SAND2011-6704C

Application of Random Projection to Seismic Inversion

Joseph Young Denis Ridzal S. Scott Collis
Bart van Bloemen Waanders

Sandia National Laboratories

josyoun@sandia.gov

SEG Annual Meeting, San Antonio, September 22, 2011

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Outline

Introduction to Random Projection

Generalization to a Finite Product of Hilbert Spaces

Application to Parameter Estimation

Computational Study

What Is Random Projection?

Let P be a set of k points in $\mathbb{R}^p$, let $\Phi \in \mathscr{L}(\mathbb{R}^p, \mathbb{R}^q)$ with $q < p$.

Formally, random projection establishes conditions when the inequality

$$(1-\epsilon)\|u-v\|_2^2 \leq \|\Phi(u-v)\|_2^2 \leq (1+\epsilon)\|u-v\|_2^2$$

holds **for all** x and y in P.

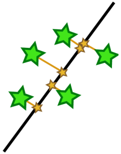
Why **random** projection?

We will see that there are classes of **random** operators Φ for which the above inequality holds with **high probability**.

An Example Application of Random Projection

- Task: Estimate pairwise distances of 5 points in $\mathbb{R}^2$.

- We can **halve** the amount of data that needs processing if we can project the data from $\mathbb{R}^2$ to $\mathbb{R}^1$ and assess the pairwise distances in that space. In order to accomplish this, we project the data onto a **random line** in $\mathbb{R}^2$.



- We only consider the pairwise distances of the projected data in $\mathbb{R}^1$.

Two Important Theorems for Random Projection: 1. Existence of Projection Operators

Theorem (Johnson-Lindenstrauss, 1984)

Let $\epsilon \in (0,1)$ be given. For every set P consisting of k points in $\mathbb{R}^p$, if q is a positive integer such that $q > O(\epsilon^{-2}\log(k))$, there exists a Lipschitz mapping $\Phi : \mathbb{R}^p \rightarrow \mathbb{R}^q$ such that

$$(1-\epsilon)\|u-v\|_2^2 \leq \|\Phi(u)-\Phi(v)\|_2^2 \leq (1+\epsilon)\|u-v\|_2^2$$

for all $u, v \in P$.

In other words, a well-behaved mapping exists so that the projected pairwise distances between points remain bounded by their original pairwise distances.

Two Important Theorems for Random Projection: 2. Construction of Random Projection Operators

Theorem (Corollary of Achlioptas, 2001, applied to pair $\{x, 0\}$)

Given $\epsilon, \beta > 0$, let

$$q_0 = \frac{4 + 2\beta}{\epsilon^2/2 - \epsilon^3/3} \log 2.$$

For integers $p > q \geq q_0$, let $\Phi \in \mathbb{R}^{q \times p}$ be a random matrix from either one of the two probability distributions:

$$\Phi_{ij} = \frac{1}{\sqrt{q}} \times \begin{cases} +1 & \text{with probability} \quad 1/2 \\ -1 & \text{"} \quad 1/2, \end{cases}$$

$$\Phi_{ij} = \frac{\sqrt{3}}{\sqrt{q}} \times \begin{cases} +1 & \text{with probability} \quad 1/6 \\ 0 & \text{"} \quad 2/3 \\ -1 & \text{"} \quad 1/6. \end{cases}$$

With probability at least $1 - 2^{-\beta}$, for all $x \in \mathbb{R}^p$

$$(1 - \epsilon)\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \epsilon)\|x\|_2^2.$$

Properties of the Construction of Random Projection Operators

- Achlioptas' random projection theorem gives sufficient (conservative) conditions on the dimension q of the projected space
- There is a large body of related work on the **reduced isometry property (RIP)** and its use in **compressed sensing**.
- See Blanchard, Cartis, Tanner:
  Compressed Sensing: How Sharp is the Restricted Isometry Property?
  SIAM Review, Vol 53, No 1, pp. 105-125.
- In the context of optimization (parameter estimation) we focus on the **application** of the inequality

  $$(1-\epsilon)\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1+\epsilon)\|x\|_2^2$$

  and **borrow** the construction procedures and probability bounds for the operator Φ from random projection.
- Constructions and bounds for Φ using the RIP may apply as well; *no connection* between our work and compressed sensing otherwise.

