A Greedy Pursuit Algorithm For Separating Signals From Nonlinear Compressive Observations

Dung Tran*, **Akshay Rangamani***, Sang (Peter) Chin*^, Trac D. Tran*

*- Electrical & Computer Engg. Department, Johns Hopkins University

^ - Department of Computer Science, Boston University

Introduction – The Unmixing Problem

- We often encounter signals that are superpositions of two (or more) components. This happens in image processing, audio processing, etc.
- The problem of separating out the components of a signal from measurements is called the Unmixing Problem, or Demixing Problem

Image credit: NASA

McCoy, Michael B., et al. "Convexity in source separation: Models, geometry, and algorithms." *IEEE Signal Processing Magazine* 31.3 (2014): 87-95.

Introduction – The Unmixing Problem

- Some related problems:
  - Morphological Components Analysis
  - Robust PCA – separation of a signal into a low rank and sparse component

Candès, Emmanuel J., et al. "Robust principal component analysis?." *Journal of the ACM (JACM)*58.3 (2011): 11.

Unmixing Problem - Formulation

- More formally, we have a signal $\boldsymbol{z} = \boldsymbol{u} + \boldsymbol{v}$ that we would like to separate into its constituent parts.
- This is ill posed in general. We usually make the assumption that the constituent signals $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^N$ have *sparse representations* in *dictionaries* $(\boldsymbol{\Psi}, \boldsymbol{\Phi})$ that are *mutually incoherent.*

$$\boldsymbol{z} = \boldsymbol{\Phi x} + \boldsymbol{\Psi y}$$

- *Mutual coherence* for two dictionaries is defined as:

$$\mu = \max_{||\boldsymbol{x}||=1, ||\boldsymbol{y}||=1} |\langle \boldsymbol{\Phi x}, \boldsymbol{\Psi y} \rangle|$$

The dictionaries are said to be *incoherent* if their mutual coherence is small.

Unmixing Problem - Formulation

- In addition to the fact that the signals are superposed, we only have access to *nonlinear*, *compressive* measurements of the superposition

$$\boldsymbol{y} = h\big(\boldsymbol{A}(\boldsymbol{u} + \boldsymbol{v})\big) + \boldsymbol{e}$$

- $\boldsymbol{A} \in \mathbb{R}^{m \times N}$ is a sensing matrix with $m \ll N$,
- $h: \mathbb{R} \to \mathbb{R}$ is a smooth, monotonic, nonlinear operator that is applied component-wise.
- Our goal is to recover $\boldsymbol{u}, \boldsymbol{v}$ from the measurements $\boldsymbol{y}$
  - We assume that we know the dictionaries $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ in which $\boldsymbol{u}$ and $\boldsymbol{v}$ are respectively sparse
  - We also assume that the sensing matrix and nonlinear operator are known

Unmixing Matching Pursuit (UnmixMP)

- We solve the following optimization problem to unmix the desired signals:

$$\min_{\boldsymbol{u},\boldsymbol{v}} f(\boldsymbol{u},\boldsymbol{v}) = \frac{1}{m}\sum_{j=1}^{m} \Gamma\left(\boldsymbol{a}_j^T(\boldsymbol{u}+\boldsymbol{v})\right) - \boldsymbol{y}_j \boldsymbol{a}_j^T(\boldsymbol{u}+\boldsymbol{v})$$

$$\text{s.t. } \|\boldsymbol{u}\|_{\mathbf{0},\boldsymbol{\Phi}} \leq k, \|\boldsymbol{v}\|_{\mathbf{0},\boldsymbol{\Psi}} \leq s$$

- Here $\Gamma(t) = \int_{-\infty}^{t} h(z)dz$, is the integral of the nonlinear link function.
- $\|\boldsymbol{u}\|_{\mathbf{0},\boldsymbol{\Phi}}$ and $\|\boldsymbol{v}\|_{\mathbf{0},\boldsymbol{\Psi}}$ are the sparsity levels of the signals in their respective dictionaries.
- The objective $f(\boldsymbol{u},\boldsymbol{v})$[1] is different from the usual squared loss function that is usually considered in signal estimation problems. A similar objective was considered in [1]

[1] Soltani, Mohammadreza, and Chinmay Hegde. "Fast algorithms for demixing sparse signals from nonlinear observations." *arXiv preprint arXiv:1608.01234* (2016).

