Arquitectura de Computadores (AC)

2º curso / 2º cuatr.

Grado Ing. Inform.

Cuaderno de prácticas. Bloque Práctico 5. Optimización de código

Estudiante (nombre y apellidos): Álvaro Vega Romero

Grupo de prácticas y profesor de prácticas: B2 – Niceto Rafael Luque Sola

Fecha de entrega: 27/05/21

Fecha evaluación en clase: 28/05/21

Antes de comenzar a realizar el trabajo de este cuaderno consultar el fichero con las normas de prácticas que se encuentra en SWAD

Denominación de marca del chip de procesamiento o procesador (se encuentra en /proc/cpuinfo): Intel(R) Core(TM) i7-1065G7 CPU @ 1.30GHz

Sistema operativo utilizado: Ubuntu 18.04 LTS

Versión de gcc utilizada: gcc (Ubuntu 7.5.0-3ubuntu1~18.04) 7.5.0

Volcado de pantalla que muestre lo que devuelve Iscpu en la máquina en la que ha tomado las medidas:

```
[AlvaroVega zuki@LAPTOP-9IDMGVGU:~] 2021-05-17 Monday
Architecture:
                     x86 64
CPU op-mode(s):
                     32-bit, 64-bit
Byte Order:
                     Little Endian
PU(s):
On-line CPU(s) list: 0-7
Thread(s) per core:
Core(s) per socket:
Socket(s):
Vendor ID:
                     GenuineIntel
CPU family:
Nodel:
Model name:
                     Intel(R) Core(TM) i7-1065G7 CPU @ 1.30GHz
Stepping:
CPU MHz:
                     1498.000
PU max MHz:
                     1498.0000
BogoMIPS:
                     2996.00
Hypervisor vendor:
                     Windows Subsystem for Linux
/irtualization type: container
                     fpu vme de pse tsc msr pae mce cx8 apic sep mtrr pge mca cmov pat pse36 clflush dts acpi mmx fxsr s
lags:
e sse2 ss ht tm pbe syscall nx pdpe1gb rdtscp lm pni pclmulqdq dtes64 est tm2 ssse3 fma cx16 xtpr pdcm pcid sse4_1 sse4
2 movbe popcnt aes xsave osxsave avx f16c rdrand hypervisor lahf_lm abm 3dnowprefetch fsgsbase tsc_adjust bmi1 avx2 sme
 bmi2 erms invpcid avx512f avx512dq rdseed adx smap avx512ifma clflushopt intel_pt avx512cd sha_ni avx512bw avx512vl av
.
512vbmi umip avx512_vbmi2 gfni vaes vpclmulqdq avx512_vnni avx512_bitalg avx512_vpopcntdq rdpid ibrs ibpb stibp ssbd
```

1. (a) Implementar un código secuencial que calcule la multiplicación de dos matrices cuadradas. Utilizar como base el código de suma de vectores de BPO. Los datos se deben generar de forma aleatoria para un número de filas mayor que 8, como en el ejemplo de BPO, se puede usar drand48()).

MULTIPLICACIÓN DE MATRICES:

CAPTURA CÓDIGO FUENTE: pmm-secuencial.c

```
/Sólo puede estar definida una de las tres constantes VECTOR_ (sólo uno de los ./
/tres defines siguientes puede estar descomentado):
/#define MATRIZ_LOCAL // descomentar para que los vectores sean variables ...
// locales (si se supera el tamaño de la pila se ...
// generará el error "Violación de Segmento")
/#define MATRIZ_GLOBAL// descomentar para que los vectores sean variables ...
// tamaño de la pila del programa)
define MATRIZ_DYNAMIC // descomentar para que los vectores sean variables ...
// dinámicas (memoria reutilizable durante la ejecución)
ifdef MATRIZ_GLOBAL
      ef MATRIZ_GLOBAL
double v1[MAX], v2[MAX], v3[MAX];
 int i;
struct timespec cgt1,cgt2; double ncgt; //para tiempo de ejecución
     (argc<2)
    printf("Faltan nº cde filas y columnas de la matriz\n");
exit(-1);
 unsigned int N = atoi(argv[1]); // Máximo N =2^32-1=4294967295 (sizeof(unsigned int) = 4 B) #lfdef MATRIZ_LOCAL
 ##Inder MARKI_LOCAL

double m1[N][N], m2[N],[N] m3[N][N]; // Tamaño variable local en tiempo de ejecución

| // disponible en C a partir de actualización C99
    ifdef MATRIZ_GLOBAL
N=MAX;
  if (N>MAX) N=MAX:
 #ifdef MATRIZ_DYNAMIC
double **m1, **m2, **m3;
m1 = (double**) malloc(N*sizeof(double*));// malloc necesita el tamaño en bytes
m2 = (double**) malloc(N*sizeof(double*)); //si no hay espacio suficiente malloc devuelve NULL
m3 = (double**) malloc(N*sizeof(double*));
   if ( (m1==NULL) || (m2==NULL) || (m3==NULL) )
      for(i = 0 ; i < N ; i++)</pre>
         m1[i] = (double*) malloc (N*sizeof(double));
m2[i] = (double*) malloc (N*sizeof(double));
m3[i] = (double*) malloc (N*sizeof(double));
          if((m1[i]==NULL) || (m2[i]==NULL) || (m3[i]==NULL))
               printf("Error en la reserva de espacio para las matrices\n");
      if (N < 9)
          for (i = 0; i < N; i++)
                for(int j = 0; j < N; j++)
                   m1[i][j] = N * 0.1 + i * 0.1;

m2[i][j] = N * 0.1 + j * 0.1;
                   m3[i][j] = 0;
          srand(time(0));
          for (i = 0; i < N; i++)
  for(int j = 0; j < N; j++)</pre>
                   m1[i][j] = rand() / ((double) rand());
m2[i][j] = rand() / ((double) rand());
m3[i][j] = 0;
      }
      clock_gettime(CLOCK_REALTIME,&cgt1);
```

