Signal Recovery with Generative Priors

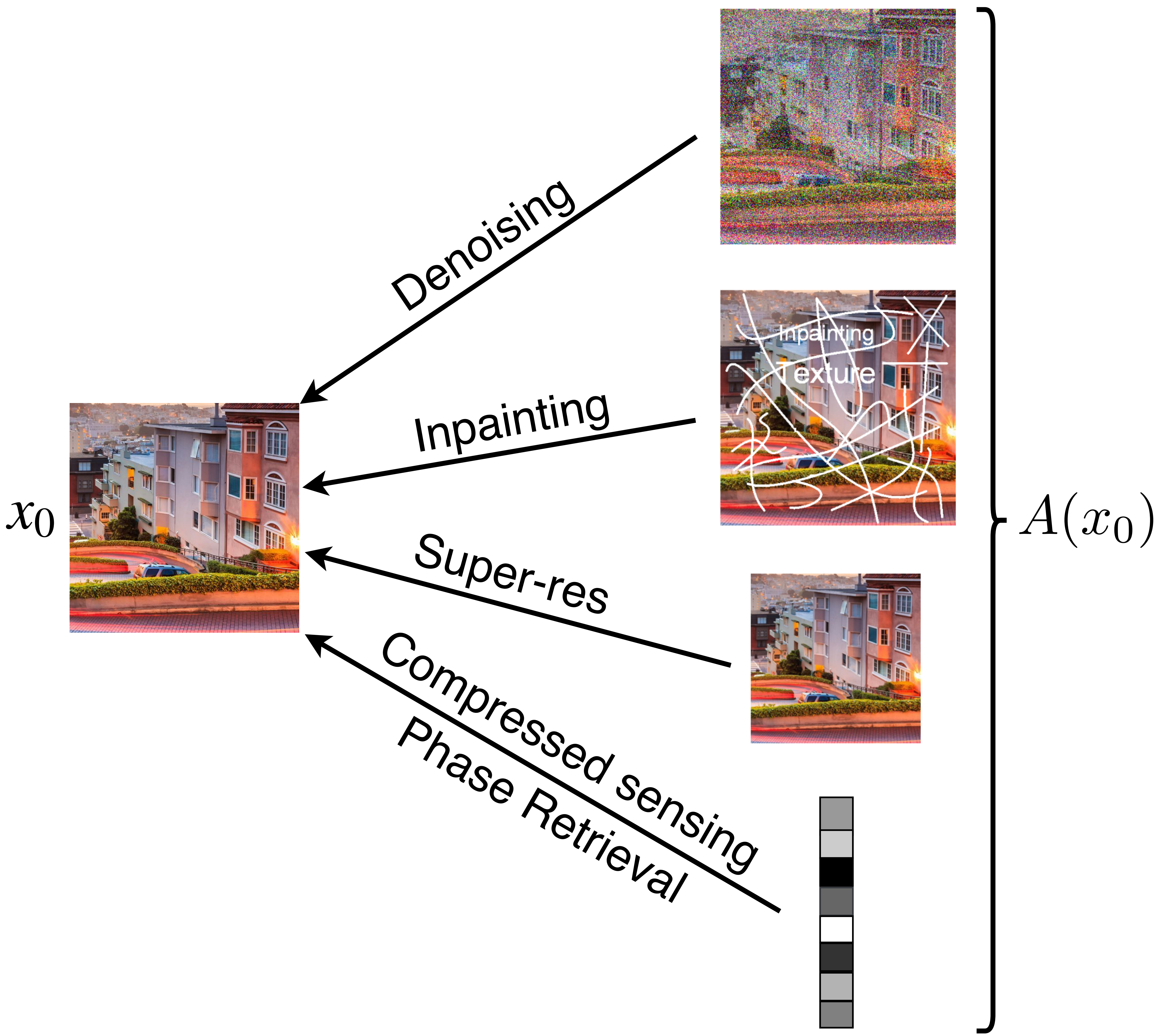
Paul Hand

Northeastern University

Collaborators: V. Voroninski, W. Huang, O. Leong, R. Heckel, J. Cocola, B. Joshi

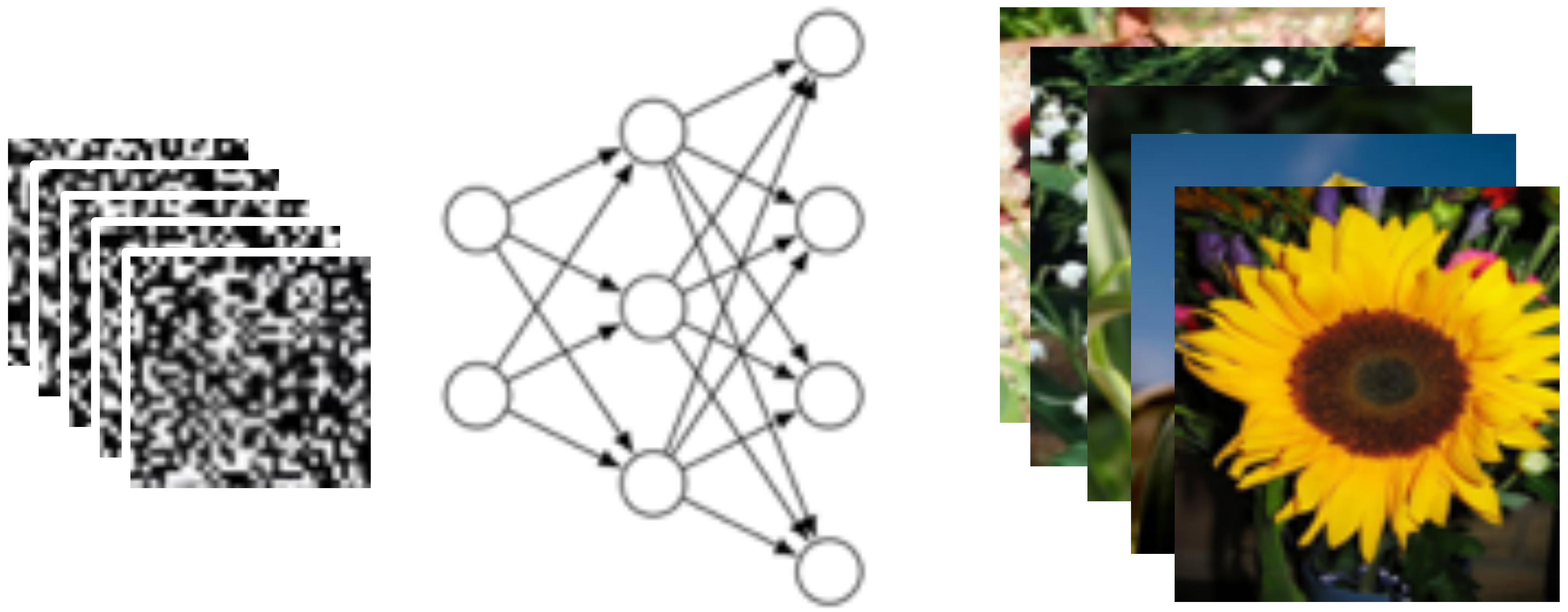
Funding: National Science Foundation

Examples of inverse problems



How can generative models used in inverse problems?

