SANDIA REPORT

SAND2013-4766
Unlimited Release
Printed June, 2013

Designing Experiments Through Compressed Sensing

Joseph Young, Denis Ridzal

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

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

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone:	(865) 576-8401
Facsimile:	(865) 576-5728
E-Mail:	reports@adonis.osti.gov
Online ordering:	http://www.osti.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd
Springfield, VA 22161

Telephone:	(800) 553-6847
Facsimile:	(703) 605-6900
E-Mail:	orders@ntis.fedworld.gov
Online ordering:	http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2013-4766
Unlimited Release
Printed June, 2013

Designing Experiments Through Compressed Sensing

Joseph Young
Numerical Analysis and Applications
Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185-1320

Denis Ridzal
Optimization and Uncertainty Quantification
Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185-1320

Abstract

In the following paper, we discuss how to design an ensemble of experiments through the use of compressed sensing. Specifically, we show how to conduct a small number of *physical* experiments and then use compressed sensing to reconstruct a larger set of data. In order to accomplish this, we organize our results into four sections. We begin by extending the theory of compressed sensing to a finite product of Hilbert spaces. Then, we show how these results apply to experiment design. Next, we develop an efficient reconstruction algorithm that allows us to reconstruct experimental data projected onto a finite element basis. Finally, we verify our approach with two computational experiments.

1 Introduction

Traditional compressed sensing studies when the following equivalence holds

$$\{\bar{x}\} = \arg\min_{x\in\mathbb{R}^p}\{\|x\|_1 : \Phi x = \Phi\bar{x}\}.$$

Here, $\bar{x} \in \mathbb{R}^p$, $\|x\|_1 = \sum_{i=1}^p |x_i|$, and $\Phi \in \mathbb{R}^{q\times p}$ where $q < p$. Note the vector $\bar{x}$ on both sides of the equality. Hence, compressed sensing studies when we can encode a vector with a sampling operator Φ in such a way that we can exactly reconstruct the element through the solution of an optimization problem. Although simple to state, the above equivalence has important implications in signal processing and other fields. Notably, Candes, Romberg, and Tao show that a signal can be sampled below the Nyquist rate, yet exactly recovered, as long as the signal is sparse in some basis and we sample the signal in a special way [2].

In the following presentation, we generalize this idea to a finite product of Hilbert spaces and consider when the related equivalence holds

$$\{\bar{x}\} = \arg\min_{x\in S^p}\{\|x\|_1 : (\Phi\otimes I)x = (\Phi\otimes I)\bar{x}\}.$$

In this case, S denotes an arbitrary Hilbert space with inner product $\langle\cdot,\cdot\rangle$ and norm $\|\cdot\|$, $S^p = S\times\cdots\times S$ designates the p-times Cartesian product of S, $\|x\|_1 = \sum_{i=1}^m \|x_i\|$, $\Phi \in \mathbb{R}^{q\times p}$ where $q < p$, and we define the Kronecker-like product $\Phi\otimes I \in \mathscr{L}(S^p, S^q)$ as

$$(\Phi\otimes I)x = \begin{bmatrix}\Phi_{11} & \dots & \Phi_{1p}\\ \vdots & \ddots & \vdots\\ \Phi_{q1} & \dots & \Phi_{qp}\end{bmatrix}\begin{bmatrix}x_1\\ \vdots\\ x_p\end{bmatrix} = \begin{bmatrix}\sum_{i=1}^p \Phi_{1i}x_i\\ \vdots\\ \sum_{i=1}^p \Phi_{qi}x_i\end{bmatrix}, \tag{1}$$

where $\mathscr{L}(S^p, S^q)$ denotes the set of bounded linear operators that map S^p to S^q.

This generalization is important for two reasons. First, it allows us to generalize compressed sensing to spaces of functions, which has implications in experiment design, parameter estimation, and PDE constrained optimization. Second, it provides a simple way to apply compressed sensing to problems that contain group sparsity. Group sparsity occurs when data can be divided into different groups and the majority of groups contain zero data. Although Turlach, Venables, and Wright previously approached group sparsity using LASSO [6], our theory provides a simple, unified framework.

We organize our presentation in the following manner. Motivated by our recent results in [9] and that of Rauhut in [5], we generalize compressed sensing to a finite product of Hilbert spaces. Then, we show how this generalization can be used to reduce the number of physical experiments required in order to generate a large collection of data. In order to reconstruct this data, we develop an algorithm based on a version of the alternating direction method of multipliers developed by Deng, Yin, and Zhang [3]. Next, using these tools, we conduct two computational experiments. In the first, we show that our methods and algorithms do extend to function spaces by encoding and reconstructing a collection of letters represented by piecewise linear functions. In the second, we simulate a sequence of physical experiments where the underlying physics are well-modeled by Poisson's equation. Then, we show that we can use compressed sensing to reduce the number of physical experiments required. Finally, we summarize our results and provide a final discussion.

2 Generalization

In the following section, we generalize compressed sensing to a finite product of Hilbert spaces. This consists of four steps. First, we discuss the *null space property* of linear operators and show that this property is necessary and sufficient for compressed sensing to occur. Second, we examine the *restricted isometry property* and its various properties. Using these properties, we show operators that possess the restricted isometry property also possess the null space property. Third, we show that if a matrix possesses the restricted isometry property, then the Kronecker-like product between the matrix and the identity operator also possesses the restricted isometry property. Finally, we explain how we do not require sparsity *per se* for compressed sensing to occur, but rather sparsity in some basis.

We begin with the definition of the null space property.

Definition 1 (Null Space Property)**.** *A linear operator $\widetilde{\Phi} \in \mathscr{L}(S^p, S^q)$ satisfies the null space property of order k when for all index sets $\alpha \subseteq \{1, \dots, p\}$ with $|\alpha| = k$ it holds*

$$\|v_\alpha\|_1 < \|v_{\alpha^c}\|_1$$

for all $v \in \mathcal{N}(\widetilde{\Phi})\backslash\{0\}$.

In the above definition, $v_\alpha \in S^p$ denotes the projection $I_\alpha v$ where $I_\alpha \in \mathscr{L}(S^p, S^p)$ and

$$[I_\alpha x]_i = \begin{cases} x_i & i \in \alpha \\ 0 & i \notin \alpha. \end{cases}$$

For future reference, we define $\widetilde{\Phi}_\alpha = \widetilde{\Phi} I_\alpha$. In addition, $\alpha^c = \{1, \dots, p\}\backslash\alpha$ denotes the complement of α and $\mathcal{N}(\widetilde{\Phi}) = \{x \in S^p : \widetilde{\Phi}(x) = 0\}$ denotes the null space of $\widetilde{\Phi}$. Next, for $x \in S^p$, we define the support as $\text{supp}(x) = \{i \in \{1, \dots, p\} : \|x_i\| \neq 0\}$. Finally, for a set α, we use the notation $|\alpha|$ to denote the cardinality of the set α.

Then, using the null space property, we prove our primary compressed sensing theorem.

Theorem 1 (Compressed Sensing)**.** *Let $\widetilde{\Phi} \in \mathscr{L}(S^p, S^q)$. Then, every k-sparse element $\bar{x} \in S^p$ satisfies*

$$\{\bar{x}\} = \arg\min_{x \in S^p} \{\|x\|_1 : \widetilde{\Phi}x = \widetilde{\Phi}\bar{x}\}$$

if and only if Φ possesses the null space property of order k.

Proof. The following proof mirrors that from Rauhut [5].

