The SEAMX Modeling Framework and Process Development for SEAM QC

William Symes, Igor Terentyev, Tetyana Vdovina

The Rice Inversion Project

20 February 2009
Agenda

Overview

Design concept for QC

SEAMX: An Acoustic Modeling Code

Numerical Issues

Qualification Results: First Pass QC

SEAMX Status, Lessons Learned, and Plans
TRIP, SEAM, and SEAMX

Phase I press releases

- SEAM Corporation announces Simulation Award
  10 February 2009

Phase I participants

The SEAM Phase I consortium consists of 24 companies from the hydrocarbon and geophysical exploration industries. Each Participant is represented on the SEAM Phase I Management Committee.

Additional industry participants are encouraged and welcome. To request a model contract, please contact SEAM.

Thank you for visiting the SEAM Phase I page!
TRIP, SEAM, and SEAMX

The question posed by SEAM:

▶ how accurate are seismic simulations, anyway?

*Irresistable!*

TRIP interaction with SEAM:

▶ develop new modeling framework to
  ▶ support all of TRIP’s inversion research
  ▶ provide benchmark for SEAM
▶ develop QC process for use by SEAM and the rest of the seismic research community

Upshot: SEAMX = new modeling package, some surprising lessons, and new research directions.
Agenda

Overview

Design concept for QC

SEAMX: An Acoustic Modeling Code

Numerical Issues

Qualification Results: First Pass QC

SEAMX Status, Lessons Learned, and Plans
Requirements for reliable multi-vendor modeling

Require that vendor traces differ by *an acceptable amount* from benchmark data. Inference: then data from various vendors can be used interchangeably for all purposes.

Implications:

- construction of *benchmark modeling code*, verification;
- spot checks of vendor shot gathers against output of benchmark code.
Proposal for Verification of Benchmark Code

For benchmark-based QC to function as intended, benchmark code must be trusted.

- open source;
- accuracy, repeatability trump speed;
- comparison with analytical solutions (homogeneous medium);
- Richardson extrapolation: *a posteriori* estimation of error without knowing the exact solution. Assume that computed data $D(\Delta t)$ differs from exact data $\bar{D}$ by $E(\Delta t) = C \Delta t^p + O(\Delta t^{p+1})$, then

$$E(\Delta t) \approx \frac{D(2\Delta t) - D(\Delta t)}{2^p - 1}.$$  

A little more work: can estimate $p$ from $D(\Delta t), D(2\Delta t), D(4\Delta t)$.

Upshot: conservative code, basic verification against analytic solutions, estimation of “typical error bars” by Richardson extrapolation using target model.
Agenda

Overview

Design concept for QC

SEAMX: An Acoustic Modeling Code

Numerical Issues

Qualification Results: First Pass QC

SEAMX Status, Lessons Learned, and Plans
Model problem and method

- Pressure-velocity formulation of the acoustic wave equation:

\[
\frac{1}{\kappa(x)} \frac{\partial p(x, t)}{\partial t} = -\nabla \cdot \mathbf{v}(x, t) + f(t, x),
\]

\[
\rho(x) \frac{\partial \mathbf{v}(x, t)}{\partial t} = -\nabla p(x, t),
\]

\(\kappa = \) bulk modulus, \(\rho = \) material density, \(f = \) source of acoustic energy (constitutive law defect).

- Numerical method: staggered FD scheme on rectangular grid

- Requirements:
  - accommodate 3D models, modern survey geometry with reasonable sample rates, respect physics
  - modular design, extensible, portable, public release
SEAMX Features

- staggered leapfrog scheme, 2nd order in time, 2k-th order in space, \( k=1, \ldots, 7 \) (Virieux 1986, Levander 1988, many others)
- PML or zero-pressure boundary conditions on any side
- model input: gridded 1D, 2D, or 3D RSF/SEP file structure, any sensible combination of velocity, bulk modulus, density, bouyancy, any axis order, scale, native/XDR floats
- data output: SU (SEGY without reel header)
- arbitrary source/receiver positions, sample rate (interpolation from comp. grid), flexible source representation
- parallelization via domain decomposition, MPI, also loop level via OpenMP
- domain decomposition \textit{computed} from stencil, \# of processes on each axis - easy extension to other models
- parallelization over shots via subclustering
- ISO C99, self-contained
- SU-style self-doc, html man pages via doxygen
Nearly PML (NPML) for 2D acoustic equation

