GANs for Compressed Sensing - Theory

CS: Let $x^* \in \mathbb{R}^n$ image observed

$A \in \mathbb{R}^{m \times n}$ observation matrix, $m<n$

$\eta \in \mathbb{R}^m$ noise/Error

$y = Ax^* + \eta$ measurements

Given: y, A Find: x^*

As $m<n$, must assume structure to decide which consistent image is most "natural"

Generative model:

You have trained $G: \mathbb{R}^k \to \mathbb{R}^n$

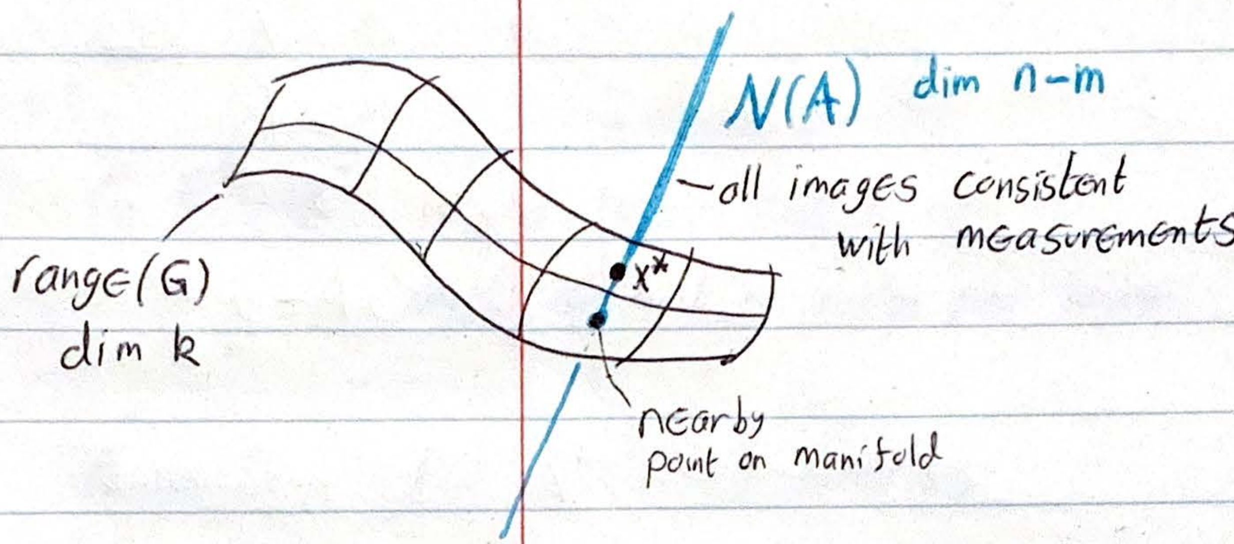
$z \mapsto x$

Such that $G(z)$ for $z \sim N(0, I_k)$ approximates sampling a natural signal distribution

Eg VAEs, GANs

Idea: Use the range of G as a proxy for what images are natural.

Visual Illustration of Geometry of GAN recovery



Recovery Algorithm:

$$\min_{z \in \mathbb{R}^k} \|AG(z) - y\| \quad \text{by gradient method}$$

(Bora et al.)

Questions:

How many measurements are needed?

Is this the right problem to solve?

Can this problem be solved?

Theory for CS w/ GAN priors

(Bora et al.)

Setup:

$G: \mathbb{R}^k \to \mathbb{R}^n$ is a d-layer ReLU net with at most n nodes per layer

$A \in \mathbb{R}^{m \times n}$ iid $N(0, 1/m)$ entries

Theorem:

Fix $x^* \in \mathbb{R}^n$. Let $y = Ax^* + \eta$. Let $m = \Omega(kd\log n)$.

