

Προγραμματισμός συστημάτων υψηλής επίδοσης

ΕΡΓΑΣΙΑ ΧΕΙΜΕΡΙΝΟΥ ΕΞΑΜΗΝΟΥ 2018-2019
ΛΙΟΥΠΗΣ ΠΑΝΑΓΙΩΤΗΣ – 1043795
ΠΑΠΑΔΟΠΟΥΛΟΣ ΠΑΝΤΕΛΗΣ – 1041854
ΚΑΤΣΟΣ ΠΑΝΑΓΙΩΤΗΣ - 1041772

Ερώτημα 1 - Convolution

Στο συγκεκριμένο υπολογιστικό πυρήνα, το κάθε thread αναλαμβάνει να υπολογίσει την νέα συνελιγμένη τιμή του μητρώου B. Τα βάρη που απαιτούνται για τον υπολογισμό αποθηκεύονται σε καταχωρητές για να εκμεταλλευτούμε την ταχύτητά τους. Για να μειωθούν οι απαιτούμενοι υπολογισμοί, δεν υπολογίζονται τα στοιχεία στην περίμετρο του μητρώου B. Όπως φαίνεται στο παρακάτω screenshot η βελτίωση στο χρόνο είναι πολύ μεγάλη ειδικά για μεγάλο μέγεθος προβλήματος. Ο πυρήνας λειτουργεί για οποιοδήποτε μέγεθος μητρώου A. Τα blocks υπολογίζονται με βάση το μέγεθος του προβλήματος που ανατίθεται στον kernel, ενώ τα threads ανά block είναι σταθερά 128.

```
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/1_2D-Convolution$ ./cpu 4096 4096
CPU Runtime: 0.162317s
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/1_2D-Convolution$ ./gpu 4096 4096
blocksPerGrid: 131072
threadsPerBlock: 128
Elapsed time : 0.003808 s
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/1_2D-Convolution$ numdiff convcpu.out convgpu.out
+++ Files "convcpu.out" and "convgpu.out" are equal
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/1_2D-Convolution$ ./cpu 1315 1453
CPU Runtime: 0.018476s
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/1_2D-Convolution$ ./gpu 1315 1453
blocksPerGrid: 14928
threadsPerBlock: 128
Elapsed time : 0.000435 s
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/1_2D-Convolution$ numdiff convcpu.out convgpu.out
+++ Files "convcpu.out" and "convgpu.out" are equal
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/1_2D-Convolution$ ./gpu 2019 723
blocksPerGrid: 11405
threadsPerBlock: 128
Elapsed time : 0.000340 s
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/1_2D-Convolution$ ./cpu 2019 723
CPU Runtime: 0.014310s
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/1_2D-Convolution$ numdiff convcpu.out convgpu.out
+++ Files "convcpu.out" and "convgpu.out" are equal
```

Ερώτημα 2 - TransNormVector

Στο ερώτημα αυτό δημιουργήσαμε 2 πυρήνες. Ο πρώτος υπολογίζει το γινόμενο $A * X$ και ο δεύτερος το γινόμενο $A^T * temp$. Αυτό έγινε γιατί την πρώτη πράξη θέλουμε να την διατρέξουν NX threads ενώ την δεύτερη NY . Έτσι, πετυχαίνουμε καλύτερη αξιοποίηση των πόρων και συνεπώς καλύτερο occupancy. Τα blocks υπολογίζονται με βάση το μέγεθος του προβλήματος που ανατίθεται στον κάθε kernel, ενώ τα threads ανά block είναι σταθερά 64. Στην περίπτωση που το μέγεθος του προβλήματος είναι πολλαπλάσιο του 512 γίνονται 128, πετυχαίνοντας έτσι καλύτερο χρόνο. Προσπαθήσαμε να βάλουμε το διάνυσμα X στην shared memory αλλά δεν βρήκαμε κάποιο δυναμικό τρόπο δέσμησης της shared memory.

```
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/2_TransNormVector$ numdiff transcpu.out transgpu.out -r 1e-15
+++ Files "transcpu.out" and "transgpu.out" are equal
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/2_TransNormVector$ ./cpu 4096 4096
CPU Runtime : 0.089835s
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/2_TransNormVector$ ./gpu 4096 4096
blocksPerGrid 32
blocksPerGrid2 32
threadsPerBlock 128
threadsPerBlock2 128
Elapsed time : 0.005927 s
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/2_TransNormVector$ numdiff transcpu.out transgpu.out -r 1e-15
+++ Files "transcpu.out" and "transgpu.out" are equal
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/2_TransNormVector$ ./cpu 8000 7000
CPU Runtime : 0.304863s
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/2_TransNormVector$ ./gpu 8000 7000
blocksPerGrid 125
blocksPerGrid2 110
threadsPerBlock 64
threadsPerBlock2 64
Elapsed time : 0.020902 s
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/2_TransNormVector$ numdiff transcpu.out transgpu.out -r 1e-15
+++ Files "transcpu.out" and "transgpu.out" are equal
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/2_TransNormVector$ ./cpu 16000 16000
CPU Runtime : 1.398628s
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/2_TransNormVector$ ./gpu 16000 16000
blocksPerGrid 250
blocksPerGrid2 250
threadsPerBlock 64
threadsPerBlock2 64
Elapsed time : 0.134928 s
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/2_TransNormVector$ numdiff transcpu.out transgpu.out -r 1e-15
+++ Files "transcpu.out" and "transgpu.out" are equal
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/2_TransNormVector$ ./cpu 16123 12977
CPU Runtime : 1.126332s
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/2_TransNormVector$ ./gpu 16123 12977
blocksPerGrid 252
blocksPerGrid2 203
threadsPerBlock 64
threadsPerBlock2 64
Elapsed time : 0.112261 s
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/2_TransNormVector$ numdiff transcpu.out transgpu.out -r 1e-15
+++ Files "transcpu.out" and "transgpu.out" are equal
```

