Large Data in MATLAB: A Seismic Data Processing Case Study

U. M. Sundar
Senior Application Engineer
Problem Statement: Scaling Up Seismic Analysis

Challenge:
- Developing a seismic analysis algorithm that can scale up to large data

Solution:
- Analyze seismic data files larger than available system memory
- Use either Kirchhoff migration or Reverse Time migration approach
- Use parallel computing or GPUs for faster processing
Challenges in working with large data

- More data than available memory
  - From multiple files
  - From large files
  - From databases
  - Generated during analysis/simulations

- MATLAB and Toolboxes works on arrays that are “in memory” only

- Multiple versions of code required for
  - In-memory vs. virtual arrays
  - Parallel computations
  - GPUs
How do I scale my algorithm for large data?

*Kirchhoff Migration*

- **Velocity Model**
- **Travel Time Field**
- **Shot Records (Field Data)**
- **Migrate(Shot, TravelTime)**
- **20 GB**
- **Reconstructed Image**

- **8 GB**
How do I scale my algorithm for large data?

*Reverse Time Migration*

- **Velocity Model**

- **Shot Records (Field Data)**

- **ForwardTime (Shot(1:t),V)**

- **FT(nz,nx,nt,nShots)** → **40 TB**

- **Correlation(FT,RT)**

- **RT(nz,nx,nt,nShots)** → **40 TB**

- **Reconstructed Image**

- **ReverseTime (Shot(t:-1:1),V)**
Solving Large Data Access Challenges

- Create custom data import classes
  - Helps reading files in parts
- Use virtual arrays to store data
  - Makes memory from the hard disk available
  - Allows storage more than that of the process memory
Running Algorithms on Large Data

- Derive the Time travel field using Ray tracing algorithm
- Use virtual memory to calculate the time section by section
- Take the shots data and travel time to solve the finite differences
- Use parallel computing to speed up calculations
Scaling MATLAB Applications

MATLAB

SIMULINK®

TOOLBOXES

BLOCKSETS

Worker

Worker

Worker

Worker

Worker

Worker

Worker

Worker
Parallel Computing Tools Address…

Long computations

- Multiple independent iterations
  
  parfor i = 1 : n
  
  % do something with i
  
  end

- Series of tasks

Task-Parallel

Data-Parallel

Large data problems
Parallel and Distributed Computing Products

Desktop Computer

Parallel Computing Toolbox

Computer Cluster

MATLAB Distributed Computing Server

Scheduler

10
What is a Graphics Processing Unit (GPU)

- Originally for graphics acceleration, now also used for scientific calculations
- Massively parallel array of integer and floating point processors
  - Typically hundreds of processors per card
  - GPU cores complement CPU cores
- Dedicated high-speed memory

* Parallel Computing Toolbox requires NVIDIA GPUs with Compute Capability 1.3 or greater, including NVIDIA Tesla 10-series and 20-series products. See [http://www.nvidia.com/object/cuda_gpus.html](http://www.nvidia.com/object/cuda_gpus.html) for a complete listing
Summary of Options for Targeting GPUs

1) Use GPU array interface with MATLAB built-in functions

2) Execute custom functions on elements of the GPU array

3) Invoke your CUDA kernels directly from MATLAB
Workflow: Scaling Up Seismic Analysis

- Get the velocity data
- Get the shot data
- Use **virtual arrays** to accommodate large data

- Derive travel time of shock from velocity data
- Solve the Shots data using finite differences
- Parallel computing with GPU and CPU

- Automatic publish
- Share MATLAB files

**Access**

Data from Multiple sources

**Explore and Create**

**Share**

Report

Application

Automate
MathWorks India – Services and Offerings

Consulting

Local website: www.mathworks.in

Technical Support India: www.mathworks.in/myservicerequests

Product Training: www.mathworks.in/training

Customer Service for non-technical questions: info@mathworks.in

Application Engineering
Training Services
Exploit the full potential of MathWorks products

Flexible delivery options:
- Public training available in several cities
- Onsite training with standard or customized courses
- Web-based training with live, interactive instructor-led courses

More than 30 course offerings:
- Introductory and intermediate training on MATLAB, Simulink, Stateflow, code generation, and Polyspace products
- Specialized courses in control design, signal processing, parallel computing, code generation, communications, financial analysis, and other areas

www.mathworks.in/training
MATLAB Central

- Community for MATLAB and Simulink users
- Over 1 million visits per month
- File Exchange
  - Upload/download access to free files including MATLAB code, Simulink models, and documents
  - Ability to rate files, comment, and ask questions
  - More than 12,500 contributed files, 300 submissions per month, 50,000 downloads per month
- Newsgroup
  - Web forum for technical discussions about MathWorks products
  - More than 300 posts per day
- Blogs
  - Commentary from engineers who design, build, and support MathWorks products
  - Open conversation at blogs.mathworks.com

Based on February 2011 data
MathWorks India Contact Details

URL: http://www.mathworks.in

E-mail: info@mathworks.in

Technical Support: www.mathworks.in/myservicerequests

Tel: +91-80-6632 6000

Fax: +91-80-6632 6010

Thank You for Attending
Talk to Us – We are Happy to Support You
Questions?