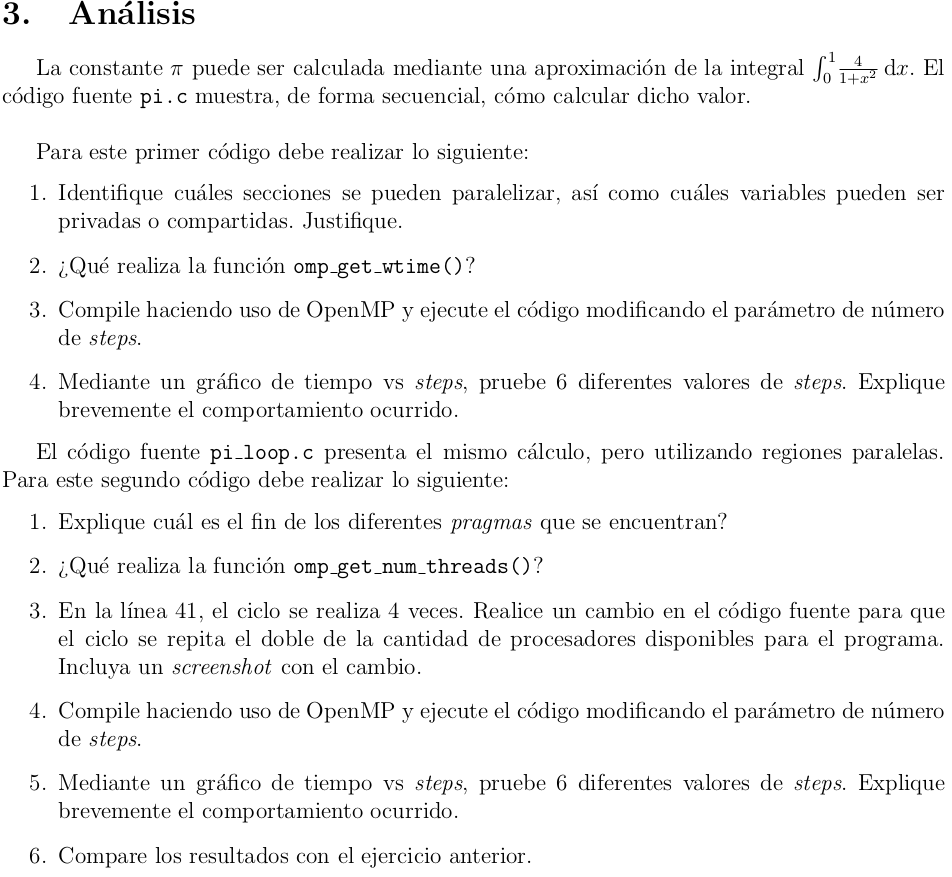
El tiempo de ejecución depende de la cantidad de cálculos realizados en cada iteración del bucle for. A medida que se aumenta la cantidad de subintervalos, se realizan más cálculos, lo que aumenta el tiempo de ejecución. El aumento no es exactamente lineal, pero sigue una escala logarítmica, lo que significa que el tiempo de ejecución aumenta aproximadamente en un orden de magnitud cada vez que se multiplica num\_steps por diez. Además de que aumenta la precisión en el cálculo.

