CSCI 551 – Numerical and Parallel Programming: FINAL PROGRAM

Improvement and Refactoring of Numerical-Parallel Program

OR

Creative Idea for Numerical-Parallel Programming

DUE: As indicated on Blackboard

Please propose a final program for a numerical problem you can solve and speed-up using parallel programming methods learned in class. ***The problem can be one of the problems from exercises #1 to #6, re-designed, re-factored, improved, with minor additional features OR it can be a creative parallel program proposal to solve a new problem***, of your own interest, that requires numerical methods and computations and can benefit from parallel programming.

Your programs should run on the ECC system, a Raspberry Pi, Jetson, or other commonly available Linux system, using Intel Parallel Studio XE compilers and tools and/or GNU open-source compilers and tools.

ECC Cluster node use policies – Based on your birthday and year, if your year of birth is even, use "o244" nodes and if you birth year is odd, use "o251" nodes. For POSIX shared memory threading (single node use), login to the node # that is the same as your birthday – e.g. for me, odd year, 14th day of month, so I would use "ssh o251-14". This should help distribute the load as we get into problems that are more CPU, I/O, and memory intensive.

Final Program Objectives:

The learning objectives for the final parallel program are:

- ***Choose a numerical and parallel problem that interests you and that is relevant to the class (studied or related to what we have studied)***, explain the problem well, and describe your proposed numerical method (algorithm), the features your program will have, and how you plan to design and program it using MPI, OpenMP, Pthreads, or some hybrid of the 3 methods (combining two or more), or some new method (e.g. CUDA).

- ***Implement a sequential program that solves your problem*** and time it accurately for variously sized problem scenarios as well as describing the algorithmic complexity.

- ***Design and implement a parallel solution*** using one of the parallel programming methods we studied on a multi-core / multi-node system that provides shared memory or distributed memory processor scaling. After making parallel, test this version still produces correct and accurate results for precision used.

- ***Show that your parallel design and implementation speeds-up your solution and that it scales*** with number of processor cores used. Carefully time, identify sections that run in parallel and those that are sequential (use code-level timestamps if more accurate than using command line "time" function – most often this is required).

- ***Provide a code walk-through of your parallel solution, demonstrate build, and run, and describe*** how it works.

1) [25 points total] ***Numerical and Parallel Problem Description*** - Please answer each of the following by indicating use and describing your particular use of the numerical method(s) and the parallel programming method(s). If you use more than one, indicated all used and describe how they are used together.

   [10 pts] ***Final program description*** in terms of major goals and objectives, and problem solved. E.g., this program speeds-up large vector matrix multiplication.

   The main goal of this project to was to implement "Moving Average Converge and Divergence" indicator or "MACD" over an equity's chart parallelly to estimate total earning over the historical data of a particular stock. MACD is an oscillator that is commonly used by the stockbroker to calculate the momentum of a stock. An oscillator a technical analysis used to calculate the high and low bands between extremes. A momentum of a stock refers to an inertia of a price trend to continue either rising or falling for a particular length of time. Momentum usually considers both price as well as the volume during its calculation. Therefore, using such an indicator we could execute trade and see the total earning over a period.

   MACD indicator line is usually calculated by subtracting the 26 days exponential moving average from 12 days exponential moving average. In order to generate a trade signal in algorithm, a signal line is also plotted. According to yahoo finance the signal line for MACD is 9 days exponential moving average of MACD indicator line. The cross over of this signal line and MACD line indicates either a buy or sell trade indicator.

   [5 pts] ***The value of your solution*** and applications of it – e.g., often used for linear systems, such as engineering problems in mechanics and chemical engineering (cite an example from class or that you research on the web).

   The MACD indicator is commonly used by a equities analyst to generate either a buy or sell trade signal for a particular trade. It is usually used in short term trade. Although there might few exceptions to this like SP500 since July of this year, where we see a weak signal,

most equities follow the MACD norm. This algorithm potentially helps traders to delegate some of their risks on such indicator to avoid huge losses.
[5 pts] What ***numerical methods and algorithms*** are used and what type of math is required? Please ***indicate and then describe*** your use and objective for using the method.

