ITERATIVE HARD THRESHOLDING AND NORMALIZED ITERATIVE HARD THRESHOLDING

MAT-314 FINAL PROJECT

DIOGO TANDETA TARTAROTTI AND HENRY DWORKIN

1. Introduction to Compressed Sensing

When we have $A \in R^{m \times n}$, $x \in R^n$, our goal is to solve the system of equations $y = Ax$. A discrete signal with n components can be measured by solving an underdetermined system of equations. However, to capture the signal's information, we don't need to make every single measurement.

In order to solve the compressed sensing problem, we must be able to recover a sparse vector $x \in R^n$ from measurements $y \in R^m$ whose entries are derived from inner product measurements. The goal of compressed sensing is to measure and compress data in a sparse vector at the same time. If $||x||_0 = k < n$ and $S = \text{supp}(x)$ then we know that we must take at least k measurements.

Therefore, the objective of compressed sensing is to find a measurement process with m measurements such that $k < m < n$. This is because we must take at least k measurements and we never need more than n measurements of $x \in \chi_n(k)$. The goal is to get m as close to k as possible, or to take the fewest number of measurements necessary to collect all the data in x.

In order to reconstruct x, we use an encoder-decoder pair. Suppose $x = \chi_n(k)$. An encoder is a matrix $A \in R^{m \times n}$, and we say that it encodes x into y when the measurements $y \in R^m, y = Ax$, capture the information content of x.

A decoder of (y, A) is an optimization method or algorithm that can recover, given the measurements y and the encoder A, the original k-sparse vector x.

Infinite solutions exist for the underdetermined system $y = Ax$. One way to choose a specific solution is to solve the least squares problem. The least squares solution, however, is typically dense, which means that most, and probably all, of its entries will be nonzero. Finding a sparse solution will allow us to simultaneously compress the signal. So, in compressed sensing, we look for a vector x that could have produced the measurements y and has the fewest possible nonzero entries.

2. Introduction to IHT and NIHT

In this paper, we address the Compressed Sensing problem of reconstructing an s-sparse vector $x \in \mathbb{C}^N$ from the measurement vector $y = Ax \in \mathbb{C}^m$ when $m << N$. We concentrate on two reconstruction techniques that enable stable and reliable sparse recovery, Iterative Hard Thresholding and Normalized Iterative Hard Thresholding.

Iterative Hard Thresholding

Date: May 2023.

Algorithm 1 Iterative Hard Thresholding (IHT)[2]

Input: A, y, k, and an error tolerance tol
Output: A k-sparse approximation $\hat{x}$ of the target signal x
Initialization: Set $j = 1$; $x^{(0)} = 0$, $r^{(0)} = y$
while $\|r\|_2 >$ tol **do**
$w^{(j)} = x^{(j-1)} + A^* r^{(j-1)}$ {Identify column of A most correlated to residual}
$T^{(j)} = \text{PrincipalSupport}_k(w^{(j)})$ {Estimate for support set}
$x^j_{T_j} = w^j_{T_j}$ and $x^j_{(T_j)^c} = 0$ {Projection onto k subspace defined by $T^{(j)}$}
$r^{(j)} = y - Ax^{(j)}$ {Update the residual}
$j = j + 1$ {Increase the iteration counter}
end while
return $\hat{x} = x^{(k)}$ {Return the k-sparse vector $x^{(k)}$} =0

Normalized Iterative Hard Thresholding

The pseudo code presented for the algorithms include the subroutines PrincipalSupport(z), returning the index set of the k largest magnitude entries in z, and Threshold(z, S), the hard thresholding operator that leaves the values in z_S unchanged and sets all other values of z to zero.

Algorithm 2 Subroutine 1: Initialization Procedure[2]

Input: A, y, k, and an error tolerance tol
Output: A k-sparse approximation $\hat{x}$ of the target signal x
Initialization and Initial Support Detection
$x_{(0)} = A^* y$ {Initial approximation}
$T_{(0)} = \text{PrincipalSupport}(x_{(0)})$ {Proxy to the support set}
$x_{(0)} = \text{Threshold}(x_{(0)}, T_{(0)})$ {Restriction to proxy support set $T_{(0)}$}
$r_{(0)} = y - Ax_{(0)}$ {Initialize residual} =0

Algorithm 3 Algorithm 1: NIHT (Normalized Iterative Hard Thresholding)[2]