1. Train generative model to output signal class:



2. Directly optimize over range of generative model via empirical risk:

$$\min_{z \in \mathbb{R}^k} \left\| A(G(z)) - A(x_0) \right\|^2$$

Bora, Jalal, Dimakis, Price 2017

Recovery Theory with Random Generative Priors for Multiple Inverse Problems

- **Compressed Sensing**
  Deterministic Conditions and Probabilistic Recovery Guarantee

- **Other Inverse Problems**
  Denoising, Compressive Phase Retrieval, Spiked Matrix Recovery

- **Advances to the Theory**
  Convolutional Structure, Expansivity

Recovery Theory with Random Generative Priors for Multiple Inverse Problems

- **Compressed Sensing**
  Deterministic Conditions and Probabilistic Recovery Guarantee

- **Other Inverse Problems**
  Denoising, Compressive Phase Retrieval, Spiked Matrix Recovery

- **Advances to the Theory**
  Convolutional Structure, Expansivity

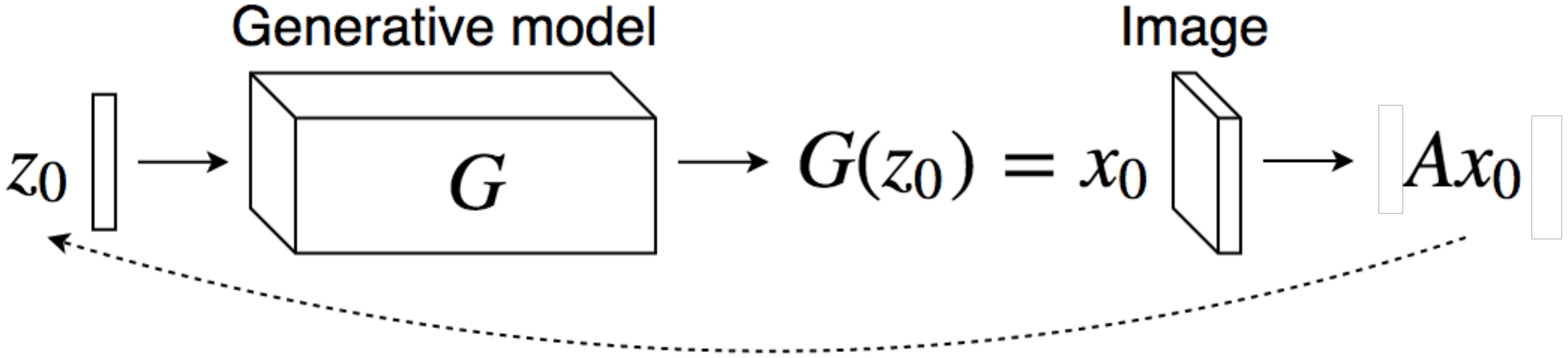
Inspiration: Recovery Theory for Compressed Sensing with Sparse Priors

Fix k-sparse vector $x_0 \in \mathbb{R}^n$.
Let $A \in \mathbb{R}^{m \times n}$ be a random gaussian matrix with $m = \Omega(k \log n)$.

$$\begin{array}{ll} \text{min} & \|x\|_1 \\ \text{s.t.} & Ax = Ax_0 \end{array} \tag{L1}$$

Theorem (Candes, Romberg, Tao. 2004. Donoho, 2004.)
The global minimizer of (L1) is x_0 with high probability.

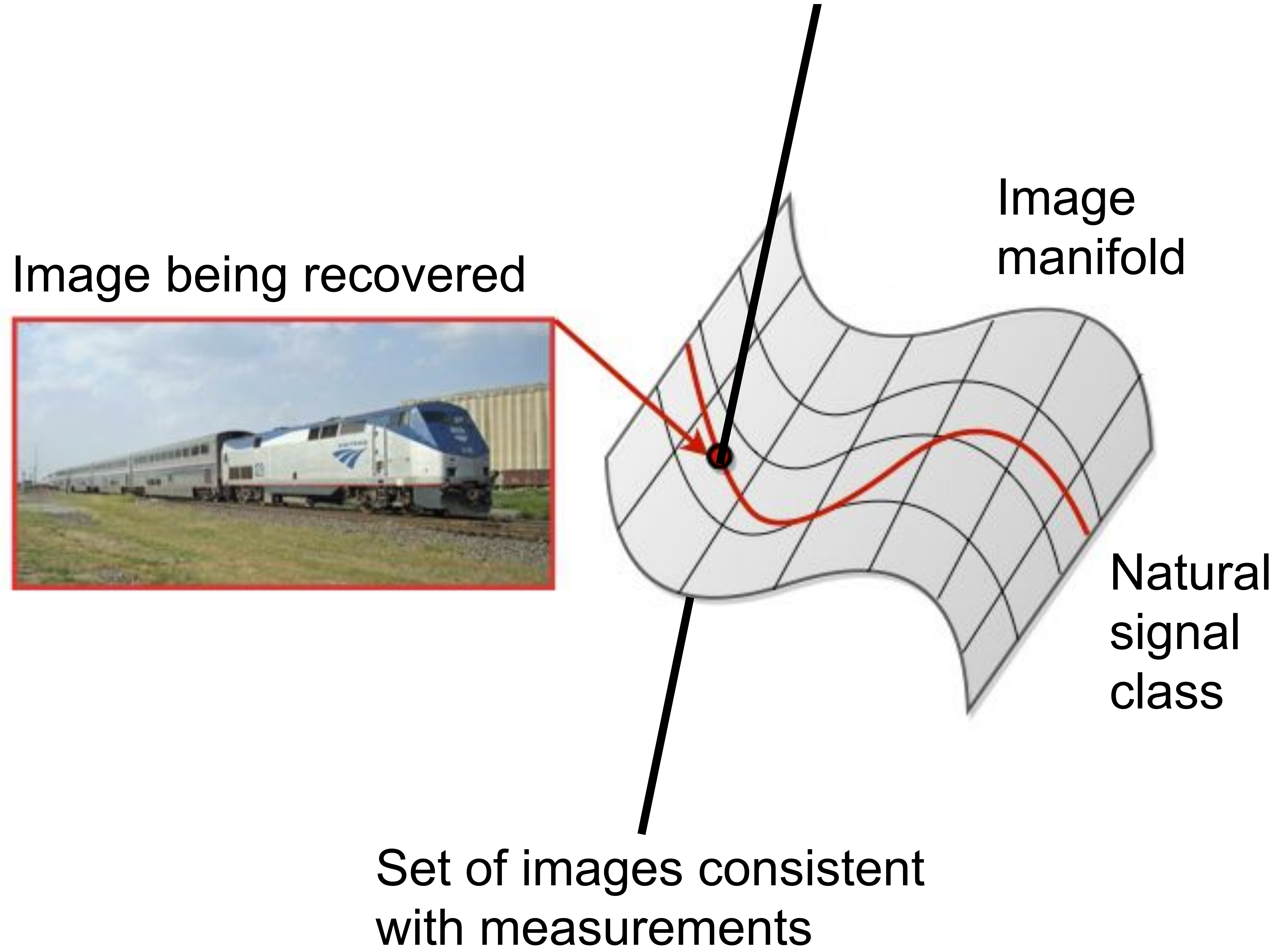
Compressed Sensing with Generative Models



$$\min_{z \in \mathbb{R}^k} \left\| AG(z) - Ax_0 \right\|^2$$

Bora, Jalal, Dimakis, Price 2017

Geometric picture of signal recovery with a low-dimensional generative prior



Expansive Gaussian Model for Generative Priors

Let: $G : \mathbb{R}^k \to \mathbb{R}^n$

$G(z) = \mathrm{relu}(W_d \dots \mathrm{relu}(W_2 \mathrm{relu}(W_1 z)) \dots)$

Given: $W_i \in \mathbb{R}^{n_i \times n_{i-1}}, A \in \mathbb{R}^{m \times n}, y := AG(z_0) \in \mathbb{R}^m$

Find: x_0

- **Expansivity**: Let $n_i > c n_{i-1} \log n_{i-1}$
- **Gaussianicity**: Let W_i and A have iid Gaussian entries.
- **Biasless**: No bias terms in G.

