Sparse and Low-Rank Matrix Recovery

Xiangyu He, Xiujia Li

9-4-2025

Brief introduction of the speakers

Xiujia Li

Master student in DCSC

Xiangyu He

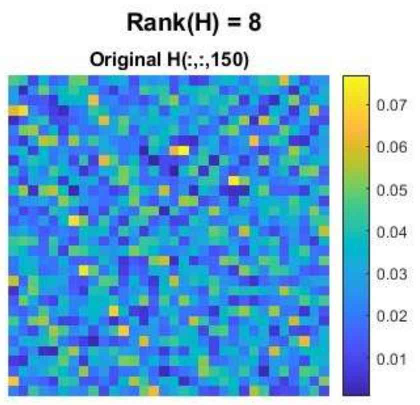
Master student in DCSC

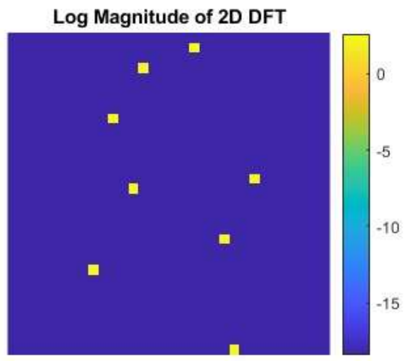
01

Introduction and Motivation

what are we trying to solve?

why is it an important problem?





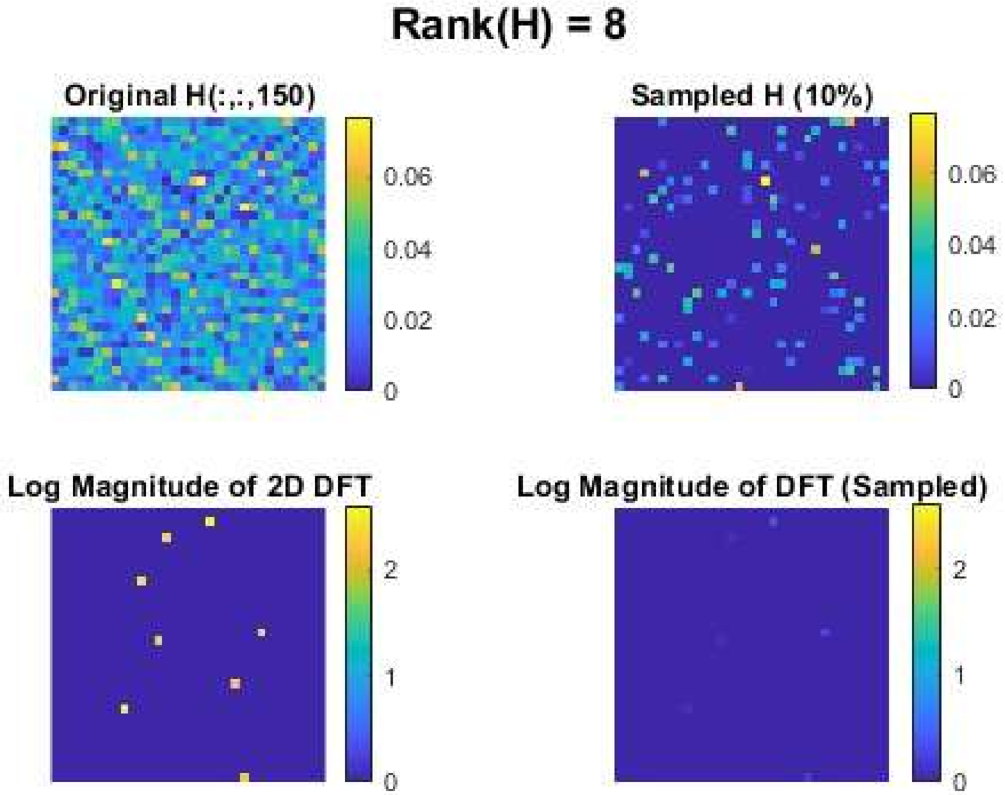
Dataset Description

Sparse and Low-Rank Matrix Recovery

Sparse in the frequency domain

Low-rank in the spatial domain

Subsampling Operator



Sparse in the frequency domain

Low-rank in the spatial domain

Why is it an important problem?

A fundamental challenge in many real-world scenarios.

- In medical imaging, such as MRI, we often want to minimize scan time.
- In wireless communications, bandwidth is limited.