Iteration: During iteration l, do
$\omega_l = \frac{\|(A^* r_{(l-1)})_{T_{l-1}}\|_2^2}{\|(A_{T_{l-1}}(A^* r_{(l-1)}))_{T_{l-1}}\|_2^2}$ {Optimal step size in the k-subspace T_{l-1}}
$v_{(l)} = x_{(l-1)} + \omega_l A^* r_{(l-1)}$ {Steepest descent step}
$T_{(l)} = \text{PrincipalSupport}(x_{(l)})$ {Proxy to the support set}
$x_{(l)} = \text{Threshold}(v_{(l)}, T_{(l)})$ {Restriction to proxy support set $T_{(l)}$}
$r_{(l)} = y - Ax_{(l)}$ {Update the residual} =0

3. Definitions

First, we will start with definitions of Asymmetric Restricted Isometry Constants defined in the Course Notes and a lemma on them.

Definition 3.1. [1] Let $A \in \mathbb{R}^{m \times n}$ has the asymmetric restricted isometry constants L_s and U_s if

$$L_s = \min_{c \geq 0} c \text{ subject to } (1-c)\|x\|_2^2 \leq \|Ax\|_2^2 \ \ \forall x \in \chi_n(s)$$

$$U_s = \min_{c \geq 0} c \text{ subject to } (1+c)\|x\|_2^2 \geq \|Ax\|_2^2 \ \ \forall x \in \chi_n(s)$$

Furthermore, the s-th order restricted isometry constant of the matrix A, δ_s, can be represented by $\delta_s = \max\{L_s, U_s\}$. For ease of notation, we are going to write all instances of $\max\{L_s, U_s\}$ as δ_s.

Lemma 3.1. *[1] Let $\Omega = \{1, ..., n\}$, and let $\lambda^{min}(B)$, $\lambda^{max}(B)$ denote the minimum and maximum eigenvalues of a matrix B. Then,*

$$1 - L_s = \min_{T \subset \Omega, |T|=s} \lambda^{min}(A_T^* A_T)$$
$$1 + U_s = \max_{T \subset \Omega, |T|=s} \lambda^{max}(A_T^* A_T)$$

4. Main Result

Theorem 4.0.1. *Suppose that the 2s-th and 3s-th order asymmetric restricted isometry constants of the matrix $A \in \mathbb{R}^{m\times n}$ satisfy*

$$(\frac{1+U_{2k}}{1-L_{2k}}) \max\{L_{3k}, U_{3k}\} = (\frac{1+U_{2k}}{1-L_{2k}})\delta_{3k} < \frac{1}{2}$$

Then, if $x \in \chi_n(k)$ is a k-sparse vector, the sequence $\{x_{(n)}\}$ defined by NIHT with $y = Ax$ converges to the vector x.
More generally, suppose S is the index set of the k largest magnitude entries of a vector $x \in \mathbb{R}^n$, and $y = Ax + e$ for some error term $e \in \mathbb{R}^n$. Let $\rho = 2(\frac{1+U_{2k}}{1-L_{2k}}) \max\{L_{3k}, U_{3k}\} = 2(\frac{1+U_{2k}}{1-L_{2k}})\delta_{3k}$ and $\beta = 2(\frac{1+U_{2k}}{1-L_{2k}})\sqrt{1+U_{2k}}$. Then,

$$\left\|x_{(n+1)} - x_S\right\|_2 \leq \rho^n \left\|x_{(0)} - x_S\right\|_2 + \sum_{i=0}^{n} \rho^i \beta \left\|e'\right\|_2 \text{ for all } n$$

To prove the main result, we need to prove a few helpful Lemmas. For the following lemmas, suppose $A \in \mathbb{R}^{m\times n}$, $x \in \chi_n(k)$, and $y = Ax$.

5. Auxiliary Lemmas

Lemma 5.1. *Suppose $L_{2k} < 1$. Then,*

$$\omega_n \leq \frac{1+U_{2k}}{1-L_{2k}}$$

Proof. Let $z = A^* r_{(n)}$. Then, from the definition of ω_n,

$$\omega_n = \frac{\|z\|_2^2}{\|Az\|_2}$$

By definition, $r_{(n)}$, $z \in \chi_n(2k)$. Then, from Definition 3.1,

$$1 - L_{2k} \leq \omega_n^{-1} \leq 1 + U_{2k}$$
$$\Longrightarrow \frac{1}{1-L_{2k}} \geq \omega_n \geq \frac{1}{1+U_{2k}}$$
$$\therefore \omega_n \leq \frac{1+U_{2k}}{1-L_{2k}}$$

□

For Lemmas 5.2, 5.3, and 5.4, suppose $T \subset \{1, 2, ...n\}$, and $|T| = s$.