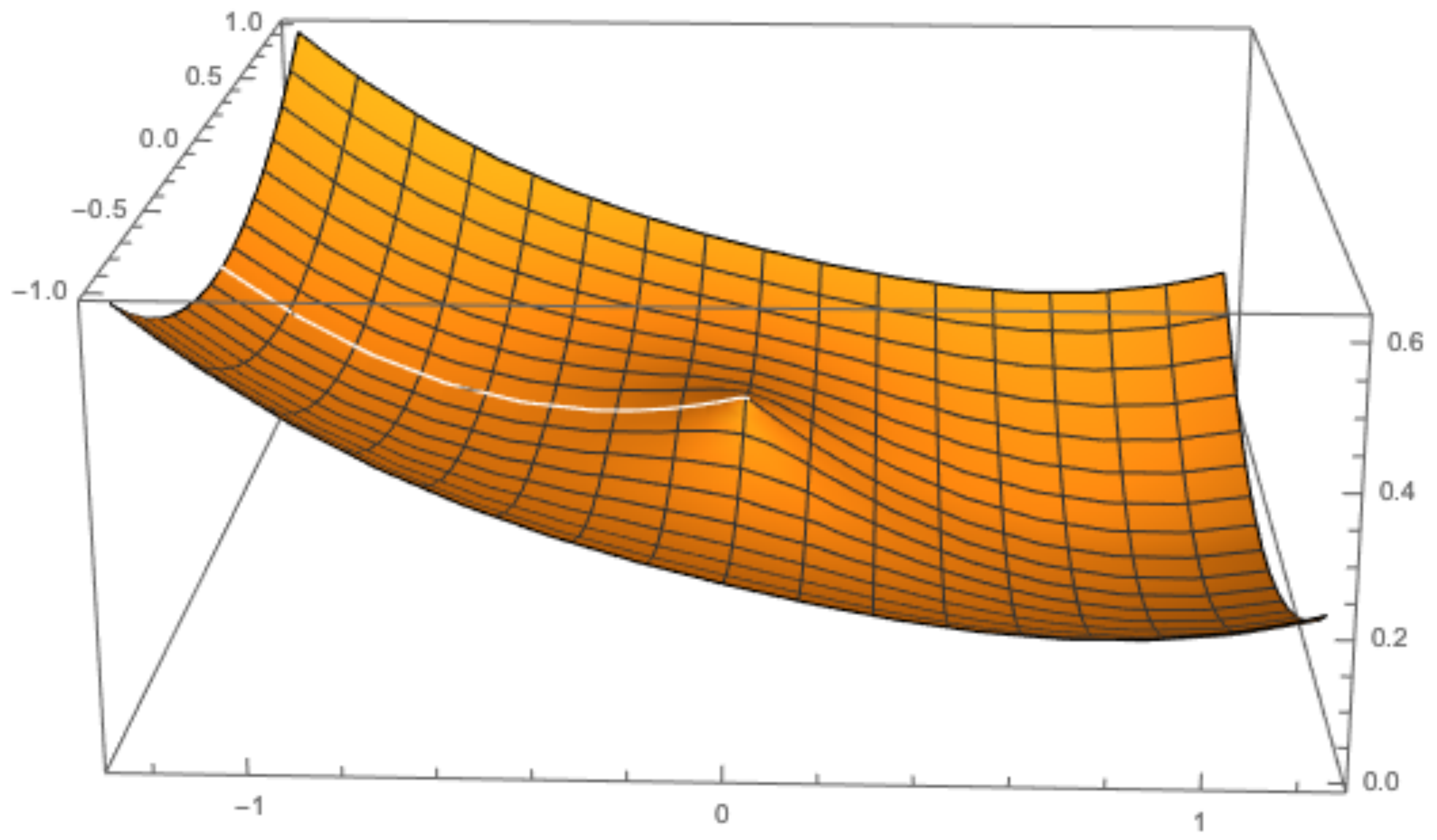
Compressive sensing with random generative prior has a provably convergent subgradient descent algorithm

Theorem (Huang, Heckel, Hand, Voroninski)

Let $d \geq 2$. If

1. $\mathcal{G}$ *is gaussian and sufficiently expansive,*
2. $m = \Omega(kd \log(\Pi_{i=1}^{d} n_i)$,
3. *measurements have sufficiently small noise*

then w.h.p. a subgradient descent with a twist converges to within the noise level of z_0.



Deterministic Condition used in Recovery Proofs

A matrix $W \in \mathbb{R}^{n\times k}$ satisfies the **Weight Distribution Condition** with constant ϵ if for all $x, y \neq 0 \in \mathbb{R}^k$,

$$\left\| \sum_{i=1}^{n} 1_{w_i\cdot x>0} 1_{w_i\cdot y>0} \cdot w_i w_i^T - Q_{x,y} \right\| \leq \epsilon, \text{ with } Q_{x,y} = \frac{\pi-\theta}{2\pi} I + \frac{\sin\theta}{2\pi} M_{\hat{x}\leftrightarrow\hat{y}}.$$

Here, w_i^T is the ith row of W; $M_{\hat{x}\leftrightarrow\hat{y}} \in \mathbb{R}^{k\times k}$ is the matrix such that $\hat{x} \mapsto \hat{y}$, $\hat{y} \mapsto \hat{x}$, and $\hat{z} \mapsto 0$ for all $z \in (\{x, y\})^\perp$; $\theta = \angle(x, y)$.

Definition 3 (Range Restricted Isometry Condition (RRIC) [52])**.** *A matrix* $\mathbf{A} \in \mathbb{R}^{m\times n}$ *satisfies the* Range Restricted Isometry Condition *with respect to* G *with constant* ϵ *if, for all* $\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \mathbf{z}_4 \in \mathbb{R}^k$*, it holds that*

$$\begin{aligned}
&\Big|\langle \mathbf{A}(G(\mathbf{z}_1) - G(\mathbf{z}_2)), \mathbf{A}(G(\mathbf{z}_3) - G(\mathbf{z}_4))\rangle \\
&\qquad - \langle G(\mathbf{z}_1) - G(\mathbf{z}_2), G(\mathbf{z}_3) - G(\mathbf{z}_4)\rangle\Big| \\
&\qquad \leq \epsilon \|G(\mathbf{z}_1) - G(\mathbf{z}_2)\|_2 \|G(\mathbf{z}_3) - G(\mathbf{z}_4)\|_2. \qquad (23)
\end{aligned}$$