In the forward direction, we assume that for any k-sparse $\bar{x} \in S^p$ we have

$$\{\bar{x}\} = \arg\min_{x \in S^p} \{\|x\|_1 : \widetilde{\Phi}x = \widetilde{\Phi}\bar{x}\}.$$

Then, for any $v \in \mathcal{N}(\Phi)\backslash\{0\}$ and any index set $\alpha \subseteq \{1, \dots, p\}$ such that $|\alpha| = k$, we have that

$$\{v_\alpha\} = \arg\min_{x \in S^p} \{\|x\|_1 : \widetilde{\Phi}x = \widetilde{\Phi}v_\alpha\}. \tag{2}$$

Since $v \in \mathcal{N}(\widetilde{\Phi})\backslash\{0\}$, we see that $0 = \Phi v = \Phi(v_\alpha + v_{\alpha^c})$, which implies $\Phi(-v_{\alpha^c}) = \Phi v_\alpha$ where $-v_{\alpha^c} \neq v_\alpha$. Hence, v_{α^c} is a feasible point to the optimization problem $\arg\min_{x \in S^p} \{\|x\|_1 : \widetilde{\Phi}x = \widetilde{\Phi}v_\alpha\}$. However, in Equation 2, we see that v_α is the *unique* optimal solution, which means that $\|v_\alpha\|_1 < \|x\|_1$ for all $x \in S^p$ such that $\tilde{\Phi}x = \tilde{\Phi}v_\alpha$. Therefore, we have that $\|-v_{\alpha^c}\|_1 = \|v_{\alpha^c}\|_1 > \|v_\alpha\|_1$, which proves the null space property of order k for $\widetilde{\Phi}$.

In the reverse direction, we assume that $\widetilde{\Phi}$ possesses the null space property of order k. Given a k-sparse vector $\bar{x} \in S^p$ and an arbitrary vector $x \in S^p$ such that $\widetilde{\Phi}x = \widetilde{\Phi}\bar{x}$ and $x \neq \bar{x}$, we define $v = \bar{x} - x$. Since $\widetilde{\Phi}x = \widetilde{\Phi}\bar{x}$ and $x \neq \bar{x}$, we see that $v \in \mathcal{N}(\widetilde{\Phi})\backslash\{0\}$. Then, for $\alpha = \text{supp}(\bar{x})$, we use the null space property of $\widetilde{\Phi}$ and notice that

$$\begin{aligned}\|\bar{x}\|_1 &= \|\bar{x} - x_\alpha + x_\alpha\|_1 \leq \|\bar{x} - x_\alpha\|_1 + \|x_\alpha\|_1 = \|\bar{x}_\alpha - x_\alpha\|_1 + \|x_\alpha\|_1 = \|v_\alpha\|_1 + \|x_\alpha\|_1 \\ &< \|v_{\alpha^c}\|_1 + \|x_\alpha\|_1 = \|-x_{\alpha^c}\|_1 + \|x_\alpha\|_1 = \|x\|_1.\end{aligned}$$

Therefore, $\bar{x}$ has a norm strictly lower than all other points and must be the unique minimizer to $\arg\min_{x\in S^p}\{\|x\|_1 : \widetilde{\Phi}x = \widetilde{\Phi}\bar{x}\}$. □

In many applications, an element $\bar{x} \in \mathcal{S}^p$ may not be inherently sparse in the product space, but a linear transformation of the element may induce sparsity. In this situation, we can still recover $\bar{x}$ with a slightly different equivalence.

Corollary 1 (Compressed Sensing with Change of Basis)**.** *Let $\widetilde{\Phi} \in \mathscr{L}(S^p, S^q)$. Let $\Psi \in \mathscr{L}(S^p, S^q)$ be invertible. Then, every k-sparse element $\Psi^{-1}\bar{x} \in S^p$ satisfies*

$$\{\bar{x}\} = \arg\min_{x\in S^p}\{\|\Psi^{-1}x\|_1 : \widetilde{\Phi}x = \widetilde{\Phi}\bar{x}\}$$

if and only if $\widetilde{\Phi}$ possesses the null space property of order k.

Proof. In the forward direction, we assume that for k-sparse $\Psi^{-1}\bar{x}$, we have

$$\{\bar{x}\} = \arg\min_{x\in S^p}\{\|\Psi^{-1}x\|_1 : \widetilde{\Phi}x = \widetilde{\Phi}\bar{x}\}.$$

Let $\bar{y} = \Psi^{-1}\bar{x}$ and $\Theta = \widetilde{\Phi}\Psi$. Then, we notice that

$$\begin{aligned}\{\bar{y}\} &= \Psi^{-1}\{\bar{x}\} \\ &= \Psi^{-1}\arg\min\{\|\Psi^{-1}x\|_1 : \widetilde{\Phi}x = \widetilde{\Phi}\bar{x}, x \in \Omega\} \\ &= \Psi^{-1}\arg\min\{\|\Psi^{-1}x\|_1 : \widetilde{\Phi}\Psi\Psi^{-1}x = \widetilde{\Phi}\bar{x}\} \\ &= \arg\min\{\|y\|_1 : \widetilde{\Phi}\Psi y = \widetilde{\Phi}\bar{x}\} \\ &= \arg\min\{\|y\|_1 : \widetilde{\Phi}\Psi y = \widetilde{\Phi}\Psi\bar{y}\} \\ &= \arg\min\{\|y\|_1 : \Theta y = \Theta\bar{y}\}.\end{aligned}$$

This states that for all k-sparse $\bar{y} \in S^p$, $\{\bar{y}\} = \arg\min\{\|y\|_1 : \Theta y = \Theta\bar{y}\}$. Hence, from Theorem 1, Θ possesses the null space property of order k. However, since Ψ is invertible,

$$\mathcal{N}(\Theta) = \mathcal{N}(\widetilde{\Phi}\Psi) = \mathcal{N}(\widetilde{\Phi}). \tag{3}$$

Therefore, $\widetilde{\Phi}$ possesses the null space property of order k.

In the reverse direction, we assume that $\widetilde{\Phi}$ possesses the null space property of order k. Since Ψ is invertible, Equation (3) shows that Θ, where $\Theta = \widetilde{\Phi}\Psi$, also possesses the nullspace property of order k. Therefore, from Theorem 1, we have that

$$\{\bar{y}\} = \arg\min\{\|y\|_1 : \Theta y = \Theta\bar{y}\}.$$

Using the same sequence of equalities from the forward direction, we have that

$$\{\bar{x}\} = \arg\min_{x\in S^p}\{\|\Psi^{-1}x\|_1 : \widetilde{\Phi}x = \widetilde{\Phi}\bar{x}\}.$$

□

Due to difficulty in proving that an arbitrary operator possesses the nullspace property, we use a related property called the restricted isometry property.

Definition 2 (Restricted Isometry Constant)**.** *For $\widetilde{\Phi} \in \mathscr{L}(S^p, S^q)$, the value*

$$\delta_k = \inf_{x \in S^p} \{\delta \in \mathbb{R} : (1-\delta)\|x\|_2^2 \leq \|\widetilde{\Phi}x\|_2^2 \leq (1+\delta)\|x\|_2^2, |\text{supp}(x)| = k\}$$

denotes the restricted isometry constant of order k.

In the above definition, $\|x\|_2 = \sqrt{\langle x, x\rangle}$ where $\langle x, y\rangle = \sum_{i=1}^p \langle x_i, y_i\rangle$. Using this definition, we say that a matrix possesses the restricted isometry property of order k when $\delta_k \in (0,1)$.

The restricted isometry constants possess several properties.

Theorem 2 (Restricted Isometry Constants)**.** *Let $\widetilde{\Phi} \in \mathscr{L}(S^p, S^q)$ have restricted isometry constants δ_k. Then,*

(a) The restricted isometry constants are ordered, $\delta_1 \leq \delta_2 \leq \delta_3 \leq \dots$.

(b) The restricted isometry constants can be redefined as

$$\delta_k = \max_{\alpha \subseteq \{1,\dots,p\}} \left\{ \|\widetilde{\Phi}_\alpha^* \widetilde{\Phi}_\alpha - I_\alpha\|_2 : |\alpha| = k \right\}.$$

(c) Let $x, y \in S^p$ be such that $\text{supp}(x) \cap \text{supp}(y) = \emptyset$ *and define* $k = |\text{supp}(x)| + |\text{supp}(y)|$. *Then,*

$$|\langle \widetilde{\Phi}x, \widetilde{\Phi}y\rangle| \leq \delta_k \|x\|_2 \|y\|_2.$$

Proof. The following proof mirrors that from Rauhut [5].

Since k-sparse elements are also $k+1$-sparse, part (a) immediately follows.

