Workshop 18 May – 19 May 2016

Sven Hammarling The University of Manchester

sven.hammarling@btinternet.com

Coverage of Workshop

- Batched BLAS
- Reproducible BLAS and reproducibility
- BLAS for different precisions
- Related Presentations
- Vendor Presentations

See: http://bit.ly/Batch-BLAS-2016 and the workshop report, MIMS EPrint: 2016.41

Batched BLAS (BBLAS)

Nine presentations related to the BBLAS, covering:

- Overview of the draft specification
- Reference implementation and testing
- Use and applications of the BBLAS
- Grouping and interleaved options for the BBLAS

BBLAS Discussion (1 of 2)

In the initial proposal, the argument *batch_opts* can take the values **BATCH_FIXED**, or **BATCH_VARIABLE**, and *batchcount* indicates the number of matrices (batches) to process. The discussion included:

- When to use **BATCH_VARIABLE**, rather than several calls to **BATCH_FIXED**
- The effect of a large value of *batchcount*
- A grouping option for *batch_opts*
- :

BBLAS Discussion (2 of 2)

- •:
- An interleaved data format
- A strided data format
- The aliasing rules
- The number of cases that need implementing
- Optional arguments (handles)
- Row stride as well as column stride (Ida and tda)

There was also discussion of BBLAS versus OpenMP.

Reproducibility

This section comprised three talks covering:

- Motivation and design for a set of Reproducible BLAS
- Reproducibility of complete programs and distributed data structures
- Projects concerned with reproducibility and replication of both software and research

BLAS for Different Precisions

This section comprised two talks covering:

- BLAS for different precisions, including reduced precision BLAS, with a suggested naming scheme
- The existing extended precision BLAS (XBLAS), together with a proposal for a slimmed down version of the XBLAS

The Related Presentations

This section comprised four talks covering:

- Bench-testing environment for automated software tuning (BEAST)
- The European Horizon 2020 project, NLAFET, concerned with parallel numerical linear algebra for future extreme scale systems.
- Communication avoiding algorithms for iterative methods. This is being investigated as part of the NLAFET project.
- Towards ATLAS 4.0. Described the work being done towards the next software release.

- Intel and MKL
- ARM and the ARM performance libraries
- NAG and the BLAS
- The MathWorks and the BLAS, including batched BLAS, the importance of reproducibility, and reduced precision BLAS
- Nvidia and the CUBLAS
- Cray and the Cray scientific and math libraries

Two of the vendors mentioned requests they have recieved for additional BLAS: Intel – integer versions of GEMM (IGEMM); NAG – multiplication of triangular matrices, which arises in computing matrix functions.

Other Issues

Is batching required for the reproducible and reduced precision BLAS?

Grateful thanks were given to the organisers, Jack and his colleagues Leighanne, Teresa and Tracy.