- Cummer 2003, Habashy et al. 2007
- 2D: 9 domains, 7 (4) variables
- 3D: 27 domains, 10 (6) variables

\[
\frac{1}{\kappa} \frac{\partial p}{\partial t} + \frac{\partial \vec{v}_x}{\partial x} + \frac{\partial \vec{v}_y}{\partial y} = 0,
\]

\[
\rho \frac{\partial v_x}{\partial t} + \frac{\partial p}{\partial x} = 0,
\]

\[
\rho \frac{\partial v_y}{\partial t} + \frac{\partial p}{\partial y} = 0,
\]

\[
\frac{\partial \bar{p}_x}{\partial t} + \sigma(x) \bar{p}_x = \frac{\partial p}{\partial t},
\]

\[
\frac{\partial \bar{p}_y}{\partial t} + \sigma(y) \bar{p}_y = \frac{\partial p}{\partial t},
\]

\[
\frac{\partial \bar{v}_x}{\partial t} + \sigma(x) \bar{v}_x = \frac{\partial v_x}{\partial t},
\]

\[
\frac{\partial \bar{v}_y}{\partial t} + \sigma(y) \bar{v}_y = \frac{\partial v_y}{\partial t}.
\]
Nearly PML (NPML) for 2D acoustic equation

Numerical Example: 3D homogeneous unbounded domain
Parallelization via Domain Decomposition

SEAMX *computes* allocation of exchange buffers, necessary exchange ops - addition of new stencil requires no new MPI code.
Parallelization via Domain Decomposition

Timing Studies: Strong Scaling

► Test Problem;
  ▶ three-dimensional problem
  ▶ 2-2 and 2-10 staggered schemes
  ▶ $480 \times 480 \times 390$ grid points and 500 time steps

► Hardware:
  ▶ SGI Altix, 384 Dual-core Intel Itanium 2 Montvale 9130M
    processors with 8 GB RAM per node
  ▶ Cray XD-1, 316 Dual-core AMD Opteron 275 (2.2GHz)
    processors with 8 GB RAM per node

► Processor Configuration
  ▶ one processor per node, i.e., # of processors = # of nodes
  ▶ one processor per core, i.e., 4 processors per node
Parallelization via Domain Decomposition

**SGI**

- Optimal
- 2–2, Config. 1
- 2–2, Config. 2
- 2–10, Config. 1
- 2–10, Config. 2

**Cray**

- Optimal
- 2–2, Config. 1
- 2–2, Config. 2
- 2–10, Config. 1
- 2–10, Config. 2
Agenda

Overview

Design concept for QC

SEAMX: An Acoustic Modeling Code

Numerical Issues

Qualification Results: First Pass QC

SEAMX Status, Lessons Learned, and Plans
Point source: \( f(t, \mathbf{x}) = w(t)\delta(\mathbf{x} - \mathbf{x}_s) \).

Adjust \( w \) to produce a given pulse \( g(t) \) at appropriate delay in homogeneous medium with velocity \( c_0 \) at offset \( r_0 \):

\[
w(t) = 4\pi c_0^2 r_0 \int_{-\infty}^{t} d\tau \ g(\tau).
\]

Simple representation of \( \delta(\mathbf{x} - \mathbf{x}_s) \): “discrete delta”.

- for \( \mathbf{x}_s \) on gridpoint, assign \( 1/(\text{cell volume}) \) to that point, zero elsewhere.
- otherwise, use adjoint linear interpolation.
Source Representation

Theory: FD field produced by “discrete delta” source is weakly convergent - averages over fixed (grid-independent) volumes converge - and at optimal order of scheme ($\sim 2$).

Practice: point values (traces) appear to converge, but slowly.

Example: 2D homogeneous square with free boundaries, comparison quasi-analytic solution courtesy T. Hagstrom.

<table>
<thead>
<tr>
<th>(\Delta x)</th>
<th>abserr</th>
<th>relerr</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.00</td>
<td>1740</td>
<td>0.201</td>
</tr>
<tr>
<td>5.00</td>
<td>911</td>
<td>0.109</td>
</tr>
<tr>
<td>2.50</td>
<td>581</td>
<td>0.067</td>
</tr>
<tr>
<td>1.25</td>
<td>340</td>
<td>0.039</td>
</tr>
</tbody>
</table>

Upshot: suboptimal convergence of traces, but probably OK because of next item...
Homogeneous or smooth models: regular mesh FD solutions converge to analytical solutions as space and time steps tend to zero in fixed proportion.

