Variational Sparse Coding with Learned Thresholding

Kion Fallah, Christopher J. Rozell

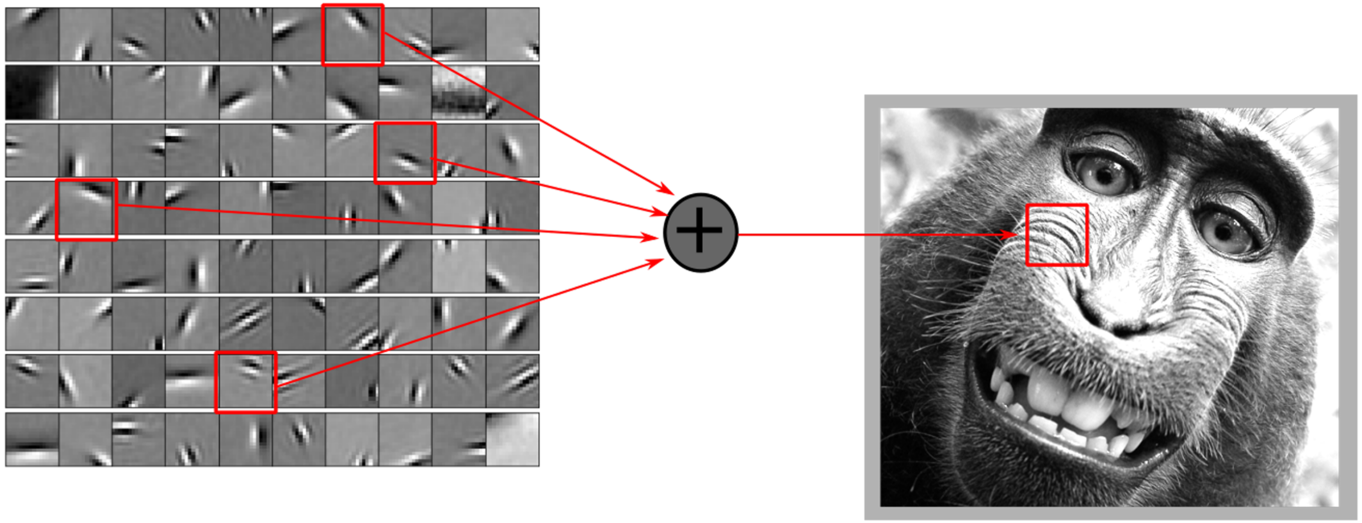
39th International Conference on Machine Learning

ICML
International Conference
On Machine Learning

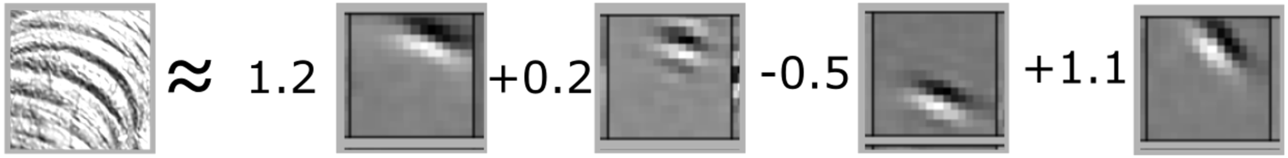
ML@GT
Machine Learning at Georgia Tech

Sparse Coding Background

- It has long been desirable to learn **efficient representations** to model natural images**.**
  - (Olshausen & Field 1996) use **sparse codes** to learn representations of images.