Lemma 5.2. *If $\lambda(A_T^* A_T)$ is an eigenvalue of $A_T^* A_T$, then*

$$\omega_n \delta_s = \omega_n \max\{U_s, L_s\} \geq |\omega_n \lambda(A_T^* A_T) - 1| = |1 - \omega_n \lambda(A_T^* A_T)|$$

Proof. Suppose $\lambda(A_T^* A_T)$ is an eigenvalue of $A_T^* A_T$. Then, $\omega_n \lambda(A_T^* A_T)$ is an eigenvalue of $\omega_n A_T^* A_T$. Let L_s' and U_s' be the asymmetric restricted isometry constants of $\omega_n A$. From Lemma 5.1, we have that

$$L_s' = \omega_n L_s$$
$$U_s' = \omega_n U_s$$

By Lemma 5.1,

$$\begin{aligned} 1 - L'_s &\le \omega_n \lambda(A_T^* A_T) \\ \implies (1 - \omega_n \lambda(A_T^* A_T)) &\le L'_s = \omega_n L_s \end{aligned}$$

Similarly from Lemma 5.1,

$$\begin{aligned} 1 + U'_s &\ge \omega_n \lambda(A_T^* A_T) \\ \iff U'_s &\ge (\omega_n \lambda(A_T^* A_T) - 1) \\ \implies \omega_n U_s &\ge (\omega_n \lambda(A_T^* A_T) - 1) \end{aligned}$$

Since $\omega_n \ge 0$, $\omega_n \max\{U_s, L_s\} \ge |\omega_n \lambda(A_T^* A_T) - 1| = |1 - \omega_n \lambda(A_T^* A_T)|$ □

Lemma 5.3. *If $\lambda(A_T^* A_T)$ is an eigenvalue of $A_T^* A_T$, $(1 - \omega_n \lambda)$ is an eigenvalue of $(I - \omega_n A_T^* A_T)$*

Proof. Suppose $\lambda(A_T^* A_T)$ is an eigenvalue of $A_T^* A_T$. Then, $\omega_n \lambda(A_T^* A_T)$ is an eigenvalue of $\omega_n A_T^* A_T$. By definition of eigenvalue, $\omega_n A_T^* A_T v = \omega_n \lambda v$ for some $v \in \mathbb{R}^n$. Then,

$$\begin{aligned} (I - \omega_n A_T^* A_T) v &= v - \omega_n \lambda v \\ &= (1 - \omega_n \lambda) v \end{aligned}$$

Therefore, if $\lambda(A_T^* A_T)$ is an eigenvalue of $A_T^* A_T$, $(1 - \omega_n \lambda)$ is an eigenvalue of $(I - \omega_n A_T^* A_T)$. □

Lemma 5.4.

$$\sigma(I - \omega_n A_T^* A_T) = |1 - \omega_n \lambda(A_T^* A_T)| \text{ for all singular values of } I - \omega_n A_T^* A_T$$

Proof.

$$\begin{aligned} (I - \omega_n A_T^* A_T)^T &= I^T - (\omega_n A_T^* A_T)^T \\ &= I - \omega_n (A_T^* A_T)^T \text{ since I is symmetric by definition} \\ &= I - \omega_n A_T^* A_T \text{ since } A_T^* A_T \text{ is symmetric by Exercise 2.4[1]} \end{aligned}$$

Hence, $I - \omega_n A_T^* A_T$ is symmetric. By exercise 2.17[1], the singular values of $I - \omega_n A_T^* A_T$ are the absolute values of the eigenvalues of $I - \omega_n A_T^* A_T$, which by Lemma 5.3 are given by $1 - \omega_n \lambda(A_T^* A_T)$. Thus,

$$\sigma(I - \omega_n A_T^* A_T) = |1 - \omega_n \lambda(A_T^* A_T)| \text{ for all singular values of } I - \omega_n A_T^* A_T$$