(b) Modificar el código (solo el trozo que calcula la multiplicación) para reducir el tiempo de ejecución. Justificar los tiempos obtenidos (usando siempre -O2) a partir de la modificación realizada. Incorporar los códigos modificados en el cuaderno.

MODIFICACIONES REALIZADAS (al menos dos modificaciones):

Modificación A) –explicación-: Cambiamos j por k para aprovechar la localidad espacial.

Modificación B) –**explicación-:** Desenrollaremos el bucle de j en 4 iteraciones.

CÓDIGOS FUENTE MODIFICACIONES

A) Captura de pmm-secuencial-modificado A.c

```
for(i=0; i<N; i++)
  for(int k = 0; k < N; k++)
    for(int j = 0; j < N; j++)
        m3[i][j] = m1[i][k] * m2[k][j];</pre>
```

Capturas de pantalla (que muestren la compilación y que el resultado es correcto):

B)

```
float tmp1, tmp2, tmp3, tmp4;
int aux;

for(i=0; i<N; i++)
{
    for(int j = 0; j < N; j++) {
        tmp1 = 0;
        tmp3 = 0;
        tmp4 = 0;

    for(int k = 0, aux = 0; aux < N/4; k+=4, aux++) {
        tmp1+=m1[i][k]*m2[k][j];
        tmp2+=m1[i][k+1]*m2[k+1][j];
        tmp3+=m1[i][k+2]*m2[k+2][j];
        tmp4+=m1[i][k+3]*m2[k+3][j];
    }

    m3[i][j]=tmp1+tmp2+tmp3+tmp4;
}
</pre>
```

```
[AlvaroVega zuki@LAPTOP-9IDMGVGU:-/bp4/ejer1] 2021-05-17 Monday
$gcc -02 -lrt -o pmm-secuencial-modificado_B pmm-secuencial-modificado_B.c
[AlvaroVega zuki@LAPTOP-9IDMGVGU:-/bp4/ejer1] 2021-05-17 Monday
$./pmm-secuencial-modificado_B 4
Tiempo(seg.):0.000001400 / Tamaño Matrices:4 / m3[0][0]=(0.640000) / / m3[N-1][N-1]=(1.960000)
m1:

0.400000 0.400000 0.400000 0.400000 0.400000
0.500000 0.500000 0.500000 0.500000 0.500000
0.700000 0.700000 0.700000 0.700000
m2:

0.400000 0.500000 0.600000 0.700000
m2:

0.400000 0.500000 0.600000 0.700000
0.400000 0.500000 0.600000 0.700000
0.400000 0.500000 0.600000 0.700000
0.400000 0.500000 0.600000 0.700000
0.400000 0.500000 0.600000 0.700000
0.400000 0.500000 0.600000 0.700000
0.400000 0.500000 0.500000 0.700000
0.400000 0.500000 0.700000
0.400000 0.500000 0.700000
0.400000 0.500000 0.700000
0.400000 0.500000 0.700000
0.400000 0.500000 0.7000000
0.400000 0.500000 0.7000000
0.400000 0.500000 0.7000000
0.400000 0.500000 0.7000000
0.400000 0.500000 0.7000000
0.400000 0.500000 0.600000 0.7000000
0.400000 0.500000 0.500000 0.7000000
0.400000 0.500000 0.500000 0.7000000
0.400000 0.500000 0.500000 0.7000000
0.400000 0.500000 0.500000 0.7000000
0.400000 0.500000 0.500000 0.7000000
0.400000 0.500000 0.500000 0.7000000
0.400000 0.500000 0.500000 0.7000000
0.400000 0.500000 0.500000 0.7000000
0.400000 0.500000 0.500000 0.7000000
0.400000 0.500000 0.500000 0.7000000
0.400000 0.500000 0.500000 0.7000000
0.400000 0.500000 0.500000 0.7000000
0.400000 0.500000 0.500000 0.7000000
0.400000 0.500000 0.500000 0.7000000
0.400000 0.500000 0.500000 0.7000000
0.400000 0.500000 0.500000 0.7000000
0.400000 0.500000 0.500000 0.7000000
0.400000 0.500000 0.500000 0.7000000
0.400000 0.500000 0.500000 0.7000000
0.400000 0.500000 0.500000 0.7000000
0.400000 0.500000 0.500000 0.7000000
0.400000 0.500000 0.500000 0.7000000
0.400000 0.500000 0.500000 0.7000000
0.400000 0.500000 0.500000 0.7000000
0.400000 0.500000 0.500000 0.7000000
0.400000 0.500000 0.500000 0.7000000
0.400000 0.500000 0.500000 0.7000000
0.
```