Improvement and Refactoring Examples (Check Used Column)

Numerical Method	**Mathematics**	**Description**	**Used**
Vector/matrix operations and transformations	Convolution and transformation	Use of 2-D convolution and transformation functions applied to images (e.g. DCT, rotation, image sharpen/blur, etc.)	
Prime number searching and testing	Prime number theorem	Use of Sieve of Eratosthenes and more advanced methods to find prime density, largest prime in range and list prime numbers in an interval	
Integration	Calculus	Use of Riemann sums, Trapezoidal, Simpson's Rule, or advanced Runge-Kutta	
Non-linear function generation (and integration)	Calculus, Accuracy and Precision	Integration of non-linear functions and sources of error compared to definite integrals	
Gaussian Elimination with Partial Pivoting, Gauss Seidel, and LU factorization or another linear systems solver	Linear systems	Solving systems of equations that describe linear systems (circuits, fluid flow and concentration, etc.) or Linear Programming (optimization), or Regression (over defined)	
Root solving with Newton Raphson, Bisection or Regula Falsi	Non-linear systems	Use of root solving algorithms to find the intercept of non-linear functions – e.g. 2 ballistic trajectories	*
Other?	Probability and Statistics, Discreate math, other?	Other mathematics and numerical methods used that you learned in this course or pre-requisite courses	
Description Please describe methods you used here.	Brute force sign change method Regula falsi method Over piece wise linear interpolation for intraday prices.		

[5 pts] What ***parallel programming methods*** are used?

Parallel Programming Methods Used (Check mark Used Column)

Parallel Programming Method	**Description**	**Used**

POSIX threads	Shared memory threading within a Linux process	
MPI	Message Passing Interfaces between Linux processes on the same node or network interconnected nodes	
OpenMP	Compiler directives to generate parallel shared memory code for specific parts of a program	*
Other	CUDA, OpenCL, hybrid combination, etc.	
Description Please describe methods you used here.		

2) [25 points total] ***Sequential solution and computation time***.

[15 pts] ***Sequential program*** (provide source and Makefile), and demonstration of build and run. Code should be well commented and readable.
After unzipping the provided folder,

- Sequential code activation is stock_macd.cpp.
  - In order to activate the sign change code, comment the top line "#define REGULA_FALSI"
  - In order to activate the regula falsi code, uncomment the top line "#define REGULA_FALSI"
- Sequential code making
  - make stock_macd
- Running the Sequential code
  - ./stock_macd

[10 pts] Use POSIX clock_gettime, ***MPI_Wtime for parallel MPI code***, or POSIX ***clock_gettime*** functions in your code to ***time and log start and complete of the program*** and ***run at least 3 times (ideally 10 or more) to get an average run time***.

The average runtime of the algorithm on o244-27 is

- Brute Force
  - 28.675091 sec = 0.4779 min
- Regula Falsi
  - 680.145505 sec = 11.3357 min

3) [25 points total] ***Parallel design and solution with computation time***.

[15 pts] ***Parallel program*** (provide source and Makefile), and demonstration of build and run. Code should be well commented and readable.

After unzipping the provided folder,

- Parallel code activation is stock_macd.cpp.
    - In order to activate the sign change code, comment the top line "#define REGULA_FALSI"
    - In order to activate the regula falsi code, uncomment the top line "#define REGULA_FALSI"
- Parallel code making
    - make stock_macd_omp
- Running the Parallel code
    - ./stock_macd_omp

[10 pts] Use POSIX clock_gettime, ***MPI_Wtime for parallel MPI***, or POSIX ***clock_gettime*** functions in your code to ***time and log start and complete of the program*** and run at least 3 times (ideally 10 or more) to get an average run time.

- Brute Force
    - 12.971248 sec = 0.2161 min
- RegulaFalsi
    - 286.340167 sec = 4.77233 min

4) [25 points total] ***Parallel speed-up analysis comparing results to Amdahl's law***.