Sparse Property of the Data

- In many cases, real-world signals exhibit sparsity in a specific domain
  - Communication signals, natural images

Low-Rank Property of the Data

- Low-rank or approximately low-rank matrices are common in data science.
  - Movie recommendation

02

Analysis Procedures and Methodologies

How are we trying to solve the problem?

Models and analysis procedures

Successive Proximal Gradient Descent

Sparse structure and sparsifying objective

$$\min_{\hat{H}} \|\mathcal{F}(\hat{H})\|_0 \quad \text{subject to} \quad \mathcal{P}_\Omega(\hat{H}) = G$$

$$\Downarrow$$

$$\min_{\hat{H}} \|\mathcal{F}(\hat{H})\|_1 \quad \text{subject to} \quad \mathcal{P}_\Omega(\hat{H}) = G$$

$$\Downarrow$$

$$\min_{\hat{H}} \|\mathcal{F}(\hat{H})\|_1 \quad \text{subject to} \quad \|\mathcal{P}_\Omega(\hat{H}) - G\|_2 \leq \epsilon$$

$$\Downarrow$$

$$\min_{\hat{H}} \|\mathcal{P}_\Omega(\hat{H}) - G\|_2^2 + \lambda\|\mathcal{F}(\hat{H})\|_1$$

Successive Proximal Gradient Descent

Low-rank structure and nuclear norm minimization

$$\min_{\hat{\mathbf{H}}} \left\|\mathcal{P}_\Omega(\hat{\mathbf{H}}) - \mathbf{y}\right\|_2^2 \quad \text{subject to} \quad \operatorname{rank}(\hat{\mathbf{H}}) = R, \mathbf{y} = \mathcal{P}_\Omega(\mathbf{H})$$

$$\Downarrow$$

$$\min_{\hat{\mathbf{H}}} \left\|\mathcal{P}_\Omega(\hat{\mathbf{H}}) - \mathbf{y}\right\|_2^2 + \lambda \left\|\hat{\mathbf{H}}\right\|_* \quad \text{subject to } \mathbf{y} = \mathcal{P}_\Omega(\mathbf{H})$$

Successive Proximal Gradient Descent

Summary

$$\min_{\hat{\mathbf{H}}} \quad \left\|\mathcal{P}_\Omega(\hat{\mathbf{H}}) - \mathbf{y}\right\|_2^2 + \lambda_1 \|\mathbf{X}\|_1 + \lambda_2 \left\|\hat{\mathbf{H}}\right\|_*$$

$$\text{subject to} \quad \mathbf{y} = \mathcal{P}_\Omega(\mathbf{H}), \quad \mathbf{X} = \mathcal{F}(\hat{\mathbf{H}})$$

$$\Downarrow$$

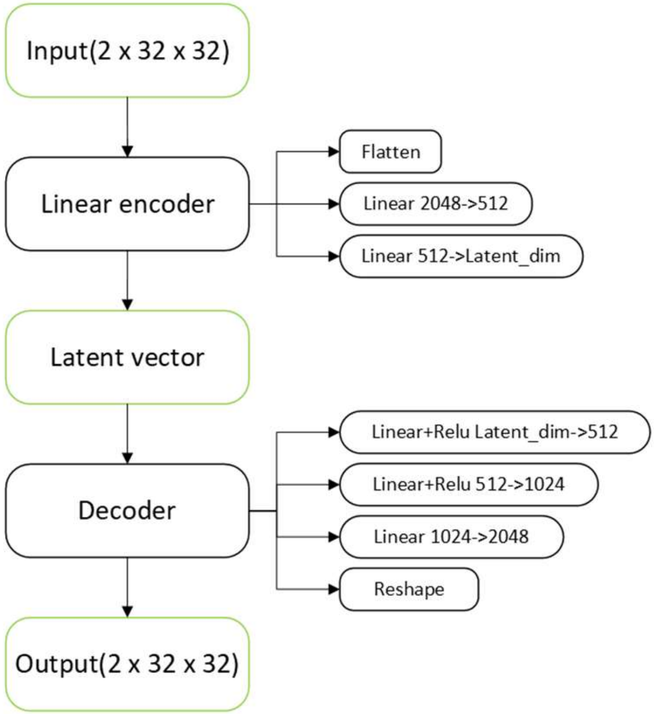
$$\mathbf{P}_{t+1} = \hat{\mathbf{H}}_t - \eta \nabla g(\hat{\mathbf{H}}_t) = \hat{\mathbf{H}}_t - 2\eta \cdot \mathcal{P}_\Omega(\hat{\mathbf{H}}_t - \mathbf{y})$$

