Constrained neural networks for inverse problems

DISC & TIAI Annual Symposium

Thursday, April 20, 2022

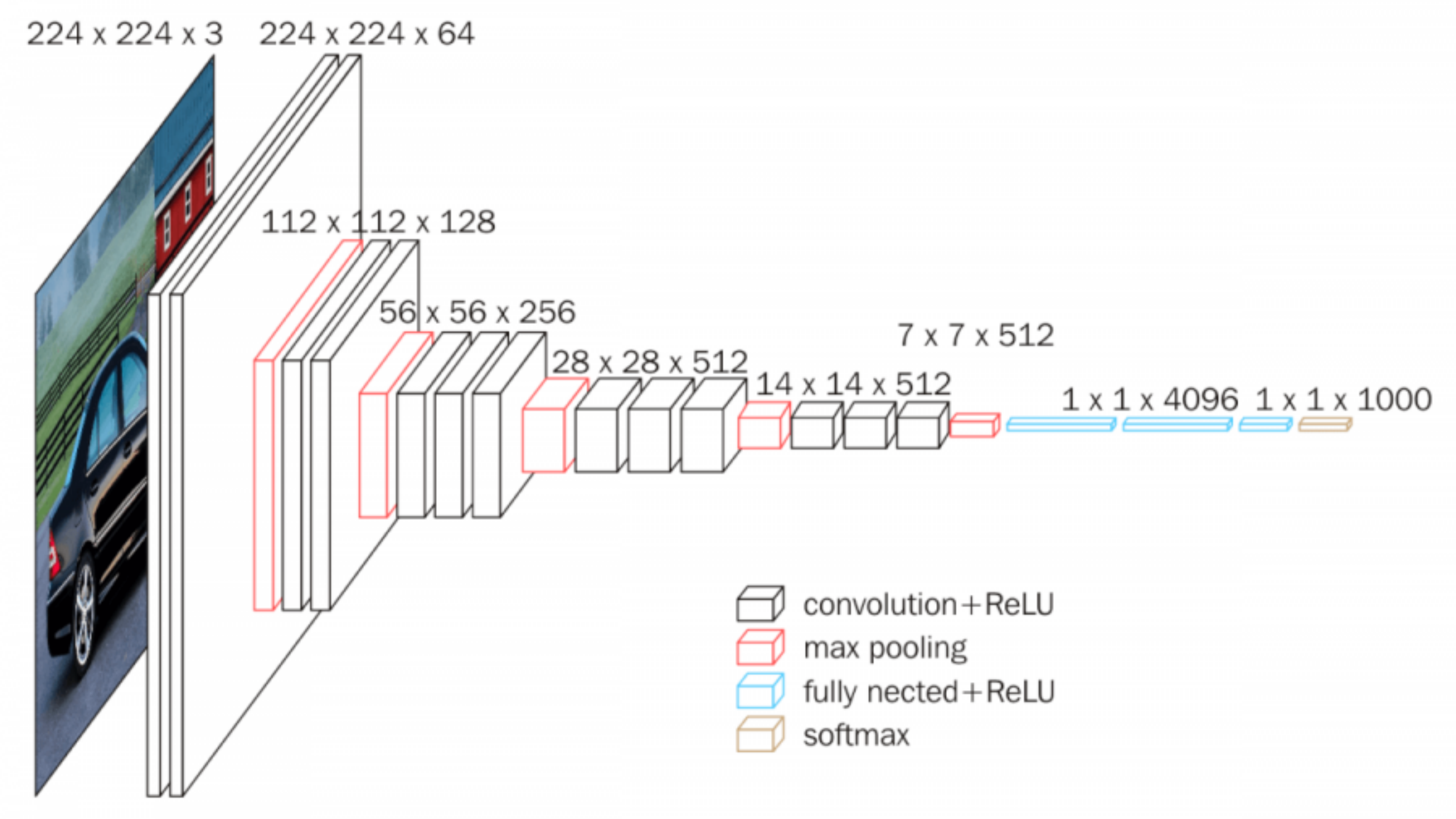
Alexander Lin
alin@seas.harvard.edu
https://sites.google.com/view/alexanderlin/