UnmixMP Algorithm

Input: Observations $\boldsymbol{y}$, sensing matrix $\boldsymbol{A}$, dictionaries $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$,
Nonlinear operator h, stopping criterion.

Output: Unmixed signals $\boldsymbol{u}$ and $\boldsymbol{v}$

Initialization: $t = 0, \;\; \Omega_u^0 = \Omega_v^0 = \emptyset, \;\; \boldsymbol{u}^0, \boldsymbol{v}^0 = \boldsymbol{0}$

while not converged:

1 $\boldsymbol{g} = \frac{1}{m}\boldsymbol{A}^T(h(\boldsymbol{A}\boldsymbol{u}^t + \boldsymbol{A}\boldsymbol{v}^t) - \boldsymbol{y})$

2 (**Selection Step**) $i_u = \operatorname{argmin}\|\text{Proj}_{\boldsymbol{\Phi}}\boldsymbol{g}\|_2, \quad i_v = \operatorname{argmin}\|\text{Proj}_{\boldsymbol{\Psi}}\boldsymbol{g}\|_2$

3 $\Omega_u^{t+1} = \Omega_u^t \cup \{i_u\}, \quad \Omega_v^{t+1} = \Omega_v^t \cup \{i_v\}$

4 **(Update Step)** $(\boldsymbol{u}^{t+1}, \boldsymbol{v}^{t+1}) = \operatorname{argmin} f(\boldsymbol{u},\boldsymbol{v})$

s.t. $\boldsymbol{u} \in \operatorname{span}\left(\boldsymbol{\Phi}_{\Omega_u^{t+1}}\right), \boldsymbol{v} \in \operatorname{span}\left(\boldsymbol{\Psi}_{\Omega_v^{t+1}}\right)$

5 $t = t + 1$

Theoretical Guarantees for UnmixMP

Definition – $(\boldsymbol{k}, \boldsymbol{s})$ **Restricted Strong Convexity / Restricted Strong Smoothness**

A function f is $(\boldsymbol{k}, \boldsymbol{s})$-RSC/RSS with parameters $m_{k,s}$ and $M_{k,s}$ if

$$m_{k,s}\left(\|\widetilde{\boldsymbol{u}} - \boldsymbol{u}\|_2^2 + \|\widetilde{\boldsymbol{v}} - \boldsymbol{v}\|_2^2\right)$$
$$\leq f(\widetilde{\boldsymbol{u}}, \widetilde{\boldsymbol{v}}) - f(\boldsymbol{u}, \boldsymbol{v}) - \langle \nabla_{\boldsymbol{u}} f(\boldsymbol{u}, \boldsymbol{v}), \widetilde{\boldsymbol{u}} - \boldsymbol{u} \rangle - \langle \nabla_{\boldsymbol{v}} f(\boldsymbol{u}, \boldsymbol{v}), \widetilde{\boldsymbol{v}} - \boldsymbol{v} \rangle$$
$$\leq M_{k,s}\left(\|\widetilde{\boldsymbol{u}} - \boldsymbol{u}\|_2^2 + \|\widetilde{\boldsymbol{v}} - \boldsymbol{v}\|_2^2\right)$$

for all $\boldsymbol{u}, \widetilde{\boldsymbol{u}} \in \boldsymbol{\mathcal{S}}_k^{\boldsymbol{u}}$ and $\boldsymbol{v}, \widetilde{\boldsymbol{v}} \in \boldsymbol{\mathcal{S}}_s^{\boldsymbol{v}}$.