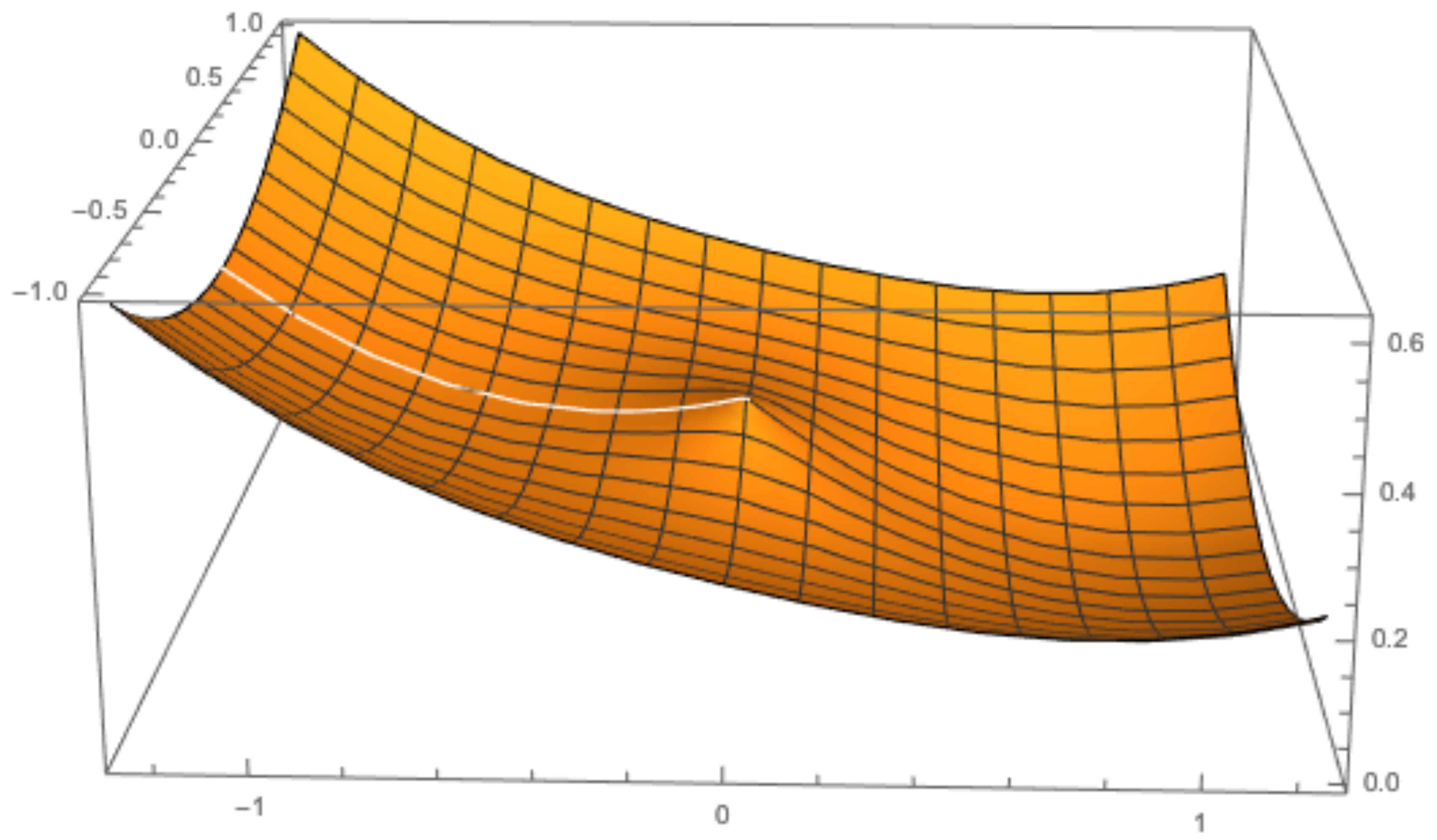
This condition states that $\mathbf{A}$ acts like an isometry when acting on pairs of secant directions (i.e., differences of two signals) with respect to the range of G.

Theorem (Huang, Hand, Heckel, Voroninski, 2017)

Assume that

1. the WDC and RRIC hold with $\epsilon \leq C/\text{poly}(d)$,
2. the measurement noise is sufficiently small.

Then a subgradient algorithm with a twist converges to within the noise level of z_0.



Weight Distribution Condition holds w.h.p. if a matrix is sufficiently expansive

Lemma: Fix ϵ. Let $W \in \mathbb{R}^{n \times k}$ have i.i.d. $\mathcal{N}(0, 1/n)$ entries. If $n > ck \log k$, then with probability at least $1 - 8ne^{-\gamma k}$, we have for all $x, y \neq 0 \in \mathbb{R}^k$,

$$\left\| \sum_{i=1}^{n} 1_{w_i \cdot x > 0} 1_{w_i \cdot y > 0} \cdot w_i w_i^T - \mathbb{E}[\cdots] \right\| \leq \epsilon$$

The constants depend polynomially on ϵ.

Range Restricted Isometry Condition holds w.h.p. if a matrix is sufficiently expansive

Lemma (Baraniuk et al. 2008)

Let $A \in \mathbb{R}^{m \times n}$ have iid $\mathcal{N}(0, 1/m)$ entries. Fix ϵ. Fix a subspace $T \subset \mathbb{R}^n$ of dimension $2k < m$. With probability at least $1 - (c_1/\epsilon)^{2k} e^{-\gamma_1 \epsilon m}$,

$$|\langle Ax, Ay\rangle - \langle x, y\rangle| \leq \epsilon \|x\|_2 \|y\|_2, \quad \forall x, y \in T$$

To establish RRIC:

- Apply lemma to all subspaces given by ReLU patterns
- $\|W_{+,x}^t A^t A W_{+,y} - W_{+,x}^t W_{+,y}\| \leq \epsilon \; \forall x, y$ whp if $m \gtrsim k \log n$

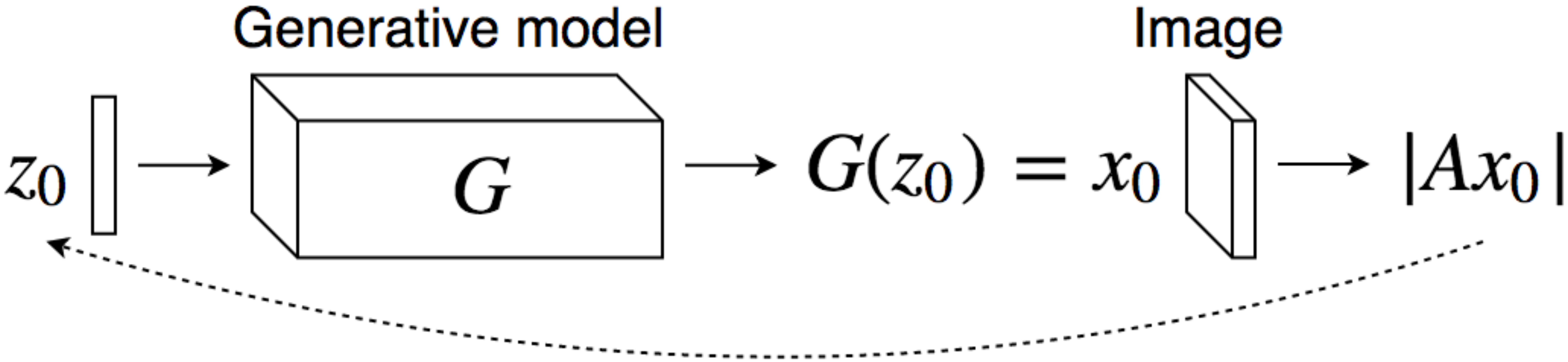
Recovery Theory with Random Generative Priors for Multiple Inverse Problems

