

SAND2011-3259C

The Application of Random Projection to Parameter Estimation

Joseph Young

Denis Ridzal

Sandia National Laboratories



May 17, 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.



Sandia
National
Laboratories





Outline

Introduction to Random Projection

Generalization to a Finite Product of Hilbert Spaces

Application to Parameter Estimation

Computational Study

What Is Random Projection?

Formally, random projection establishes conditions when the following inequality holds

$$(1 - \epsilon)\|x - y\| \leq \|\Phi(x - y)\| \leq (1 + \epsilon)\|x - y\|$$

Traditionally, $x, y \in \mathbb{R}^n$ and $\Phi \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$.

If random projection involves two inequalities, why do we call it *random projection*?

Origin of the Name *Random Projection*

- Imagine a situation where we have 5 points in $\mathbb{R}^2$ where we are interested less in the data and more in the pairwise distances between data points.



- We halve the amount of data that we need to process if we can project our data from $\mathbb{R}^2$ into $\mathbb{R}$. In order to accomplish this, we project the data onto a random plane in $\mathbb{R}^2$.



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



Two Important Theorems for Random Projection: I

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 $f : \mathbb{R}^p \rightarrow \mathbb{R}^q$ such that

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

for all $u, v \in P$.

This states that a well behaved mapping exists so that the pairwise distances between points remains bounded.

Two Important Theorems for Random Projection: II

Theorem (Achlioptas, 2001)

Let $\epsilon \in (0, 1)$ and $\beta > 0$ be given. For every set P consisting of k points in $\mathbb{R}^P$, if q is a positive integer such that

$$q > \frac{4 + 2\beta}{\epsilon^2/2 - \epsilon^3/3} \log k,$$

and $\Phi \in \mathbb{R}^{q \times P}$ is chosen where either

$$\Phi_{ij} = \begin{cases} +1 & \text{with probability } 1/2 \\ -1 & \text{" } 1/2 \end{cases} \quad \text{or} \quad \Phi_{ij} = \begin{cases} +1 & \text{with probability } 1/6 \\ 0 & \text{" } 2/3 \\ -1 & \text{" } 1/2 \end{cases},$$

then we have with probability $1 - k^{-\beta}$ that

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

for all $u, v \in P$.



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 as $Y^p = \underbrace{Y \times \cdots \times Y}_p$. Then, we define the 2-norm of y in Y^p as

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

Note, the underlying Hilbert space is general, e.g. $L^2(\Omega)$ or $H^1(\Omega)$ for some domain Ω are fine. Next, let us define a projection operator on the product space, $\Phi \in \mathcal{L}(Y^p, Y^q)$, so that

$$\Phi = \begin{bmatrix} \phi_{11}I & \cdots & \phi_{1p}I \\ \vdots & \ddots & \vdots \\ \phi_{q1}I & \cdots & \phi_{qp}I \end{bmatrix}.$$

In other words, $\Phi = \phi \otimes I$ where $\phi \in \mathbb{R}^{q \times p}$.

Generalization of Random Projection to a Product of Hilbert Spaces

Theorem (Young and Ridzal, 2011)

Let $\epsilon \in (0, 1)$ and $\beta > 0$ be given. For every set P consisting of k points in Y^P , if q is a positive integer such that

$$q > \frac{4 + 2\beta}{\epsilon^2/2 - \epsilon^3/3} \log k,$$

and $\Phi = \phi \otimes I$ where $\phi \in \mathbb{R}^{q \times P}$ is chosen where either

$$\phi_{ij} = \begin{cases} +1 & \text{with probability } 1/2 \\ -1 & \text{" } 1/2 \end{cases} \quad \text{or } \phi_{ij} = \begin{cases} +1 & \text{with probability } 1/6 \\ 0 & \text{" } 2/3 \\ -1 & \text{" } 1/2 \end{cases},$$

then we have with probability $1 - k^{-\beta}$ that

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

for all $u, v \in P$.



