

1 Task 1: Heat2D on MPI

Initial remark: An attempt at solving the full task including subtasks c) and d) was made but time was not sufficient to eradicate bugs in subsections c) and d) for multigrid. Hence, the code is submitted as-is with partially functioning multigrid kernels which are deactivated by setting the gridCount to 1.

1.1 Implementation

The original code was refactored and extended to include the following:

- storing all grids in `double*` pointers instead of the original cumbersome `double**` pointers.
- extending `gridLevel` to hold MPI-specific information
- extending each level's grids to include halos for MPI communication
- introducing a new `world` struct for each MPI rank to hold grid-specific information
- using `MPI_Cart_create` and associated MPI calls to dynamically create a cartesian grid

1.2 Results

The CPU code was run as-is for a baseline reference. The MPI code was run for a single rank and then for 1, 2 and attempted 4 full nodes. Unfortunately, the production version of the code without any print statements slowing down the execution was only submitted on the morning of May 31, 2019 after loss of Euler priority privileges. This meant that 4 node jobs could no longer be submitted and also that all but one of those jobs were still pending the evening of May 31 after having sat in the queue for 10+ hours. Switching all jobs over to Daint at that point in time was not realistic, but limited functional but non-representative experiments on a smaller grid were conducted on my own machine. All outputs obtained are given in appendix A.

1.3 Discussion

All runs exhibit the same L2Norm as required in the task specification. The strong scaling speed-up and efficiency are observed as follows.

1.3.1 Jacobi smoothing

- 24 threads: Speed-up xxx, efficiency xxx%
- 48 threads: Speed-up xxx, efficiency xxx%
- 96 threads: Speed-up xxx, efficiency xxx%

1.3.2 Residual calculation

- 24 threads: Speed-up xxx, efficiency xxx%
- 48 threads: Speed-up xxx, efficiency xxx%
- 96 threads: Speed-up xxx, efficiency xxx%

1.3.3 L2Norm calculation

- 24 threads: Speed-up xxx, efficiency xxx%
- 48 threads: Speed-up xxx, efficiency xxx%
- 96 threads: Speed-up xxx, efficiency xxx%

1.3.4 Overall

- 24 threads: Speed-up xxx, efficiency xxx%
- 48 threads: Speed-up xxx, efficiency xxx%
- 96 threads: Speed-up xxx, efficiency xxx%

As explained above, the only experiments that could be completed were on my own machine with a smaller grid, where where the Jacobi smoothing and the residual calculation show a linear speedup and almost perfect efficiency with increasing number of ranks, while the L2Norm kernel's efficiency decreases as ranks are added due to the collective operation of exchanging local L2Norms.

Looking at the overall efficiency it would be expected that communication costs increase and efficiency decreases as we add ranks on multiple nodes. Optimizing communication could help with this to some extent, but ultimately the algorithm as-is suffers from having to perform an `MPI_Allreduce` for each rank to decide when to stop iterations.

2 Task 2: Communication-Tolerant Programming

2.1 Part a)

2.1.1 Implementation

Using OpenMP, the first hybrid model can be implemented by adding the line `#pragma omp parallel for collapse(3)` before calculating the Jacobi stencil. This can be done because the loops are perfectly nested and no data race is created. Further, as discussed on Piazza, the grid parameter was changed to 768 to allow for integer divisibility by the required number of ranks.

2.1.2 Results

Strangely, the L2 Norms achieved were somewhat inconsistent even though the directive of adapting the grid size to 768 given on Piazza was adhered to. This applies also to the given pure MPI code which was run otherwise unmodified. The pure MPI code was run as-is for a baseline reference on 1, 2 and 4 full nodes. The hybrid MPI/OpenMP code was run for a partial load on 1, 2 and 4 nodes and then for 1, 2 and 4 full nodes. All outputs are given in appendix B.

2.1.3 Discussion

Looking at the pure implementation, network communication becomes evident as expected when we increase from 2 to 4 nodes. For the hybrid implementation, it was attempted to unveil communication cost by running a partial load distributed across several nodes. The increasing significant `MPI_Waitall` times when the isolated MPI ranks are distributed across several nodes reflect those relevant communication cost. Running the hybrid implementation on full nodes in an attempt to reduce the requirement for communication initially shows no improvement on 2 full nodes probably as the reduced communication between ranks is compensated by increasing message size as the ranks' grid fractions are larger in this case. However, when running the hybrid model on 4 full nodes, the `MPI_Waitall` reduce by a factor of almost two as the benefits of less communication per se coincides with reduced message size, hinting that this benefit would likely increase when adding further nodes.