□

Lemma 5.5. *If $T \subset \{1, 2, ...n\}$, and $|T| = s$, then*

$$\|I - \omega_n A_T^* A_T\|_2 \le \omega_n \delta_s$$

Proof. By Theorem 2.3.2[1],

$$\begin{aligned} \|I - \omega_n A_T^* A_T\|_2 &= \max \sigma(I - \omega_n A_T^* A_T) \\ &= \max |1 - \omega_n \lambda(A_T^* A_T)| \text{ by Lemma 5.4} \\ &\le \omega_n \max\{U_s, L_s\} = \omega_n \delta_s \text{ by Lemma 5.2} \end{aligned}$$

□

Lemma 5.6. *Suppose $u, v \in \mathbb{R}^n$ with $S = supp(u)$, $T = supp(v)$, and $|S \cup T| = s$. Then,*

$$\langle (I - \omega_n A^* A) u, v \rangle \le \omega_n \delta_s \|u\|_2 \|v\|_2$$

Proof. Following a very similar logic to Lemma 9.2[1], let $Q = S \cup T$. From the definition of Q and exercise 7.1[1], we get $\langle u, v \rangle = \langle u_Q, v_Q \rangle$, $\|u_Q\|_2 = \|u\|_2$, $\|v_Q\|_2 = \|v\|_2$, $Au = A_Q u_Q$, and $Av = A_Q v_Q$. Then,

$$\begin{aligned} \langle (I - \omega_n A^* A) u, v \rangle &= \langle u, v \rangle - \langle \omega_n A^* A u, v \rangle \text{ by linearity of inner product} \\ &= \langle u, v \rangle - \langle \omega_n A u, A v \rangle \\ &= \langle u_Q, v_Q \rangle - \langle \omega_n A_Q u_Q, A_Q v_Q \rangle \end{aligned}$$

$$
\begin{aligned}
&= \left\langle (I - \omega_n A_Q^* A_Q) u_Q, v_Q \right\rangle \\
&\leq \left\| (I - \omega_n A_Q^* A_Q) u_Q \right\|_2 \|v_Q\|_2 \text{ by the Cauchy-Schwarz inequality} \\
&\leq \left\| (I - \omega_n A_Q^* A_Q) \right\|_2 \|u_Q\|_2 \|v_Q\|_2 \\
&\leq \omega_n \delta_s \|u\|_2 \|v\|_2 \text{ by Lemma 5.5}
\end{aligned}
$$

□

Now, we can prove Theorem 4.0.1. The proof will follow a similar outline to the proof for IHT in [3, Theorem 3].

6. Proof of Main Result

Setup. Let $A \in \mathbb{R}^{m \times n}$, and $x \in \chi_n(k)$. Suppose $y = Ax + e$, where $e \in \mathbb{R}^n$ is an error term, and A has the asymmetric restricted isometry constants L_k, U_k,L_{2k}, U_{2k}, L_{3k}, and U_{3k} . Let S be the index set of the k largest magnitude entries in x. Suppose $L_{2k} < 1$. Let ω_n, $r_{(n)}$, and the sequence $\{x_{(n)}\}$ be defined as in the pseudocode for NIHT.

Proof of Theorem 4.0.1. First, let $e' = Ax_{\overline{S}} - e$ and define

$$
\begin{aligned}
v_{(n+1)} &\equiv x_{(n)} + \omega_n A^*(y - Ax_{(n)}) \\
&= x_{(n)} + \omega_n A^* A(x_S - x_{(n)}) + x_{(n)} + \omega_n A^*(Ax_{\overline{S}} - e) \\
&= (I - \omega_n A^* A)(x_{(n)} - x_S) + x_S + \omega_n A^* e'
\end{aligned}
$$