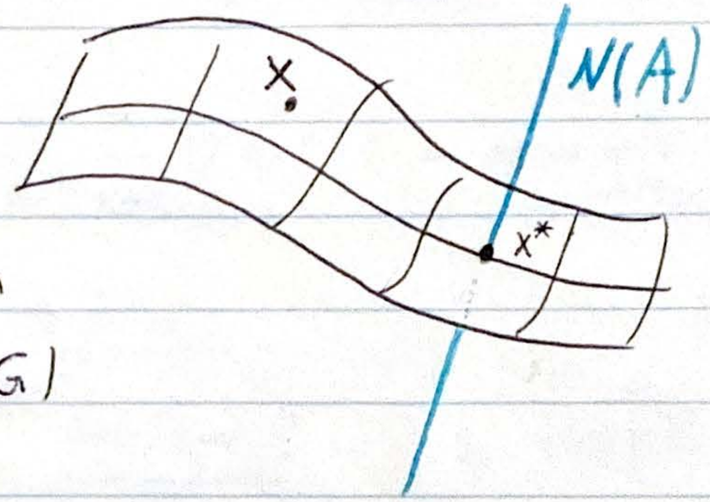
Suppose $\|AG(\hat{z}) - y\| \leq \varepsilon + \min_{z} \|AG(z) - y\|$

Then with probability $1 - e^{-\Omega(m)}$

$$\|G(\hat{z}) - x^*\| \leq \underbrace{6 \min_{z^* \in \mathbb{R}^k} \|G(z^*) - x^*\|}_{\text{representation error}} + \underbrace{3\|\eta\|}_{\text{noise}} + \underbrace{2\varepsilon}_{\text{error from optimiz.}}$$

Recovery based on Set-Restricted Eigenvalue Condition

To guarantee injectivity, want $N(A)$ to be away from directions between pairs of points in range(G)



Defn:

$A \in \mathbb{R}^{m \times n}$ satisfies S-REC(S, γ) if $\forall\, x_1, x_2 \in S$

$$\|A(x_1 - x_2)\| \geq \gamma \|x_1 - x_2\|$$

ie $N(A)$ is away from secant lines within S

Lemma: Let $y = Ax^* + \eta$.

Suppose A satisfies S-REC(S, γ) w/ prob $1-p$

and for fixed $x \in \mathbb{R}^n$, $\|Ax\| \leq 2\|x\|$ w/ prob $1-p$

and $\|A\hat{x} - y\| \leq \min\limits_{x \in S} \|y - Ax\| + \varepsilon$ for $\hat{x} \in S$

then w/ prob $1-2p$

$$\|\hat{x} - x^*\| \leq \left(\frac{4}{\gamma} + 1\right) \min_{x \in S} \|x^* - x\| + \frac{2\|\eta\| + \varepsilon}{\gamma}$$

Proof:

Fix x^* – true image

$\bar{x} = \underset{x \in S}{\operatorname{argmin}} \|x^* - x\|$ – closest img in S

$\hat{x}$ st $\|A\hat{x} - y\| \leq \min_{x \in S} \|y - Ax\| + \varepsilon$ – approximate optimizer

$$
\begin{aligned}
\|\hat{x} - x^*\| &\leq \|x^* - \bar{x}\| + \|\bar{x} - \hat{x}\| \\
&\leq \cdots \quad \frac{\|A(\bar{x} - \hat{x})\|}{\gamma} && \text{by S-REC} \\
&\leq \cdots + \frac{\|A\bar{x} - y\| + \|A\hat{x} - y\|}{\gamma} \\
&\leq \cdots + \frac{2\|A\bar{x} - y\| + \varepsilon}{\gamma} && \text{by approx minimization by } \hat{x} \\
&\leq \cdots + \frac{2\|A\bar{x} - Ax^*\| + 2\|Ax^* - y\| + \varepsilon}{\gamma} \\
&\leq \cdots + \frac{2\|A(\bar{x} - x^*)\| + 2\|\eta\| + \varepsilon}{\gamma} \\
&\leq \cdots + \frac{4\|\bar{x} - x^*\| + 2\|\eta\| + \varepsilon}{\gamma} && \text{by assump. on A} \\
&\leq \left(1 + \frac{4}{\gamma}\right)\|x^* - \bar{x}\| + \frac{2\|\eta\| + \varepsilon}{\gamma} \qquad \blacksquare
\end{aligned}
$$

Random matrices satisfy S-REC

To show A w/ $N(0, 1/m)$ entries satisfies S-REC for G given by ReLU nets, we need two technical results