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 $A \in \mathcal{L}(U, \mathcal{L}(Y))$ and $B \in \mathcal{L}(Y)$ be two differential operators so that $A(u) + B$ is invertible for some set of parameters $u \in U$. For example, in Poisson's equation, we have that

$$A(u) = -\nabla \cdot (u \nabla y) \quad B = 0.$$

Then, let us define the block operator $A_p(u) \in \mathcal{L}(U, \mathcal{L}(Y^p))$ where

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

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

Application of Random Projection to Parameter Estimation

Consider the parameter estimation problem

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

Let us rewrite this in block form as

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

This is known as the full-space formulation. In the reduced-space formulation, we solve

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

Application of Random Projection to Parameter Estimation

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

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

This tells us that the objective value from the projected problem lies within a constant factor of the objective value from the unprojected problem. Further, we have that

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

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)

Let us define the unprojected and projected objective function as $J : U \rightarrow \mathbb{R}$ and $J_\Phi : U \rightarrow \mathbb{R}$, respectively, where

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

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

and Φ is chosen to be either the binary or ternary operator from above parametrized by β and ϵ . In addition, let $u_k, u_{k+1} \in U$ be iterates such that $J_\Phi(u_{k+1}) < C J_\Phi(u_k)$, where $C = (1 - \epsilon)/(1 + \epsilon)$. Then, we have that $J(u_{k+1}) < J(u_k)$ with probability $1 - 2^{-\beta}$.

This tells us that if our algorithm reduces the objective value quickly enough in the projected problem, then we achieve monotonic decrease in the unprojected objective value. This is a **conservative** estimate.



Introduction to Random Projection

Generalization to a Finite Product of Hilbert Spaces

Application to Parameter Estimation

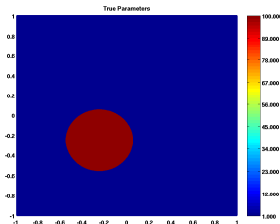
Computational Study

Application to Poisson's Equation

In the following set of examples, we solve the problem

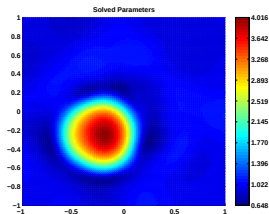
$$\arg \min_{y \in Y^p, u \in L^2(\Omega)} \left\{ \frac{1}{2} \|y - d\|^2 : -\nabla \cdot (u \nabla y_i) = b_i, i = 1, \dots, p \right\}$$

where $Y = \{y \in H^1(\Omega) : \gamma_0^{\partial\Omega} y = 0\}$ and $d, b \in L^2(\Omega)^p$ are both given. We generate the data based on a true set of parameters,

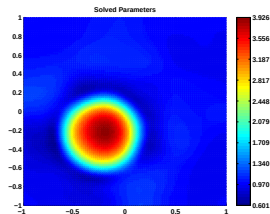


As sources, we use Gaussians with variance .001 and amplitude 10 spaced evenly along unit square centered at zero. Finally, we solve the PDE using a Galerkin finite element method with piecewise linear elements.

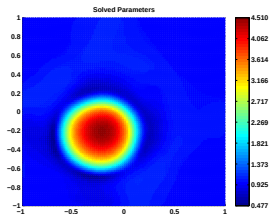
Steepest Descent, 50 Iterations, 20 Sources