Note that the k-sparse vector $x_{(n+1)}$ is a better k-term approximation to $v_{(n+1)}$ than x_S. Therefore,

$$
\begin{aligned}
&\left\| v_{(n+1)} - x_{(n+1)} \right\|_2^2 \leq \left\| v_{(n+1)} - x_S \right\|_2^2 \\
\iff &\left\| (v_{(n+1)} - x_S) + (x_S - x_{(n+1)}) \right\|_2^2 \leq \left\| v_{(n+1)} - x_S \right\|_2^2 \\
\iff &\left\| (v_{(n+1)} - x_S) \right\|_2^2 + 2\langle (v_{(n+1)} - x_S), (x_S - x_{(n+1)}) \rangle + \left\| (x_S - x_{(n+1)}) \right\|_2^2 \leq \left\| v_{(n+1)} - x_S \right\|_2^2 \\
\implies &2\langle (v_{(n+1)} - x_S), (x_S - x_{(n+1)}) \rangle + \left\| (x_{(n+1)} - x_S) \right\|_2^2 \leq 0 \\
\implies &\left\| (x_{(n+1)} - x_S) \right\|_2^2 \leq 2\langle (v_{(n+1)} - x_S), (x_{(n+1)} - x_S) \rangle
\end{aligned}
$$

Let $Q = \text{supp}(x) \cup \text{supp}(x_{(n)}) \cup \text{supp}(x_{(n+1)})$.

$$
\begin{aligned}
2\langle (v_{(n+1)} - x_S), (x_{(n+1)} - x_S) \rangle &= 2\langle (I - \omega_n A^* A)(x_{(n)} - x_S) + \omega_n A^* e', x_{(n+1)} - x_S \rangle \\
&= 2\langle (I - \omega_n A^* A)(x_{(n)} - x_S), x_{(n+1)} - x_S \rangle + 2\langle \omega_n A^* e', x_{(n+1)} - x_S \rangle \\
&= 2\langle (I - \omega_n A^* A)(x_{(n)} - x_S), x_{(n+1)} - x_S \rangle + 2\omega_n \langle e', A(x_{(n+1)} - x_S) \rangle
\end{aligned}
$$

By the Cauchy-Schwarz inequality and Lemma 5.6

$$
\leq 2\omega_n \delta_{3k} \left\| x_{(n)} - x_S \right\|_2 \left\| x_{(n+1)} - x_S \right\|_2 + 2\omega_n \|e'\|_2 \left\| A(x_{(n+1)} - x_S) \right\|_2
$$

By Definition 3.1

$$
\begin{aligned}
&2\omega_n \delta_{3k} \left\| x_{(n)} - x_S \right\|_2 \left\| x_{(n+1)} - x_S \right\|_2 + 2\omega_n \|e'\|_2 \left\| A(x_{(n+1)} - x_S) \right\|_2 \\
\leq &2\omega_n \delta_{3k} \left\| x_{(n)} - x_S \right\|_2 \left\| x_{(n+1)} - x_S \right\|_2 + 2\omega_n \|e'\|_2 \sqrt{1 + U_{2k}} \left\| x_{(n+1)} - x_S \right\|_2
\end{aligned}
$$

Hence,

$$
\begin{aligned}
&\left\| (x_{(n+1)} - x_S) \right\|_2^2 \leq 2\omega_n \delta_{3k} \left\| x_{(n)} - x_S \right\|_2 \left\| x_{(n+1)} - x_S \right\|_2 + 2\omega_n \|e'\|_2 \sqrt{1 + U_{2k}} \left\| x_{(n+1)} - x_S \right\|_2 \\
\implies &\left\| (x_{(n+1)} - x_S) \right\|_2 \leq 2\omega_n \delta_{3k} \left\| x_{(n)} - x_S \right\|_2 + 2\omega_n \|e'\|_2 \sqrt{1 + U_{2k}} \\
\implies &\left\| (x_{(n+1)} - x_S) \right\|_2 \leq 2(\frac{1 + U_{2k}}{1 - L_{2k}}) \delta_{3k} \left\| x_{(n)} - x_S \right\|_2 + 2(\frac{1 + U_{2k}}{1 - L_{2k}}) \|e'\|_2 \sqrt{1 + U_{2k}} \text{ by Lemma 5.1}
\end{aligned}
$$

Let $\rho = 2(\frac{1+U_{2k}}{1-L_{2k}})\delta_{3k}$ and $\beta = 2(\frac{1+U_{2k}}{1-L_{2k}})\sqrt{1+U_{2k}}$. Then,

$$\begin{aligned}\left\|x_{(n+1)} - x_S\right\|_2 &\leq \rho \left\|x_{(n)} - x_S\right\|_2 + \beta \left\|e'\right\|_2 \\ &\leq \rho^n \left\|x_{(0)} - x_S\right\|_2 + \sum_{i=0}^{n} \rho^i \beta \left\|e'\right\|_2\end{aligned}$$