- **Compressed Sensing**
  Deterministic Conditions and Probabilistic Recovery Guarantee

- **Other Inverse Problems**
  Denoising, Compressive Phase Retrieval, Spiked Matrix Recovery

- **Advances to the Theory**
  Convolutional Structure, Expansivity

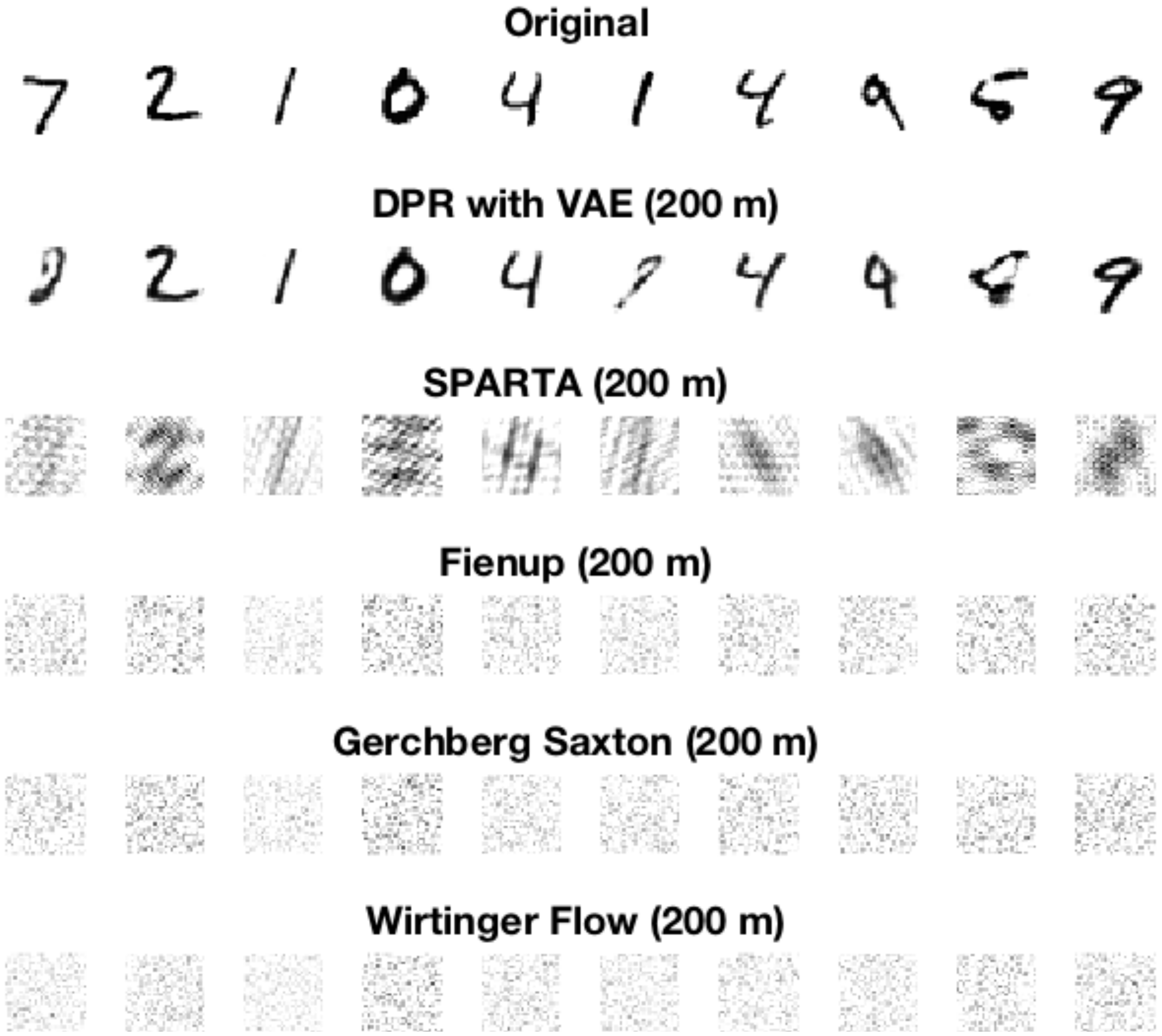
Compressive Phase Retrieval with Generative Models



$$\min_{z\in\mathbb{R}^k} \left\| |AG(z)| - |Ax_0| \right\|^2$$

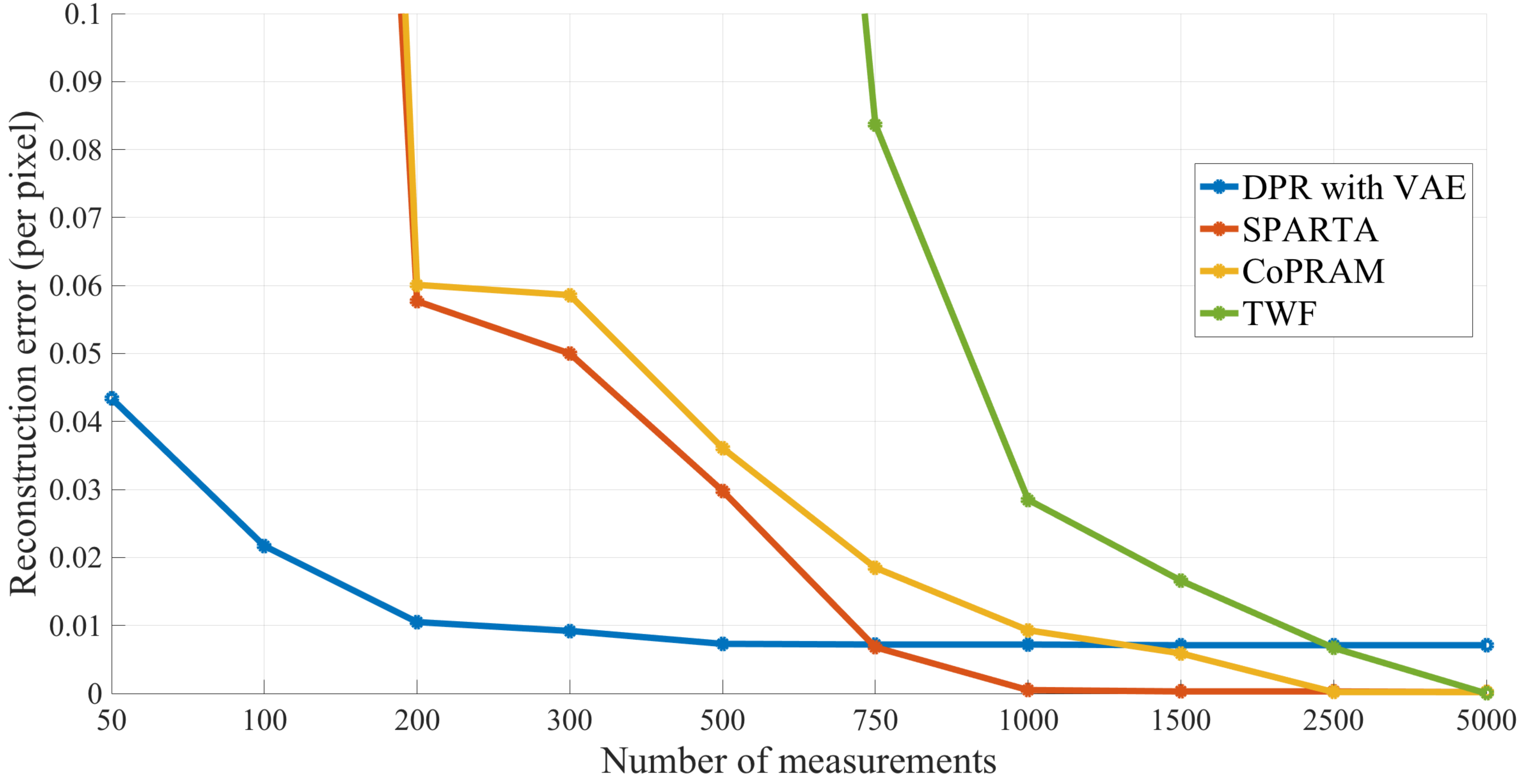
Hand, Leong, Voroninski 2018

Compressive Phrase Retrieval on MNIST



Hand, Leong, Voroninski 2018

Deep phase retrieval can outperform sparse phase retrieval in the low measurement regime