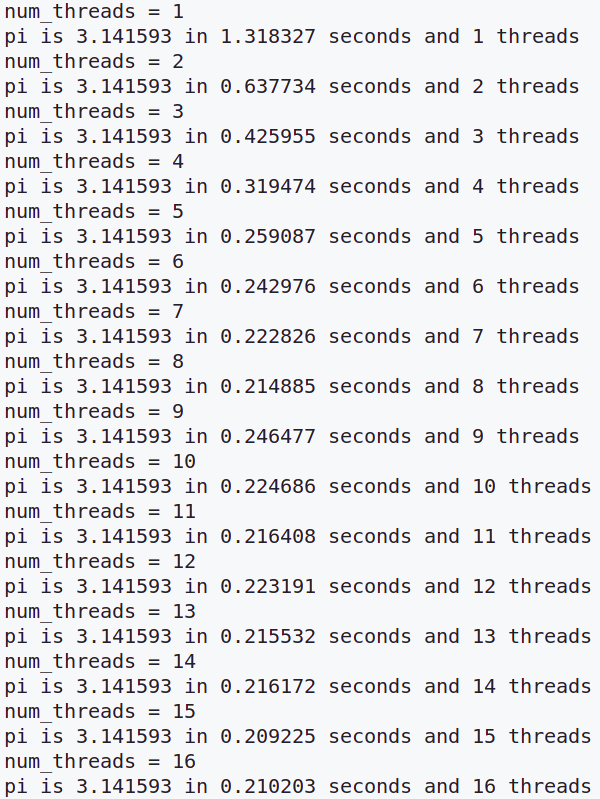


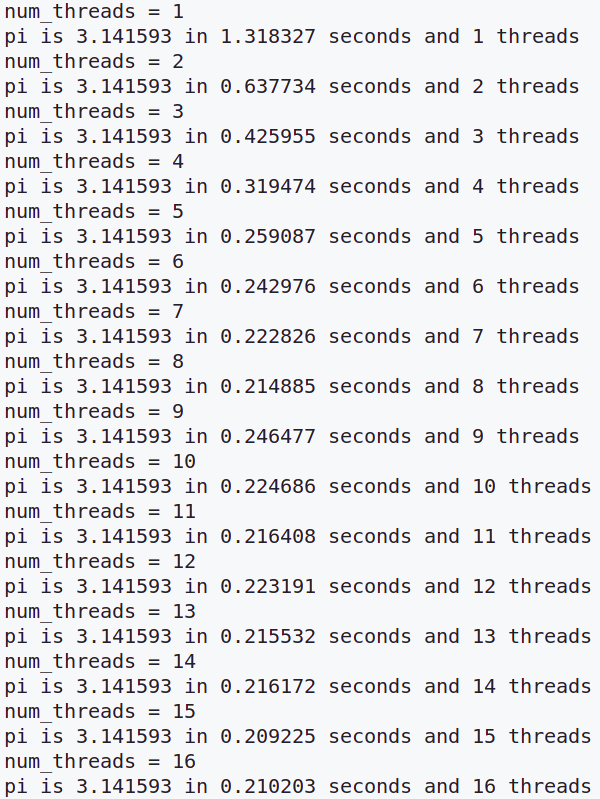
* #pragma omp parallel: Se utiliza para crear un equipo de hilos y la región paralela. Cada hilo en el equipo ejecuta el mismo código. Como se observa, dentro de esta directiva, se pueden utilizar otras directivas para controlar el comportamiento del equipo de hilos.
* #pragma omp single: Se utiliza para garantizar que solo se ejecute una vez la impresión del número de hilos que se están ejecutando en el equipo.
* #pragma omp for: Esta directiva se utiliza para dividir un bucle en varias iteraciones y distribuir las iteraciones a diferentes hilos para su ejecución. Cada hilo ejecuta una parte diferente del bucle. En este caso se emplea para paralelizar el bucle que calcula la suma de la serie para estimar el valor de Pi.
* #reduction(+:sum): Se utiliza dentro de la cláusula for para garantizar que la actualización de sum sea segura, Lo que ocurre aquí es que cada hilo mantiene una copia privada de la variable compartida y actualiza esta copia local en su sección de código. Al final de la sección paralela, todas las copias locales se combinan en una única variable global utilizando una operación específica de reducción (en este caso la suma). El resultado final se almacena en una variable privada.
* private(x): Se utiliza dentro de la cláusula for para declarar la variable x como privada para cada hilo. Esto significa que cada hilo tendrá su propia copia local de la variable x, en lugar de compartir una variable global.