Here $\boldsymbol{\mathcal{S}}_k^{\boldsymbol{u}}$ $(\boldsymbol{\mathcal{S}}_s^{\boldsymbol{v}})$ are unions of subspaces spanned by all subsets of columns of $\boldsymbol{\Phi}$ $(\boldsymbol{\Psi})$ of size k (s)

Theoretical Guarantees for UnmixMP

Theorem 1 (Convergence of UnmixMP).

Suppose f is $(\boldsymbol{k}, \boldsymbol{s})$-RSC/RSS with parameters $m_{k,s}$ and $M_{k,s}$. Let $(\boldsymbol{u}^*, \boldsymbol{v}^*)$ be optimal solution of our optimization problem. Then under some mild condition on $m_{k,s}$ and $M_{k,s}$, the following holds:

$$\left\|\boldsymbol{u}^{t+1} - \boldsymbol{u}^*\right\|_2 + \left\|\boldsymbol{v}^{t+1} - \boldsymbol{v}^*\right\|_2 \leq \eta^t \left(\left\|\boldsymbol{u}^0 - \boldsymbol{u}^*\right\|_2 + \left\|\boldsymbol{v}^0 - \boldsymbol{v}^*\right\|_2\right) + C$$

with convergence rate $\eta < 1$. Here, C is a small quantity depending on the sparsities of the optimal signals $(\boldsymbol{u}^*, \boldsymbol{v}^*)$, m (the number of measurements), and the noise level.

Theoretical Guarantees for UnmixMP

Theorem 2 (Sample complexity).

Let the elements of $\boldsymbol{A} \in \mathbb{R}^{m \times N}$ be drawn from a zero mean Gaussian distribution. Assume that the absolute value of the derivative of h is bounded and the constituent dictionaries $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ are sufficiently *mutually incoherent*.

If the number of measurements $m = \mathcal{O}\left((s+k)\log\frac{N}{s+k}\right)$, the assumptions of **Theorem 1** hold with high probability.

Experimental Results – Signal Model

- We generate signals $\boldsymbol{u}$ and $\boldsymbol{v}$ which are s-sparse in the DCT and identity dictionaries, respectively. We then generate *nonlinear compressive* measurements $\boldsymbol{y}$ according to:

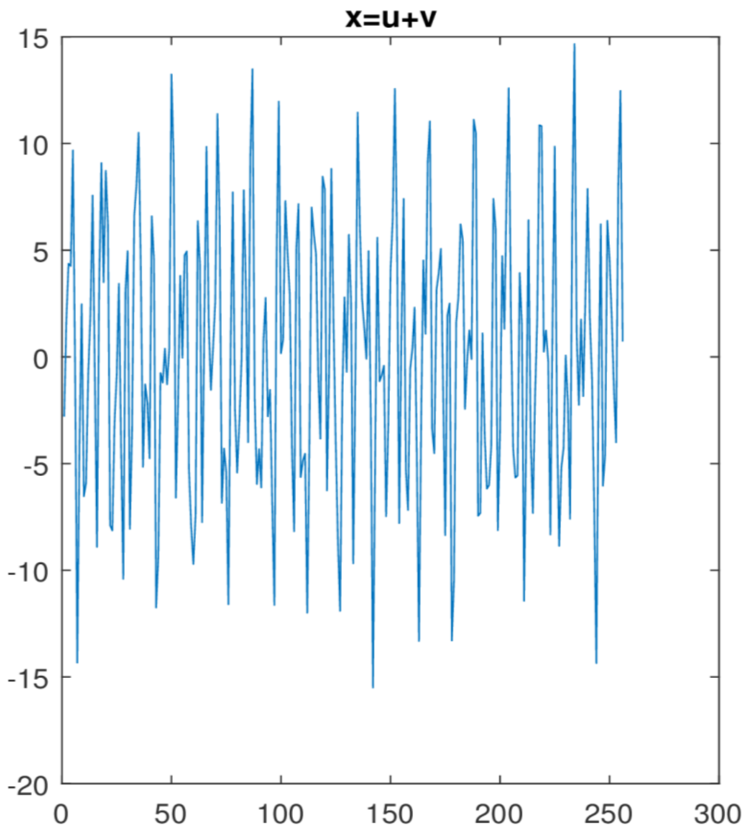
$$\boldsymbol{y} = h\big(\boldsymbol{A}(\boldsymbol{u} + \boldsymbol{v})\big)$$