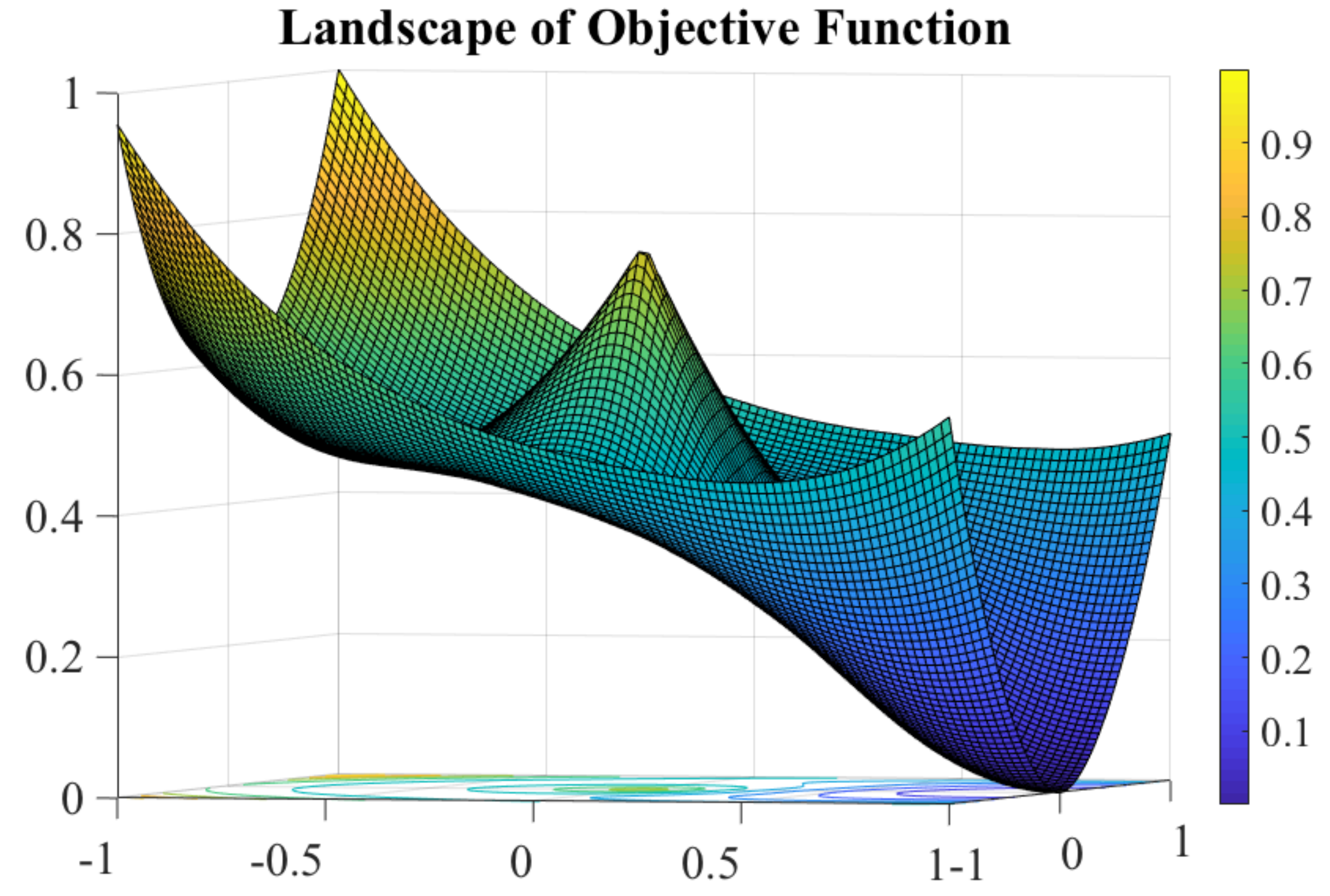
Hand, Leong, Voroninski 2018

Compressive phase retrieval from generic measurements is possible at optimal sample complexity

1. $\#$ measurements $= \Omega(k)$, up to log factors
2. network layers are sufficiently expansive
3. A and weights of G have i.i.d. Gaussian entries

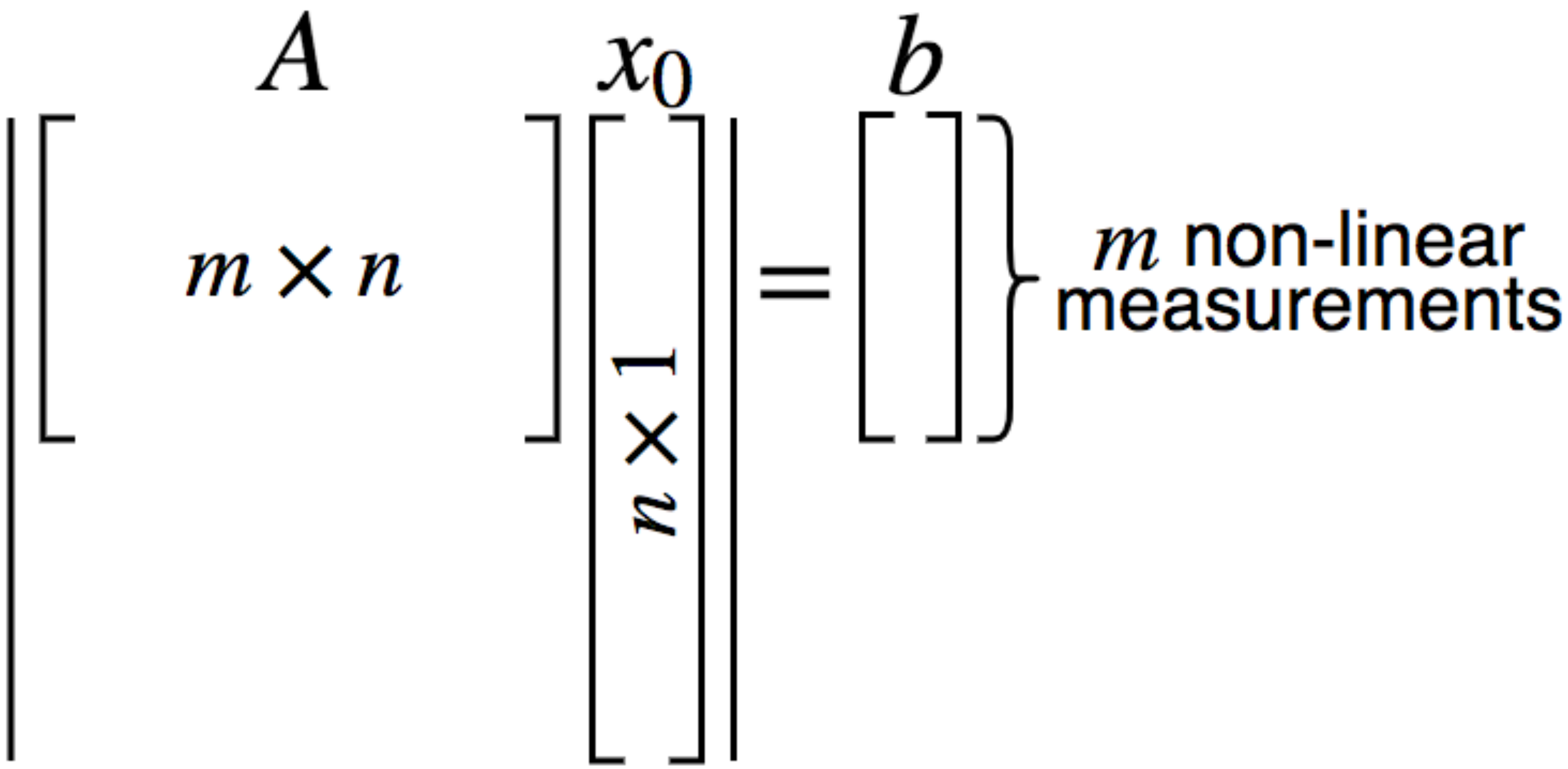
Theorem (Hand, Leong, Voroninski)