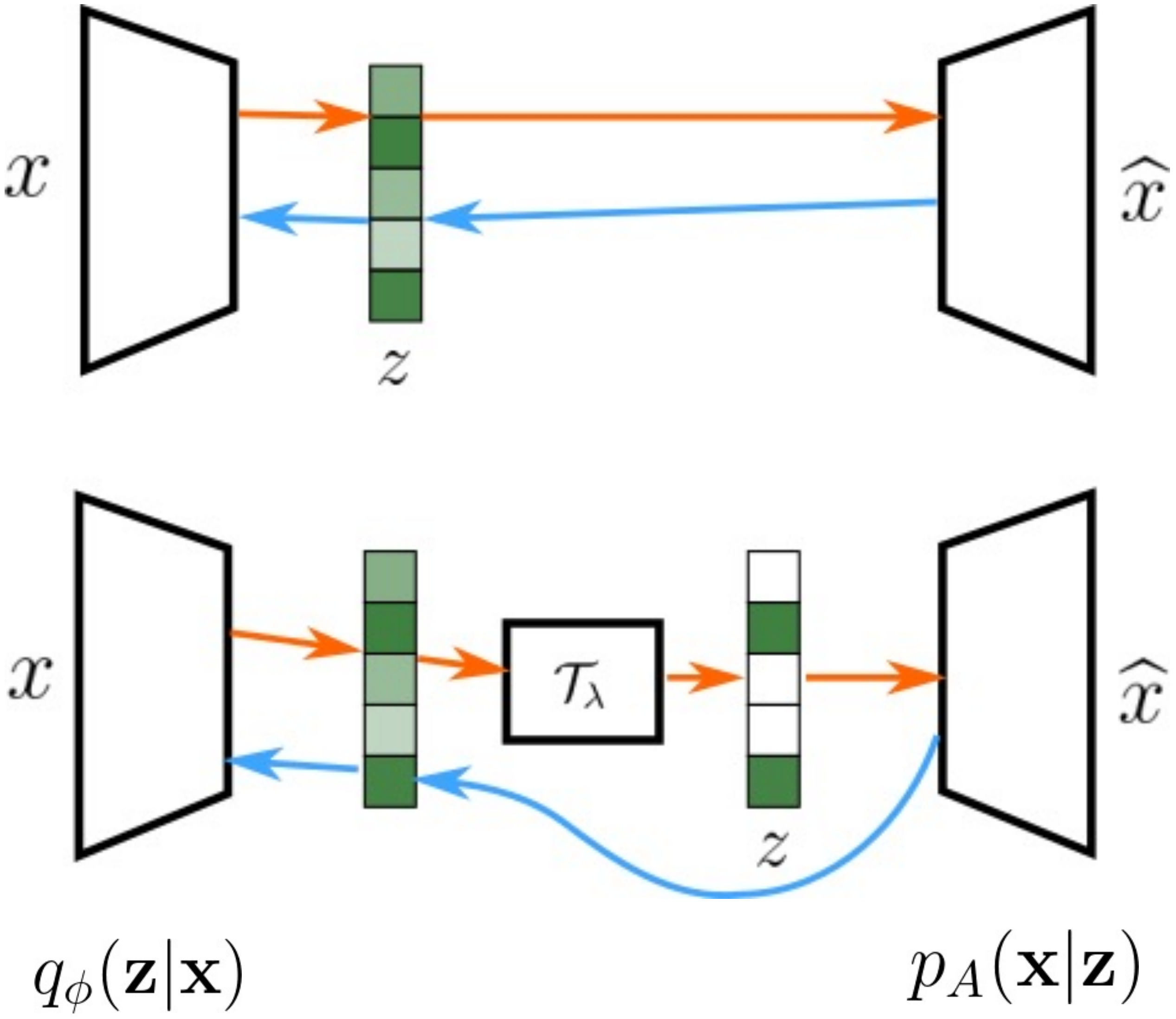


- Traditional sparse coding schemes rely on **iterative** optimization procedures for **inference.**
  - Computational bottleneck.



$$\mathbf{x} = \sum_{i=1}^{n} \mathbf{a}_i \mathbf{z}_i + \epsilon$$

Variational Sparse Coding



- **Black-box variational inference (BBVI)** allows one to **bypass expensive inference procedures** with a learned inference network.
- We propose an approach that applies a **shifted soft-threshold** of samples from a base distribution to incorporate sparsity.
- Orange arrows depict forward pass; blue arrows depict backwards pass.
  - Straight-through estimator to avoid numerical instability.

Details on Reparameterization

- Pass gradient directly from decoder to the inference network.
  - sg[] is identity in forward pass with partial derivative equal to zero.
  - Allow inference network to improve when a feature is not activated.

$$\mathbf{s}^k \sim q_\phi(\mathbf{s} \mid \mathbf{x}^k)$$

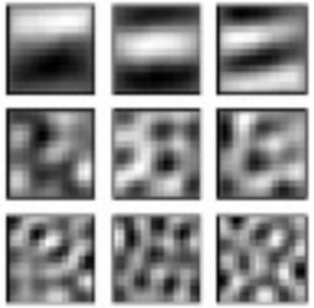
$$\widetilde{\mathbf{z}}^k = \mathbf{s}^k + \mathcal{T}_{\boldsymbol{\lambda}^k}\left(\mathrm{sg}\left[\mathbf{s}^k\right]\right) - \mathrm{sg}\left[\mathbf{s}^k\right]$$

- Alternative sampling motivated by (Olshausen & Field 1996) to pick best support for sparse codes.