Introduction to Random Projection

Generalization to a Finite Product of Hilbert Spaces

Application to Parameter Estimation

Computational Study

Generalization to a Finite Product of Hilbert Spaces

In order to fix notation, we denote a finite product of some Hilbert space Y by $Y^p = \underbrace{Y \times \cdots \times Y}_{p}$. Then, we define the 2-norm of y in Y^p as

$$\|y\|_2 = \sqrt{\sum_{i=1}^{p} \|y_i\|^2}.$$

Note, the underlying Hilbert space is general, e.g. $L^2(\Omega)$ or $H^1(\Omega)$ for some domain Ω are fine.

Next we define a projection operator on the product space, $\Phi \otimes I \in \mathscr{L}(Y^p, Y^q)$, so that $\Phi \in \mathbb{R}^{q \times p}$ and

$$\Phi \otimes I = \begin{bmatrix} \Phi_{11} I & \ldots & \Phi_{1p} I \\ \vdots & \ddots & \vdots \\ \Phi_{q1} I & \ldots & \Phi_{qp} I \end{bmatrix}.$$

Generalization to a Finite Product of Hilbert Spaces

Theorem (Young and Ridzal, 2011)

Given $\epsilon, \beta > 0$*, let*

$$q_0 = \frac{4 + 2\beta}{\epsilon^2/2 - \epsilon^3/3} \log 2.$$

For integers $p > q \geq q_0$*, let* $\Phi \in \mathbb{R}^{q \times p}$ *be a random matrix from either one of the two probability distributions:*

$$\Phi_{ij} = \frac{1}{\sqrt{q}} \times \begin{cases} +1 & \text{with probability} \quad 1/2 \\ -1 & \text{"} \quad 1/2, \end{cases}$$

$$\Phi_{ij} = \frac{\sqrt{3}}{\sqrt{q}} \times \begin{cases} +1 & \text{with probability} \quad 1/6 \\ 0 & \text{"} \quad 2/3 \\ -1 & \text{"} \quad 1/6. \end{cases}$$

Let Y *denote a Hilbert space where* Y^p *denotes the p-times Cartesian product. With probability at least* $1 - 2^{-\beta}$*, for all* $y \in Y^p$

$$(1 - \epsilon)\|y\|_2^2 \leq \|(\Phi \otimes I)y\|_2^2 \leq (1 + \epsilon)\|y\|_2^2.$$

Introduction to Random Projection

Generalization to a Finite Product of Hilbert Spaces

Application to Parameter Estimation

Computational Study

Application of Random Projection to Parameter Estimation

Let $A \in \mathscr{L}(U, \mathscr{L}(Y))$ and $B \in \mathscr{L}(Y)$ be two differential operators so that $A(u) + B$ is invertible for some set of parameters $u \in U$.
A typically represents a spatial differential operator while B is a temporal differential operator. For example, for the wave equation, we have that

$$A(u) = -\nabla \cdot (u \nabla y) \qquad B = y_{tt}.$$

We define the block operator $A_p(u) \in \mathscr{L}(U, \mathscr{L}(Y^p))$ where

$$A_p(u) = \begin{bmatrix} A(u) & & \\ & \ddots & \\ & & A(u) \end{bmatrix},$$

and we define $B_p \in \mathscr{L}(Y^p)$ analogously.

Application of Random Projection to Parameter Estimation

Consider the parameter estimation problem

$$\arg \min_{u\in U, y \in Y^p} \left\{ \frac{1}{2} \sum_{i=1}^{p} \|y_i - d_i\|_2^2 : (A(u) + B)y_i = b_i, i = 1, \ldots, p \right\}.$$

Rewrite in block form as

$$\arg \min_{u\in U, y \in Y^p} \left\{ \frac{1}{2} \|y - d\|_2^2 : (A_p(u) + B_p)y = b \right\}.$$

This is known as the full-space formulation.