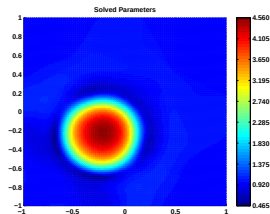
20:1



20:2

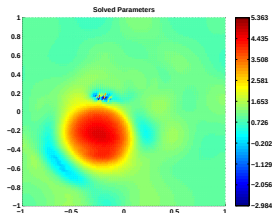


20:5

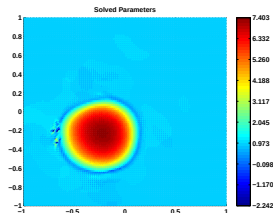


None

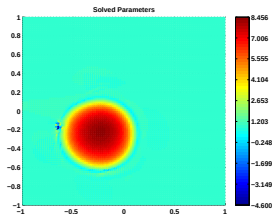
Gauss-Newton, 50 Iterations, 20 Sources



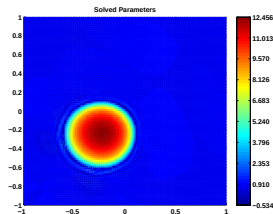
20:1



20:2

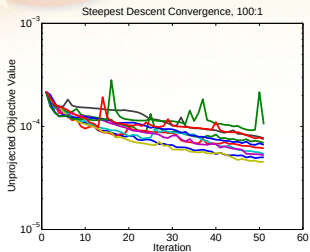


20:5

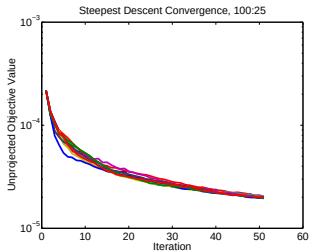


None

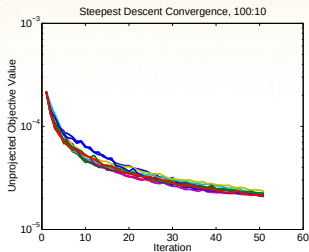
Steepest Descent Convergence, Binary Encoding



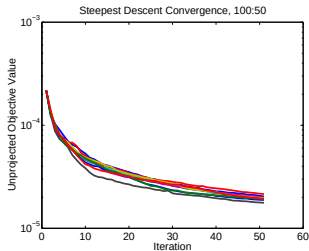
100:1



100:25

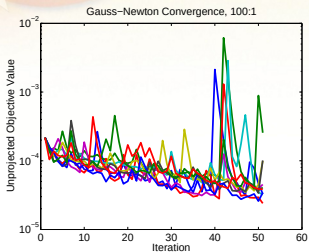


100:10

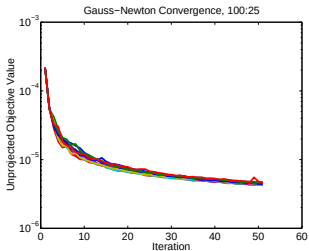


100:50

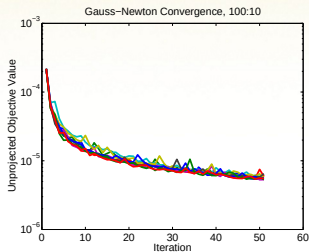
Gauss-Newton Convergence, Binary Encoding



