Instance-Optimal Compressed Sensing via Posterior Sampling

Ajil Jalal Sushrut Karmalkar Alex Dimakis Eric Price

UT Austin

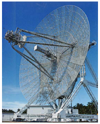
Compressed Sensing

- Want to recover a signal (e.g. an image) from noisy measurements.

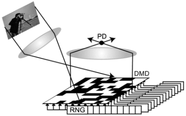
Compressed Sensing

- Want to recover a signal (e.g. an image) from noisy measurements.

Medical Imaging



Astronomy



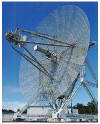
Single-Pixel Camera

Oil Exploration

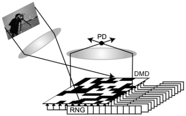
Compressed Sensing

- Want to recover a signal (e.g. an image) from noisy measurements.

Medical Imaging



Astronomy



Single-Pixel Camera

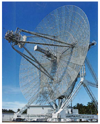
Oil Exploration

- *Linear* measurements: see $y = Ax^* + \text{noise}$, for $A \in \mathbb{R}^{m \times n}$.

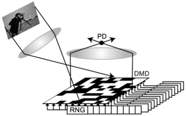
Compressed Sensing

- Want to recover a signal (e.g. an image) from noisy measurements.

Medical Imaging



Astronomy



Single-Pixel Camera

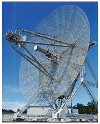
Oil Exploration

- *Linear* measurements: see $y = Ax^* + \text{noise}$, for $A \in \mathbb{R}^{m \times n}$.
- How many measurements m to recover x^*?

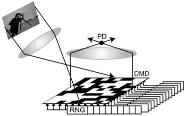
Compressed Sensing

- Want to recover a signal (e.g. an image) from noisy measurements.

Medical Imaging



Astronomy

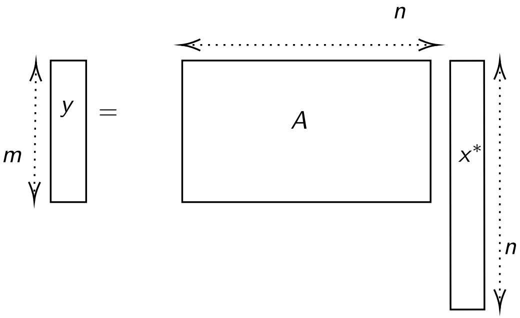


Single-Pixel Camera

Oil Exploration

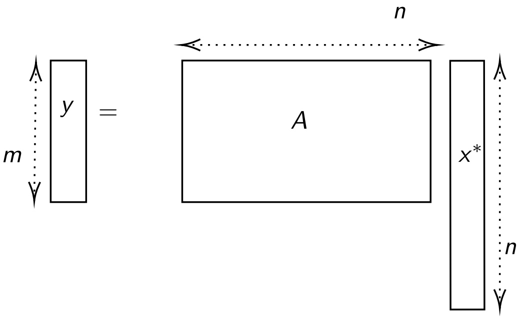
- *Linear* measurements: see $y = Ax^* + \text{noise}$, for $A \in \mathbb{R}^{m \times n}$.
- How many measurements m to recover x^*?
- Algorithm to recover x^*?

Compressed Sensing



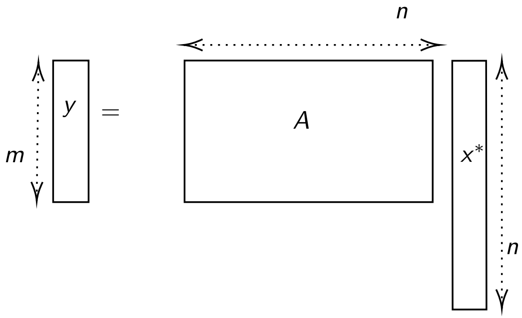
- Given linear measurements $y = Ax^*$, for $A \in \mathbb{R}^{m\times n}$.
- How many measurements m to learn the signal x^*?