Assuming the measurements are accurate, namely $e' = 0$, and noticing x is k-sparse,

$$\begin{aligned}\rho^n \left\|x_{(0)} - x_S\right\|_2 + \sum_{i=0}^{n} \rho^i \beta \left\|e'\right\|_2 &= \rho^n \left\|x_{(0)} - x\right\|_2 \\ \therefore \ \left\|x_{(n+1)} - x\right\|_2 &\leq \rho^n \left\|x_{(0)} - x\right\|_2\end{aligned}$$

In this case, $\rho < 1$ is a sufficient condition for the convergence of the series $\{x_n\}$ to x.

$$\begin{aligned}\rho &< 1 \\ \iff 2(\frac{1+U_{2k}}{1-L_{2k}})\delta_{3k} &< 1 \\ \implies (\frac{1+U_{2k}}{1-L_{2k}})\delta_{3k} &< \frac{1}{2}\end{aligned}$$

□

7. Empirical Testing Results

We compared our implementations of IHT and NIHT with J.D Blanchard's implementation of OMP, using a testing suite provided by him. Overall, our results for NIHT were similar to OMP, with NIHT having a slight edge for $0.9 \leq \delta \leq 1.0$, while IHT was worse than both NIHT and OMP.

The following were the plotted results for $n = 1024$.