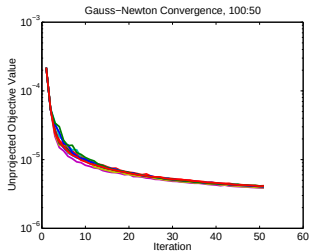
100:1



100:25

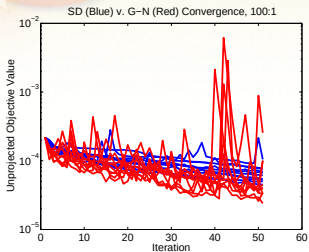


100:10

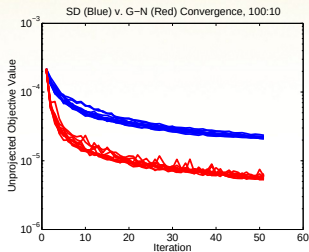


100:50

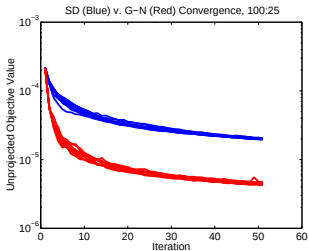
SD v. GN, Binary Encoding



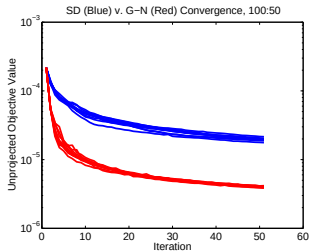
100:1



100:10

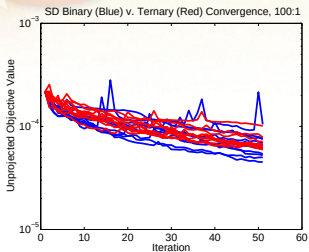


100:25

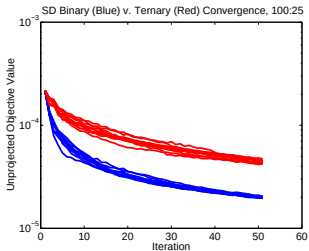


100:50

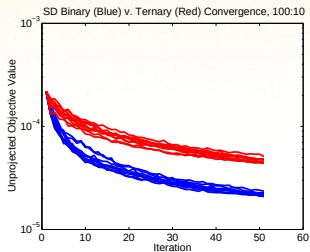
Steepest Descent, Binary v. Ternary Encoding



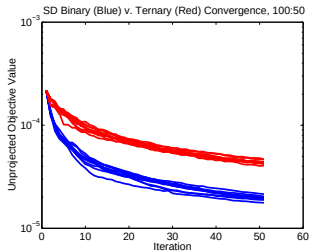
100:1



100:25

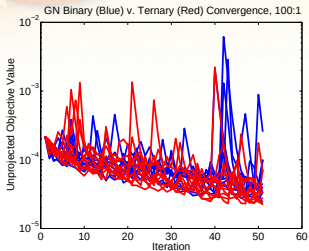


100:10

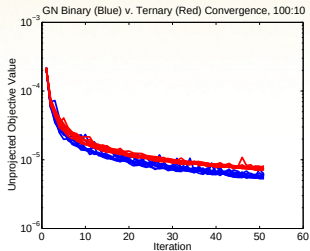


100:50

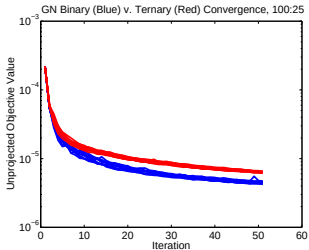
Gauss-Newton, Binary v. Ternary Encoding



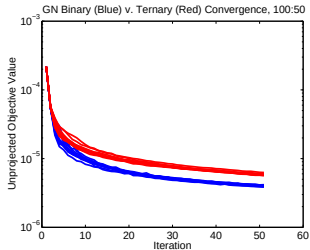
100:1



100:10

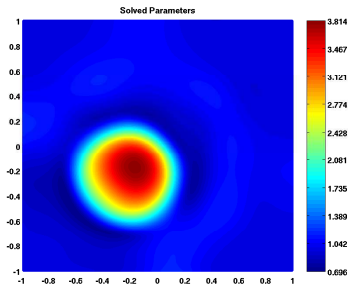


100:25

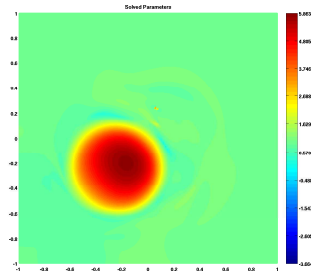


100:50

Medium Scale Study, SD v. G-N, 500000 Elements



Encoding 10:1
Iterations 50
Time 500s



Encoding 10:1
Iterations 30
Time 2000s

Comment: As we have seen before, Gauss-Newton violates the physical bounds more frequently than steepest descent. However, it matches the peak parameter value more closely.

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.
- Computational results verify the method works on a model problem.