Rate of convergence is determined by truncation error.

Staggered finite differences of order 2 in time and 2k in space:
- the error is eventually proportional to $\Delta t^2$,
- spatial order controls numerical dispersion on coarser grids.
FD: Homogeneous/smooth vs. Heterogeneous Media

Brown ’84, Gustafsson & Mossberg ’04, Gustafsson & Wahlund ‘04

➤ Discontinuous models (interfaces or worse):

\[
\text{Error} = \text{component 1} + \text{component 2}
\]

➤ Component 1:
  ➤ corresponds to the truncation error of homogeneous problem
  ➤ is responsible for numerical “grid” dispersion
  ➤ can be controlled by application of higher order methods

➤ Component 2:
  ➤ stems from misalignment of grid and material interfaces
  ➤ is of the first order and insensitive to scheme order
  ➤ related concept: stairstep diffraction

Once the dispersion error (component 1) is controlled, interface misalignment error (component 2) remains and becomes dominant.
Interface Error: Dome Model

- Sampled in 3D from analytical specs
- 2nd order in time, 10th order in space staggered grid Taylor series stencil, $0.2\Delta t_{\text{max}}$
- Point source calibrated to produce unit amplitude Ricker pulse at specified distance; position: near the center of the model, 40 m depth
- 301 traces at 20 m group interval, 20 m depth, 3 s.

**Velocity field**

**Density field**
Interface Error: Dome Model

Shot gathers: 20m, 10m, and 5m grids.

Trace 100 (offset 280 m), $\Delta x = 5m$ (black), 10m (blue), 20m (red)
Interface Error: Dome Model

Trace 100 (offset 280 m), $\Delta x = 5$ m (black), 10m (blue), 20m (red)
Trace 100 (offset 280 m), $\Delta x = 5\text{m} \text{ (black), 10m (blue), 20m (red)}$
Homogeneous case! Trace 100 (offset 280 m), $\Delta x = 5\text{m}$ (black), 10m (blue), 20m (red)
Trace 100 (offset 280 m) for $\Delta x = 5$ m (black) vs. error estimated from $\Delta x = 5$ m, 10 m (red). Note that after 1.5 s, error appears to be 100%!
Effect on Richardson Extrapolation

Richardson estimates of relative RMS errors in various windows:

- 0.7-1.1 s: 20%
- 1.1-1.3 s: 51%
- 1.5-1.7 s: 88%
- 1.9-2.1 s: 120% (!)

Conclusion: interface location error accumulates as wave passes through interface; after encountering a few interfaces, computed wave may be 100% in error!
Agenda

Overview

Design concept for QC

SEAMX: An Acoustic Modeling Code

Numerical Issues

Qualification Results: First Pass QC

SEAMX Status, Lessons Learned, and Plans
Target pulse *shape* specified, but not amplitude.

Error assessment: scale each trace to best RMS fit normalized free space solution.

Compute max and average RMS errors.

Because traces scaled individually to maximize fit, this is an *underestimate* of actual error.
## Homogenous medium test

Maximum relative RMS errors over all traces:

<table>
<thead>
<tr>
<th>Vendor</th>
<th>Max RMS Err</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.100</td>
</tr>
<tr>
<td>2</td>
<td>0.047</td>
</tr>
<tr>
<td>3</td>
<td>0.046</td>
</tr>
<tr>
<td>4</td>
<td>0.047</td>
</tr>
<tr>
<td>5</td>
<td>0.047</td>
</tr>
<tr>
<td>6</td>
<td>0.475</td>
</tr>
<tr>
<td>7</td>
<td>1.606</td>
</tr>
<tr>
<td>8</td>
<td>0.052</td>
</tr>
</tbody>
</table>

Vendors 2-5 and 8 passed this test (so did we).
Half Model Experiments

Half model: 14 km × 14 km × 7.5 km subcube of SEAM model, centered at source (x,y). Extracted by window (and subsample, for 20m) by SU::subset from SEAM 07/08 data files (.vodat). Output 12 km × 12 km × 8 s subset of SEAM shot1 traces, centered at shot1 source (sx=15km, sy=17km), also inline and crossline through source position sampled at 20 m group interval. Source: SEAM 05/08 wavelet.