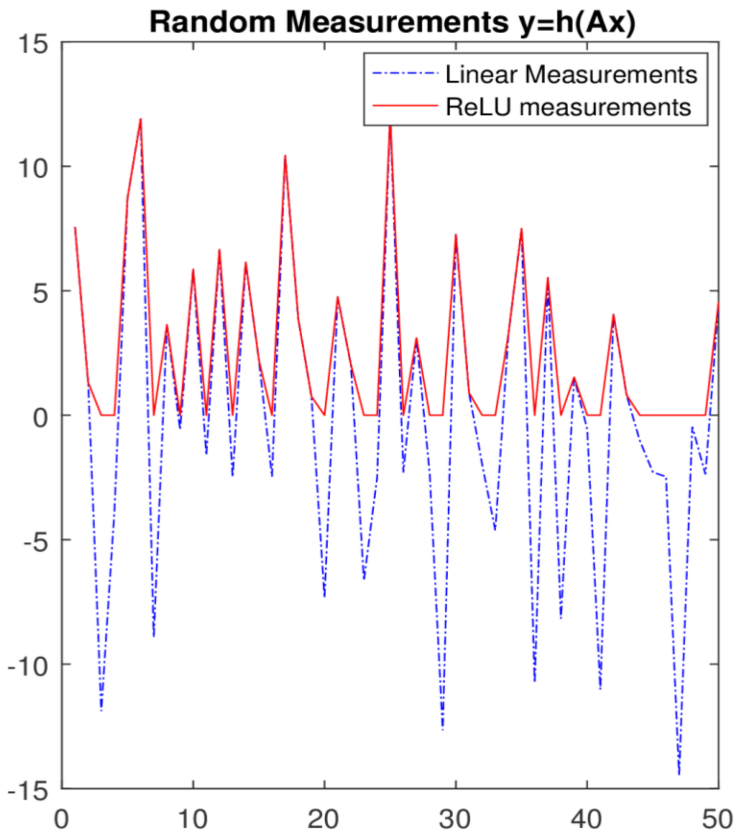
- The sensing matrix $\boldsymbol{A}$ is a random Gaussian matrix. We experiment with two different choices for the nonlinear operator h – the sigmoid function and the ReLU function.