For part (b), notice that we can rewrite our definition of a restricted isometry constant as

$$\begin{aligned}
\delta_k &= \inf_{x \in S^p} \{\delta \in \mathbb{R} : (1-\delta)\|x\|_2^2 \leq \|\widetilde{\Phi}x\|_2^2 \leq (1+\delta)\|x\|_2^2, |\text{supp}(x)| = k\} \\
&= \inf_{x \in S^p} \{\delta \in \mathbb{R} : |\|\widetilde{\Phi}x\|_2^2 - \|x\|_2^2| \leq \delta\|x\|_2^2, |\text{supp}(x)| = k\} \\
&= \inf_{x \in S^p} \{\delta \in \mathbb{R} : |\|\widetilde{\Phi}_\alpha x\|_2^2 - \|I_\alpha x\|_2^2| \leq \delta\|I_\alpha x\|_2^2, \alpha \subseteq \{1,\dots,p\}, |\alpha| = k\} \\
&= \inf_{x \in S^p} \{\delta \in \mathbb{R} : |\|\widetilde{\Phi}_\alpha x\|_2^2 - \|I_\alpha x\|_2^2| \leq \delta\|x\|_2^2, \alpha \subseteq \{1,\dots,p\}, |\alpha| = k\} \qquad (4)
\end{aligned}$$

where the last equality follows since $\|I_\alpha x\|_2 \leq \|x\|_2$ and for all $x \in S^p$ there exists $y \in S^p$ such that $I_\alpha x = y$. Then, we see that

$$\begin{aligned}
\delta_k =& \text{Equation (4)} \\
=& \inf_{x \in S^p} \{\delta \in \mathbb{R} : |\langle \widetilde{\Phi}_\alpha x, \widetilde{\Phi}_\alpha x\rangle - \langle I_\alpha x, I_\alpha x\rangle| \leq \delta\langle x, x\rangle, \alpha \subseteq \{1,\dots,p\}, |\alpha| = k\} \\
=& \inf_{x \in S^p} \{\delta \in \mathbb{R} : |\langle \widetilde{\Phi}_\alpha x, \widetilde{\Phi}_\alpha x\rangle - \langle I_\alpha^* I_\alpha x, x\rangle| \leq \delta\langle x, x\rangle, \alpha \subseteq \{1,\dots,p\}, |\alpha| = k\} \\
=& \inf_{x \in S^p} \{\delta \in \mathbb{R} : |\langle \widetilde{\Phi}_\alpha x, \widetilde{\Phi}_\alpha x\rangle - \langle I_\alpha x, x\rangle| \leq \delta\langle x, x\rangle, \alpha \subseteq \{1,\dots,p\}, |\alpha| = k\} \\
=& \inf_{x \in S^p} \left\{\delta \in \mathbb{R} : \left| \frac{\langle (\widetilde{\Phi}_\alpha^* \widetilde{\Phi}_\alpha - I_\alpha)x, x\rangle}{\langle x, x\rangle} \right| \leq \delta, \alpha \subseteq \{1,\dots,p\}, |\alpha| = k \right\} \\
=& \max_{\alpha \subseteq \{1,\dots,p\}} \left\{ \inf_{x \in S^p} \left\{ \delta \in \mathbb{R} : \left| \frac{\langle (\widetilde{\Phi}_\alpha^* \widetilde{\Phi}_\alpha - I_\alpha)x, x\rangle}{\langle x, x\rangle} \right| \leq \delta \right\} : |\alpha| = k \right\} \\
=& \max_{\alpha \subseteq \{1,\dots,p\}} \left\{ \|\widetilde{\Phi}_\alpha^* \widetilde{\Phi}_\alpha - I_\alpha\|_2 : |\alpha| = k \right\}.
\end{aligned}$$

In part (c), let $x, y \in S^p$ where $\alpha = \text{supp}(x)$, $\beta = \text{supp}(y)$, and $\alpha \cap \beta = \emptyset$. Due to the disjointness of their supports, we see that $\langle x, y\rangle = 0$. Keeping this in mind, we have that

$$
\begin{aligned}
|\langle \widetilde{\Phi}x, \widetilde{\Phi}y\rangle| &= |\langle \widetilde{\Phi}_{\alpha\cup\beta}x, \widetilde{\Phi}_{\alpha\cup\beta}y\rangle| \\
&= |\langle \widetilde{\Phi}_{\alpha\cup\beta}x, \widetilde{\Phi}_{\alpha\cup\beta}y\rangle - \langle I_{\alpha\cup\beta}x, I_{\alpha\cup\beta}y\rangle| \\
&= |\langle \widetilde{\Phi}_{\alpha\cup\beta}x, \widetilde{\Phi}_{\alpha\cup\beta}y\rangle - \langle I^*_{\alpha\cup\beta}I_{\alpha\cup\beta}x, y\rangle| \\
&= |\langle \widetilde{\Phi}^*_{\alpha\cup\beta}\widetilde{\Phi}_{\alpha\cup\beta}x, y\rangle - \langle I_{\alpha\cup\beta}x, y\rangle| \\
&= |\langle (\widetilde{\Phi}^*_{\alpha\cup\beta}\widetilde{\Phi}_{\alpha\cup\beta} - I_{\alpha\cup\beta})x, y\rangle| \\
&\leq \|\widetilde{\Phi}^*_{\alpha\cup\beta}\widetilde{\Phi}_{\alpha\cup\beta} - I_{\alpha\cup\beta}\|_2\|x\|_2\|y\|_2 \\
&\leq \delta_k\|x\|_2\|y\|_2.
\end{aligned}
$$

□

Now, we connect the restricted isometry property with the null space property with the following theorem

Theorem 3 (Restricted Isometry Implies Null Space)**.** *If the restricted isometry constant of an operator $\Phi \in \mathscr{L}(S^p, S^q)$ satisfies*

$$\delta_{2k} < \frac{1}{3},$$

then Φ satisfies the null space property of order k.

Proof. The proof mirrors that of Rauhut [5].

Let $v \in \mathcal{N}(\Phi)\backslash\{0\}$ and $\alpha_i \subseteq \{1, \dots, p\}$ a collection of index sets that partition $\{1, \dots, p\}$ so that v_{α_0} contains k elements of largest norm, v_{α_1} contains the next k largest elements, and so on. In other words,

$$
\begin{aligned}
\alpha_i &= \arg\max_{\alpha\subseteq\beta_i}\{\|v_\alpha\|_1 : |\alpha| = \min\{k, |\beta_i|\}\} \\
\beta_0 &= \{1, \dots, p\} \\
\beta_i &= \{1, \dots, p\}\backslash\bigcup_{j=0}^{i-1}\alpha_j.
\end{aligned}
$$

Note, the $\min\{k, |\beta_i|\}$ term insures that we grab k elements at a time until we have fewer than k elements. Then, we take the remaining elements. In this way, $v = \sum_{i\geq 0} v_{\alpha_i}$ and since $v \in \mathcal{N}(\Phi)$, $\Phi v_{\alpha_0} = \Phi(\sum_{i\geq 1} -v_{\alpha_i})$.

Next, since Φ possesses the restricted isometry property, we have that

$$
\begin{aligned}
\|v_{\alpha_0}\|_2^2 &\leq \frac{1}{1-\delta_{2k}}\|\Phi v_{\alpha_0}\|_2^2 \\
&= \frac{1}{1-\delta_{2k}}\langle \Phi v_{\alpha_0}, \Phi v_{\alpha_0}\rangle \\
&= \frac{1}{1-\delta_{2k}}\left\langle \Phi v_{\alpha_0}, \Phi\left(\textstyle\sum_{i\geq 1} -v_{\alpha_i}\right)\right\rangle \\
&= \frac{1}{1-\delta_{2k}}\sum_{i\geq 1}\langle \Phi v_{\alpha_0}, \Phi(-v_{\alpha_i})\rangle \\
&\leq \frac{\delta_{2k}}{1-\delta_{2k}}\sum_{i\geq 1}\|v_{\alpha_0}\|_2\| - v_{\alpha_i}\|_2
\end{aligned}
$$

where the last line results from Theorem 2(c). Therefore,

$$\|v_{\alpha_0}\|_2 \leq \frac{\delta_{2k}}{1-\delta_{2k}}\sum_{i\geq 1}\|v_{\alpha_i}\|_2. \tag{5}$$