Figure 1. Recovery Plot for OMP

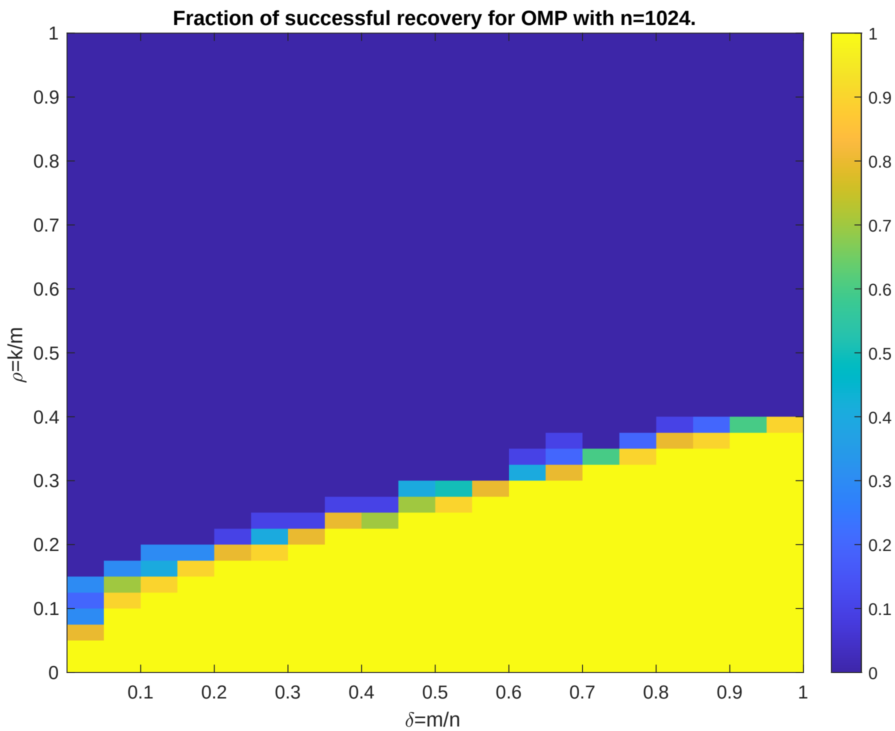


FIGURE 2. Recovery Plot for IHT

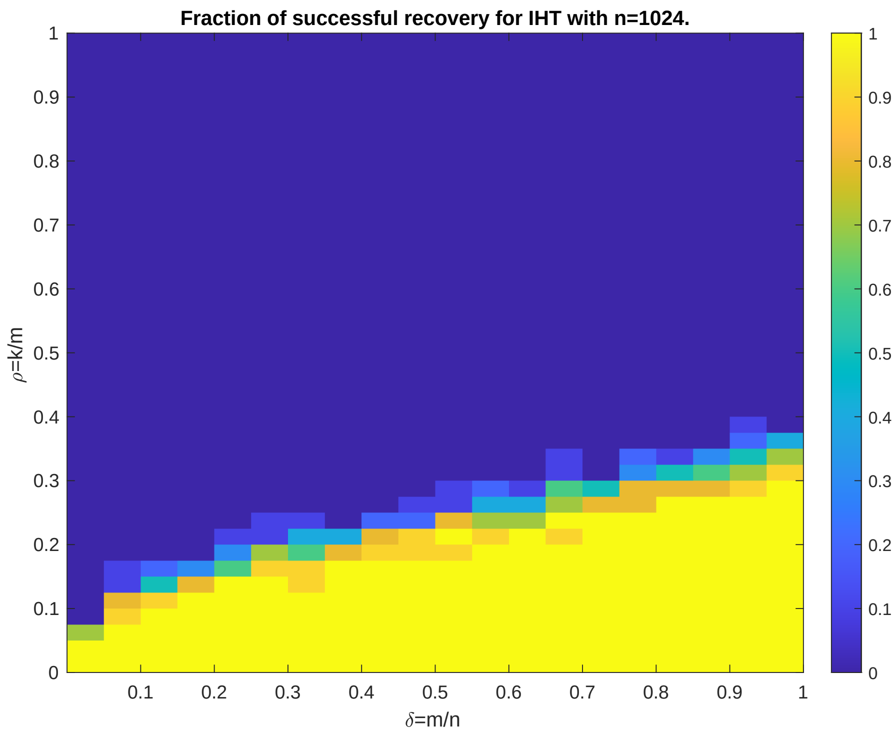
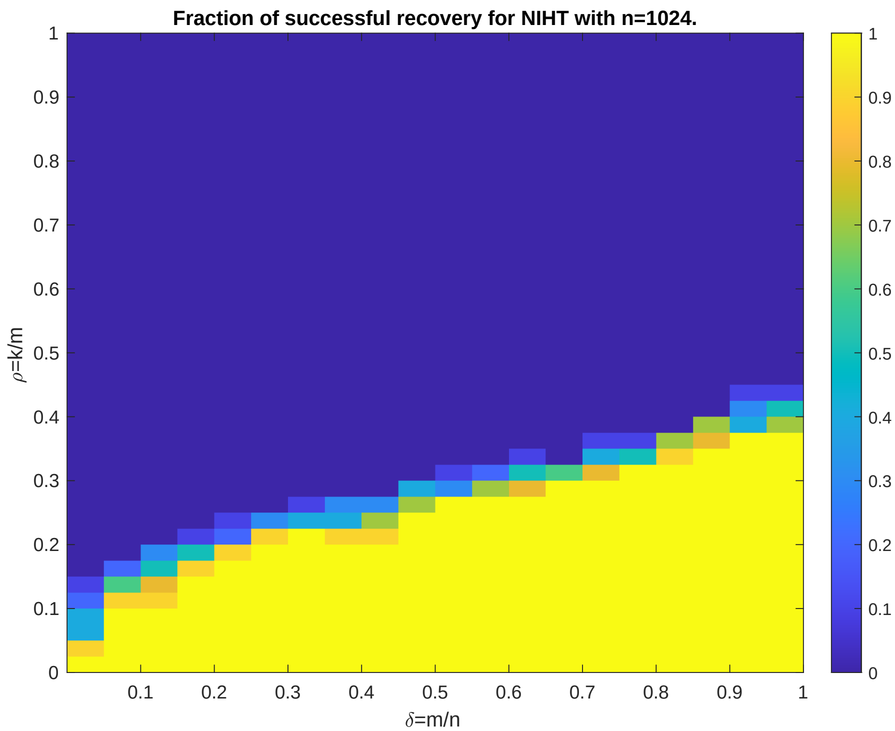


FIGURE 3. Recovery Plot for NIHT



REFERENCES

[1] J.D. Blanchard. "Compressed Sensing Primer: an exploratory introduction". In: *Course Notes* (2023).

[2] Jeffrey D. Blanchard and Jared Tanner. “Performance comparisons of greedy algorithms in compressed sensing”. In: *Numerical Linear Algebra with Applications* 22.2 (2015), pp. 254–282. DOI: https://doi.org/10.1002/nla.1948. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/nla.1948. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/nla.1948.

[3] S. Foucart. “Sparse Recovery Algorithms: Sufficient Conditions in terms of Restricted”. In: *Approximatio Theory XIII: San Antonio, Springer Proceedings in Mathematics* (2010).