Compressed Sensing with Approximate Priors via Conditional Resampling

Ajil Jalal, Sushrut Karmalkar, Alex Dimakis, Eric Price

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:

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

Compressed Sensing + Generative Priors

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

  \[ \|x^* - \hat{x}\| \leq \min_{x \in \mathcal{X}} \|x^* - x\| + \|n\| + \epsilon \]

- Small \( k \) = error stops improving after some point because of limited model capacity.

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 \( \mathcal{W}_2(P,R) \leq \epsilon \).
- Goal: estimate \( \hat{x} \) such that with probability 0.97,

  \[ \|x^* - \hat{x}\| \leq \epsilon. \]  

Upper bound: \( m = O(\log \text{Cov}_{x,0,0.1}(R)) \) suffices for (*)

Lower bound: \( m = \Omega(\log \text{Cov}_{x,0,0.1}(R)) \) 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 \( p(y|x) \), estimate is \( \hat{x} \), such that:

  \[ p(y|x) \propto p(\hat{x}|y)p(y|\hat{x}). \]

- Langevin dynamics:

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

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

References