Now, we must bound $\|v_{\alpha_i}\|_2$ by $\|v_{\alpha_{i-1}}\|_1$. Since we partition the index set $\{1,\dots,p\}$ so that $v_{\alpha_{i-1}}$ contains elements with larger norm than v_{α_i}, we have that for all $j \in \alpha_i$, $\|v_j\| \leq \frac{1}{k}\|v_{\alpha_{i-1}}\|_1$. In other words, every element in v_{α_i} has smaller norm than the average norm of the elements in $v_{\alpha_{i-1}}$. Then, we notice that

$$\begin{aligned}
\|v_{\alpha_i}\|_2 =& \sqrt{\sum_{j\in\alpha_i}\|v_j\|^2} \leq \sqrt{\sum_{j\in\alpha_i}\left(\frac{1}{k}\|v_{\alpha_{i-1}}\|_1\right)^2}\\
\leq& \sqrt{k\left(\frac{1}{k}\|v_{\alpha_{i-1}}\|_1\right)^2} = \frac{1}{\sqrt{k}}\|v_{\alpha_{i-1}}\|_1.
\end{aligned}$$

Therefore,

$$\|v_{\alpha_i}\|_2 \leq \frac{1}{\sqrt{k}}\|v_{\alpha_{i-1}}\|_1. \tag{6}$$

In addition, we must bound $\|v_\alpha\|_1$ by $\|v_\alpha\|_2$ for an arbitrary index set α. For any index set $\alpha \subseteq \{1,\dots,p\}$, we see that

$$\left(\frac{1}{|\alpha|}\|v_\alpha\|_1\right)^2 = \left(\frac{1}{|\alpha|}\sum_{j\in\alpha}\|v_j\|\right)^2 \leq \frac{1}{|\alpha|}\sum_{j\in\alpha}\|v_j\|^2 = \frac{1}{|\alpha|}\|v_\alpha\|^2$$

where the inequality results from Jensen's inequality. Hence,

$$\|v_\alpha\|_1 \leq \sqrt{|\alpha|}\|v_\alpha\|_2. \tag{7}$$

Putting everything together, for an arbitrary index set $\alpha \subseteq \{1,\dots,p\}$ such that $|\alpha| = k$, we see that

$$\begin{aligned}
\|v_\alpha\|_1 \leq& \|v_{\alpha_0}\|_1 && \alpha_0 \text{ contains indices with largest norm}\\
\leq& \sqrt{k}\|v_{\alpha_0}\|_2 && \text{Inequality (7)}\\
\leq& \sqrt{k}\left(\tfrac{\delta_{2k}}{1-\delta_{2k}}\right)\sum_{i\geq 1}\|v_{\alpha_i}\|_2 && \text{Inequality (5)}\\
\leq& \left(\tfrac{\delta_{2k}}{1-\delta_{2k}}\right)\sum_{i\geq 1}\|v_{\alpha_{i-1}}\|_1 && \text{Inequality (6)}\\
\leq& \left(\tfrac{\delta_{2k}}{1-\delta_{2k}}\right)\|v\|_1 && \text{Summing over } i\\
<& \tfrac{1}{2}\|v\|_1 && \text{Assumption that } \delta_{2k} < \tfrac{1}{3}\\
=& \tfrac{1}{2}(\|v_\alpha\|_1 + \|v_{\alpha^c}\|)
\end{aligned}$$

Therefore,

$$\|v_\alpha\|_1 < \|v_{\alpha^c}\|_1,$$

which proves the null space property of $\widetilde{\Phi}$ of order k. □

At this point, we turn our attention to the so called *sensing* or *encoding* operator $\widetilde{\Phi} \in \mathscr{L}(S^p, S^q)$. In traditional compressed sensing, a large amount of work has been done in showing that random encoding matrices possess the restricted isometry property with high probability. For example, Candes, Romberg, and Tao [2] as well as Donoho [4] show that matrices with elements that are normally distributed with mean zero and variance one have restricted isometry property with high probability. Baraniuk, Davenport, DeVore, and Wakin prove a similar result for binary encoding operators [1]. Rather than duplicate these proofs, we show that the operator $\Phi \otimes I \in \mathscr{L}(S^p, S^q)$ possesses the restricted isometry property whenever $\Phi \in \mathbb{R}^{q\times p}$ does. In order to accomplish this, we require two intermediate results.

We begin by showing that the norm of $\Phi \otimes I$ is bounded by Φ. Notationally, let $\|\Phi \otimes I\|_2$ denote the operator norm of $\Phi \otimes I$ corresponding to the 2-norm in the product space Y^q and $\|\Phi\|_2$ denote the operator norm of Φ corresponding to the 2-norm in Euclidean space. Then, we note that

Lemma 1 (Norm Bound of Kronecker-like Product)**.** *For $\Phi \in \mathbb{R}^{p\times q}$, $I \in \mathscr{L}(Y)$, and $\Phi \otimes I \in \mathscr{L}(Y^p, Y^q)$, we have that*

$$\|\Phi \otimes I\|_2 \leq \|\Phi\|_2. \tag{8}$$

Proof. The proof is completed as Lemma 2.1 of [9]. □

Next, we recall a result that states that the geometry in $\mathbb{R}^p$ differs only slightly from the geometry in Y^p.

Lemma 2 (Geometry of a Product Space)**.** *Given $\Phi \in \mathbb{R}^{q\times p}$ where $p > q$ and some $y \in Y^p$, there exists an $x \in \mathbb{R}^p$ such that*

$$\|x\|_2 = \|y\|_2 \text{ and } \|\Phi x\|_2 = \|(\Phi \otimes I)y\|_2.$$

and $\operatorname{supp}(x) = \operatorname{supp}(y)$*.*

Proof. The proof is completed as Theorem 2.2 of [9]. □

Theorem 4 (Restricted Isometry of Kronecker-like Products)**.** *Let $\Phi \in \mathbb{R}^{p\times q}$ possess restricted isometry constants δ_k and $\Phi \otimes I \in \mathscr{L}(S^p, S^q)$ possess restricted isometry constants $\hat{\delta}_k$. Then, $\hat{\delta}_k \leq \delta_k$.*

Proof. By definition, we have that

$$\begin{aligned}\hat{\delta}_k &= \inf\{\delta \in \mathbb{R} : (1-\delta)\|y\|_2^2 \leq \|(\Phi \otimes I)y\|_2^2 \leq (1+\delta)\|y\|_2^2, y \in S^p, |\operatorname{supp}(y)| = k\} \\ &= \inf\{\delta \in \mathbb{R} : (1-\delta)\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1+\delta)\|x\|_2^2, x \in \mathbb{R}^p, \\ &\qquad \|x\|_2 = \|y\|_2, \|\Phi x\|_2 = \|\Phi y\|_2, \operatorname{supp}(x) = \operatorname{supp}(y), y \in S^p, |\operatorname{supp}(x)| = k\}\end{aligned} \tag{9}$$

where the last equality arises from Theorem 2. Next, we eliminate the references to y, which increases the number of inequality constraints on x and, hence, increases the infimum. This gives

$$\begin{aligned}\hat{\delta}_k &= \text{Equation (9)} \\ &\leq \inf\{\delta \in \mathbb{R} : (1-\delta)\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1+\delta)\|x\|_2^2, x \in \mathbb{R}^p, |\operatorname{supp}(x)| = k\} \\ &= \delta_k.\end{aligned}$$

□

3 Application to Experiment Design

In the following section, we discuss the application of the theory in the previous section to experiment design. Specifically, we demonstrate how to use compressed sensing in order to reduce the number of *physical experiments* required in order to obtain a larger set of experimental data.