In the reduced-space formulation, we solve

$$\arg \min_{u\in U} \left\{ \frac{1}{2} \|(A_p(u) + B_p)^{-1} b - d\|_2^2 \right\}.$$

Application of Random Projection to Parameter Estimation

A direct application of random projection to the reduced objective function from parameter estimation gives

$$(1-\epsilon)\|(A_p(u)+B_p)^{-1}b-d\|_2^2 \leq$$

$$\|(\Phi\otimes I)\left((A_p(u)+B_p)^{-1}b-d\right)\|_2^2 \leq$$

$$(1+\epsilon)\|(A_p(u)+B_p)^{-1}b-d\|_2^2$$

This tells us that the objective value for the projected problem can be **bounded tightly** by the objective value for the original problem.

Further, we have that

$$\|(\Phi\otimes I)\left((A_p(u)+B_p)^{-1}b-d\right)\|_2^2 = \|(A_q(u)+B_q)^{-1}(\Phi\otimes I)b-(\Phi\otimes I)d\|_2^2$$

This tells us that we require

q PDE solves in the projected problem rather than p.

Application of Random Projection to Parameter Estimation

Theorem (Young and Ridzal, 2011)

Given $\beta > 0$ *and* $\epsilon \in (0, 1)$ *choose* q *and* $\Phi \in \mathbb{R}^{q \times p}$ *according to Achlioptas. Let us define the original and projected objective functions as* $J : U \to \mathbb{R}$ *and* $J_\Phi : U \to \mathbb{R}$*, respectively, where*

$$J(u) = \frac{1}{2} \|(A_p(u) + B_p)^{-1} b - d\|_2^2$$

$$J_\Phi(u) = \frac{1}{2} \|(A_q(u) + B_q)^{-1} (\Phi \otimes I) b - (\Phi \otimes I) d\|_2^2 \,.$$

In addition, let $u, s \in U$ *be such that* $J_\Phi(u + s) < C J_\Phi(u)$*, where* $C = (1 - \epsilon)/(1 + \epsilon)$*.*

Then, we have that $J(u + s) < J(u)$ **with probability** $1 - 2^{-\beta}$**.**

This tells us that if an optimization algorithm reduces the projected objective value *quickly enough*, we can guarantee monotonic decrease in the original objective value.

This gives a mathematical foundation for random phase encoding from Krebs, et. al. *Fast full wave seismic inversion using encoded sources* Geophysics, Vol 74, No 6, pp. 177-188.

Introduction to Random Projection

Generalization to a Finite Product of Hilbert Spaces

Application to Parameter Estimation

Computational Study

Random Projection Operators

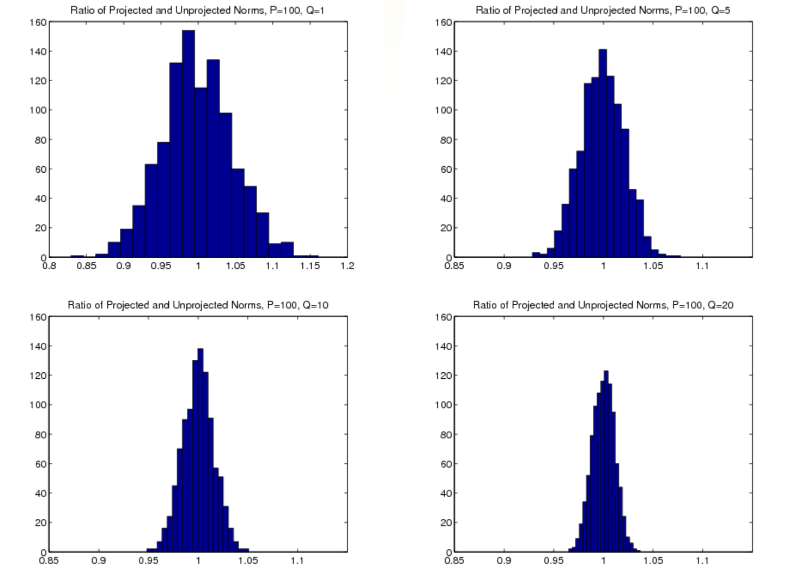
In our first experiment, we examine the bound

