Compressed Sensing with Approximate Priors via Conditional Resampling

Ajil Jalal, Sushrut Karmalkar, Alex Dimakis, Eric Price

NEURAL INFORMATION PROCESSING SYSTEMS

Compressed Sensing - Introduction

- Goal: Estimate a signal $x^* \in \mathbb{R}^n$ from a linear system $y = Ax^* + \eta$.
- Let $A \in \mathbb{R}^{m \times n}$. How many measurements are needed? Naively $m \geq n$, else underdetermined; multiple x possible.
- But not all x are plausible/natural:

5 MB

36 MB

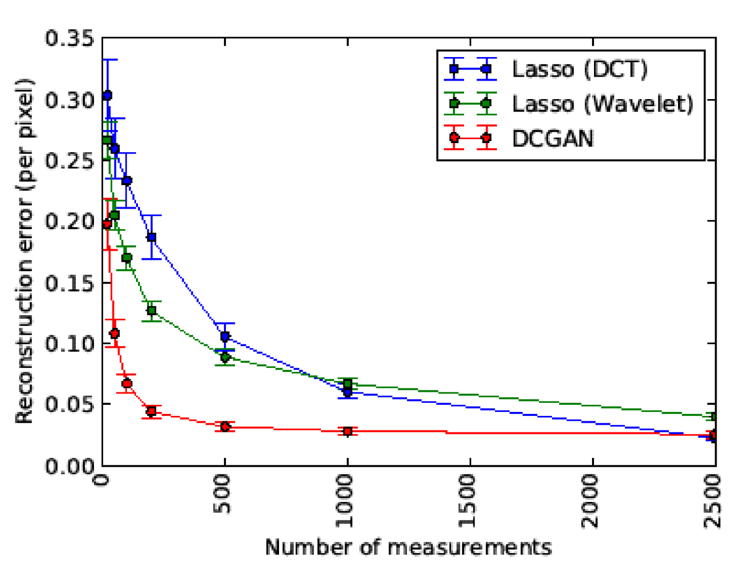
- [Candes-Romberg-Tao '06]: Possible to recover x^* if it is approximately sparse and A is Gaussian.

Compressed Sensing + Generative Priors

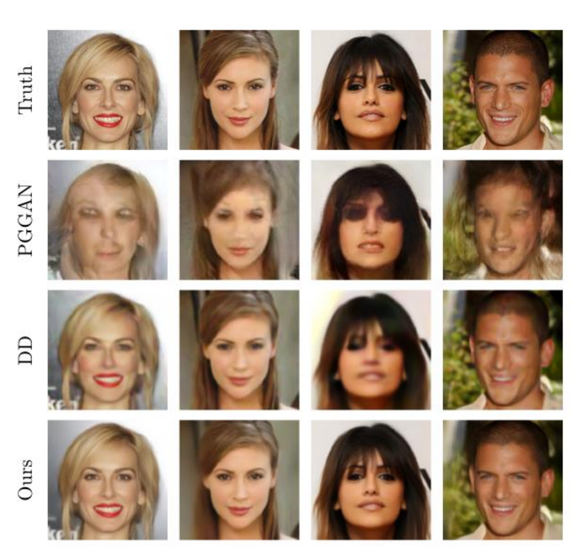
- [Bora-Jalal-Price-Dimakis '17]: When $G: \mathbb{R}^k \to \mathbb{R}^n$ is a d −layered neural network, A is i.i.d. Gaussian with $m = O(kd \log n)$ rows, then gradient descent finds $\hat{x}$ satisfying

$$||x^* - \hat{x}|| \lesssim \min_{x \in \text{range}(G)} ||x^* - x|| + ||\eta|| + \varepsilon$$

- Small $k \Rightarrow$ error stops improving after some point because of limited model capacity.



Bora et. al.'17



Asim et. al.'20

- [Asim-Daniels-Leong-Ahmed-Hand '20]: Use bijective $H: \mathbb{R}^n \to \mathbb{R}^n$, and use the distribution of the generative model.
- When $k = n$, prior work doesn't explain why compression is possible. Asim et al. analyze a Gaussian with eigenvalue decay.