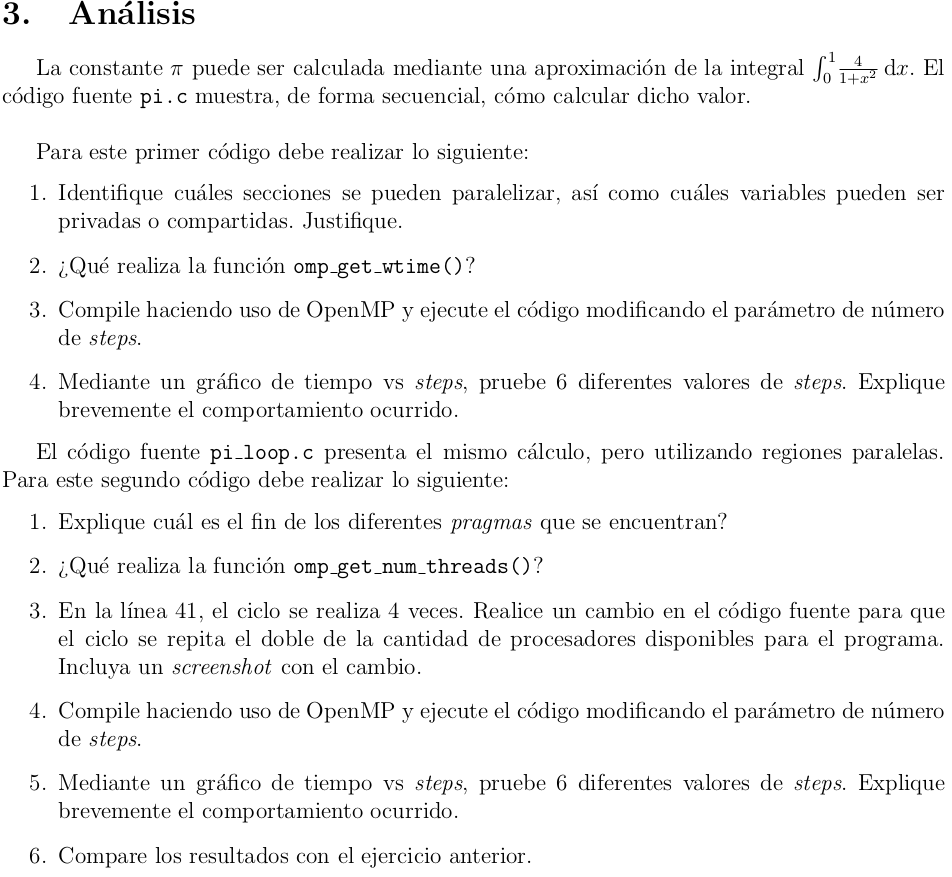


Según [6], es una función de la API de OpenMP que se utiliza para obtener el número de hilos en ejecución en un equipo en particular. En el código dado, se llama dentro de pragma omp single para imprimir el número de hilos en ejecución en ese momento. Por lo que, la llamada a omp\_get\_num\_threads() se realiza solo una vez y se imprime el número de hilos que se utilizarán para la próxima sección paralela omp for.



