

## Problem 2

The most important issue here is that gpuarray  $a$  is being written to while some threads are still reading from it. This causes the wrong data to be read and produces wrong results. Also, access to global could be reduced by placing chunks of  $a$  into shared memory per block. I have written two different implementations, one with `constdouble * a`, and the other with  $a$  being written into shared memory. The performance doesn't seem to change, which makes sense, since the access protocols for a read-only variable is different than a readwrite variable. Figure 1 shows that the computed derivative matches the actual derivative perfectly. The absolute error is 0.0124%. Execution time is 261  $\mu$ seconds.

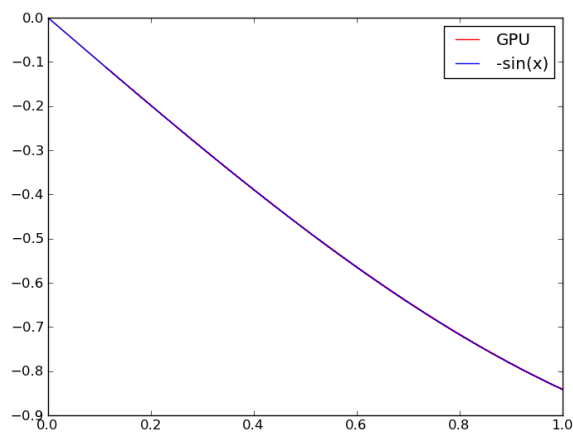


Figure 1: Performance of the second order central difference kernel.