PARALLEL PROGRAMMING IN MATLAB

Forest Danford
CS 620
2/11/16
MATLAB

• "high-level language and interactive environment for numerical computation, visualization, and programming"
• Dynamic, weakly typed, scripting language
• Runs on the JVM (mostly true)
• Can directly write java code and access java libraries
• Can interface with Fortran, C, C++, Perl, Java, Python
  • Users have written custom interfaces too, allowing you to control programs like ImageJ from MATLAB
• Excellent code analysis, incredibly informative error messages, and very very convenient debugging
• Matrix/array is the basic building block of almost everything
The Evolution of Parallel MATLAB

• **1995** - “Why there isn’t parallel MATLAB – our experience has made us skeptical”¹
  
  • Distributed memory model of most contemporary parallel computers = incompatible with MATLAB memory model
  
  • MATLAB spent small portion of execution time on tasks that could be automatically parallelized
  
  • Not enough potential customers to justify a significant development effort
The Evolution of Parallel MATLAB

- **2003** – “Parallel MATLAB: Doing It Right”\(^2\)
  - 27 parallel MATLAB projects identified
  - Leads to development of Interactive Supercomputing's StarP
- **2007** – “The situation is very different today” \(^3\)
  - MATLAB evolved from simple “Matrix Laboratory” into a mature technical computing environment…much more than numerical linear algebra
  - Today’s microprocessors have two or four computational cores
Key features of parallel MATLAB

- Easy learning curve going from serial to parallel
- Excellent documentation and support
  - 600+ page Parallel Computing Toolbox™ User's Guide
- Parallel for loop (parfor)
  - Fork/Join model with incredibly low tangling (literally a one line change)
- Integrated CUDA support
- Seamless transition and integration from local parallel code to running on a compute cluster or grid computing service
- Interactive and batch execution
- Distributed arrays and SPMD
- Integrated MPI for message passing

• http://www.mathworks.com/products/parallel-computing/
Case study

• Empirically investigate Girko’s circular law
  • Eigenvalues of random $N \times N$ square matrix, whose elements come from normal distribution, all lie inside circle radius $\sqrt{N}$

• To the code…
Types of parallel computing in MATLAB

• Implicit parallelism
  • Multithreading

• Explicit parallelism
  • Parallel for loops
  • SPMD
  • Distributed arrays
  • Message passing

• Nearly everything is hidden by software abstraction
  • Pros/Cons
Multithreading

• Many core algorithms in MATLAB are multithreaded and user gets benefit with 0 tangling (also requires no extra knowledge)
• Linear algebra and numerical functions such as fft, \ (mldivide), eig, svd, and sort are multithreaded
• Intel Math Kernel Library integrated - includes multithreaded versions of the BLAS (Basic Linear Algebra Subroutines)
• MATLAB elementary function library, which includes exponential and trigonometric functions, is multithreaded
• Many functions in Image Processing Toolbox™ are multithreaded
• Multithreaded computations are on by default since Release 2008a

• To the code…
parfor

- MATLAB client (main interactive window) coordinates with workers (i.e. separate instances of MATLAB that are headless)
- Iterations are divided into contiguous chunks in reverse loop order from the indices
  - Similar indices get computed in the same chunk and by the same worker
  - Load balancing?
- Data on which parfor operates is sent from the client to workers
- Results are sent back to the client and aggregated

- To the code…
SPMD

• Single Program Multiple Data
  • Indicates there is a single code for execution by each parallel core, but each core has its own on which data to operate
  • MATLAB imperative – spmd-end block (just like parfor)

• Slightly more tangling than parfor, along with need for increased knowledge and different thinking (load balancing)


• Data stored on worker’s heap is accessible by client

• To the code…

• Parfeval is also an alternative [http://www.mathworks.com/help/distcomp/parfeval.html](http://www.mathworks.com/help/distcomp/parfeval.html)
Distributed and Codistributed Arrays

• Again, software abstracts the “heavy lifting” to distribute portions of the array to different workers or replicate the same array on all workers
• Can distribute array upon construction (nice if array too large to fit into a single compute node’s memory)
• Keywords/class overloads many stock MATLAB functions
  • Will change class of returned object
• Many functions are polymorphic and accept dense, sparse, and distributed arrays
• [http://www.mathworks.com/help/distcomp/distribute-arrays.html#br9_n8c](http://www.mathworks.com/help/distcomp/distribute-arrays.html#br9_n8c)
• To the code…
Message Passing

- Most manual (by far) of paradigms this paper covers
- Makes use of the MPI library
- Used inside SPMD block
- As we have seen this necessitates careful thinking
- MATLAB provides some nice abstractions like gop to ease burden of synchronization on developer

- To the code…
Batch on UA’s HPC

- To PuTTY
  - login.hpc.arizona.edu
  - 2 step authentication
  - PBS imperatives
  - qsub shell script
  - How to check status of job
  - Use top to look at other tasks on compute node job got assigned to
  - Error logs and out logs
Questions
References

4. Luszczek, Piotr. "Parallel programming in MATLAB."  