$$(1-\epsilon)\|y\|_2^2 \leq \|(\Phi \otimes I)y\|_2^2 \leq (1+\epsilon)\|y\|_2^2$$

on a 30×30 triangular grid Ω where $y \in L^2(\Omega)$. We use Achlioptas' binary operator

$$\Phi_{ij} = \frac{1}{\sqrt{q}} \times \begin{cases} +1 & \text{with probability} \quad 1/2 \\ -1 & \quad\quad\quad '' \quad\quad\quad\quad 1/2 \end{cases}$$

$\|(\boldsymbol{\Phi} \otimes \mathbf{I})\mathbf{y}\|_2^2 / \|\mathbf{y}\|_2^2$



Application to the Acoustic Wave Equation

In the following set of examples, we solve the problem

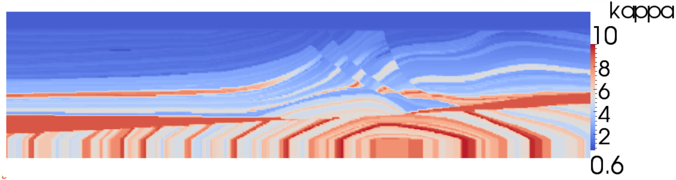
$$\arg \min_{(p,v)\in Y^{N_s},(\rho,K)\in U^2} \left\{ \frac{1}{2}\|\Psi(p,v)-(p,v)^{true}\|^2 : (A_{N_s}(\rho,K)+B_{N_s})(p,v)=s \right\}$$

where

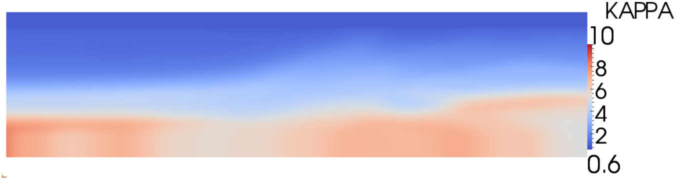
$p \equiv$ Pressure
$v \equiv$ Velocity
$\rho \equiv$ Density
$K \equiv$ Bulk Modulus
$(A(\rho,K)+B)(p,v) \equiv$ First order decomposition of the acoustic wave equation
$N_s \equiv$ Number of sources
$s \equiv$ Source that is Gaussian in space and a Ricker wavelet in time
$\Psi \equiv$ Projection for receivers
$(p,v)^{true} \equiv$ Pressures and velocities generated from **Marmousi2**

True and Initial Starting Models

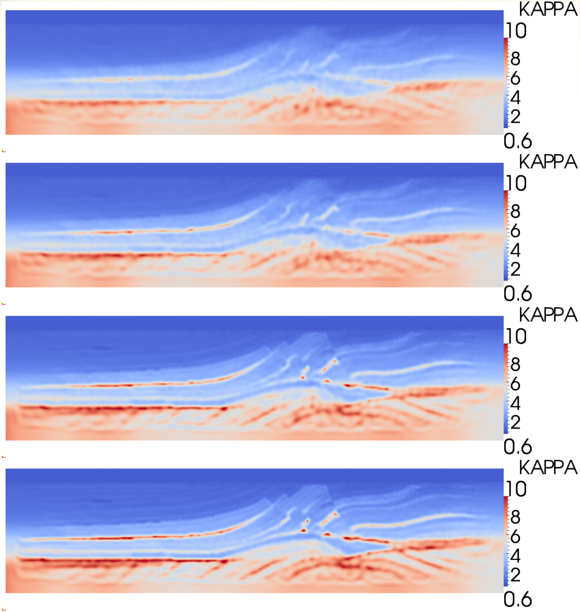
True solution:



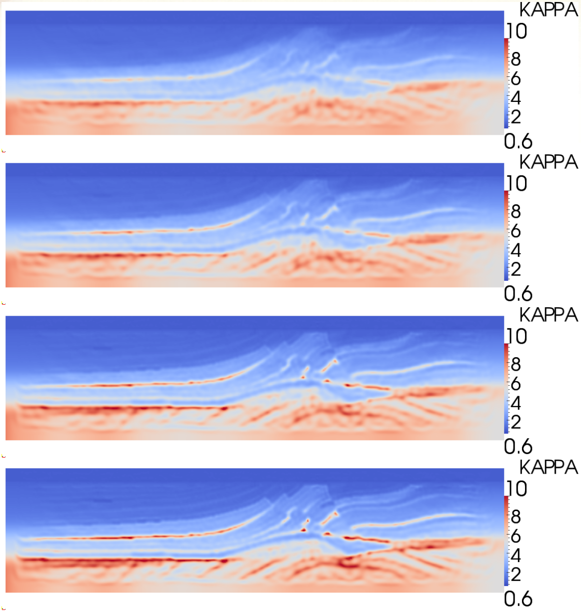
Initial solution for inversion:



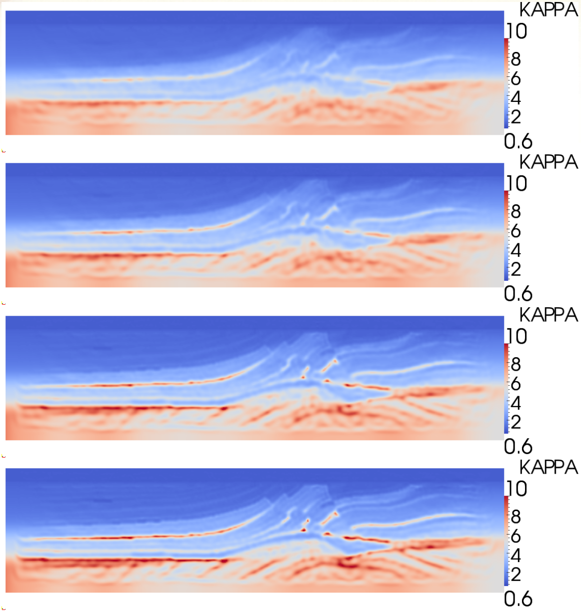
Steepest Descent, 1 Encoded Experiment, 20, 40, 80, 160 Iter



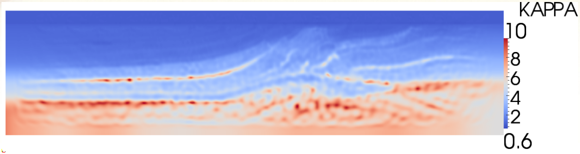
Steepest Descent, 2 Encoded Experiments, 20, 40, 80, 160 Iter

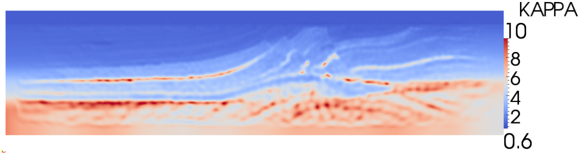


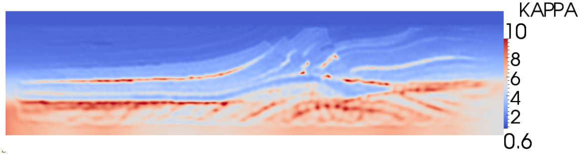
Steepest Descent, 4 Encoded Experiments, 20, 40, 80, 160 Iter