Consider a physical system that can be well modeled by the linear PDE $(A(u) + B)y_i = b_i$. Here, $A \in \mathscr{L}(U, \mathscr{L}(Y))$ defines the spatial differential operator, $B \in \mathscr{L}(Y)$ denotes the temporal differential operator, U denotes a Hilbert space of possible parameters, Y denotes a Hilbert space of possible states, u is the parameter that governs the behavior of the system, y_i gives the state or result of the i-th physical experiment, and b_i describes the forcing term or source given by the i-th experiment. For example, in Poisson's equation, $A(u)y = -\nabla \cdot (u\nabla y)$ and $B = 0$. Hence, the effect of p physical experiments can be modeled by the system of differential equations

$$(A(u) + B)y_i = b_i \quad i = 1, \dots, p.$$

For ease in notation later, we can rewrite this as the block system

$$\left(\begin{bmatrix} A(u) & & \\ & \ddots & \\ & & A(u) \end{bmatrix} + \begin{bmatrix} B & & \\ & \ddots & \\ & & B \end{bmatrix} \right) \begin{bmatrix} y_1 \\ \vdots \\ y_p \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_p \end{bmatrix},$$

which we abbreviate as $(A_p(u) + B_p)y = b$. Next, let us assume that there exists sparsity in the data in some basis. Specifically, we assume there exists sparsity in the *product space*, which does not necessarily imply sparsity in the state space. In order to manipulate sparsity in the product space, let $\Psi \in \mathbb{R}^{p\times p}$ and $\Psi \otimes I \in \mathscr{L}(Y^p)$ denote the change of basis operator where $\Psi \otimes I$: sparse $\to$ dense and $\Psi^{-1} \otimes I$: dense $\to$ sparse. Finally, let $\Phi \in \mathbb{R}^{q\times p}$ denote a matrix that possesses the restricted isometry property of order k.

Next, we notice the following. From Theorems 1 and 4 and Corollary 1, we have that

$$\{(A_p(u) + B_p)^{-1}b\} = \arg\min_{y\in Y^p} \{\|(\Psi^{-1} \otimes I)y\|_1 : (\Phi \otimes I)y = (\Phi \otimes I)(A_p(u) + B_p)^{-1}b\}.$$

In other words, we can use compressed sensing to recover our experimental data $(A_p(u) + B_p)^{-1}b$, which is known to be sparse through the transformation $\Psi^{-1} \otimes I$. However, due to the structure of $\Phi \otimes I$, we see that

$$\begin{aligned}
(\Phi \otimes I)(A_p(u) + B_p) &= \begin{bmatrix} \Phi_{11}I & \dots & \Phi_{1p}I \\ \vdots & \ddots & \vdots \\ \Phi_{q1}I & \dots & \Phi_{qp}I \end{bmatrix} \underbrace{\begin{bmatrix} A(u)+B & & \\ & \ddots & \\ & & A(u)+B \end{bmatrix}}_{p} \\
&= \begin{bmatrix} \sum_{i=1}^{p} \Phi_{1i}(A(u)+B) \\ \vdots \\ \sum_{i=1}^{p} \Phi_{qi}(A(u)+B) \end{bmatrix} = \begin{bmatrix} (A(u)+B)\sum_{i=1}^{p} \Phi_{1i}I \\ \vdots \\ (A(u)+B)\sum_{i=1}^{p} \Phi_{qi}I \end{bmatrix} \\
&= \underbrace{\begin{bmatrix} A(u)+B & & \\ & \ddots & \\ & & A(u)+B \end{bmatrix}}_{q} \begin{bmatrix} \Phi_{11}I & \dots & \Phi_{1p}I \\ \vdots & \ddots & \vdots \\ \Phi_{q1}I & \dots & \Phi_{qp}I \end{bmatrix} \\
&= (A_q(u) + B_q)(\Phi \otimes I)\,.
\end{aligned}$$

In other words, we have a pseudo-commutative relationship where

$$(\Phi \otimes I)(A_p(u) + B_p) = (A_q(u) + B_q)(\Phi \otimes I).$$

This allows us to rewrite the compressed sensing reconstruction problem as

$$\{(A_p(u) + B_p)^{-1}b\} = \arg\min_{y \in Y^p} \{\|(\Psi^{-1} \otimes I)y\|_1 : (\Phi \otimes I)y = (A_q(u) + B_q)^{-1}(\Phi \otimes I)b\}.$$

In other words, we can generate p pieces of data by conducting q experiments followed by a compressed sensing reconstruction problem. The q experiments consist of the same experiment setup, but where each of the q experiments consists of different linear combinations of the original p forcing terms or sources.

4 Finite Element Based Reconstruction

In this section, we discuss algorithms for solving the compressed sensing reconstruction problem. Since our eventual goal is the application of this process to experiment design, we focus on formulations that can be directly applied to finite element discretizations.

In this example, we assume that our experimental data can be accurately represented as functions in the Sobolov space $H^1(\Omega)$. In order to reconstruct our experiment data, we must solve the problem

$$\{d\} = \arg \min_{y \in H^1(\Omega)^p} \{\|(\Psi^{-1} \otimes I)y\|_1 : (\Phi \otimes I)y = \widetilde{d}\}$$

where $\widetilde{d} \in H^1(\Omega)^q$ denotes the compressed experimental data. If we use a finite element representation y^h of y, this becomes

$$\arg \min_{y^h \in \mathbb{R}^{mp}} \left\{ \sum_{i=1}^{p} \|\pi_i[(\Psi^{-1} \otimes I)y^h]\|_M : (\Phi \otimes I)y^h = \widetilde{d}^h \right\}$$

where m denotes the number of finite element coefficients, $\pi_i x = x_{1+(i-1)m:im}$, $\otimes$ denotes the Kronecker product, rather than our previous Kronecker-like product, $I \in \mathbb{R}^{m \times m}$ denotes the identity matrix, and M denotes the mass matrix where $M_{ij} = \int_\Omega b_i^h b_j^h$ for some basis functions $b_i \in H^1(\Omega)$. Let $M = LL^T$ be the Choleski factorization of M. Then, we notice that

$$\begin{aligned}
&\min_{y^h \in \mathbb{R}^{mp}} \left\{ \sum_{i=1}^{p} \|\pi_i[(\Psi^{-1} \otimes I)y^h]\|_M : (\Phi \otimes I)y^h = \widetilde{d}^h \right\} \\
= &\min_{y^h \in \mathbb{R}^{mp}} \left\{ \sum_{i=1}^{p} \|L^T \pi_i[(\Psi^{-1} \otimes I)y^h]\|_2 : (\Phi \otimes I)y^h = \widetilde{d}^h \right\} \\
= &\min_{y^h \in \mathbb{R}^{mp}} \left\{ \sum_{i=1}^{p} \|\pi_i[(\Psi^{-1} \otimes L^T)y^h]\|_2 : (\Phi \otimes I)y^h = \widetilde{d}^h \right\} \\
= &\min_{\substack{y^h \in \mathbb{R}^{mp}, \\ w \in \mathbb{R}^{mp}}} \left\{ \sum_{i=1}^{p} \|\pi_i w\|_2 : (\Phi \otimes I)y^h = \widetilde{d}^h, w = (\Psi^{-1} \otimes L^T)y^h \right\} \\
= &\min_{w \in \mathbb{R}^{mp}} \left\{ \sum_{i=1}^{p} \|\pi_i w\|_2 : (\Phi \otimes I)(\Psi \otimes L^{-T})w = \widetilde{d}^h \right\} \\
= &\min_{w \in \mathbb{R}^{mp}} \left\{ \sum_{i=1}^{p} \|\pi_i w\|_2 : (\Phi\Psi \otimes L^{-T})w = \widetilde{d}^h \right\} \\
= &\min_{\substack{z \in \mathbb{R}^{p}, \\ w \in \mathbb{R}^{mp}}} \left\{ \sum_{i=1}^{p} z_i : \ (\Phi\Psi \otimes L^{-T})w = \widetilde{d}^h, (z_i, \pi_i w) \succeq_Q 0 \right\}
\end{aligned} \tag{10}$$

where $(z_i, \pi_i w) \succeq_Q 0$ can also be written as $z_i \geq \|\pi_i w\|_2$ and denotes a second order cone constraint. Using this final formulation, we reconstruct the state with

$$y^h = (\Psi \otimes L^{-T})w.$$

Together, this gives a second-order cone program. Note, this differs from traditional compressed sensing, which requires the solution to a linear program in order to recover a sparse solution.

Although we can solve the above formulation using an interior point method, such as one provided by SDPT3 [7], the size of the problem quickly becomes intractable. For example, consider an experimental where we conduct 32 encoded experiments and attempt to reconstruct 256 pieces of data. In order to discretize the problem, we sample the data on a 32×32 grid and represent the result using piecewise linear elements.