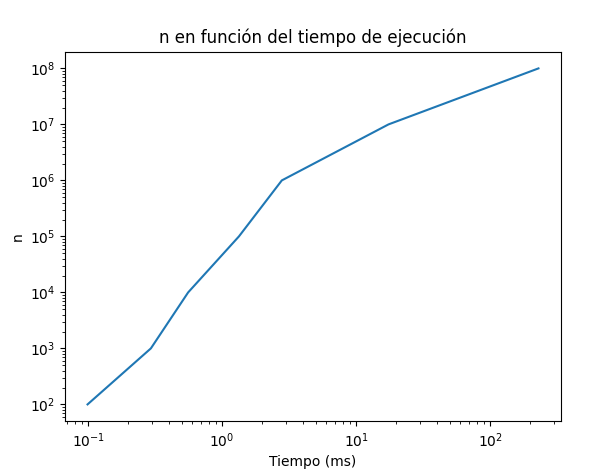






**Muestras:**

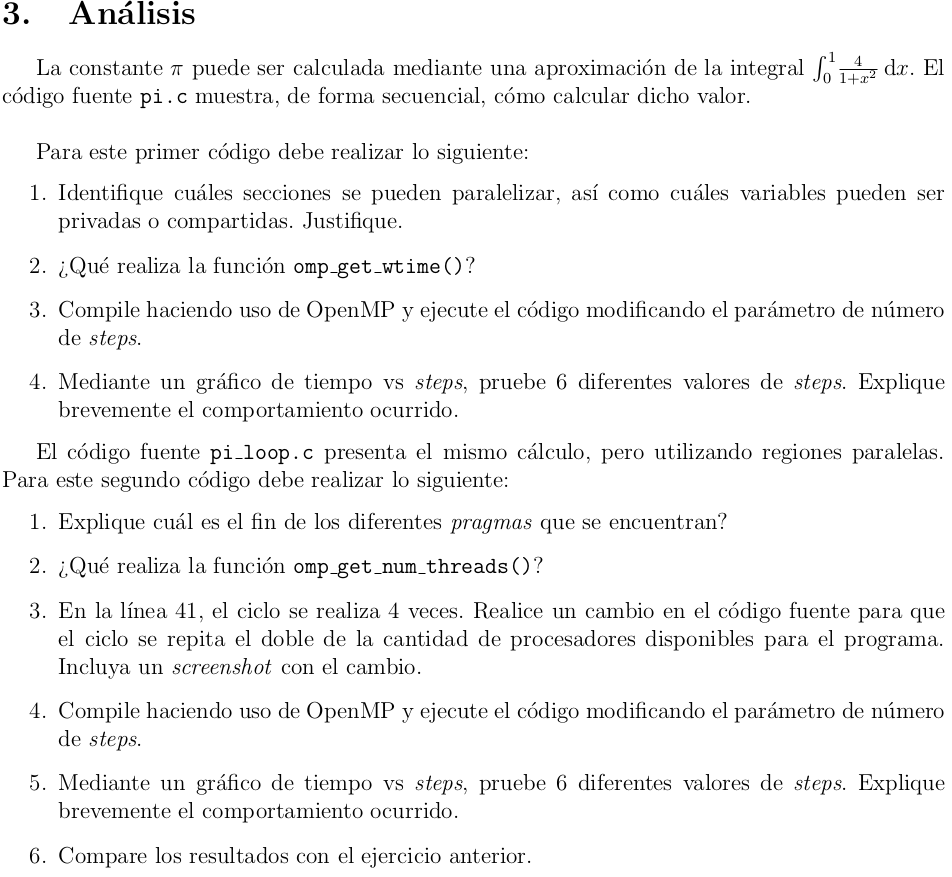
1. pi with 100 steps is 3.141601 in 0.000099 seconds and 16 threads
2. pi with 1000 steps is 3.141593 in 0.000294 seconds and 16 threads
3. pi with 10000 steps is 3.141593 in 0.000558 seconds and 16 threads
4. pi with 100000 steps is 3.141593 in 0.001336 seconds and 16 threads
5. pi with 1000000 steps is 3.141593 in 0.002792 seconds and 16 threads
6. pi with 10000000 steps is 3.141593 in 0.017465 seconds and 16 threads
7. pi with 100000000 steps is 3.141593 in 0.229951 seconds and 16 threads



A medida que aumenta el número de subintervalos, el tiempo de ejecución aumenta de manera significativa. En general, el tiempo de ejecución aumenta proporcionalmente al número de subintervalos. Lo cual se debe a que el cálculo de cada subintervalo es independiente del cálculo de los demás subintervalos, lo que permite un procesamiento en paralelo eficiente utilizando OpenMP.