TIEMPOS: (N=10000)

Modificación	Iodificación Breve descripción de las modificaciones	
Sin modificar	-	1.5645
Modificación A)	Cambio de índices entre j y k	0.6216
Modificación B)	Desenrollado de bucle k en 4 iteraciones	1.6712

```
[AlvaroVega zuki@LAPTOP-9IDMGVGU:~/bp4/ejer1] 2021-05-17 Monday
$./pmm-secuencial 1000
Tiempo(seg.):1.564528000 / Tamaño Matrices:1000 / m3[0][0]=(10233.195015) / / m3[N-1][N-1]=(8804.898990
)[AlvaroVega zuki@LAPTOP-9IDMGVGU:~/bp4/ejer1] 2021-05-17 Monday
$./pmm-secuencial-modificado_A 1000
Tiempo(seg.):0.621638900 / Tamaño Matrices:1000 / m3[0][0]=(10554.297259) / / m3[N-1][N-1]=(26062.342575
)[AlvaroVega zuki@LAPTOP-9IDMGVGU:~/bp4/ejer1] 2021-05-17 Monday
$./pmm-secuencial-modificado_B 1000
Tiempo(seg.):1.671282400 / Tamaño Matrices:1000 / m3[0][0]=(14962.193359) / / m3[N-1][N-1]=(11262.910156
)[AlvaroVega zuki@LAPTOP-9IDMGVGU:~/bp4/ejer1] 2021-05-17 Monday
```

COMENTARIOS SOBRE LOS RESULTADOS Y JUSTIFICACIÓN DE LAS MEJORAS EN TIEMPO:

Se puede ver a vista de los resultados, que merece la pena optimizar el código. Con un simple cambio de cambiar j por k en un índice, (para aprovechar la localidad espacial) podemos obtener una gran mejora. También (aunque en mi caso probablemente por un error acumulado no) se debería ver una mejora al desenrollar el bucle en 4 iteraciones. Por lo tanto, merece mucho la pena optimizar el código para optimizar el rendimiento

2. (a) Usando como base el código de BPO, generar un programa para evaluar un código de la Figura 1. M y N deben ser parámetros de entrada al programa. Los datos se deben generar de forma aleatoria para valores de M y N mayores que 8, como en el ejemplo de BPO.

CÓDIGO FIGURA 1:

CAPTURA CÓDIGO FUENTE: figura1-original.c

Figura 1. Código C++ que suma dos vectores. My N deben ser parámetros de entrada al programa, usar valores mayores que 1000 en la evaluación.

(b) Modificar el código C (solo el trozo a evaluar) para reducir el tiempo de ejecución. Justificar los tiempos obtenidos (usando siempre -O2) a partir de la modificación realizada. En las ejecuciones de evaluación usar valores de N y M mayores que 1000. Incorporar los códigos modificados en el cuaderno.

MODIFICACIONES REALIZADAS (al menos dos modificaciones):

Modificación A) – explicación: Juntar los dos for anidados en un for donde i vaya de 2 en 2

Modificación B) -explicación-: Juntar los dos for anidados en un for donde desenrolemos de 2 en 2 u otro número

Otras posibles mejoras que podríamos hacer pero no he implementado son:

Desplazar 1 bit a la izq cuando multiplicamos por dos

En vez de usar if else, usar el operador?

