Ordinary RK4 does 4 function evaluations per step, whereas the modified one does 11. As a result, where I did 200 steps for original RK4, I did 73 for the modified one. It turns out that the modified RK4 gave us about 3.4 times higher error on average for the same number of function evaluations. See file p1_out.txt for the output of my program.