A Task 1 Euler outputs

Listing 1: Task 2: Collected Euler outputs.

```

CPU:                1 node(s) / 1 thread(s)
N/A (Job still in Euler queue after 12h approaching deadline on May 31)
MPI:                1 node(s) / 1 thread(s)
N/A (Job still in Euler queue after 12h approaching deadline on May 31)
MPI:                1 node(s) / 24 thread(s)
N/A (Job still in Euler queue after 12h approaching deadline on May 31)
MPI:                2 node(s) / 48 thread(s)
MPI num_procs: 48
MPI dims_x: 8
MPI dims_y: 6

```

	Time (s)	Grid0	Total
Smoothing	478.754		478.754
Residual	138.568		138.568
Restriction	0.000		0.000
Prolongation	0.000		0.000
L2Norm	341.930		341.930
Total	959.252		959.252

```

Running Time          : 20.232s
MPI L2Norm: 70812.9164
MPI:                4 node(s) / 96 thread(s)
N/A (Job no longer accepted on Euler w/ standard user priority on May 31)

```

B Task 2 Euler outputs

Listing 2: Task 2: Collected Euler outputs.

```
Pure MPI: 1 node(s) / 24 * 1 = 24 thread(s): FULL
```

```
Execution Times:  
Compute:      1.0350 s  
MPI_Irecv:    0.0003 s  
MPI_Isend:    0.0021 s  
Packing:      0.0000 s  
Unpacking:    0.0000 s  
MPI_Waitall:  0.4012 s
```

```
Total Time:   1.4423 s  
L2 Norm:      0.9811283499
```

```
Hybrid MPI/OpenMP: 1 node(s) / 4 * 6 = 24 thread(s): FULL
```

```
[eu-c7-104-11:09836] SETTING BINDING TO CORE  
[eu-c7-104-11:09836] MCW rank 0 bound to socket 0[core 0[hwt 0-1]], socket 0[core 1[hwt 0-1]], socket  
→ 0[core 2[hwt 0-1]], socket 0[core 3[hwt 0-1]], socket 0[core 4[hwt 0-1]], socket 0[core 5[hwt  
→ 0-1]]: [BB/BB/BB/BB/BB/BB/.../.../.../.../.../.../.../.../.../...]  
[eu-c7-104-11:09836] MCW rank 1 bound to socket 0[core 6[hwt 0-1]], socket 0[core 7[hwt 0-1]], socket  
→ 0[core 8[hwt 0-1]], socket 0[core 9[hwt 0-1]], socket 0[core 10[hwt 0-1]], socket 0[core 11[  
→ hwt 0-1]]: [.../.../.../.../.../BB/BB/BB/BB/BB/BB][.../.../.../.../.../.../.../.../.../...]  
[eu-c7-104-11:09836] MCW rank 2 bound to socket 1[core 12[hwt 0-1]], socket 1[core 13[hwt 0-1]],  
→ socket 1[core 14[hwt 0-1]], socket 1[core 15[hwt 0-1]], socket 1[core 16[hwt 0-1]], socket 1[  
→ core 17[hwt 0-1]]: [.../.../.../.../.../.../.../.../.../.../.../BB/BB/BB/BB/BB/BB/.../.../.../.../  
[eu-c7-104-11:09836] MCW rank 3 bound to socket 1[core 18[hwt 0-1]], socket 1[core 19[hwt 0-1]],  
→ socket 1[core 20[hwt 0-1]], socket 1[core 21[hwt 0-1]], socket 1[core 22[hwt 0-1]], socket 1[  
→ core 23[hwt 0-1]]: [.../.../.../.../.../.../.../.../.../.../.../.../.../.../.../.../BB/BB/BB/BB/BB/BB]
```

```
Execution Times:  
Compute:      6.8342 s  
MPI_Irecv:    0.0003 s  
MPI_Isend:    0.0041 s  
Packing:      0.0000 s  
Unpacking:    0.0000 s  
MPI_Waitall:  2.2758 s
```

```
Total Time:   9.1238 s  
L2 Norm:      0.9887095465
```

```
Hybrid MPI/OpenMP: 1 node(s) / 2 * 12 = 24 thread(s): FULL
```

```
[eu-c7-104-10:46038] SETTING BINDING TO CORE  
[eu-c7-104-10:46038] MCW rank 0 bound to socket 0[core 0[hwt 0-1]], socket 0[core 1[hwt 0-1]], socket  
→ 0[core 2[hwt 0-1]], socket 0[core 3[hwt 0-1]], socket 0[core 4[hwt 0-1]], socket 0[core 5[hwt  
→ 0-1]], socket 0[core 6[hwt 0-1]], socket 0[core 7[hwt 0-1]], socket 0[core 8[hwt 0-1]],  
→ socket 0[core 9[hwt 0-1]], socket 0[core 10[hwt 0-1]], socket 0[core 11[hwt 0-1]]: [BB/BB/BB/  
→ BB/BB/BB/BB/BB/BB/BB/BB][.../.../.../.../.../.../.../.../.../.../.../.../.../.../.../...]
```