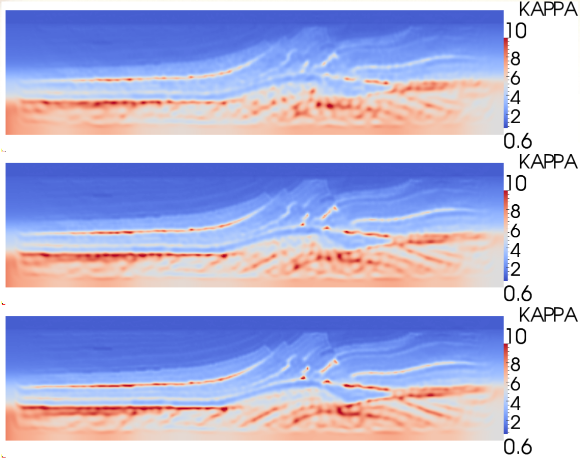
Gauss-Newton, 1 Encoded Experiment, 20, 40, 80 Iter



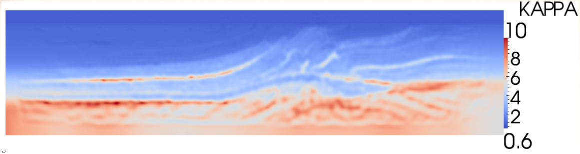


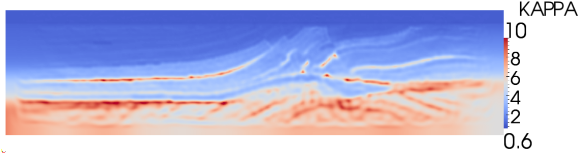


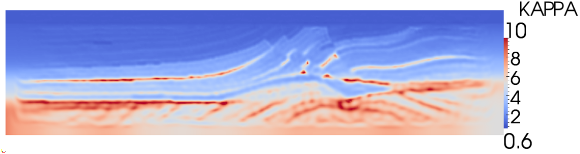
Gauss-Newton, 2 Encoded Experiments, 20, 40, 80 Iter



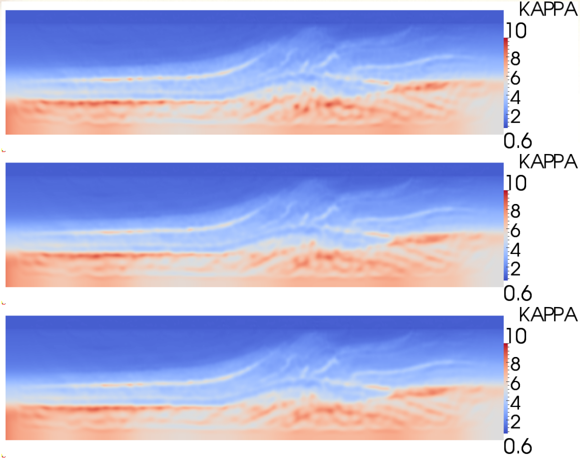
Gauss-Newton, 4 Encoded Experiments, 20, 40, 80 Iter



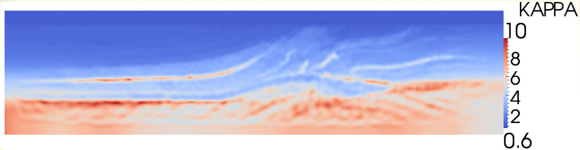


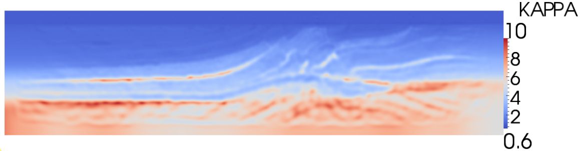


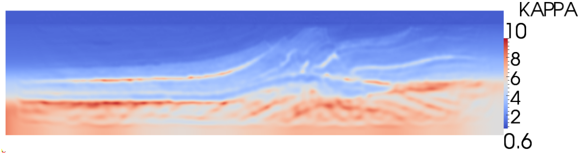
Steepest Descent, 20 Iter, 1, 2, 4 Encoded Exp.