$$h(x) = \frac{1}{1 + e^{-x}}$$

$$h(x) = \max(0, x)$$

Experimental Results – Signal Model



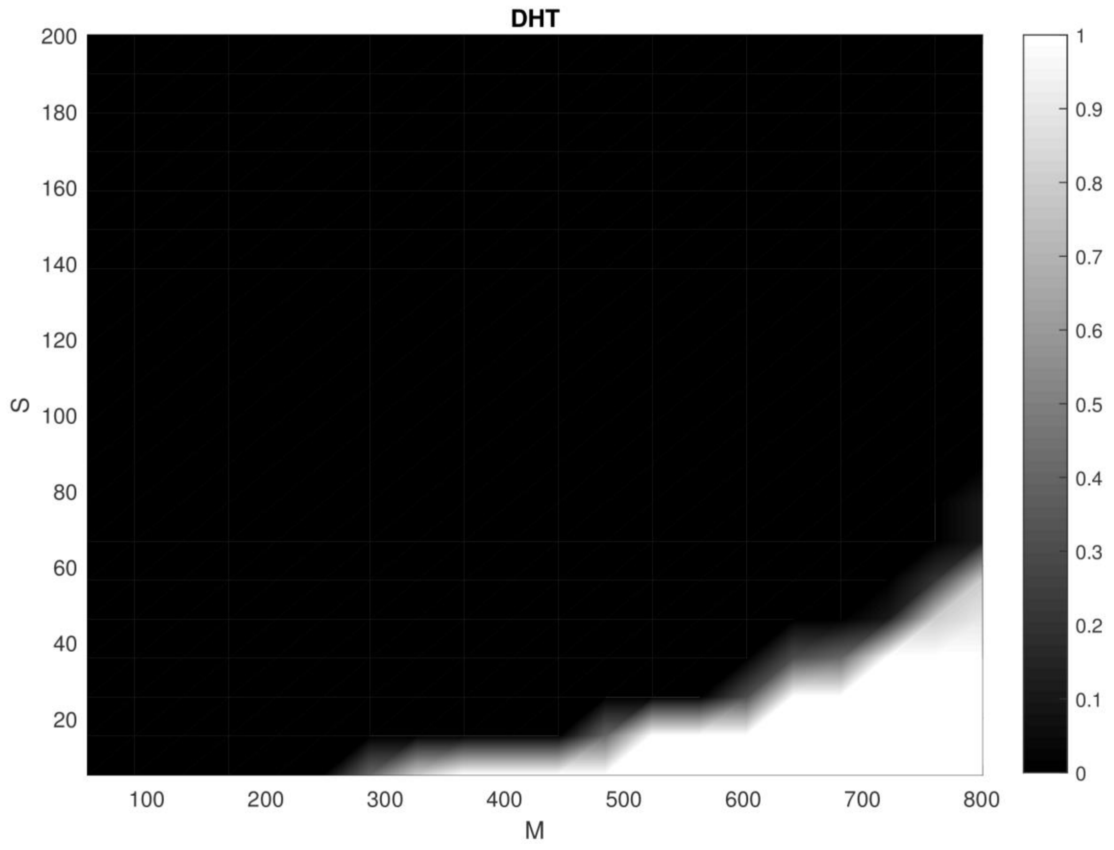


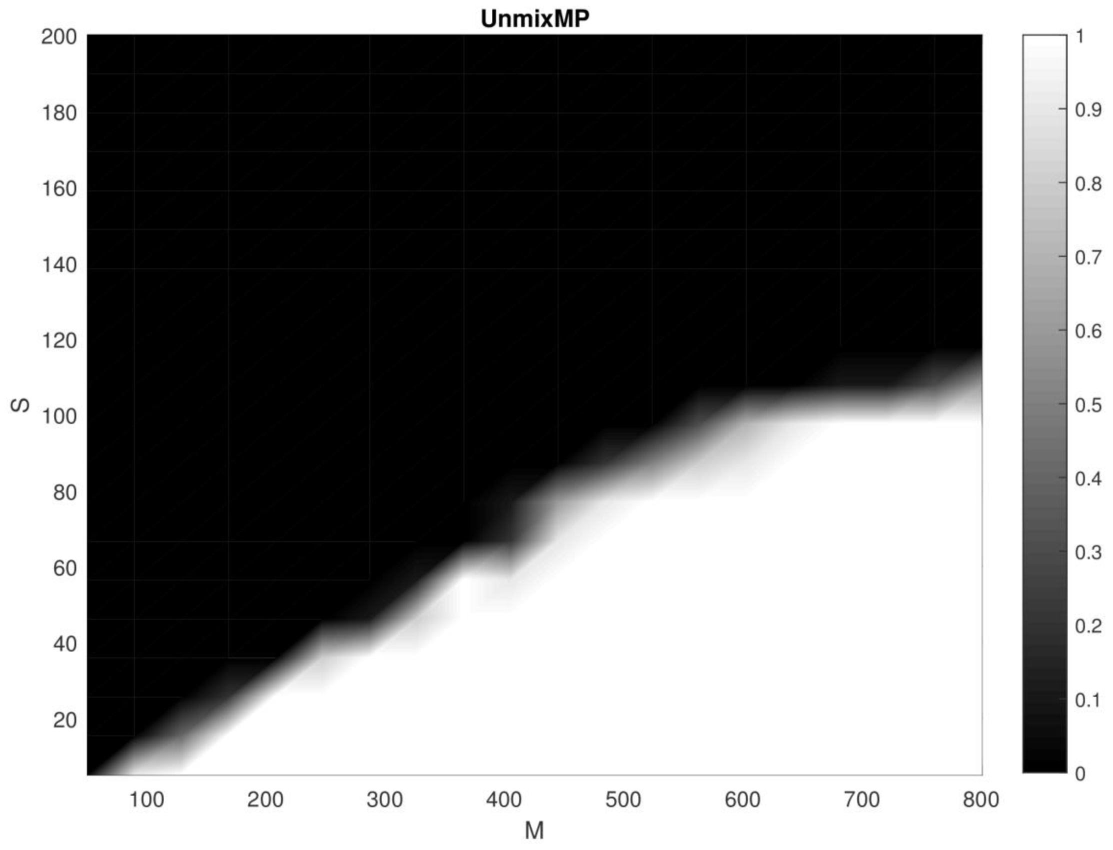
Experimental Results

- We compare the performance of UnmixMP to another algorithm that uses a similar framework – DHT [1] (Demixing with Hard Thresholding).