Ερώτημα 3 – Covariance

Στο ερώτημα αυτό δημιουργήσαμε 3 πυρήνες. Ο πρώτος υπολογίζει τον μέσο όρο ανά στήλη, χρησιμοποιώντας 128 threads ανά block και block ανάλογα με τον αριθμό των γραμμών του μητρώου. Ο δεύτερος αφαιρεί από το κάθε στοιχείο του μητρώου την αντίστοιχη μέση τιμή της στήλης στην οποία ανήκει. Χρησιμοποιεί 128 threads ανά block και block ανάλογα με τον αριθμό των στηλών του μητρώου. Ο τρίτος υπολογίζει το αποτέλεσμα της πράξης του μητρώου που προέκυψε με το ανάστροφο του. Χρησιμοποιεί 128 threads ανά block και block ανάλογα με τον αριθμό των γραμμών του μητρώου. Όπως φαίνεται και στα παρακάτω screenshots υπάρχει μεγάλη βελτίωση στον χρόνο για οποιαδήποτε διάσταση μητρώου.

```
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/3_Covariance$ ./cpu 2831 1761
CPU Runtime: 55.690250s
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/3_Covariance$ ./gpu 2831 1761
blocksPerGrid 23
blocksPerGrid2 14
blocksPerGrid3 23
Elapsed time : 2.331822 s
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/3_Covariance$ numdiff covcpu.out covgpu.out -r 1e-11
+++ Files "covcpu.out" and "covgpu.out" are equal
```

```
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/3_Covariance$ ./cpu 512 512
CPU Runtime: 0.250855s
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/3_Covariance$ ./gpu 512 512
blocksPerGrid 5
blocksPerGrid2 5
blocksPerGrid3 5
Elapsed time : 0.112279 s
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/3_Covariance$ numdiff covcpu.out covgpu.out -r 1e-11
+++ Files "covcpu.out" and "covgpu.out" are equal
```

```
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/3_Covariance$ ./cpu 1024 1024
CPU Runtime: 2.867621s
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/3_Covariance$ ./gpu 1024 1024
blocksPerGrid 9
blocksPerGrid2 9
blocksPerGrid3 9
Elapsed time : 0.447269 s
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/3_Covariance$ numdiff covcpu.out covgpu.out -r 1e-11
+++ Files "covcpu.out" and "covgpu.out" are equal
```

```
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/3_Covariance$ ./cpu 2048 2048
CPU Runtime: 34.652091s
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/3_Covariance$ ./gpu 2048 2048
blocksPerGrid 17
blocksPerGrid2 17
blocksPerGrid3 17
Elapsed time : 1.881622 s
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/3_Covariance$ numdiff covcpu.out covgpu.out -r 1e-11
+++ Files "covcpu.out" and "covgpu.out" are equal
```

```
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/3_Covariance$ ./cpu 2500 2500
CPU Runtime: 67.343295s
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/3_Covariance$ ./gpu 2500 2500
blocksPerGrid 20
blocksPerGrid2 20
blocksPerGrid3 20
Elapsed time : 2.865641 s
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/3_Covariance$ numdiff covcpu.out covgpu.out -r 1e-10
+++ Files "covcpu.out" and "covgpu.out" are equal
```

```

panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/3_Covariance$ ./gpu 631 1513
blocksPerGrid 5
blocksPerGrid2 12
blocksPerGrid3 5
Elapsed time : 0.385342 s
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/3_Covariance$ ./cpu 631 1513
CPU Runtime: 1.201286s
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/3_Covariance$ numdiff covcpu.out covgpu.out -r 1e-12

+++ Files "covcpu.out" and "covgpu.out" are equal
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/3_Covariance$ ./cpu 1024 512
CPU Runtime: 1.109667s
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/3_Covariance$ ./gpu 1024 512
blocksPerGrid 9
blocksPerGrid2 5
blocksPerGrid3 9
Elapsed time : 0.237136 s
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/3_Covariance$ numdiff covcpu.out covgpu.out -r 1e-12

+++ Files "covcpu.out" and "covgpu.out" are equal
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/3_Covariance$ ./cpu 1831 761
CPU Runtime: 6.944864s
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/3_Covariance$ ./gpu 1831 761
blocksPerGrid 15
blocksPerGrid2 6
blocksPerGrid3 15
Elapsed time : 0.628307 s
panagiotis@panagiotis-MS-7B78:~/Desktop/Cuda/Code/3_Covariance$ numdiff covcpu.out covgpu.out -r 1e-12

+++ Files "covcpu.out" and "covgpu.out" are equal

```

Ο έλεγχος της ορθότητας των αποτελεσμάτων γίνεται σε όλες τις περιπτώσεις με την συνάρτηση numdiff αποθηκεύοντάς τα σε αρχείο και για τον cpu και για την gpu. Η ακρίβεια των αποτελεσμάτων στο πρώτο ερώτημα είναι μέχρι το 15^ο ψηφίο, στο δεύτερο μέχρι το 11^ο ψηφίο και στο τρίτο μέχρι το 12^ο ψηφίο. Η μικρή διαφορά μεταξύ των αποτελεσμάτων οφείλεται στη διαφορετική αριθμητική και λογική μονάδα που χρησιμοποιεί ο cpu-gpu καθώς και σε διάφορες στρογγυλοποιήσεις.

SPECS:

- CPU: AMD Ryzen 2700x (8 core-16 thread) @4.0 GHz
- RAM: 16GB DDR4 3400MHz
- GPU: NVIDIA GTX 970 4GB