CÓDIGOS FUENTE MODIFICACIONES

A) Captura figura1-modificado_A.c

```
for (ii=0; ii<M;ii++)
{
    X1=0; X2=0;
    for(i=0; i<N;i+=1)
    {
        X1+=2*s[i].a+ii;
        X2+=3*s[i].b-ii;
    }
    if (X1<X2)
        R[ii]=X1;
    else
        R[ii]=X2;
}</pre>
```

Capturas de pantalla (que muestren la compilación y que el resultado es correcto):

```
[AlvaroVega zuki@LAPTOP-9IDMGVGU:~/bp4/ejer2] 2021-05-17 Monday
5./figura1-modificado_A 1000 1000
Fiempo(seg.):0.001948300 / Tamaño Vectores:N: 1000, M: 1000
/ R[0] = 6480 / R[N-1] = 4293982297
```

B)

```
for (ii=0; ii<M;ii++)
{
    X1=0; X2=0;
    for(i=0; i<N;i+=2)
    {
        X1+=2*s[i].a+ii;
        X2+=3*s[i].b-ii;
        X1+=2*s[i+1].a+ii;
        X2+=3*s[i+1].b-ii;
    }
    if (X1<X2)
        R[ii]=X1;
    else
        R[ii]=X2;
}</pre>
```

```
[AlvaroVega zuki@LAPTOP-9IDMGVGU:~/bp4/ejer2] 2021-05-17 Monday
$./figura1-modificado_B 1000 1000
Tiempo(seg.):0.001433900 / Tamaño Vectores:N: 1000, M: 1000
/ R[0] = 9352 / R[N-1] = 4293983044
```

TIEMPOS:

Modificación	dificación Breve descripción de las modificaciones	
Sin modificar	-	0.119748600
Modificación A)	Fusionar bucles	0.078078000
Modificación B)	Fusionar bucles y desenrolar de 2 en 2	0.077411500

COMENTARIOS SOBRE LOS RESULTADOS Y JUSTIFICACIÓN DE LAS MEJORAS EN TIEMPO:

Volvemos a observar como podemos mejorar los resultados de los programas que hagamos mediante pequeños cambios. Estas mejoras en estos casos se producen al no usar recursos en hacer dos bucles idénticos y en desenrolar ese bucle funcionado de 2 en 2 iteraciones.

3. El benchmark Linpack ha sido uno de los programas más ampliamente utilizados para evaluar las prestaciones de los computadores. De hecho, se utiliza como base en la lista de los 500 computadores más rápidos del mundo (el Top500 Report). El núcleo de este programa es una rutina que opera con flotantes de doble precisión denominada DAXPY (*Double precision- real Alpha X Plus Y*) que multiplica un vector por una constante y los suma a otro vector (Lección 3/Tema 1):

for
$$(i=0;i< N;i++)$$
 $y[i]= a*x[i] + y[i];$

Generar los programas en ensamblador para cada una de las siguientes opciones de optimización del compilador: -O0, -Os, -O2, -O3. Explique las diferencias que se observan en el código justificando al mismo tiempo las mejoras en velocidad que acarrean. Incorporar los códigos al cuaderno de prácticas y destacar las diferencias entre ellos. Sólo se debe evaluar el tiempo del núcleo DAXPY. N deben ser parámetro de entrada al programa.

CAPTURA CÓDIGO FUENTE: daxpy.c

Tiempos ejec.	-O0	-Os	-O2	-03
Longitud	0.08400790	0.02256300	0.022479700	0.0238578
vectores=23Millone s(23000000)				