Our Questions

- **Question 1:** For general distributions, how do we formalize the number of measurements needed to compress the distribution?
- **Question 2:** What algorithm can recover signals using this sample complexity?

Our Goal

- $x^* \sim R, \;\; y = Ax^* + \eta, \;\; A \;\&\; \eta$ are i.i.d. Gaussian.
- Have access to distribution P, such that $\mathcal{W}_2(P, R) \leq \varepsilon$.
- Goal: estimate $\hat{x}$ such that with probability 0.97,

$$||x^* - \hat{x}|| \lesssim ||\eta|| \approx \varepsilon. \;\; (*)$$

Our Results

- New complexity measure: $B_2(x, \varepsilon)$ is the ε −radius ℓ_2 ball around x, and R is a probability distribution. Then:

$$\text{Cov}_{\varepsilon,\delta}(R) \coloneqq \min\left\{ k : R\left[\bigcup_{i=1}^{k} B_2(x_i, \varepsilon)\right] \geq 1 - \delta, x_i \in \mathbb{R}^n \right\}.$$

- Upper bound: $m = O(\log \text{Cov}_{\varepsilon,0.01}(R))$ suffices for $(*)$.
- Lower bound: $m = \Omega(\log \text{Cov}_{5\varepsilon,0.1}(R))$ is necessary for $(*)$.
- Optimal algorithm: *Conditional resampling* is optimal.
- Instance Optimality: Lower bound holds for *any* distribution R, and not just for particular hard distributions.
- Distributional robustness: Algorithm can tolerate mismatch between R and P.

Our Algorithm

- Given measurements y, density p over images, measurement likelihood $\pi(y|x)$, estimate is $\hat{x}$, such that:

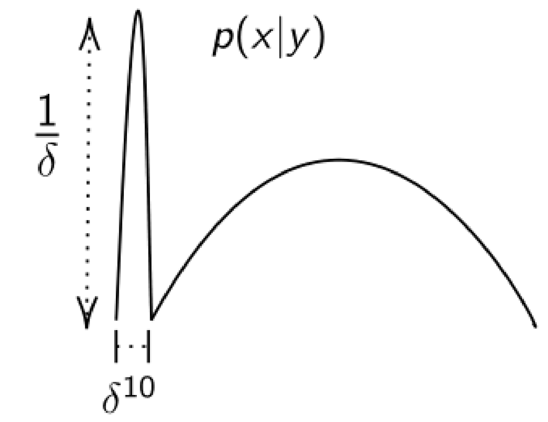
$$p(\hat{x}|y) \propto p(\hat{x})\pi(y|\hat{x}).$$

- Langevin dynamics:

$$\hat{x}_{t+1} \leftarrow \hat{x}_t + \beta_t \nabla_{\hat{x}_t} \log p(\hat{x}_t|y) + N(0, 2\beta_t).$$

- We use annealed Langevin dynamics [Song & Ermon].

MAP vs. Conditional Resampling



- MAP works for "nice'' distributions.
- MAP picks narrow peak.
- Conditional resampling will pick wide peak.

Qualitative Results: Inpainting

MAP (Asim et al '20):