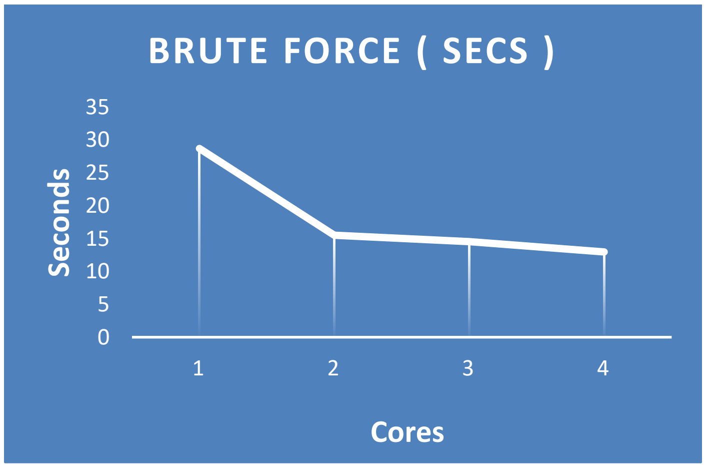
[10 pts] ***Determine parameters for Amdahl's Law***: sequential portion of your code **(1-P)**, parallel portion **P** (note that you can use "gcc -S" to generate instructions to count for best results or count C/C++ statements for a more approximate % of each) and then decide upon a range for number of processor cores **S** (must be at least 1 and 2, but ideally also 4 and 8 if possible). Document this in the table below.

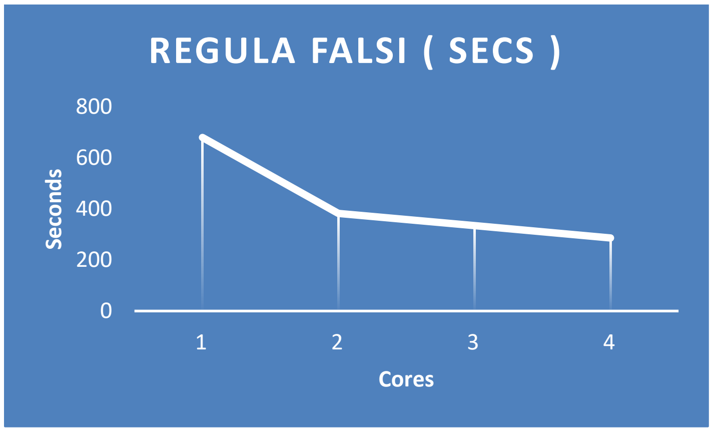
Amdahl's Law parameter	**How obtained?**	**Description**
Sequential portion (% of total)	Timing Serial	Brute Force = 26.99% Regula Falsi = 26.26%
Parallel portion (% of total)	Timing Parallel	Brute Force = 73.01% Regula Falsi = 73.74%
Number of shared memory cores used and	Lscpu/htop	4 cores used for scaling

type		
Number of nodes used in MPI distributed program X # of cores per node	Na	Na
Final value used for S, the scaling factor	Time serial/Time Parallel	Brute Force = 2.21066x Regula Falsi = 2.37530x

[15 pts] ***Plot Amdahl's Law ideal speed-up and your actual speed-up*** based on timings from problem 3 for at least two data points or more. Describe how well your solution compares to the ideal potential speed-up.

Cores	Brute Force (secs)	Regula falsi (secs)
1	28.7044	680.145505
2	15.5187	382.73566
3	14.5359	334.305006
4	12.971248	286.340167





5) [100 points total] ***Code walk-through video or ZOOM live presentation [15 to 30 minutes in length, no longer than 40 minutes maximum]***.

Present your parallel code and demonstrate build and run followed by a walk-through line by line to describe how it works. You can record this as a video (see instructions for Windows and Macintosh recording). Make sure this is high quality and that it is complete, correct, consistent and clear.