In this case, we have $m = (32+1)^2 = 1089$, $p = 256$, and $q = 32$, which yields a second order cone program with 279873 variables, 34848 linear equality constraints, and 256 second order cone constraints of size 1090.

In order to more efficiently compute a solution to the discretized problem, we apply a variant of the alternating direction method of multipliers (ADM) developed by Deng, Yin, and Zhang [3], which generalizes the algorithm developed by Yang and Zhang [8] to problems with group sparsity. In the ADM algorithm, we reformulate our compressed sensing problem as

$$\begin{aligned}
&\min_{y^h \in \mathbb{R}^{mp}} \left\{ \sum_{i=1}^{p} \|\pi_i[(\Psi^{-1} \otimes I)y^h]\|_M : (\Phi \otimes I)y^h = \widetilde{d}^h \right\} \\
= &\min_{\substack{y^h \in \mathbb{R}^{mp} \\ z \in \mathbb{R}^{mp}}} \left\{ \sum_{i=1}^{p} \|z\|_{I\otimes M} : z = (\Psi^{-1} \otimes I)y^h, (\Phi \otimes I)y^h = \widetilde{d}^h \right\}.
\end{aligned}$$

Then, we find a solution by minimizing the augmented Lagrangian

$$\begin{aligned}
L_A(y^h, z, \lambda_1, \lambda_2) = \; & \|z\|_{I\otimes M} \\
& -\lambda_1^T (I \otimes M)(z - (\Psi^{-1} \otimes I)y^h) \\
& +\tfrac{\beta_1}{2} \|z - (\Psi^{-1} \otimes I)y^h\|_{I\otimes M}^2 \\
& -\lambda_2^T (I \otimes M)((\Phi \otimes I)y^h - \widetilde{d}^h) \\
& +\tfrac{\beta_2}{2} \|(\Phi \otimes I)y^h - \widetilde{d}^h\|_{I\otimes M}^2,
\end{aligned}$$

with respect to y^h and z. Here, λ_1, λ_2 denote an estimate of the Lagrange multipliers, and β_1, β_2 denote penalty parameters. Rather than minimize over y^h and z simultaneously, we alternate between fixing z and solving for y^h and fixing y^h and solving for z. Hence, for a fixed z, notice that

$$\begin{aligned}
&\arg \min_{y^h \in \mathbb{R}^{mp}} L_A(y^h, z, \lambda_1, \lambda_2) \\
= &\arg \min_{y^h \in \mathbb{R}^{mp}} \; -\lambda_1^T (I \otimes M)(-(\Psi^{-1} \otimes I)y^h) \\
&\quad +\beta_1(-z^T (I \otimes M)(\Psi^{-1} \otimes I)y^h + \tfrac{1}{2}(y^h)^T (\Psi^{-1} \otimes I)^T (I \otimes M)(\Psi^{-1} \otimes I)y^h) \\
&\quad -\lambda_2^T (I \otimes M)((\Phi \otimes I)y^h) \\
&\quad +\beta_2(\tfrac{1}{2}(y^h)^T (\Phi \otimes I)^T (I \otimes M)(\Phi \otimes I)y^h - (y^h)^T (\Phi \otimes I)^T (I \otimes M)\widetilde{d}^h) \\
= &\arg \min_{y^h \in \mathbb{R}^{mp}} \; \lambda_1^T (\Psi^{-1} \otimes M)y^h \\
&\quad +\beta_1(-z^T (\Psi^{-1} \otimes M)y^h + \tfrac{1}{2}(y^h)^T (\Psi^{-T}\Psi^{-1} \otimes M)y^h) \\
&\quad -\lambda_2^T (\Phi \otimes M)y^h \\
&\quad +\beta_2(\tfrac{1}{2}(y^h)^T (\Phi^T\Phi \otimes M)y^h - (y^h)^T (\Phi^T \otimes M)\widetilde{d}^h).
\end{aligned}$$

This minimum can be found analytically by $y^h = A^{-1}b$ where

$$\begin{aligned}
A &= (\beta_1 \Psi^{-T}\Psi^{-1} + \beta_2 \Phi^T \Phi) \otimes M \\
b &= (\Psi^{-T} \otimes M)(\beta_1 z - \lambda_1) + (\Phi^T \otimes M)(\beta_2 \widetilde{d}^h + \lambda_2).
\end{aligned}$$

In a similar manner, for a fixed y^h, notice that

$$
\begin{aligned}
&\arg\min_{z\in\mathbb{R}^{mp}} L_A(y^h, z, \lambda_1, \lambda_2) \\
=&\arg\min_{z\in\mathbb{R}^{mp}} \quad \|z\|_{I\otimes M} \\
&\qquad\qquad -\lambda_1^T(I\otimes M)z \\
&\qquad\qquad +\beta_1(\tfrac{1}{2}z^T(I\otimes M)z - z^T(I\otimes M)(\Psi^{-1}\otimes I)y^h) \\
=&\arg\min_{z\in\mathbb{R}^{mp}} \|z\|_{I\otimes M} + \beta_1\left(\frac{1}{2}z^T(I\otimes M)z - z^T(\Psi^{-1}\otimes M)y^h - \frac{1}{\beta_1}z^T(I\otimes M)\lambda_1\right) \\
=&\arg\min_{z\in\mathbb{R}^{mp}} \|z\|_{I\otimes M} + \frac{\beta_1}{2}\left\|(I\otimes M)z - (\Psi^{-1}\otimes M)y^h - \frac{1}{\beta_1}(I\otimes M)\lambda_1\right\|_2^2 \\
=&\arg\min_{z\in\mathbb{R}^{mp}} \|z\|_{I\otimes M} + \frac{\beta_1}{2}\left\|z - (\Psi^{-1}\otimes I)y^h - \frac{1}{\beta_1}\lambda_1\right\|_{I\otimes M}^2.
\end{aligned}
$$

Again, we can find the minimum analytically through the following procedure. Let

$$r = (\Psi^{-1}\otimes I)y^h - \frac{1}{\beta_1}\lambda_1.$$

Then, for $i = 1 \to p$, set

$$\pi_i z = \begin{cases} 0 & \alpha < \epsilon \\ \max\{\alpha - \frac{1}{\beta_1}, 0\}\frac{\pi_i r}{\alpha} & \text{otherwise} \end{cases}$$

where $\alpha \leftarrow \|\pi_i r\|_M$ and $\epsilon > 0$ is small. We summarize the entire ADM algorithm in Algorithm 1. Experimentally, we have found that setting $\beta_1 = \frac{3}{\text{mean}(|b|)}$, $\beta_2 = \frac{.3}{\text{mean}(|b|)}$, $\epsilon = 10^{-8}$, stopping tolerance $= 10^{-6}$, and max iteration $= 2000$ where mean$(|b|)$ denotes the average of the absolute values of the elements in b provides effective results.

Algorithm 1: Primal ADM algorithm for reconstructing compressed experimental data that has been represented with finite elements

1. Initialize $z \in \mathbb{R}^{mp}$, $\lambda_1 \in \mathbb{R}^{mp}$, $\lambda_2 \in \mathbb{R}^{mq}$, $\beta_1, \beta_2 > 0$, $\gamma_1, \gamma_2 > 0$, $\epsilon > 0$, max iteration, and stopping tolerance.