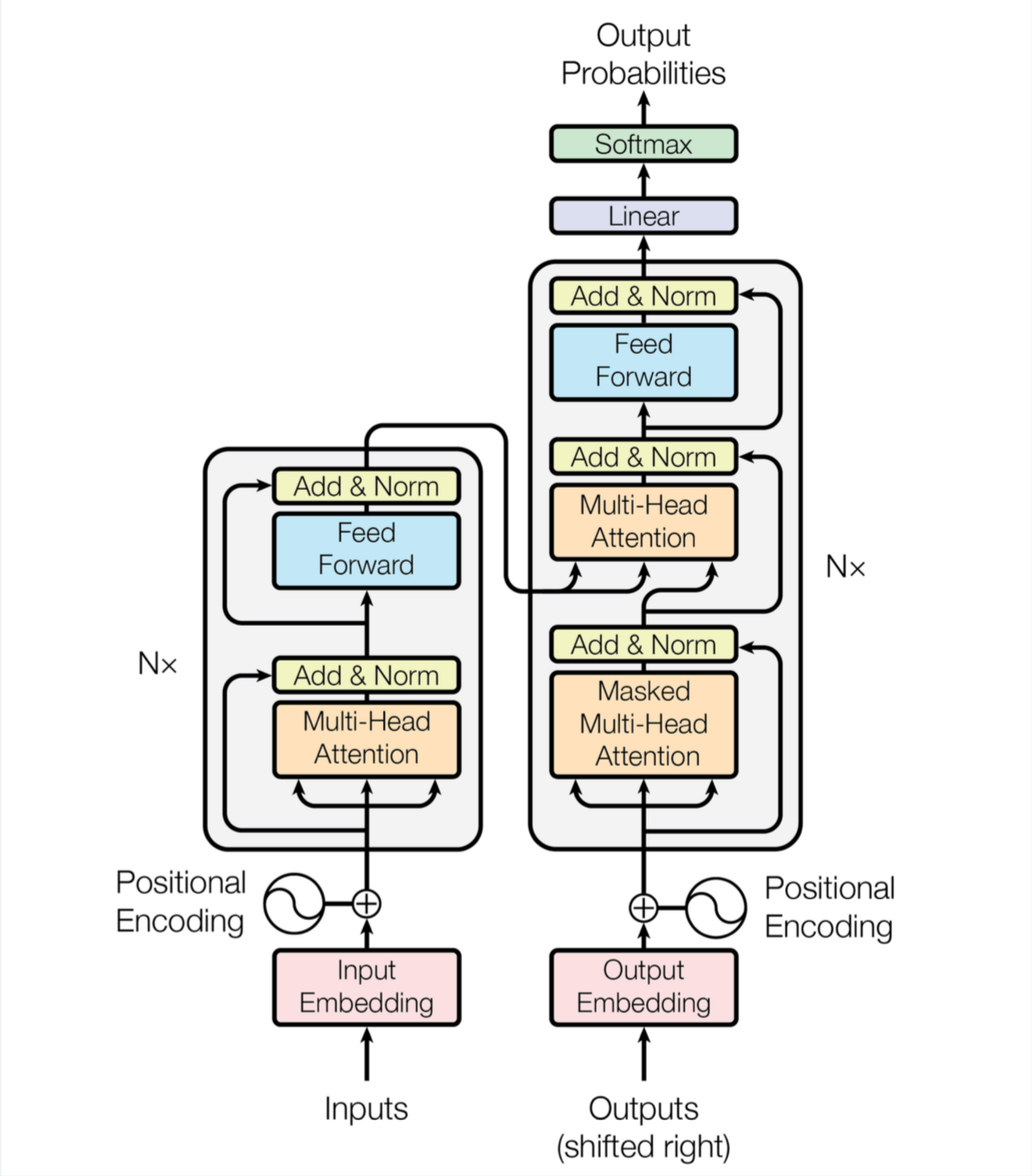
Emmanouil Theodosis
etheodosis@seas.harvard.edu
manosth.github.io/

Motivation

Interpretability

Deep learning is empirical and hard to reason





Crafting representations

How do we instill domain knowledge?

ReLU vs sigmoid vs tanh vs ....

Analysis is usually post-hoc.