Compressed Sensing



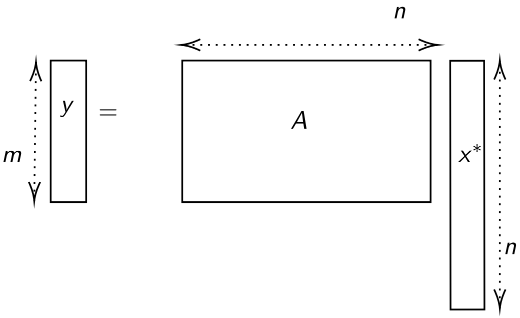
- Given linear measurements $y = Ax^*$, for $A \in \mathbb{R}^{m \times n}$.
- How many measurements m to learn the signal x^*?
  - Naively: $m \geq n$.

Compressed Sensing



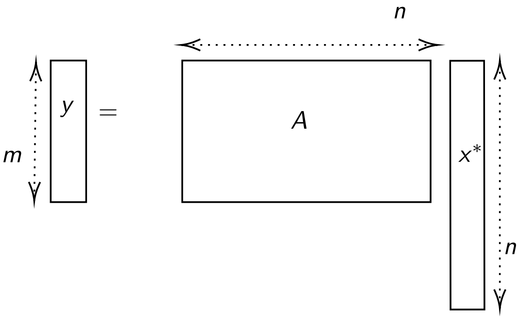
- Given linear measurements $y = Ax^*$, for $A \in \mathbb{R}^{m \times n}$.
- How many measurements m to learn the signal x^*?
  - Naively: $m \geq n$.
  - If $m < n$, underdetermined and multiple possible solutions.

Compressed Sensing



- Given linear measurements $y = Ax^*$, for $A \in \mathbb{R}^{m \times n}$.
- How many measurements m to learn the signal x^*?
  - Naively: $m \geq n$.
  - If $m < n$, underdetermined and multiple possible solutions.
  - With additional assumptions on x^*, A, it is possible to recover x^*.

Compressed Sensing



- Given linear measurements $y = Ax^*$, for $A \in \mathbb{R}^{m \times n}$.
- How many measurements m to learn the signal x^*?
  - Naively: $m \geq n$.
  - If $m < n$, underdetermined and multiple possible solutions.
  - With additional assumptions on x^*, A, it is possible to recover x^*.
- Candes-Romberg-Tao 2006: If x^* is $k-$sparse in some basis, and A is Gaussian i.i.d., then $m = O\left(k \log n\right)$ suffices to recover x^*.

Generative Models

BigGAN (Brock et al)

- Many natural ways to use neural networks
- This work will focus on *generative models* (Goodfellow et al, Kingma & Welling, Dinh et al).
- Want to model a distribution $\mathcal{D}$ of images.
- Function $G : \mathbb{R}^k \to \mathbb{R}^n$.
- When $z \sim \mathcal{N}(0, I_k)$, then ideally $G(z) \sim \mathcal{D}$.

Generative Models

BigGAN (Brock et al)

- Many natural ways to use neural networks
- This work will focus on *generative models* (Goodfellow et al, Kingma & Welling, Dinh et al).
- Want to model a distribution $\mathcal{D}$ of images.
- Function $G : \mathbb{R}^k \to \mathbb{R}^n$.
- When $z \sim \mathcal{N}(0, I_k)$, then ideally $G(z) \sim \mathcal{D}$.
- Bora-Jalal-Price-Dimakis 2017:
  - if $G : \mathbb{R}^k \to \mathbb{R}^n$ is a $d-$layered ReLU neural network.
  - if x^* lies close to the range of G.
  - $m = O(kd \log n)$ Gaussian measurements suffice.
  - For general $L-$Lipshitz G, $m = O(k \log L)$ suffices.
  - Algorithm for recovery MAP / Maximum Likelihood.

Going beyond $k \ll n$

- $k \ll n$ is good for theory, but hurts empirical performance.