- We recover the unmixed signals $\hat{\boldsymbol{u}}$ and $\hat{\boldsymbol{v}}$ using UnmixMP and DHT, and measure their quality using normalized ℓ_2 error.

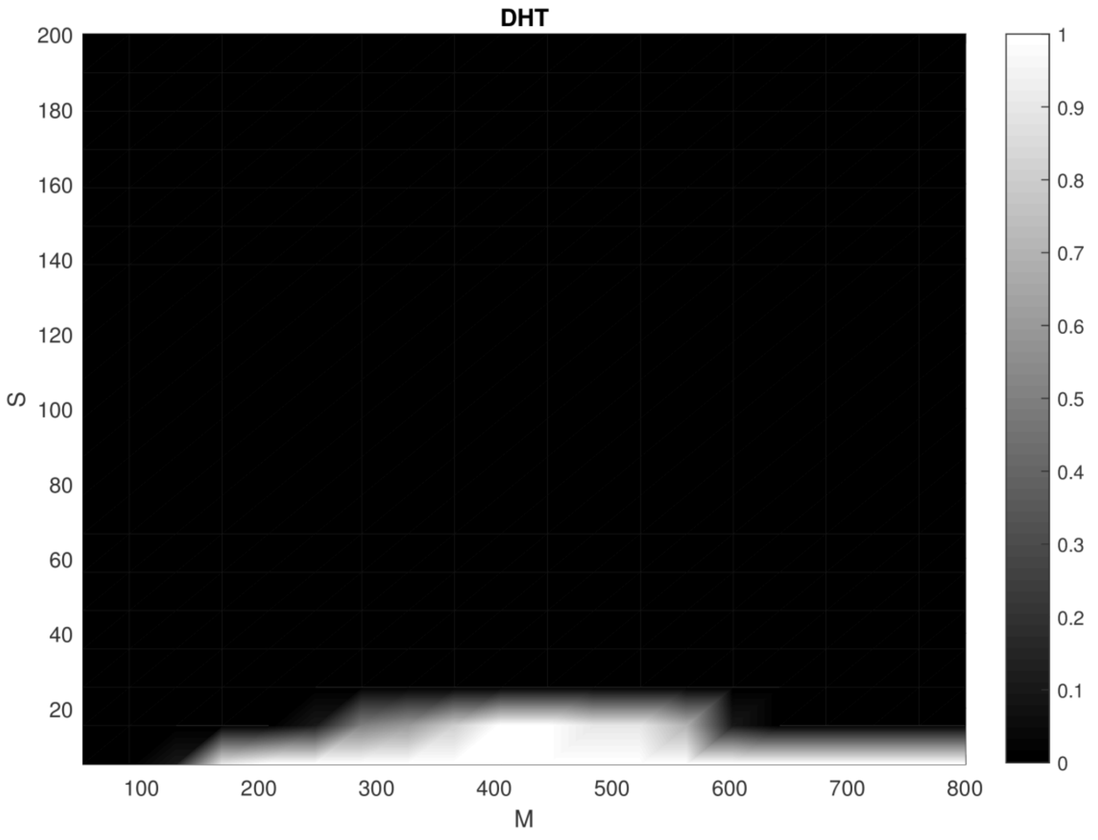
[1] Soltani, Mohammadreza, and Chinmay Hegde. "Fast algorithms for demixing sparse signals from nonlinear observations." *arXiv preprint arXiv:1608.01234* (2016).