CAPTURAS DE PANTALLA (que muestren la compilación y que el resultado es correcto):

```
[AlvaroVega zuki@LAPTOP-9IDMGVGU:~/bp4/ejer3] 2021-05-17 Monday

$gcc -02 -lrt -o daxpy02 daxpy.c
[AlvaroVega zuki@LAPTOP-9IDMGVGU:~/bp4/ejer3] 2021-05-17 Monday

$gcc -01 -lrt -o daxpy01 daxpy.c
[AlvaroVega zuki@LAPTOP-9IDMGVGU:~/bp4/ejer3] 2021-05-17 Monday

$gcc -0s -lrt -o daxpy0s daxpy.c
[AlvaroVega zuki@LAPTOP-9IDMGVGU:~/bp4/ejer3] 2021-05-17 Monday

$gcc -00 -lrt -o daxpy00 daxpy.c
```

```
[AlvaroVega zuki@LAPTOP-9IDMGVGU:~/bp4/ejer3] 2021-05-18 Tuesday
$./daxpy03 23000000
Tiempo(seg.):0.023857800
                                 / Tamaño Vectores:N: 23000000, / y[0] = 11.112781 / y[N-1] = 20.202529
[AlvaroVega zuki@LAPTOP-9IDMGVGU:~/bp4/ejer3] 2021-05-18 Tuesday
$./daxpy02 23000000
Tiempo(seg.):0.022479700
                                / Tamaño Vectores:N: 23000000, / y[0] = 31.475705 / y[N-1] = 5.322294
[AlvaroVega zuki@LAPTOP-9IDMGVGU:~/bp4/ejer3] 2021-05-18 Tuesday
$./daxpy0s 23000000
Tiempo(seg.):0.022563000
                                 / Tamaño Vectores:N: 23000000, / y[0] = 2208.752252 / y[N-1] = 24.489891
[AlvaroVega zuki@LAPTOP-9IDMGVGU:~/bp4/ejer3] 2021-05-18 Tuesday
$./daxpy00 23000000
Tiempo(seg.):0.084007900
                                / Tamaño Vectores:N: 23000000, / y[0] = 226.896647 / y[N-1] = 9.961027
```

COMENTARIOS QUE EXPLIQUEN LAS DIFERENCIAS EN ENSAMBLADOR:

- O0: El compilador no realiza ninguna optimización del Código, por tanto, suele tener el peor resultado en las ejecuciones aunque mejor tiempo de compilación.
- Os: Optimiza el que más el Código, ocupando menos instrucciones y ejecutándose en un muy buen tiempo.
- O2: Es la opción que hemos usado durante todo el curso, y como bien podemos suponer, la mejor en términos generales. Intenta aumentar el rendimiento del Código sin comprometer el tamaño ni gastar ,ucho tiempo de compilación.
- O3: Obtenemos la mayor optimización posible del Código, aunque estas optimizaciones sean caras a nivel de recursos, provocando que realmente no optimice mucho a nivel a cantidad de instrucciones. Por lo general, el tiempo de ejecución suele ser el mejor de todos.

CÓDIGO EN ENSAMBLADOR (no es necesario introducir aquí el código como captura de pantalla, ajustar el tamaño de la letra para que una instrucción no ocupe más de un renglón):

(PONER AQUÍ SÓLO LA ZONA DEL CÓDIGO ENSAMBLADOR DONDE ESTÁ EL CÓDIGO EVALUADO, USE COLORES PARA DESTACAR LAS DIFERENCIAS)

	daxpyO0.s		daxpyOs.s			daxpyO2.s			daxpyO3.s		
.L6:		0.s -84(%rbp) .L5 -84(%rbp), %eax 0(,%rax,8), %rdx -64(%rbp), %rax %rdx, %rax (%rax), %xmm1 -76(%rbp), %xmm0 -84(%rbp), %eax 0(,%rax,8), %rdx -72(%rbp), %rax %rdx, %rax (%rax), %xmm2 %xmm2, %xmm0 -84(%rbp), %eax 0(,%rax,8), %rdx -64(%rbp), %rax %rdx, %rax %xmm1, %xmm0 %xmm0, (%rax) \$1, -84(%rbp) -84(%rbp), %eax %eax, -80(%rbp) .L6 -32(%rbp), %rax %rax, %rsi \$0, %edi	movsd .L5: (%rbx,%ra:	.LC1(%rip xorl cmpl jbe movsd (%r14,%ra mulsd addsd (%rbx,%ra movsd		.L5:	movsd xorl .p2align 4,, .p2align 3 movsd (%r12,%rax mulsd addsd 0(%r13,%ra	.LC1(%rip), %xmm1 %eax, %eax 10	movq .L6:	## ## ## ## ## ## ## ## ## ## ## ## ##	
										movslq %eax, %rcx movsd .LC1(%rip), %xmm0 movsd (%r12,%rcx,8), %xmm1 leaq 0(%rbp,%rcx,8), %rdx addl \$1, %eax mulsd %xmm0, %xmn cmpl %eax, %r14d	

Cuaderno de prácticas de Arquitectura de Computadores, Grado en Ingeniería Informática

	.L9:	leaq O(%rbp,%ra addsd movsd	(%rdx), %xmm1 %xmm1, (%rdx) .L9 x,8), %xmm0 ax,8), %rdx (%rdx), %xmm0 %xmm0, (%rdx)
	1	leaq xorl	32(%rsp), %rsi %edi, %edi