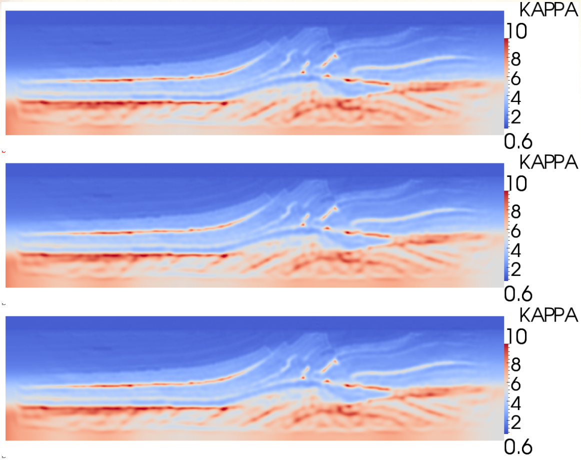
Steepest Descent, 40 Iter, 1, 2, 4 Encoded Exp.



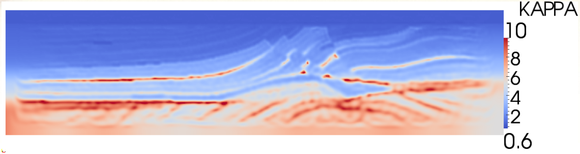


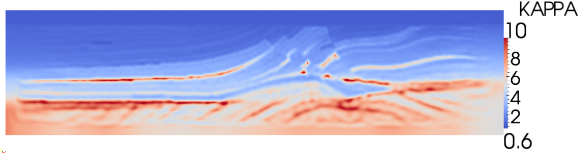


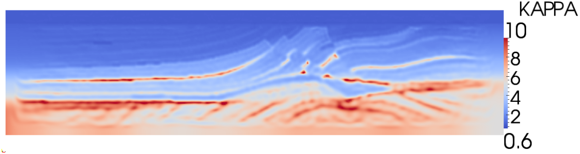
Steepest Descent, 80 Iter, 1, 2, 4 Encoded Exp.



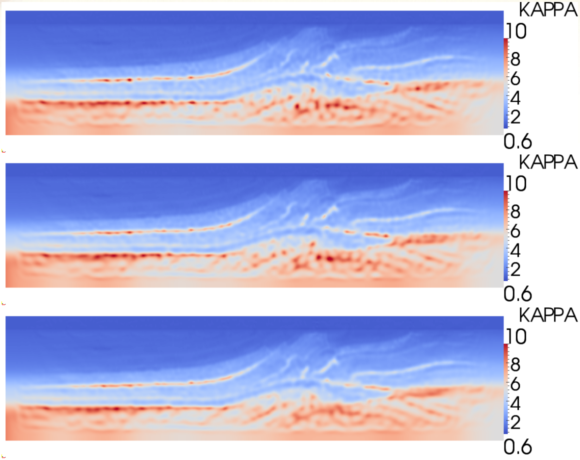
Steepest Descent, 160 Iter, 1, 2, 4 Encoded Exp.



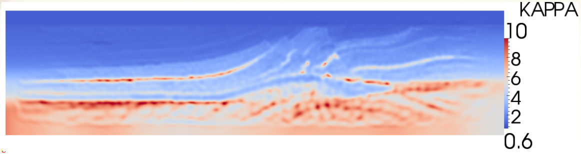


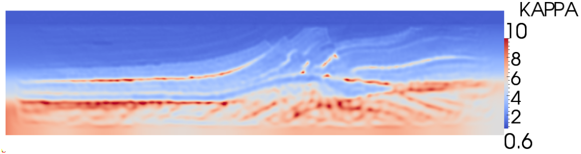


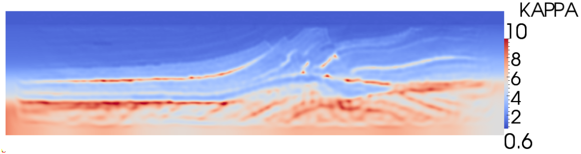
Gauss-Newton, 20 Iter, 1, 2, 4 Encoded Exp.