$$\text{argmin}_{z \in \mathbb{R}^n} ||y - AG(z)||^2 + \gamma||z||^2.$$

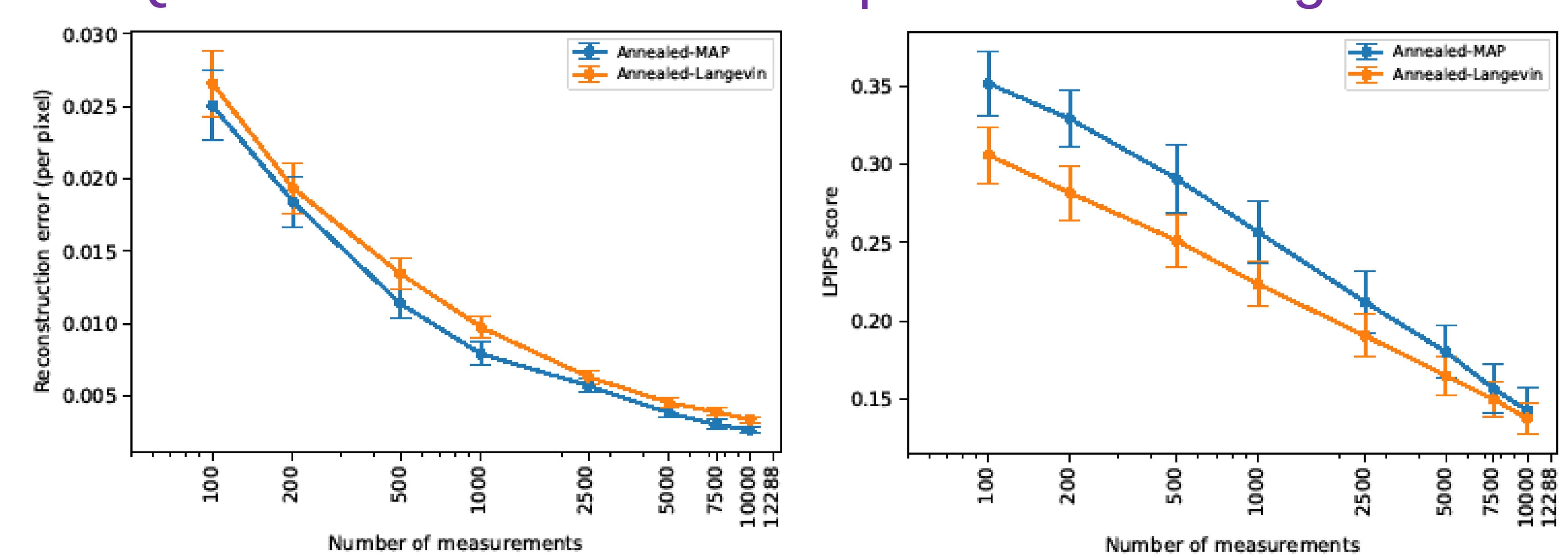
Langevin sampling:

$$z_{t+1} \leftarrow z_t - \beta_t \nabla_{z_t}(||y - AG(z_t)||^2 / 2\sigma_t^2 + ||z_t||^2/2) + N(0, 2\beta_t),$$
$$\hat{x} = G(z_T)$$

(a) Original (b) Input (c) Reconstructions by MAP and Langevin

MAP produces *one* image with small $||z||$, while Langevin produces diverse images with $||z||^2 \approx n$. Analogous to points close to origin having large density in a high dimensional Gaussian, yet atypical.

Quantitative Results: Compressed Sensing



Comparison of our algorithm with Asim et al'20. Left column is MSE, and right column is LPIPS scores between reconstruction and ground truth. LPIPS is a measure of how perceptually distant two images are. Our algorithm has no statistically significant difference in comparison to MAP, but produces perceptually closer images.

References

Candès, E. J., Romberg, J., & Tao, T. (2006). Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on information theory*, *52*(2), 489-509.

Bora, A., Jalal, A., Price, E., & Dimakis, A. G. (2017, July). Compressed Sensing using Generative Models. In *International Conference on Machine Learning* (pp. 537-546).

Asim, M., Daniels, M., Leong, O., Ahmed, A., & Hand, P. (2020, November). Invertible generative models for inverse problems: mitigating representation error and dataset bias. In *International Conference on Machine Learning* (pp. 399-409). PMLR.

Song, Y., & Ermon, S. (2019). Generative modeling by estimating gradients of the data distribution. In *Advances in Neural Information Processing Systems* (pp. 11918-11930).

Wu, Y. (2011). *Shannon theory for compressed sensing*. Princeton University.