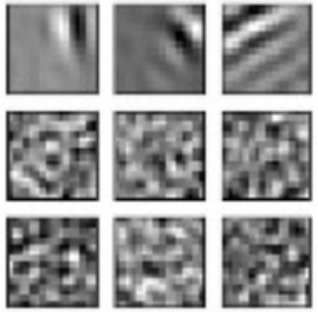
$$\widehat{\mathcal{L}}^k_{max} = \max_j \mathcal{L}^{j,k}$$

Results on Linear Sparse Coding

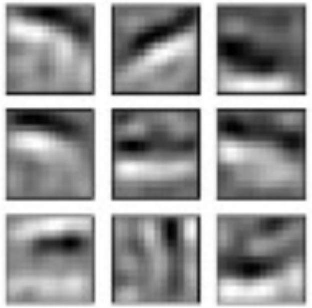
Method/Prior Distribution	Validation Loss J=1	Validation Loss J=20
FISTA (baseline)	1.01E+02	–
Gaussian	1.35E+03	1.35E+03
Laplacian	5.96E+02	5.79E+02
Spike-and-slab	2.52E+02	2.39E+02
Thresholded Gaussian	2.32E+02	2.30E+02
Thresholded Gaussian+Gamma	2.85E+02	2.70E+02
Thresholded Laplacian	**1.98E+02**	**1.94E+02**
Thresholded Laplacian+Gamma	2.23E+02	2.11E+02



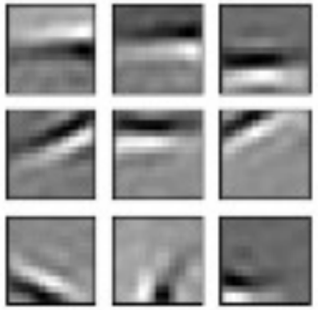
[1] Gaussian, J=1



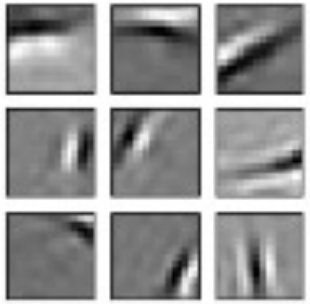
[2] Laplacian, J=1



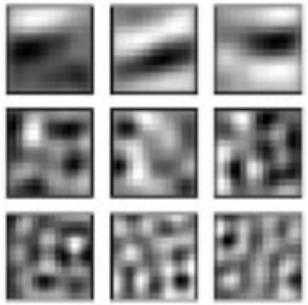
[3] Spike-and-slab, J=1



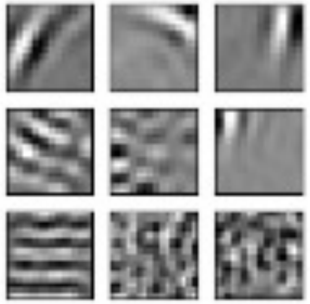
(Ours) Thresh Laplacian, J=1



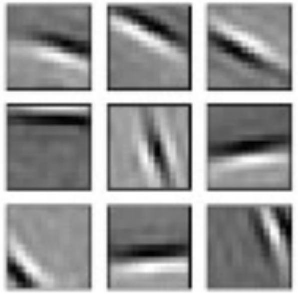
FISTA



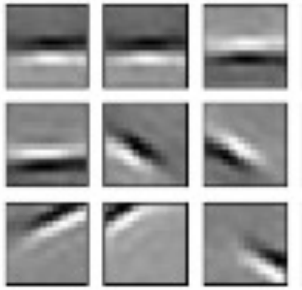
[1] Gaussian, J=20



[2] Laplacian, J=20



[3] Spike-and-slab, J=20



(Ours) Thresh Laplacian, J=20

[1] D. Kingma & M. Welling, *Auto-Encoding Variational Bayes (2013).*
[2] G. Barello, A. Charles, & J. Pillow, *Sparse-Coding Variational Auto-Encoders (2018).*
[3] F. Tonolini et al., *Variational Sparse Coding (2020).*

Estimating Linear Dictionary for Visualizing Features

(a) Gaussian

(b) Thresholded Gaussian+Gamma

First row: Recovered dictionary entries with highest magnitude.
Bottom row: Random dictionary entries.

- Train sparse VAE on CelebA image dataset with 512 features, 10% sparsity, and J=10 max ELBO samples.
- Recent sparse coding work shows a dictionary can be recovered using the sparse codes [1, 2].
- Dictionaries recovered from sparse VAEs better resemble independent components of input data.

[1] Isley et al., *Deciphering subsampled data: adaptive compressive sampling as a principle of brain communication (2010)*.
[2] K. Fallah***,** A. Willats*, N. Liu, & C. Rozell, *Learning sparse codes from compressed representations with biologically plausible local wiring constraints (2020).*