Parallel algorithms for accurate sum and dot product on GPU

Tomohiro SUZUKI*
Interdisciplinary Graduate School of Medical and Engineering,
University of Yamanashi

Accurate summation and dot product algorithms of floating point numbers using error-free transformations ware proposed by Ogita et al.[3] The computed results are as accurate as if computed in K-fold working precision, $K \geq 2$. They require only addition, subtraction and multiplication of floating point numbers in the same working precision as the given data. Since the dot product is an elementary operation in the numerical linear algebra, highly accurate computations for large arrays are frequently required in many applications. However accurate sum and dot product algorithms seem inherently sequential.

To execute the accurate dot product (and sum) efficiently in the SMP environment, a parallel algorithm was proposed[4]. This algorithm is executed according to the following procedures: the array is divided into suitable number of segments, all CPUs process the sequential algorithm for each segment, one CPU processes the sequential algorithm once again gathering all the intermediate results.

From the different viewpoint of the parallelism, we develop an efficient data parallel algorithm applying the *all-prefix-sums* operation[1] (or the parallel reduction, more simply). This is suitable for multicore processors, especially GPUs. This implementation using NVIDIA CUDA[2] with GeForce GTX 285 is tens of times faster than the naive sequential algorithm with a 2.66GHz Intel Core i7 processor.

References

[1] M. Harris, S. Sgengupta and J. D. Owens, *Parallel Prefix Sum (Scan) with CUDA*, In GPU Gems 3, edited by H. Nguyen, pp. 851 – 876, Addison-Wesley.

[2] NVIDIA *CUDA Programing Guide 2.2.1* (2009), `http://www.nvidia.com/object/cuda_get.html`.

[3] T. Ogita, S. M. Rump and S. Oishi, *Accurate sum and dot product*, SIAM Journal on Scientific Computing (SISC), 26(6), pp. 1955 – 1988, (2005).

[4] N. Yamanaka, T. Ogita, S. M. Rump and S. Oishi, *A parallel algorithm for accurate dot product*, Parallel Computing, 34, pp. 392 – 410, (2008).

*stomo@yamanashi.ac.jp