The objective function has a strict descent direction in latent space outside of two small neighborhoods of the minimizer and a negative multiple thereof, with high probability.



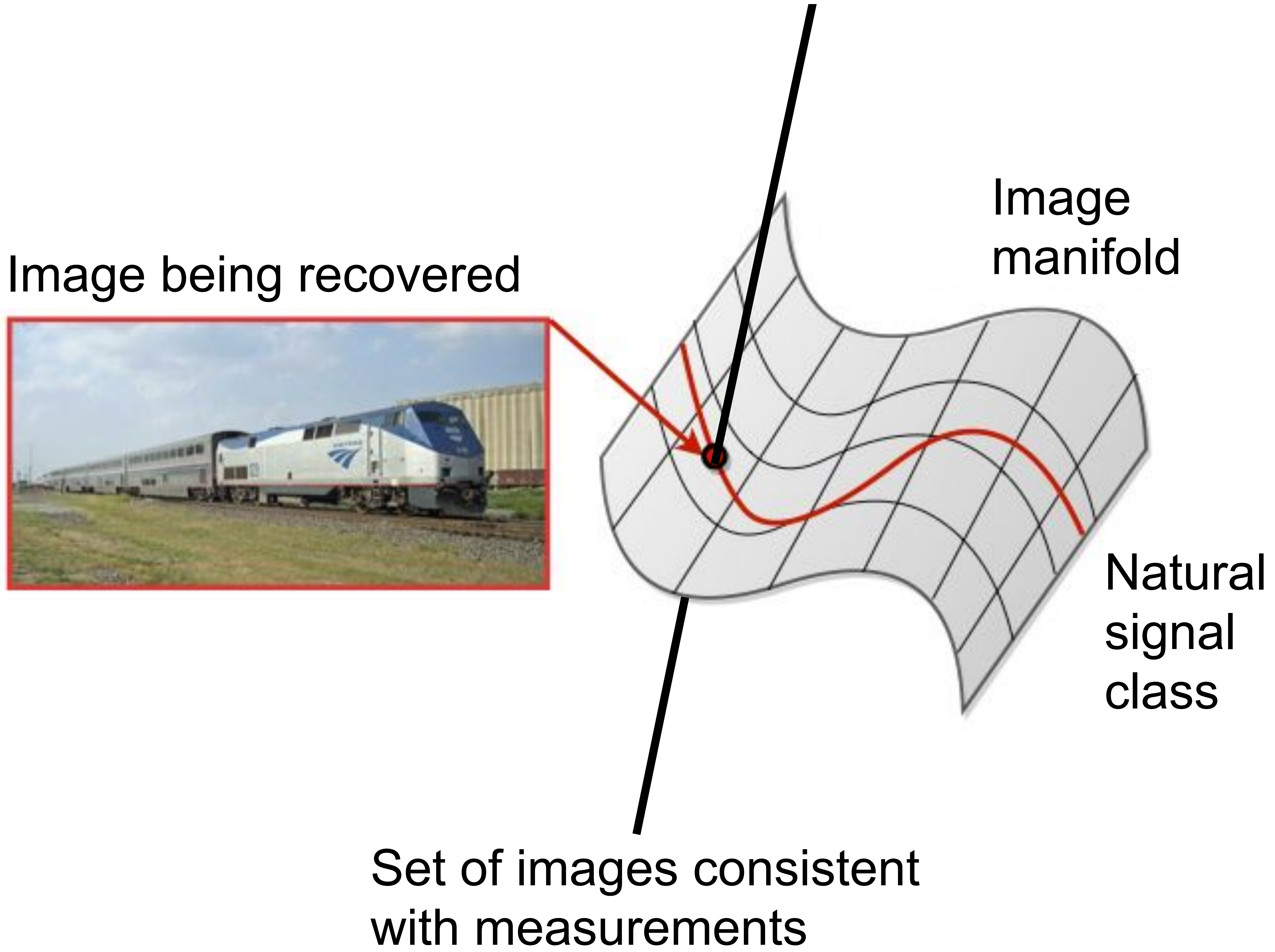
Hand, Leong, Voroninski 2018

Sparsity appears to fail in Compressive Phase Retrieval



Open problem: there is no known efficient algorithm to recover s-sparse x_0 from $O(s)$ generic measurements

Geometric picture of signal recovery with a low-dimensional generative prior



Information and Inference: A Journal of the IMA (2020) **00**, 1–35
doi:10.1093/imaiai/iaaa011

Rate-optimal denoising with deep neural networks

REINHARD HECKEL[†]
Department of Electrical and Computer Engineering, Rice University, Houston, Texas
[†]Corresponding author. Email: rh43@rice.edu

WEN HUANG
School of Mathematical Sciences, Xiamen University, Xiamen, China

PAUL HAND
Department of Mathematics and Khoury College of Computer Science, Northeastern University, Boston, Massachusetts

AND

VLADISLAV VORONINSKI
Helm.ai, Menlo Park, California USA

[Received on 6 October 2019; revised on 31 March 2020; accepted on 2 April 2020] *entropy* MDPI

Article

No Statistical-Computational Gap in Spiked Matrix Models with Generative Network Priors †

Jorio Cocola [1,*], Paul Hand [1,2] and Vladislav Voroninski [3]

SPIKED WIGNER MATRIX RECOVERY WITH A DEEP GENERATIVE PRIOR

Let: $G : \mathbb{R}^k \to \mathbb{R}^n$ generative network.
Let: $y_\star = G(x_\star)$ for some unknown $x_\star \in \mathbb{R}^k$.

Given: G.
Given: Noisy matrix $B = y_\star {y_\star}^T + \sigma\mathcal{H} \in \mathbb{R}^{n\times n}$, with $\mathcal{H}$ from a Gaussian Orthogonal Ensemble.

Estimate: $y_\star$.