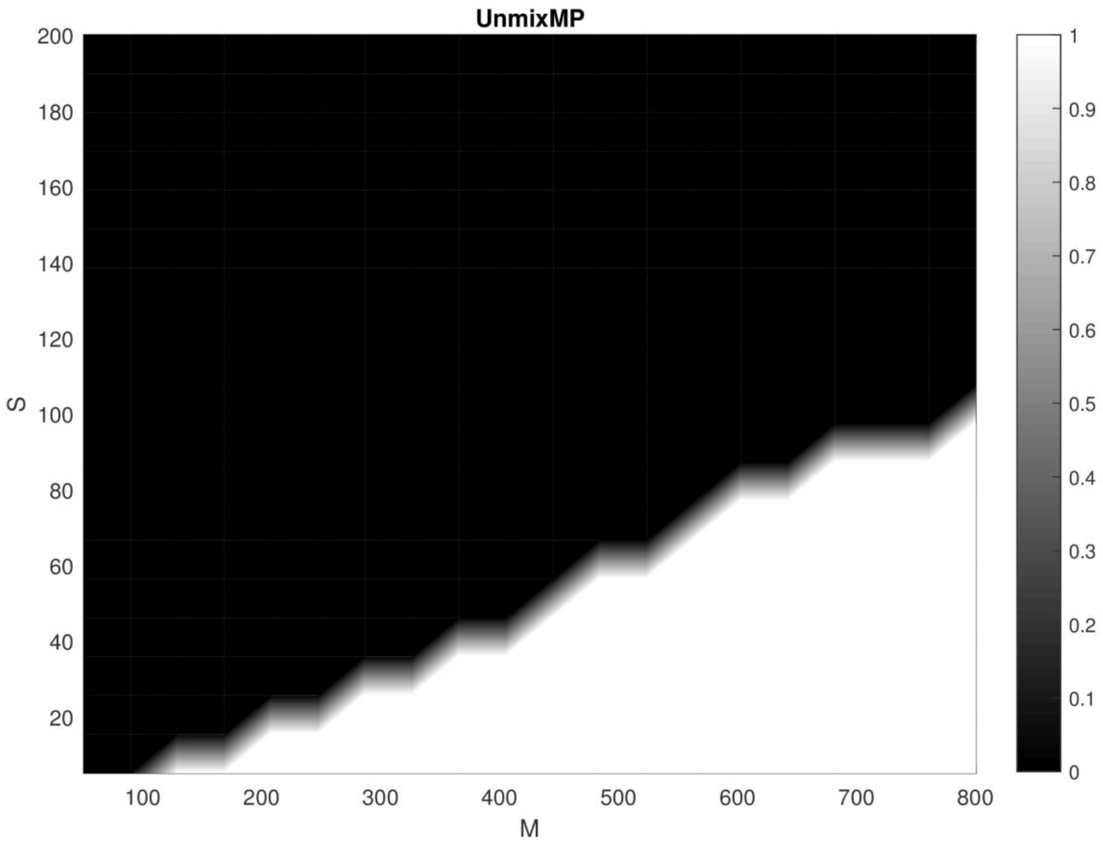
Experimental Results – Sigmoid Nonlinearity





Experimental Results – ReLU Nonlinearity





Conclusions

- We propose UnmixMP, a new greedy pursuit algorithm to separate signals from *nonlinear, compressive* measurements.
- We prove that UnmixMP converges linearly to the optimal solution, and also derive bounds on its sample complexity.
- We support these theoretical results with experimental validation, and improve upon results using DHT (Demixing with Hard Thresholding).
- Even though our convergence results require the nonlinear function to be smooth, we are still able to recover signals from non-smooth measurements like ReLU.