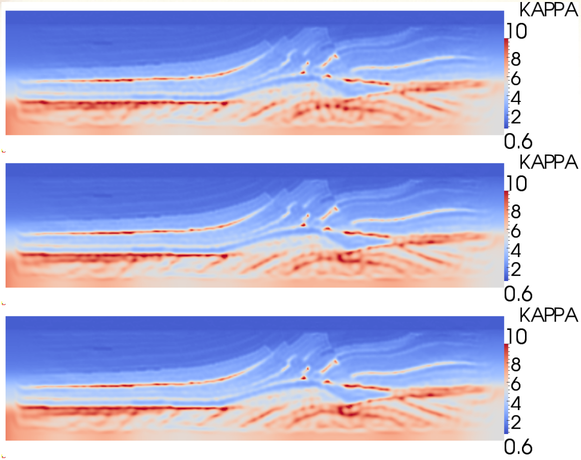
Gauss-Newton, 40 Iter, 1, 2, 4 Encoded Exp.



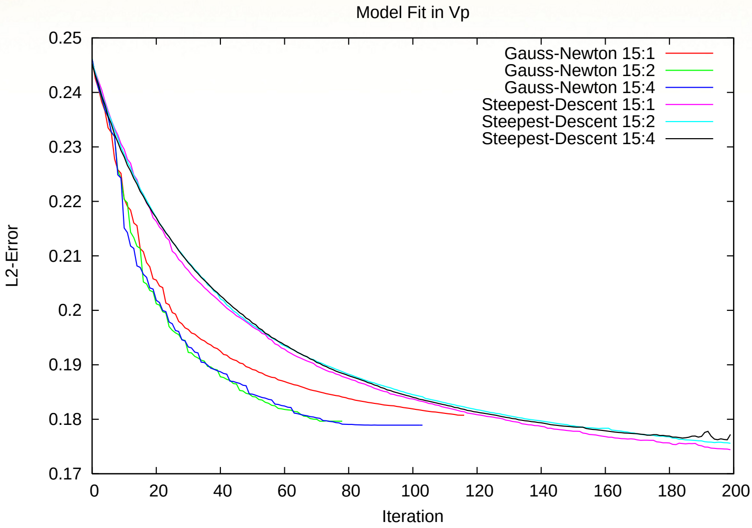




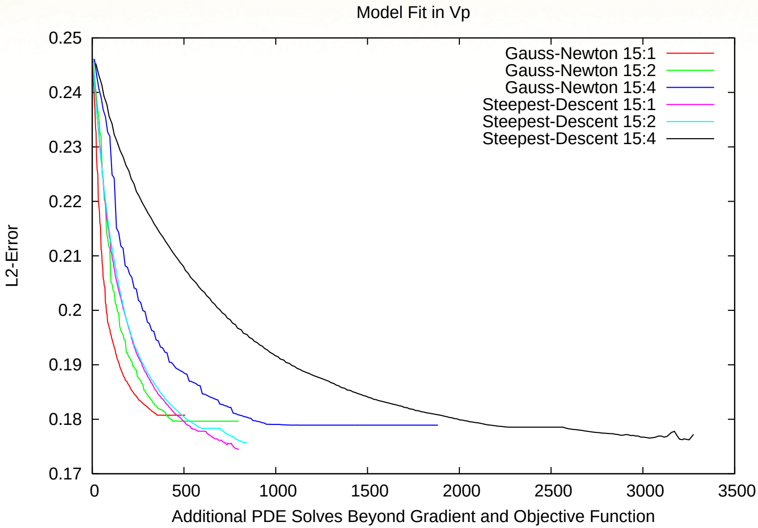
Gauss-Newton, 80 Iter, 1, 2, 4 Encoded Exp.



Model Fit in *Vp* Per Iteration



Model Fit in *Vp* Per Extra PDE Solve



Conclusion

- Random projection extends to a finite product of Hilbert spaces.
- Using random projection, we can reduce the number of PDE solves required for a reduced-space approach to parameter estimation.
- Random projection provides a mathematical foundation for random phase encoding.
- Computational results verify the method works on the acoustic wave equation.

- Future work: Experimentation with different random encoding strategies.