2. **for** $i = 1 \to$ max iteration **do**

   (a) $y^h_{old} \leftarrow y^h$

   (b) $y^h \leftarrow A^{-1}b$ where

   $$A = (\beta_1 \Psi^{-T}\Psi^{-1} + \beta_2 \Phi^T \Phi) \otimes M$$
   $$b = (\Psi^{-T} \otimes M)(\beta_1 z - \lambda_1) + (\Phi^T \otimes M)(\beta_2 \widetilde{d}^h + \lambda_2)$$

   (c) $r \leftarrow (\Psi^{-1} \otimes I)y^h + \frac{1}{\beta_1}\lambda_1$

   (d) **for** $i = 1 \to p$ **do**

      i. $\alpha \leftarrow \|\pi_i r\|_M$

      ii. $\pi_i z = \begin{cases} 0 & \alpha < \epsilon \\ \max\{\alpha - \frac{1}{\beta_1}, 0\}\frac{\pi_i r}{\alpha} & \text{otherwise} \end{cases}$

   (e) $\lambda_1 \leftarrow \lambda_1 - \gamma_1\beta_1(z - (\Psi^{-1} \otimes I)y^h)$

   (f) $\lambda_2 \leftarrow \lambda_2 - \gamma_2\beta_2((\Phi \otimes I)y^h - \widetilde{d}^h)$

   (g) **if** $\|y^h - y^h_{old}\|_{I\otimes M} <$ (stopping tolerance)$\|y^h_{old}\|_{I\otimes M}$ **then** stop

5 Computational Study

In the following section, we conduct two computational studies. In the first, we verify that the ADM algorithm derived above provides fast, accurate solutions to the compressed sensing reconstruction problem. In the second, we demonstrate that compressed sensing can be successfully integrated into the experiment design process.

In the following experiment, we encode and recover four different letters represented by piecewise linear functions. Specifically, we represent the letters using a piecewise linear finite element basis on a 20×20 regular mesh composed of triangles over the domain $\Omega = [-1, 1] \times [-1, 1]$. These functions can be seen in Figure 1. Next, we generate a collection of $p = 50$ functions and store the result in $\widetilde{y}$. Each of the functions in

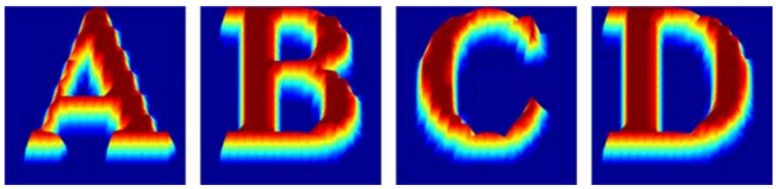

Figure 1. Letters A, B, C, and D represented as piecewise linear functions. These functions serve as the true set of data in our first computational experiment.

$\widetilde{y}$ is zero except for four different, randomly determined elements, which contain the letters generated above. Then, we generate a random change of basis operator $\Psi \in \mathbb{R}^{p\times p}$ whose elements are normally distributed with mean 0 and variance 1. Using this operator, we generate a dense set of data $\bar{y} = (\Psi \otimes I)\widetilde{y}$. Next, we generate a random encoding operator $\Phi \in \mathbb{R}^{q\times p}$ where $q = 15$ and the entries of Φ are normally distributed with mean 0 and variance 1. This allows us to generate our compressed data as $\widetilde{d} = (\Phi \otimes I)\bar{d}$. Finally, we solve the compressed sensing reconstruction problem described in Section 4. We do so in two different ways. First, we apply SDPT3 to the cone formulation in (10). Second, we use the ADM algorithm described in Algorithm 1 where we have set the stopping tolerance to 10^{-8}. We summarize the relative error in the reconstruction and running time of the algorithms in Table 1 and show the letters reconstructed by the ADM algorithm after we apply $\Psi^{-1} \otimes I$ to the result in Figure 2. As we can see, both SDPT3 and the ADM

	Relative Error	Iterations	Time (HH:MM:SS)
SDPT3	1.93e-08	13	(04:27:12)
ADM	1.01e-07	185	(00:00:23)

Table 1. Relative error, number of iterations, and running time of SDPT3 and the ADM algorithm applied to the letters example. In this experiment, we calculate the relative error by $\|y_{recon} - y_{true}\|_{I\otimes M}/(1 + \|y_{true}\|_{I\otimes M})$. As we can see, both algorithms produce accurate reconstructions of the encoded data, but ADM produces the reconstruction much faster than SDPT3.

algorithm produce accurate reconstructions of the true data. Nonetheless, the ADM algorithm produces a solution far more quickly than SDPT3.

In the second computational experiment, we simulate the experiment design process. In order to accomplish this, we generate experimental data $\widetilde{d}$ according to Poisson's equation, $-\nabla \cdot (u\nabla y) = b$. In order to generate this data, we use a piecewise linear finite element basis on a 128×128 regular mesh composed of tri-

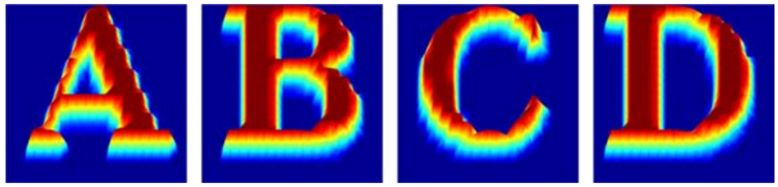

Figure 2. Letters reconstruction using ADM. The result is virtually identical to the true set of letters.

angles over the domain $\Omega = [-1, 1] \times [-1, 1]$. For the parameters u, we use a true set of parameters shown in Figure 3. For the sources b, we use 256 Gaussians with variance 0.01 and a base amplitude 10 spaced evenly

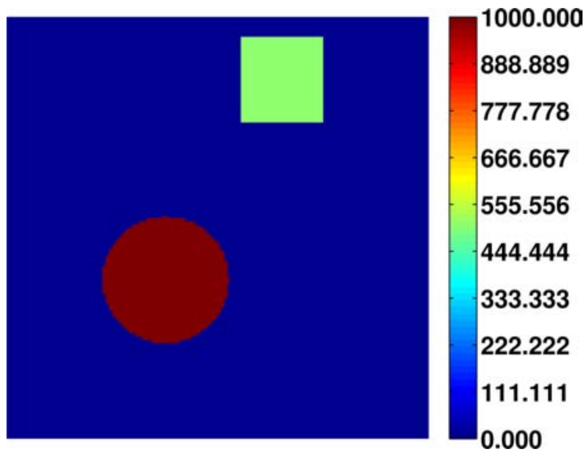

Figure 3. The true parameter for the experiment design study. The parameter equals 1000 within the circle of radius 0.3 centered at $(-0.25, -0.25)$, 500 within the square with side length .4 centered at $(0.3, 0.7)$, and 1 everywhere else.

on a square grid over the domain $[-.25, .25] \times [-.25, .25]$. Then, we generate $q = 32$, 64, or 128 experimental pieces of data by solving Poisson's equation where we use all q sources simultaneously where the amplitude of the ith source in the jth experiment has been scaled by Φ_{ij} where Φ_{ij} is normally distributed with mean 0 and variance 1. Next, we sample this data at the nodal points of a 32×32 regular mesh composed of triangles over the same domain. This gives us our sampled, encoded experimental data $\widetilde{d}$. For the change of basis operator, we use a discrete Haar wavelet transform in two dimensions where the level of the coarsest scale is 1, $\Psi \in \mathbb{R}^{256\times 256}$. Note, this is the discretized transform for 16×16 images where the image has been reshaped into a 256×1 vector. Finally, we apply ADM to reconstruct experimental data that would have been generated by running all $p = 256$ experiments individually. We summarize the relative errors, with respect to the $[L^2(\Omega)]^p$ norm, in this reconstruction in Table 2 and show a visual reconstruction of two different experiments in Figures 4 and 5. As we can see, the compressed sensing reconstruction generates data whose relative error varies between 2% and 9%. The amount of error is lower when we conduct a larger number of encoded experiments. The reason that we do not achieve an exact reconstruction is that we lack perfect sparsity in the true experimental data. In order to see that, we show the sparsity structure of the data in Figures 6 and 7. Nonetheless, in spite of the lack of exact sparsity, the reconstructed data is still relatively accurate.

Encoded Experiments (q)	Relative Error	ADM Iterations
128	2.29e-02	475
64	4.17e-02	545
32	9.27e-02	551

Table 2. Relative error and number of iterations ADM used when applied to the experiment design example. Here, we calculate the relative error by $\|d_{recon} - d_{sampled}\|_{I\otimes M}/(1 + \|d_{sampled}\|_{I\otimes M})$ where $d_{sampled}$ is the data generated by running each experiment individually followed by sampling the result on the same grid used for the compressed sensing reconstruction. As we can see, the amount of error in the reconstructed solution varies from 2% to 9% where the error is smaller when conduct a larger number of encoded experiments. Nonetheless, we see that it is possible to reduce the number of physical experiments by 87.5% and still generate data that contains less than 10% additional error than running the experiments separately.