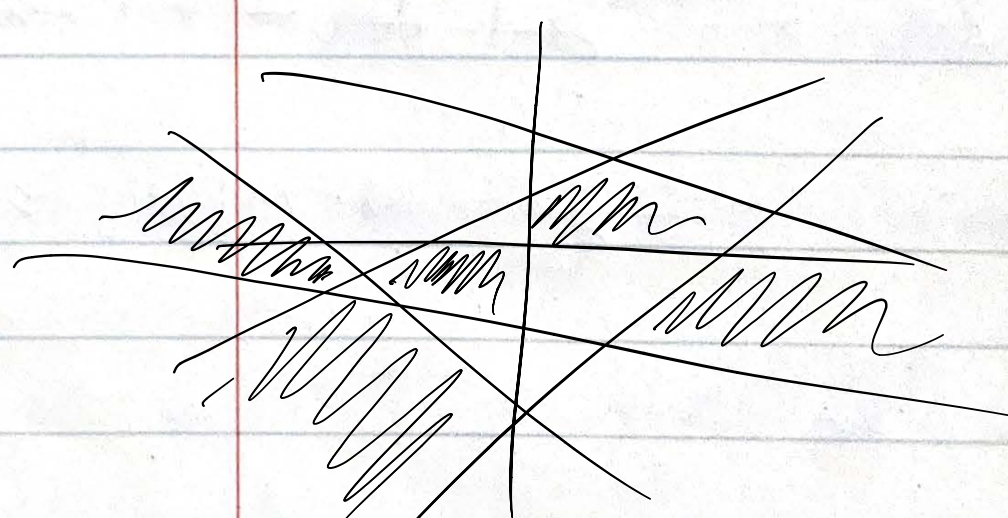
<u>Theorem</u>: Let $A \in \mathbb{R}^{m \times k}$ have iid $N(0,1)$ entries. $\forall\, t \geq 0$, with prob at least $1 - 2e^{-t^2/2}$,

$$\sqrt{m} - \sqrt{k} - t \leq \sigma_{min}(A) \leq \sigma_{max}(A) \leq \sqrt{m} + \sqrt{k} + t$$

See Vershynin "Introduction to the non-asymptotic analysis of random matrices"

States: tall random Gaussian matrices are approximate isometries

<u>Theorem</u>: If $\mathbb{R}^k$ is partitioned by c hyperplanes, the number of partition pieces is $O(c^k)$



Lemma: Let $G: \mathbb{R}^k \to \mathbb{R}^n$ be a d-layer NN w/ each layer a linear transformation followed by a ReLU. Suppose G has at most n nodes per layer.
If $m = \Omega(\frac{1}{\alpha^2} kd \log n)$ then $A \in \mathbb{R}^{m \times n}$ w/ iid $N(0, 1/m)$ entries satisfies S-REC$(G(\mathbb{R}^k), 1-\alpha)$ w/ prob $1 - e^{-\Omega(\alpha^2 m)}$

Proof: First layer has at most n^k diff. linear pieces
Applying to d layers, at most n^{kd} linear pieces

So $G(\mathbb{R}^k)$ lives in union of n^{kd} subspaces of dim k. Thus $\{x_1 - x_2 \mid x_{1,2} \in G(\mathbb{R}^k)\}$ lives in union of n^{2kd} subspaces of dim 2k.

On each subspace A satisfies S-REC w/ param $1-\alpha$ w/ prob $1 - e^{-\Omega(\alpha^2 m)}$ if $m = \Omega(k/\alpha^2)$

Prob that S-REC holds on all subspaces is at least

$$1 - n^{2kd} e^{-\Omega(\alpha^2 m)}$$

which is $1 - e^{-\Omega(\alpha^2 m)}$ if $m = \Omega(\frac{kd \log n}{\alpha^2})$.

When can one solve the optimization?

$$\min_{z \in \mathbb{R}^k} \|A\,G(z) - y\|$$

Finding minimizer for nonconvex problems is NP-hard in general.

Can provably solve this problem under a random model for G.

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

$$G(z) = W_d \cdots \mathrm{relu}(W_2 \,\mathrm{relu}(W_1 z)) \qquad W_i \in \mathbb{R}^{n_i \times n_{i-1}}$$

fully connected   no bias terms

Assume:
- $n_i > c\, n_{i-1} \log n_{i-1}$ expansivity
- W_i has iid $N(0,1)$ entries   Gaussian
- $m > c\, k\, d \log(n_1 n_2 \cdots n_d)$ $\Omega(k)$ measurements.

Then w/ high prob a gradient algorithm will converge to z^* if $y = A\,G(z^*)$.

(Hand + Voroninski, Huang et al.)