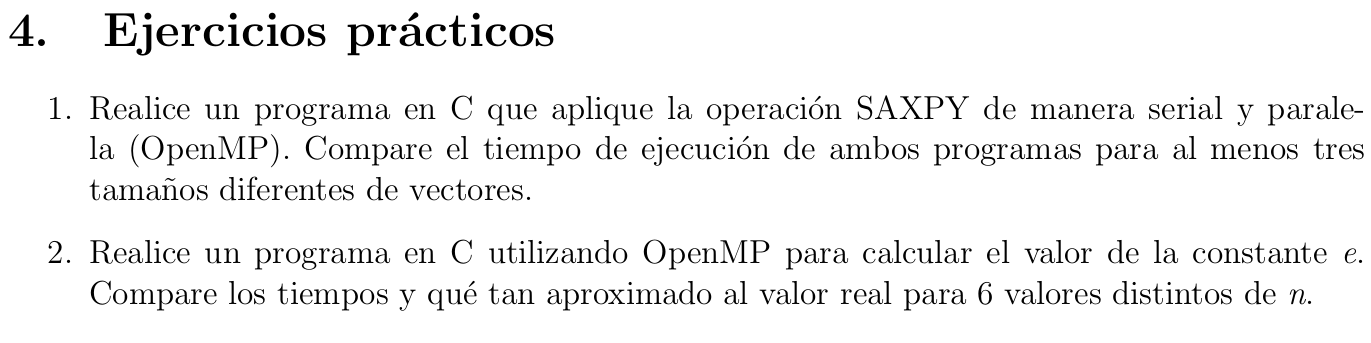
Se observa que el tiempo de ejecución aumenta a medida que se aumenta la precisión de la aproximación (al utilizar mayor número de subintervalos). No obstante, el aumento en el tiempo de ejecución no es lineal debido a la sobrecarga de administración de subprocesos, en algún punto se limita.

En definitiva, el comportamiento esperado de los resultados es que el tiempo de ejecución aumente proporcionalmente con el número de pasos y disminuya a medida que se aumenta el número de hilos utilizados. Sin embargo, hay una sobrecarga asociada con la paralelización que puede afectar el tiempo de ejecución y limita el beneficio del paralelismo.

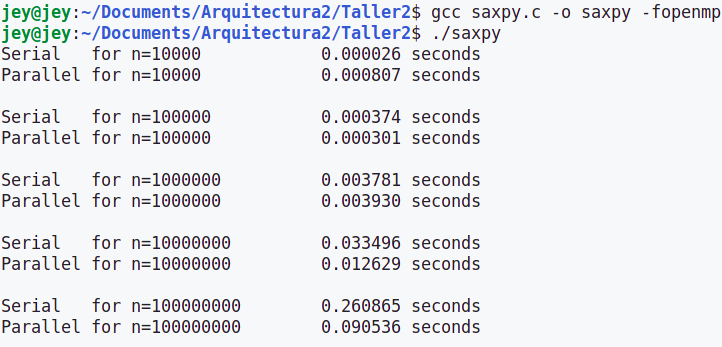


La comparación de los resultados en serie y en paralelo muestra que el tiempo de ejecución para el mismo valor de num\_steps es significativamente más corto en la versión en paralelo. Esto se debe a que el uso de múltiples hilos permite que los cálculos se realicen en paralelo, lo que reduce el tiempo total de ejecución.

También se observa que a medida que aumenta el valor de num\_steps, los tiempos de ejecución de las versiones serial y paralela también aumentan, pero la versión paralela sigue siendo significativamente más rápida.