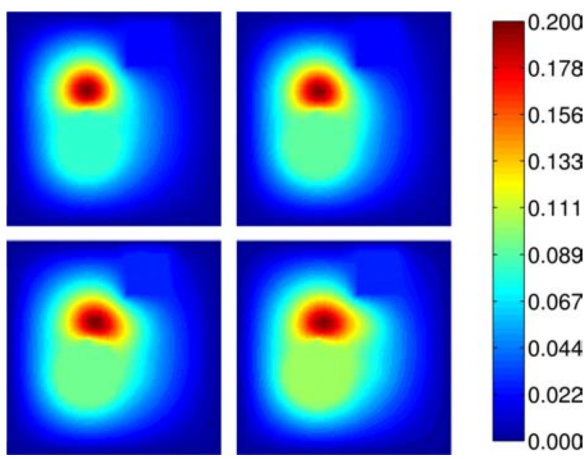

Figure 4. Result generated by the 241st experiment. The top left plot denotes the result of sampling the experiment directly. The top right plot shows the result of the compressed sensing reconstruction where $q = 128$; bottom left, $q = 64$; bottom right, $q = 32$. The reconstructions reproduce the experimental result relatively accurately where we notice less error when conducting a larger number of encoded experiments.

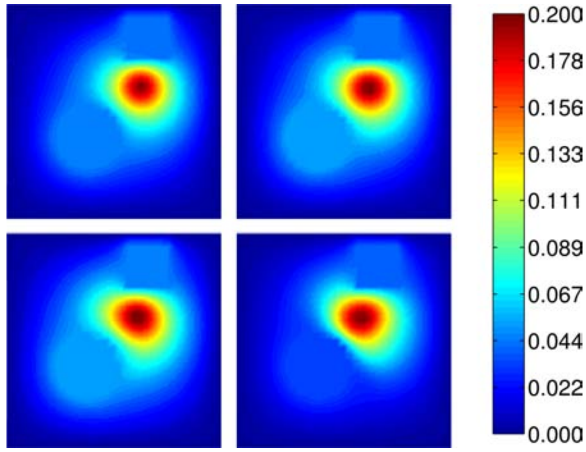

Figure 5. Result generated by the 256th experiment. The top left plot denotes the result of sampling the experiment directly. The top right plot shows the result of the compressed sensing reconstruction where $q = 128$; bottom left, $q = 64$; bottom right, $q = 32$. The reconstructions reproduce the experimental result relatively accurately where we notice less error when conducting a larger number of encoded experiments.

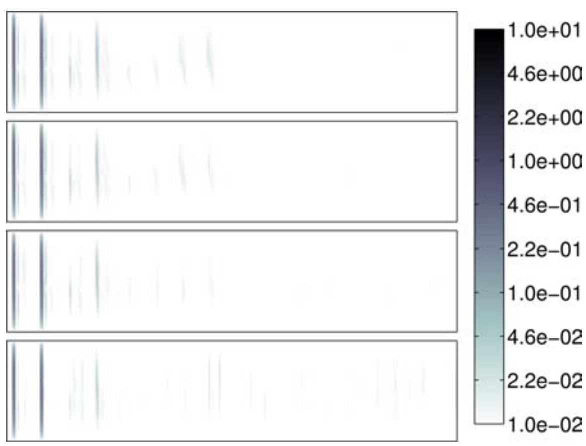

Figure 6. Sparsity structure of the experimental data. Each plot denotes the magnitudes of the coefficients of the finite element representation of the data after applying the change of basis operator Ψ^{-1}. In addition, each column of the plot denotes a different experiment. The top plot denotes the data when sampled directly; second plot, q=128; third plot, q=64; fourth plot, q=32. As we can see, there are four dominant components in the product space (the two dark stripes denote two experiments each) in addition to several minor components. In each case, the compressed sensing reconstruction correctly identifies the dominant components, but does not exactly reproduce all of the minor components.

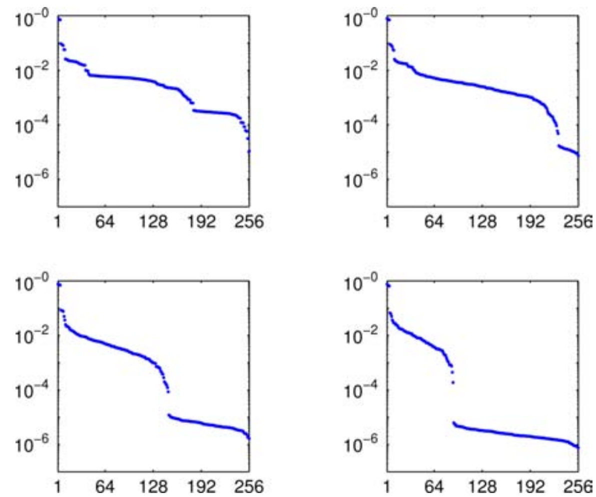

Figure 7. Sparsity structure of the experimental data. Each plot contains the sorted magnitudes of the individual experiments after applying the change of basis operator Ψ^{-1}. The top left plot denotes the data when sampled directly; top right, q=128; bottom left, q=64; bottom right, q=32. Here, we see that each set of data contains four dominant components whereas the remaining components have magnitudes that decrease rapidly.

6 Conclusion

In the preceding discussion, we have discussed how to apply compressed sensing to experiment design. We accomplish this by first extending the theory of compressed sensing to a finite product of Hilbert spaces. This allows a clean extension of the theory of compressed sensing to group sparsity as well as collections of functional representations of experimental data. Next, we have shown how to use these results to conduct a small number of physical experiments and then reconstruct a larger collection of experimental data. In order to reconstruct this data, we have presented an efficient reconstruction algorithm for this generalization based on the alternating direction method of multipliers. Then, using these tools, we have demonstrated that compressed sensing can be used in an experimental context to reduce the number of physical experiments conducted by nearly 90% while simultaneously preserving the fidelity of the data.

Several theoretical questions remain open. In traditional compressed sensing, an alternative reconstruction algorithm based on a second-order cone program can be used when it is known that the data lacks exact sparsity. This formulation must still be generalized to the product-space setting. In addition, in order for this method to be an effective tool for experiment design, a good sparsifying operator must be known. Although there has been much work in finding a sparse basis for images or acoustic signals, there has been less work on the application of these bases for experiment design. In the future, we look forward to addressing both of these questions.

References

[1] Richard Baraniuk, Mark Davenport, Ronald DeVore, and Michael Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28(3):253–263, 2008.

[2] Emmanuel J. Candes, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489 – 509, February 2006.

[3] Wei Deng, Wotao Yin, and Yin Zhang. Group sparse optimization by alternating direction method. Technical Report TR11-06, Rice University, 2011.

[4] David L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory,*, 52(4):1289 –1306, April 2006.

[5] Holger Rauhut. Compressive sensing and structured random matrices. In Massimo Fornasier, editor, *Theoretical Foundations and Numerical Methods for Sparse Recovery*, volume 9 of *Radon Series on Computational and Applied Mathematics*, pages 1–92. De Gruyter, 2010.

[6] Berwin A. Turlach, William N. Venables, and Stephen J. Wright. Simultaneous variable selection. *Technometrics*, 47(3), 2005.

[7] R.H. Tütüncü, K. C. Toh, and M. J. Todd. Solving semidefinite-quadratic-linear programs using SDPT3. *Mathematical Programming*, 95:189–217, 2003. 10.1007/s10107-002-0347-5.

[8] Junfeng Yang and Yin Zhang. Alternating direction algorithms for ℓ_1-problems in compressive sensing. *SIAM Journal on Scientific Computing*, 33(1):250–278.

[9] Joseph Young and Denis Ridzal. An application of random projection to parameter estimation in partial differential equations. *SIAM Journal on Scientific Computing*, 34(4), 2012.

DISTRIBUTION:

1	MS 0899	RIM-Reports Management, 9532 (electronic copy)
1	MS 0899	Technical Library, 9536 (electronic copy)

Sandia National Laboratories