To estimate $y_\star$, [8] proposes to find the latent code $\hat{x}$ that minimizes the reconstruction error

$$\tilde{x} = \arg\min_{x\in\mathbb{R}^x} f_{\text{spiked}}(x) := \frac{1}{2}\|M - G(x)G(x)^T\|_F^2,$$
$$y_\star \approx G(\tilde{x}),$$

where

- in the spiked Wishart model $M = B^T B/N - \sigma^2 I_n$;
- in the spiked Wigner model $M = B$.

Cocola, Hand, Voroninski 2021. Improved in Cocola 2022.

Recovery Theory with Random Generative Priors for Multiple Inverse Problems

- **Compressed Sensing**
  Deterministic Conditions and Probabilistic Recovery Guarantee

- **Other Inverse Problems**
  Denoising, Compressive Phase Retrieval, Spiked Matrix Recovery

- **Advances to the Theory**
  Convolutional Structure, Expansivity

Invertibility of Convolutional Generative Networks from Partial Measurements

Fangchang Ma*
MIT
fcma@mit.edu

Ulas Ayaz*
MIT
uayaz@mit.edu
uayaz@lyft.com

Sertac Karaman
MIT
sertac@mit.edu

2.1 Notation and Assumptions

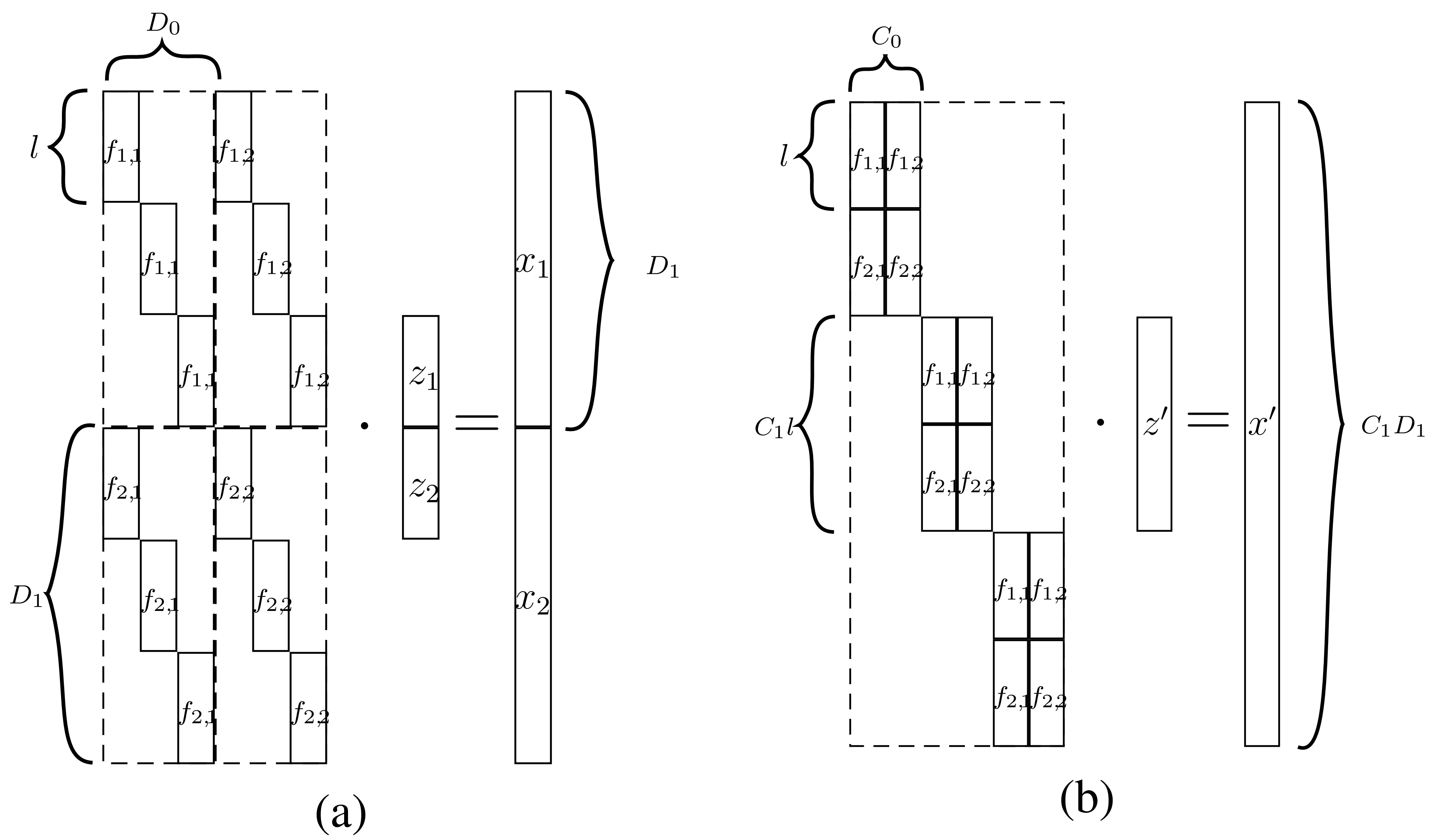


Figure 2: Illustration of a single transposed convolution operation. $f_{i,j}$ stands for i^{th} filter kernel for the j^{th} input channel. z and x denote the input and output signals, respectively. (a) The standard transposed convolution represented as linear multiplication. (b) With proper row and column permutations, the permuted weight matrix has a repeating block structure.

Weakening Expansivity Assumption

- $n_i \geq n_{i-1} \log n_{i-1} \operatorname{poly}(d)$ Hand and Vorninski, 2019
- $n_i \geq n_{i-1} \operatorname{poly}(d)$ Daskalakis et al., 2020
- $n_i \geq 5^i k \operatorname{poly}(d)$ Joshi et al., 2021
- $n_i \geq k \operatorname{poly}(d)$ Cocola, 2022

Signal Recovery Under Generative Priors

- There is a recovery theory for generative priors for multiple inverse problems
- Generative priors may outperform sparsity priors for a variety of problems
- Generative priors could provide tighter representations of natural images
- Generative priors can be optimally exploited for some nonlinear problems

References

Deep Compressive Sensing

- Bora et al. 2017 - *ICML*
- Hand and Voroninski 2019 - *IEEE Trans IT*
- Huang et al. 2021 - *J. Fourier Anal. App.*

Other Inverse Problems

- Denoising:
  Heckel et al. 2020 - *Information and Inference*
- Phase Retrieval:
  Hand, Leong, and Voroninski 2018 - *NeurIPS*
- Spiked Matrix Recovery:
  Aubin et al. 2020 - *IEEE Trans IT*
  Cocola et al. 2020 - *NeurIPS, Entropy*
- Blind Demodulation:
  Hand and Joshi 2018 - *NeurIPS*

Generalization of Assumptions

- Convolutional Generators
  Ma, Ayaz, and Karaman 2018 - *NeurIPS*
- Better Expansivity Condition
  Daskalakis et al. 2020 - *NeurIPS*
- Non-Expansive Priors with Different Algorithm
  Joshi et al. 2021 - *NeurIPS*
- Non-Expansive Priors
  Cocola *2022 - NeurIPS*

Review Articles

- Lucas et al. 2018 - *IEEE Sig. Proc. Mag.*
- Ongie et al. 2020 - *IEEE JSAIT*
- Scarlett et al*. 2022 - IEEE JSAIT*