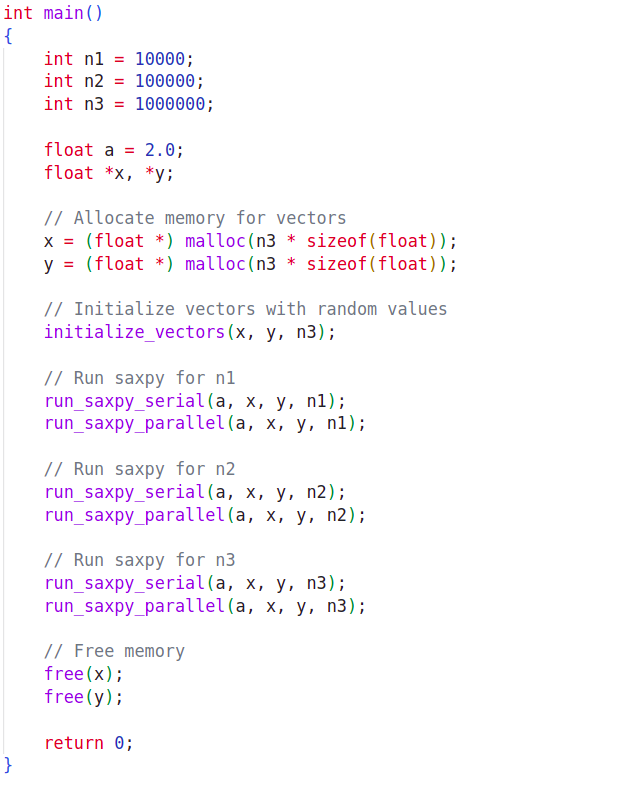
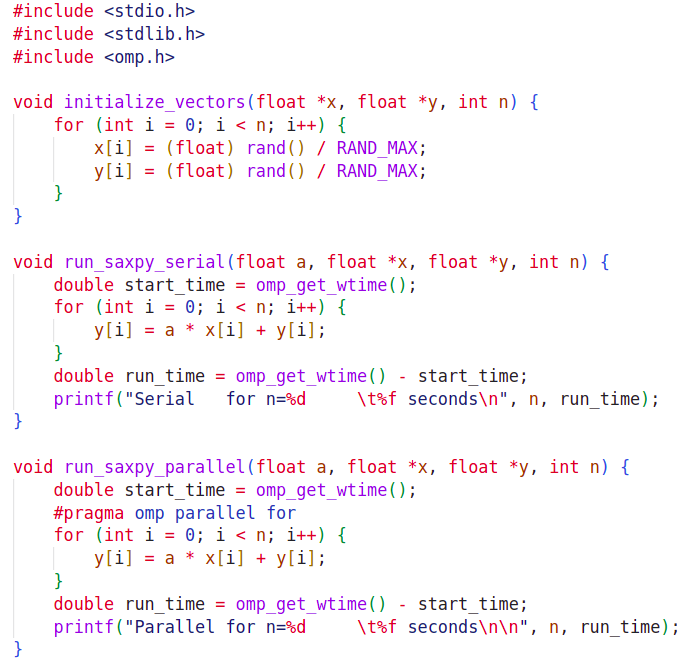


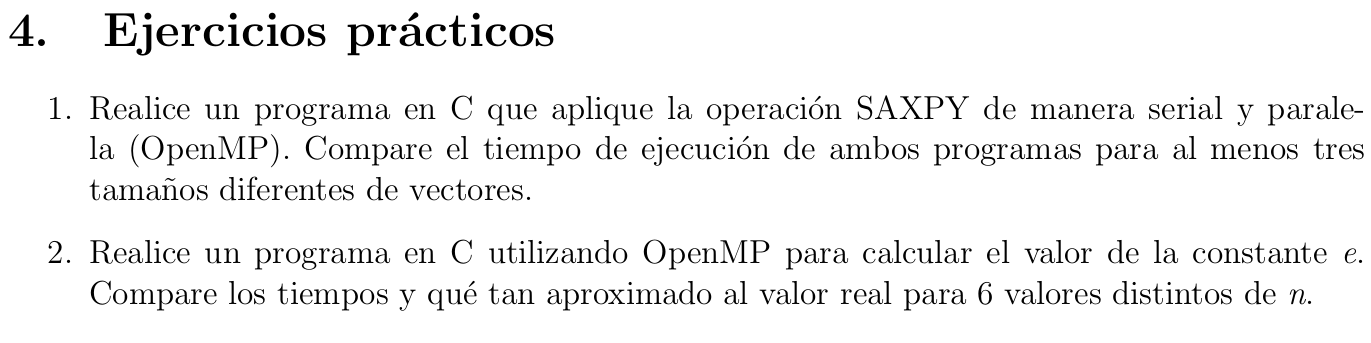
Implementación de la operación SAXPY fue tomada de la sección 2.2.2 de [8].