Going beyond $k \ll n$

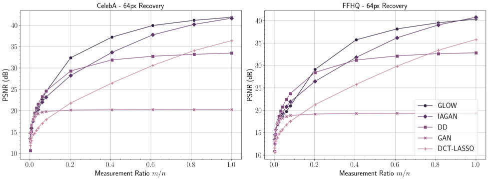
- $k \ll n$ is good for theory, but hurts empirical performance.
- Asim-Daniels-Leong-Ahmed-Hand 2020:
  invertible neural networks $H : \mathbb{R}^n \to \mathbb{R}^n$ give superior performance.

Going beyond $k \ll n$

- $k \ll n$ is good for theory, but hurts empirical performance.
- Asim-Daniels-Leong-Ahmed-Hand 2020:
  invertible neural networks $H : \mathbb{R}^n \to \mathbb{R}^n$ give superior performance.
- Existing theory says $m = \Omega(n)$ measurements are necessary, but $m \ll n$ works in practice.

Going beyond $k \ll n$

- $k \ll n$ is good for theory, but hurts empirical performance.
- Asim-Daniels-Leong-Ahmed-Hand 2020: *invertible* neural networks $H : \mathbb{R}^n \to \mathbb{R}^n$ give superior performance.
- Existing theory says $m = \Omega(n)$ measurements are necessary, but $m \ll n$ works in practice.



MSE (Asim et. al. '20)

Reconstructions(Asim et. al. '20)

Our Questions

Question 1

For general distributions, how do we formalize the number of measurements needed to compress the distribution?

Our Questions

Question 1

For general distributions, how do we formalize the number of measurements needed to compress the distribution?

Question 2

Does there exist an algorithm that can recover images using this sample complexity?

Theoretical Results

- $x^* \sim R$, $\|x^*\| \leq r$.

Theoretical Results

- $x^* \sim R$, $\|x^*\| \leq r$.
- See $y = Ax^* + \eta$, with η i.i.d. Gaussian and $\mathbb{E}[\|\eta\|^2] = \varepsilon^2$.

Theoretical Results

- $x^* \sim R$, $\|x^*\| \leq r$.
- See $y = Ax^* + \eta$, with η i.i.d. Gaussian and $\mathbb{E}[\|\eta\|^2] = \varepsilon^2$.
- P: distribution of generative model.

Theoretical Results

- $x^* \sim R$, $\|x^*\| \leq r$.
- See $y = Ax^* + \eta$, with η i.i.d. Gaussian and $\mathbb{E}[\|\eta\|^2] = \varepsilon^2$.
- P: distribution of generative model.

$\mathrm{Cov}_{\varepsilon,\delta}(R) :=$ smallest number of ε-radius balls covering $1 - \delta$ of R.

Theoretical Results

- $x^* \sim R$, $\|x^*\| \leq r$.
- See $y = Ax^* + \eta$, with η i.i.d. Gaussian and $\mathbb{E}[\|\eta\|^2] = \varepsilon^2$.
- P: distribution of generative model.

$\mathrm{Cov}_{\varepsilon,\delta}(R) :=$ smallest number of ε-radius balls covering $1-\delta$ of R.

Upper bound + Robustness

If A is Gaussian, and $P\&R$ are ε-close in Wasserstein distance, then $\widehat{x} \sim P(x|y)$ and $m = O(\log \mathrm{Cov}_{\varepsilon,\delta}(R))$ gets $\|x^* - \widehat{x}\| \leq O(\varepsilon)$ with probability $1 - 3\delta$.

Theoretical Results

- $x^* \sim R$, $\|x^*\| \le r$.
- See $y = Ax^* + \eta$, with η i.i.d. Gaussian and $\mathbb{E}[\|\eta\|^2] = \varepsilon^2$.
- P: distribution of generative model.

$\mathrm{Cov}_{\varepsilon,\delta}(R) :=$ smallest number of ε-radius balls covering $1-\delta$ of R.

Upper bound + Robustness

If A is Gaussian, and $P \& R$ are ε-close in Wasserstein distance, then $\widehat{x} \sim P(x|y)$ and $m = O(\log \mathrm{Cov}_{\varepsilon,\delta}(R))$ gets $\|x^* - \widehat{x}\| \le O(\varepsilon)$ with probability $1 - 3\delta$.