$$\begin{aligned}\hat{\mathbf{H}}_{t+1} &= \operatorname{prox}_{\eta h}(\mathbf{P}_{t+1}) \\ &= \arg\min_{\hat{\mathbf{H}}} \left\{ \frac{1}{2\eta} \left\|\hat{\mathbf{H}} - \mathbf{P}_{t+1}\right\|_F^2 + \lambda \left\|\hat{\mathbf{H}}\right\|_* \right\}\end{aligned}$$

$$u_{\text{opt}}[i] = \begin{cases} w[i] - \eta\lambda, & \text{if } w[i] > \eta\lambda \\ w[i] + \eta\lambda, & \text{if } w[i] < -\eta\lambda \\ 0, & \text{otherwise} \end{cases}$$

Autoencoder

The matrix H is low-rank and sparse in the 2D DFT domain. This structure matches well with the strengths of an autoencoder.



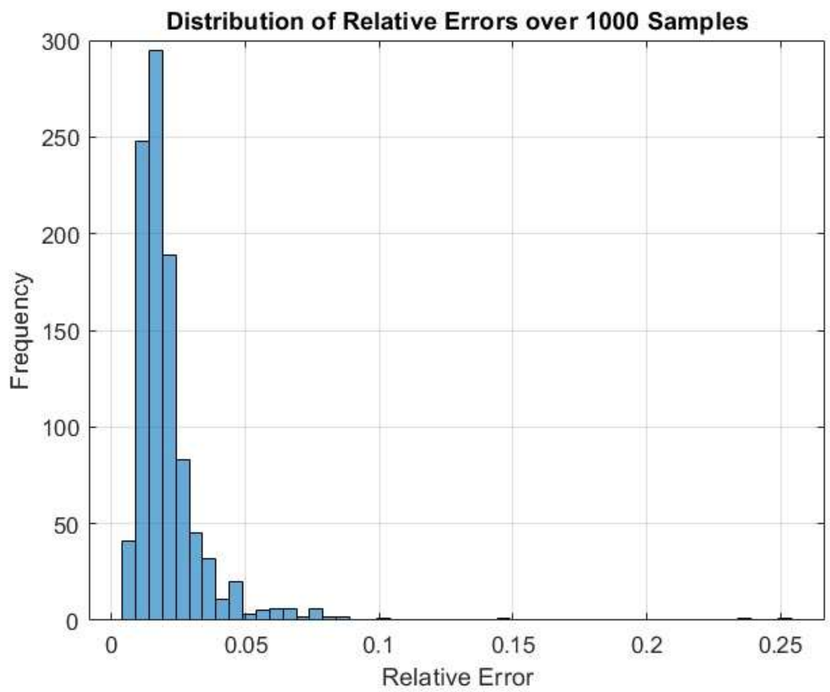
03

Results

Successive PGD and autoencoder

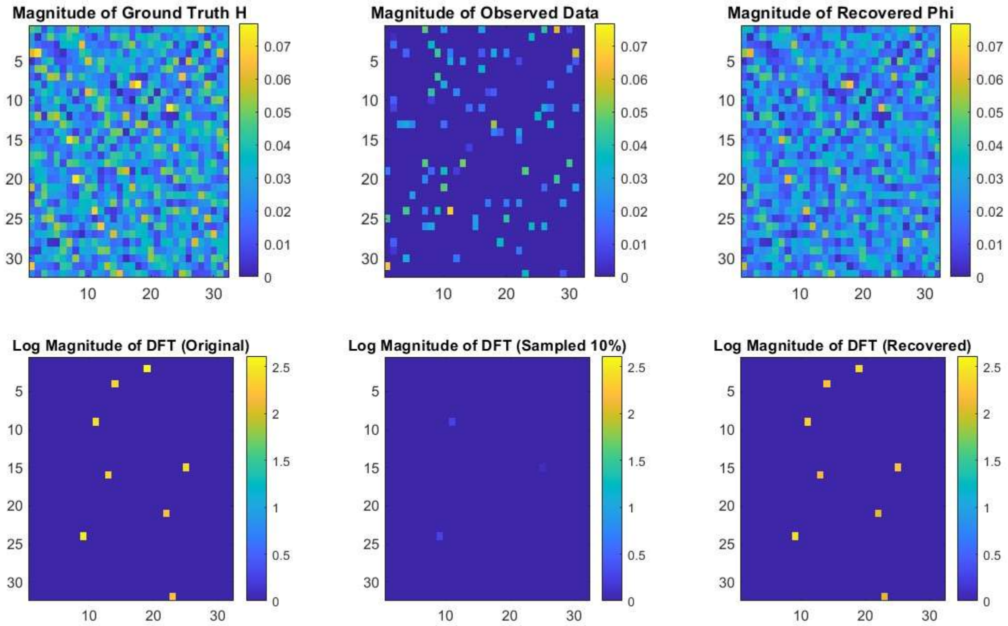
Successive Proximal Gradient Descent

After careful hyperparameter tuning of $\lambda_1, \lambda_2, \eta$



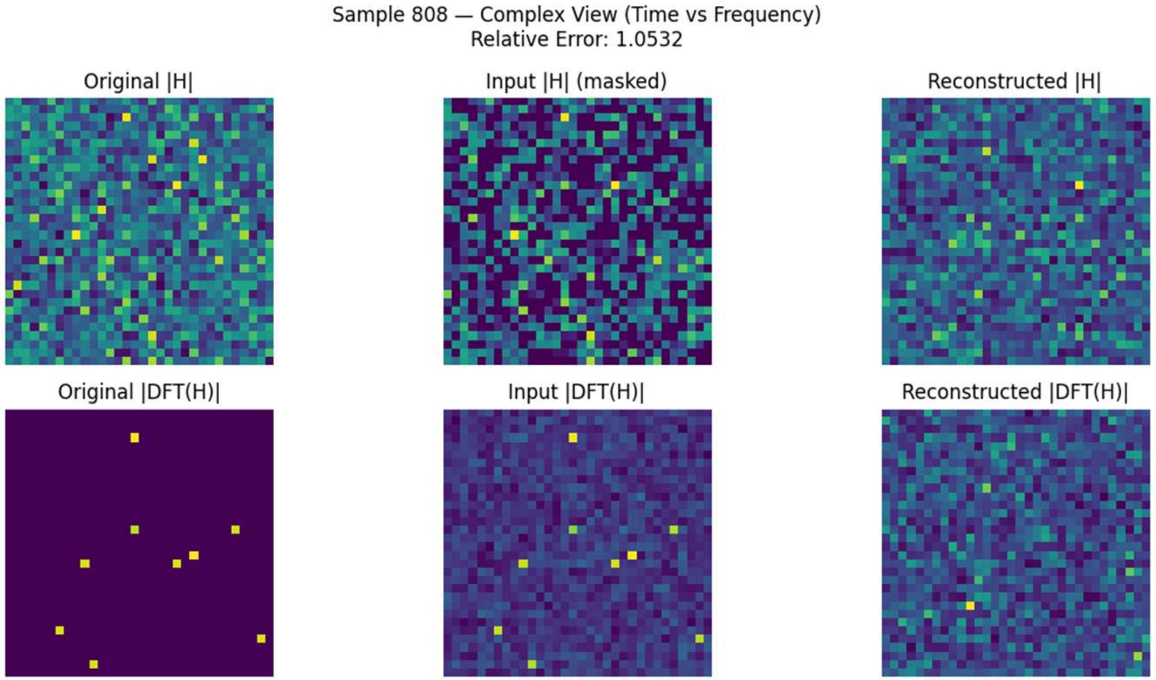
Successive Proximal Gradient Descent

After careful hyperparameter tuning of $\lambda_1, \lambda_2, \eta$



Autoencoder

Poor performance due to limited training data



04

Discussions and Conclusions

Successive PGD and autoencoder

Discussions

A fundamental challenge in many real-world scenarios.

- Successive PGD method consistently achieves low relative reconstruction.
- The autoencoder performs significantly worse, which is due to limited data set.

Successive PGD

- It requires much smaller training
- It can directly exploit known structural priors
    - Manual tuning of hyperparameters

Autoencoder

- It can learn complex nonlinear structures from data
    - Availability of large and diverse training datasets

Conclusions

In a setting with limited data set like ours, classical model-based methods such as successive PGD clearly outperform data-driven approaches.

However, with access to a much larger dataset, the autoencoder may be able to match or even surpass successive PGD, provided it learns the relevant structures effectively.

Thank you for your attention.