Resources

Dr. Siewert
https://www.fidelity.com/viewpoints/active-investor/how-to-use-macd

https://investopedia.com/terms/o/oscillator.asp#:~:text=An%20oscillator%20is%20a%20technical,term%20overbought%20or%20oversold%20conditions

https://www.investopedia.com/terms/m/momentum.asp

https://finance.yahoo.com/news/trading-trends-macd-030000467.html

https://www.investopedia.com/terms/m/macd.asp

https://www.iiis.org/CDs2013/CD2013SCI/SCI_2013/PapersPdf/SA695UD.pdf

https://www.investopedia.com/articles/trading/04/012804.asp

https://finance.yahoo.com/quote/BTC-USD?p=BTC-USD&.tsrc=fin-srch

Overall, provide a well-documented professional report of your findings, output, and tests so that it is easy for a colleague (or instructor) to understand what you have done, what worked, what did not and why (even if you can't complete to your satisfaction). Provide clear instructions on how to run your programs, including command line arguments required and screenshots demonstrating use and test cases you used to verify your parallel and sequential programs.

Include any design files or log files, C/C++ source code you write (or modify) and Makefiles needed to build your code. I will look at your report first, so it must be well written and clearly address each problem providing clear and concise responses and example results (e.g. summary analysis and clearly boxed mathematical answers) to receive credit, but I will look at your log files, code and test results as well if I have questions.

Report file MUST be separate from the ZIP file with code and other supporting materials.

<u>Rubric for Scoring for scale 0…10 (adjust for 0…5 and other scales with 0.0 to 1.0)</u>

Score	**Description of reporting and code quality**
0	No answer, no work done
1	Attempted and some work provided, incomplete, does not build, no Makefile
2	Attempted and partial work provided, but unclear, Makefile, but builds and runs with errors
3	Attempted and some work provided, but unclear, build warnings, runs with no apparent error, but not correct or does not terminate
4	Attempted and more work provided, but unclear, build warnings, runs with no apparent error, but not correct or does not terminate
5	Attempted and most work provided, but unclear, build warnings, runs with no apparent error, but not correct or does not terminate
6	Complete answer, but does not answer question well and code build and run has warnings and does not provide expected results
7	Complete, mostly correct, average answer to questions, with code that builds and runs with average code quality and overall answer clarity
8	Good, easy to understand and clear answer to questions, with easy to read code that builds and runs with no warnings (or errors), completes without error, and provides a credible result
9	Great, easy to understand and insightful answer to questions, with easy to read code that builds and runs cleanly, completes without error, and provides an excellent result
10	Most complete and correct - best answer and code given in the current class

<u>Grading Checklist for Rubric</u>

[25 points] ***Program introduction, significance, and description***:

Problem	Score	Possible	Comments
Goals and objectives with problem statement and solution proposed		10	
Value of your solution (why others should "care")		5	
Numerical methods used		5	
Parallel programming methods used		5	
TOTAL		25	

[25 points] ***Sequential solution and computation time***:

Problem	Score	Possible	Comments
Sequential solution		15	
Time measurement for sequential solution		10	
TOTAL		25	

[25 points] ***Parallel design and solution with computation time***:

Problem	Score	Possible	Comments
Parallel solution		15	
Time measurement for parallel solution		10	
TOTAL		25	

[25 points] ***Parallel speed-up analysis***:

Problem	Score	Possible	Comments
Estimation of Amdahl's law parameters for your parallel solution		10	
Comparison of linear, Amdahl's law and your actual speed-up		15	
TOTAL		25	

[100 points] ***Code Demonstration and Walkthrough***:

Aspect	**Score**	**Possible**	**Comments**
Professionalism		5	
Quality of code submitted – used for walk-through (builds without errors, runs well, commented well, readable, modularized with functions, etc.)		10	
Technical content - (error free, correct use of terminology, concise, but complete)		5	
Interest - motivated and interested in topic, shows passion for work done, concept is clear		5	
Build demonstration – parallel and sequential		10	
Run demonstration – parallel and sequential		10	
Demonstration of speed-up and scaling - of parallel implementation and comparison to sequential		10	
Source code description - function by function and line-by-line as needed		10	
Description of speed-up attained - compared to ideal, original sequential program and Amdahl's law		10	
Speed-up and scaling result - explanation of methods used and whether more improvement is possible		10	
Numerical method(s) used - how well they are explained and used		10	
Significance of solution – why the problem is relevant, of interest to others, and how this solution helps solve the problem		5	
TOTAL		100	