Lower bound

For *arbitrary* A, any algorithm that achieves (ε, δ)-recovery requires

$$m \ge \Omega\left(\frac{\log \mathrm{Cov}_{C\varepsilon, C\delta}(R)}{\log(r\|A\|_\infty/\varepsilon)}\right)$$

for some constant $C > 0$.

Instance optimality

- Existing lower bounds in compressed sensing consider *worst-case* distributions.

Instance optimality

- Existing lower bounds in compressed sensing consider *worst-case* distributions.
- Typical example: for a uniform distribution over $k-$sparse vectors in $\mathbb{R}^n$, you need $m = \Omega(k \log(n/k))$ for successful recovery.

Instance optimality

- Existing lower bounds in compressed sensing consider *worst-case* distributions.
- Typical example: for a uniform distribution over $k-$sparse vectors in $\mathbb{R}^n$, you need $m = \Omega(k \log(n/k))$ for successful recovery.
- Our upper and lower bounds are tight upto constants for *any distribution*.

Instance optimality

- Existing lower bounds in compressed sensing consider *worst-case* distributions.
- Typical example: for a uniform distribution over $k-$sparse vectors in $\mathbb{R}^n$, you need $m = \Omega(k \log{(n/k)})$ for successful recovery.
- Our upper and lower bounds are tight upto constants for *any distribution*.

Gaussian measurements are good for any distribution

For any distribution of x^*, if there exists any matrix A and an algorithm that uses m measurements and achieves $\|x^* - x'\| \leq \varepsilon$ with probability $1-\delta$, then posterior sampling with $O\left(m \log \frac{r\|A\|}{\varepsilon}\right)$ *Gaussian* measurements satisfies $\|x^* - \widehat{x}\| \leq O(\varepsilon)$ with probability $1 - O(\delta)$.

Instance optimality

- Existing lower bounds in compressed sensing consider *worst-case* distributions.
- Typical example: for a uniform distribution over $k-$sparse vectors in $\mathbb{R}^n$, you need $m = \Omega(k \log(n/k))$ for successful recovery.
- Our upper and lower bounds are tight upto constants for *any distribution*.

Gaussian measurements are good for any distribution

For any distribution of x^*, if there exists any matrix A and an algorithm that uses m measurements and achieves
$\|x^* - x'\| \leq \varepsilon$ with probability $1 - \delta$, then posterior sampling with $O\left(m \log \frac{r\|A\|}{\varepsilon}\right)$ *Gaussian* measurements satisfies $\|x^* - \widehat{x}\| \leq O(\varepsilon)$ with probability $1 - O(\delta)$.

Sampling + Gaussian measurements is robust to incorrect prior

If $x^* \sim R$ but your prior is $P \neq R$, the guarantee still holds if P and R are ε-close in Wasserstein distance.

Experiments - Compressed Sensing

Implement Posterior Sampling via Annealed Langevin Dynamics

Figure: FFHQ dataset using NCSNv2 model (Song & Ermon 2020).

Experiments - Inpainting

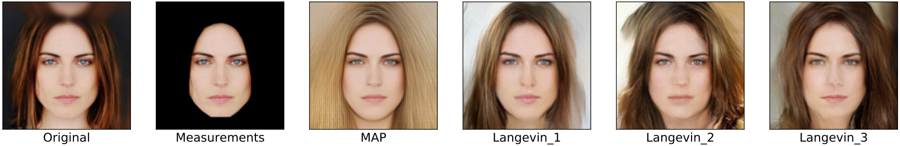


Figure: CelebA dataset using GLOW model (Kingma & Dhariwal 2018).

- MAP estimate is averaged-out, similar to eigenfaces.
- Posterior sampling is more realistic & can be sampled multiple times for diversity.

Thank you!

- Code & models:
  `https://github.com/ajiljalal/code-cs-fairness`
- Related paper on fairness: Fairness for Image Generation with Uncertain Sensitive Attributes (ICML 2021)