10 m grid: run on TACC-Ranger, 8×8×4 domains, 7 hrs.

20 m grid: run on CAAM-Obelix 2×2×2 domains, 13 hrs.
Half Model Experiments

Velocity \((x,z)\) slice \(y=17\text{km}\)
Half Model Experiments
Density \((x,z)\) slice \(y=17\text{km}\)
Half Model Experiments

Inline $gy=17\text{km}$, $dgx=20\text{m}$
Half Model Experiments

Crossline gx=15km, dgy=20m
Trace at (12km,17km) - Vendors 3 (blue), 4 (green), 5 (yellow), 8 (black), SEAMX (red).
Half Model Experiments

Trace at (12km,17km) - Vendors 3 (blue), 4 (green), 5 (yellow), 8 (black), SEAMX (red).
Half Model Experiments

Trace at (12km,17km) - Vendors 3 (blue), 4 (green), 5 (yellow), 8 (black), SEAMX (red).
Half Model Experiments

Trace at (12km,17km) - Vendors 3 (blue), 4 (green), 5 (yellow), 8 (black), SEAMX (red).
Relative RMS Errors in each window vs. Vendor 5 trace, compared with Richardson error estimates from 20 m and 10 m SEAMX runs, assuming 1st order convergence:

<table>
<thead>
<tr>
<th>window</th>
<th>1.7-2.3 s</th>
<th>2.4-3.4 s</th>
<th>5.5-6.5 s</th>
<th>7.0-8.0 s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vendor 3</td>
<td>0.79</td>
<td>0.16</td>
<td>0.38</td>
<td>0.47</td>
</tr>
<tr>
<td>Vendor 4</td>
<td>0.79</td>
<td>0.10</td>
<td>0.35</td>
<td>0.42</td>
</tr>
<tr>
<td>Vendor 8</td>
<td>0.77</td>
<td>0.44</td>
<td>0.80</td>
<td>0.77</td>
</tr>
<tr>
<td>SEAMX</td>
<td>0.77</td>
<td>0.30</td>
<td>0.55</td>
<td>0.63</td>
</tr>
<tr>
<td>Rich. Err</td>
<td>0.06</td>
<td>0.31</td>
<td>0.61</td>
<td>0.62</td>
</tr>
</tbody>
</table>

Upshot: variance between results roughly \(\sim\) predicted error in SEAMX 10 m trace.
Full Model Experiments

10 m grid version of full SEAM model, full SEAM qualification geometry, 05/08 wavelet.

Grid volume: \( nx=2800 \times ny=2800 \times nz=1500 \) \( \times \) \( nt=20000 \).

40 hrs on 900 PEs.

Analysis in progress.

Thanks: Scott Morton
Trace at (12km,17km) - Vendors 2 (blue), 3 (purple), 4 (green), 5 (black), 8 (yellow), SEAMX (red).
Agenda

Overview

Design concept for QC

SEAMX: An Acoustic Modeling Code

Numerical Issues

Qualification Results: First Pass QC

SEAMX Status, Lessons Learned, and Plans
Current projects

- i/o server to replace serial disk accesses with message-passing - partially tested
- portable performance tuning, vector/SSE instr. in inner loops - prelim tests promising
- bandwidth-optimized stencil coefficients - appears to be straightforward
- finish documentation - is it ever finished?
- complete test suite - exists, but needs grooming and regression harness

Release acoustic app under GPL, article to Geophysics software section - Q2 09?
What We Have Learned

▶ Quantitative assessment of large-scale complex simulations appears to be feasible.
▶ Basic FD not very accurate (interface misalignment effect, 1st order error) - cf. other talks this AM.
▶ Simple “discrete delta” source modeling not very accurate. With homog. zone around source, can use *time tube* construction - optimal order convergence for point sources (prototype demo’d by TWV in spring 08). Without homog. near-source, other ideas needed (grid refinement?)
▶ Current package attains most design goals.
Where We Go From Here

- implement checkpointing design [needed both for fault-tolerance and for RTM, inversion apps]
- validate framework design by adding new models, schemes: other acoustic methods, *elastic* modeling
- investigate accuracy enhancement methods based on FEM (other talks this AM)
- coupling to inversion framework via TSOpt package, described in PM - RTM, WI, DS