Se observa que el tiempo de ejecución serial, tiene un comportamiento lineal respecto al número de iteraciones, ahora bien, para el tiempo paralelo, se nota un decremento del tiempo de ejecución alrededor del n = 100000, lo cual indica que alrededor de este valor hay un rendimiento óptimo de los recursos del sistema, o bien la administración de los hilos no se sobrecarga y por ende hay un tiempo de ejecución excelente. Al aumentar o disminuir las iteraciones, tomando este punto como referencia, se ve cómo el tiempo empieza a aumentar.

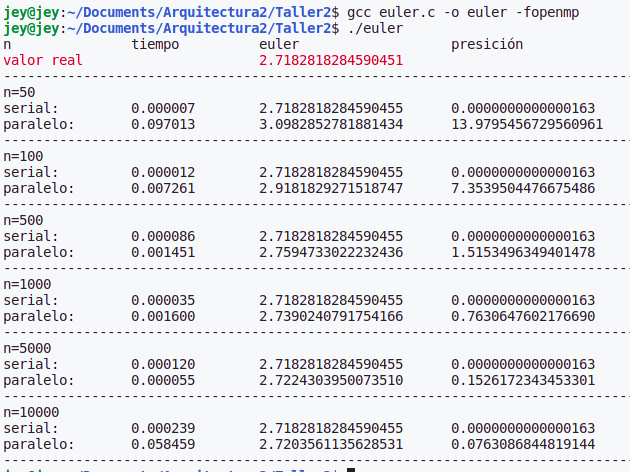
**Código**





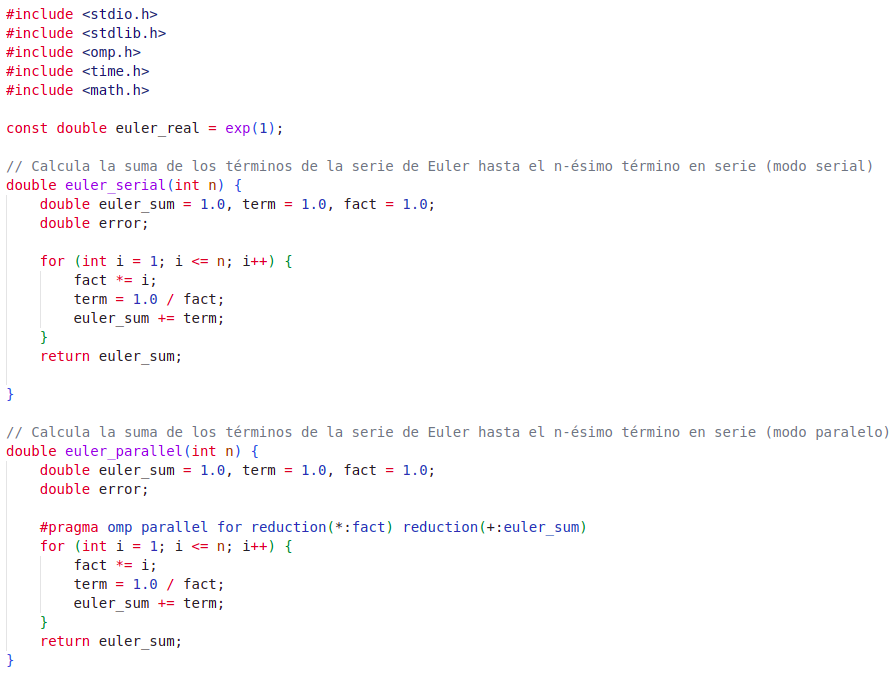
Para calcular el valor de la constante e, podemos utilizar la siguiente serie:

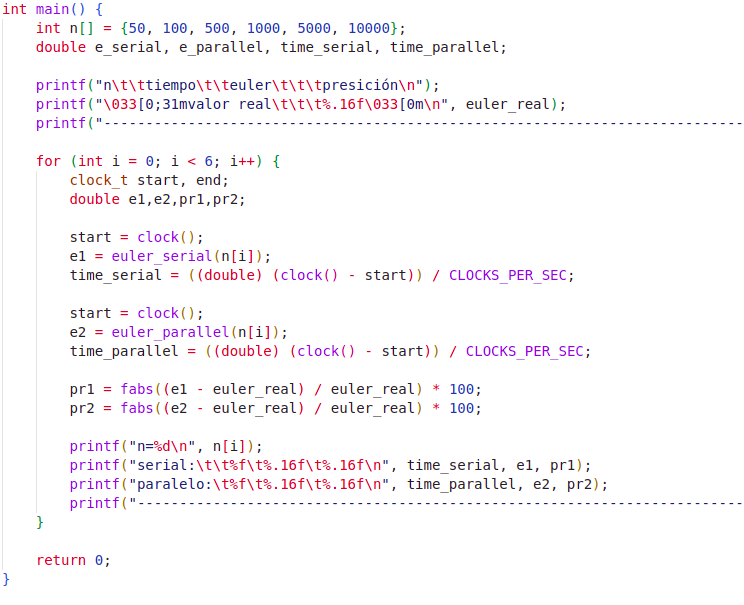
e = 1 + 1/1! + 1/2! + 1/3! + ...



Se observa que la precisión de la ejecución serial permanece igual para todas las ejecuciones, mientras que la precisión paralela es más exacta, la cantidad de error disminuye a medida que aumenta el n. Respecto a los tiempos, se nota que para los valores cercanos al n=5000, los resultados de paralelización son buenos, los valores por debajo y por arriba de esta marca, muestran que existen limitaciones por la saturación. Y como se espera para el tiempo de la ejecución serial, va aumentando a medida que se incrementan las iteraciones.

**Código**





**References**

[1] “What is OpenMP?,” *www.tutorialspoint.com*.  
 <https://www.tutorialspoint.com/what-is-openmp>

[2] “Intro to Parallel Programming with OpenMP,” Bowdoin.edu, 2023.

[https://tildesites.bowdoin.edu/%7Eltoma/teaching/cs3225-GIS/fall17/Lectures/openmp.html#:%7E:text=When%20run%2C%20an%20OpenMP%20program‌](https://tildesites.bowdoin.edu/%7Eltoma/teaching/cs3225-GIS/fall17/Lectures/openmp.html#:%7E:text=When%20run%2C%20an%20OpenMP%20program%E2%80%8C).

[3] “OpenMP Tutorial - Critical, Atomic and Reduction,” *Ive Moved*, Oct. 23, 2011.

<https://michaellindon.github.io/lindonslog/programming/openmp/openmp-tutorial-critical-atomic-and-reduction/index.html>.

‌[4] “Reduction Clauses and Directives,” *www.openmp.org*.  
 <https://www.openmp.org/spec-html/5.0/openmpsu107.html>.

[5] “The Flush Operation,” *www.openmp.org*.  
 <https://www.openmp.org/spec-html/5.0/openmpsu12.html>.

‌[6] “omp\_get\_num\_threads,” www.openmp.org.  
 <https://www.openmp.org/spec-html/5.0/openmpsu111.html>.

[7] “omp\_get\_wtime,” *www.openmp.org*.  
 <https://www.openmp.org/spec-html/5.0/openmpsu160.html>.

[8] ‌M. J. Quinn, "Parallel Programming in C with MPI and OpenMP," McGraw-Hill  
 Education, 2004, ISBN: 978-0071232654.

‌