Optimization-informed deep learning

Linear generative model

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$$

Inverse problem

$$\boldsymbol{x}^* = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{y}$$

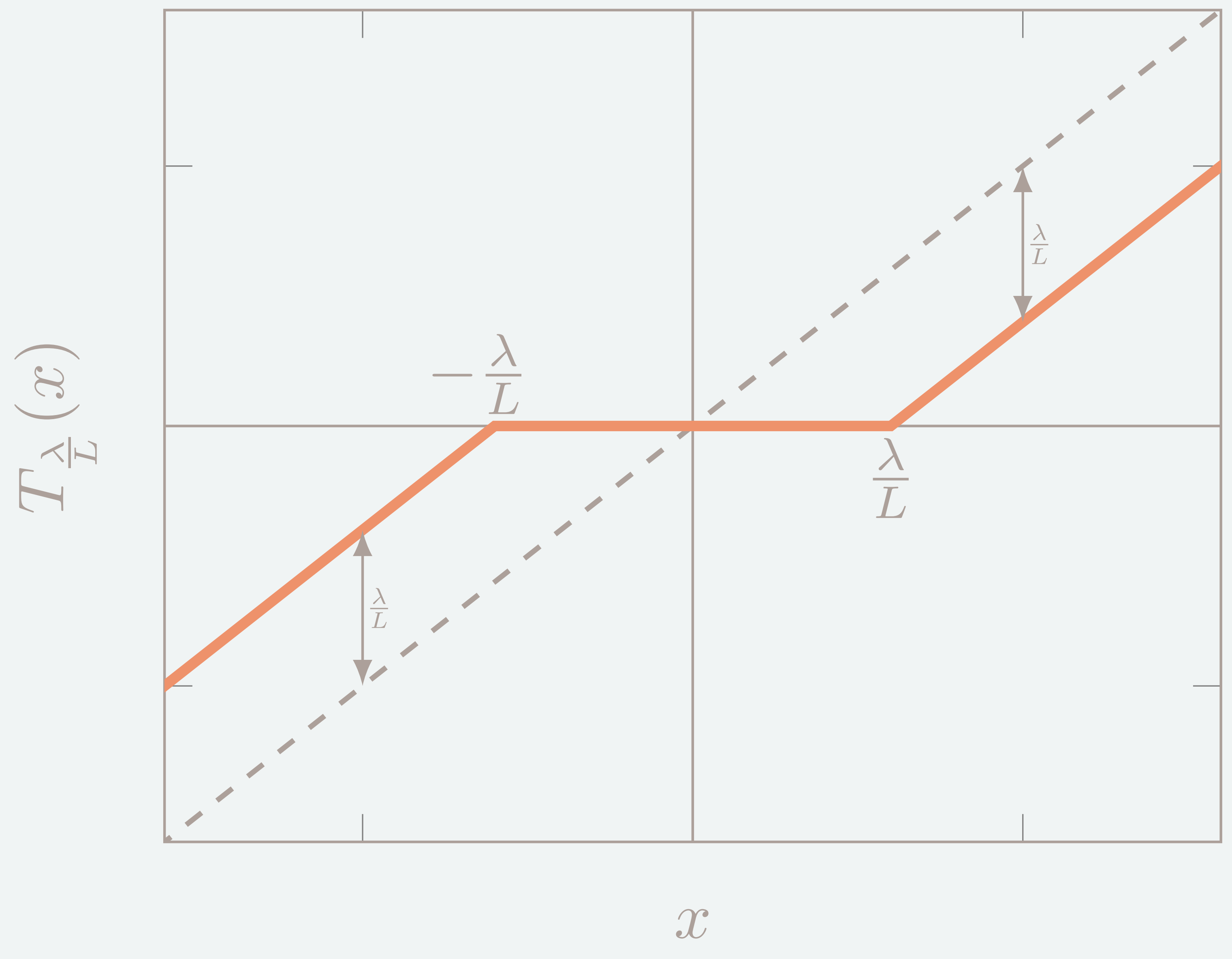
In many real world problems, $\boldsymbol{A}$ is over-complete, e.g.

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \quad \text{s.t.} \quad \boldsymbol{x} \text{ sparse}$$

$$\min_{\boldsymbol{x}} ||\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}||_2^2 + \lambda||\boldsymbol{x}||_1$$

Iterative soft thresholding

$$\boldsymbol{x}^{(l+1)} \leftarrow T_{\frac{\lambda}{L}}(\boldsymbol{x}^{(l)} + \frac{1}{L}\boldsymbol{A}^T(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}^{(l)}))$$



From sparsity to group sparsity

Motivation

Sparsity is separable, but lacks structure that is useful for clustering.

Signal model

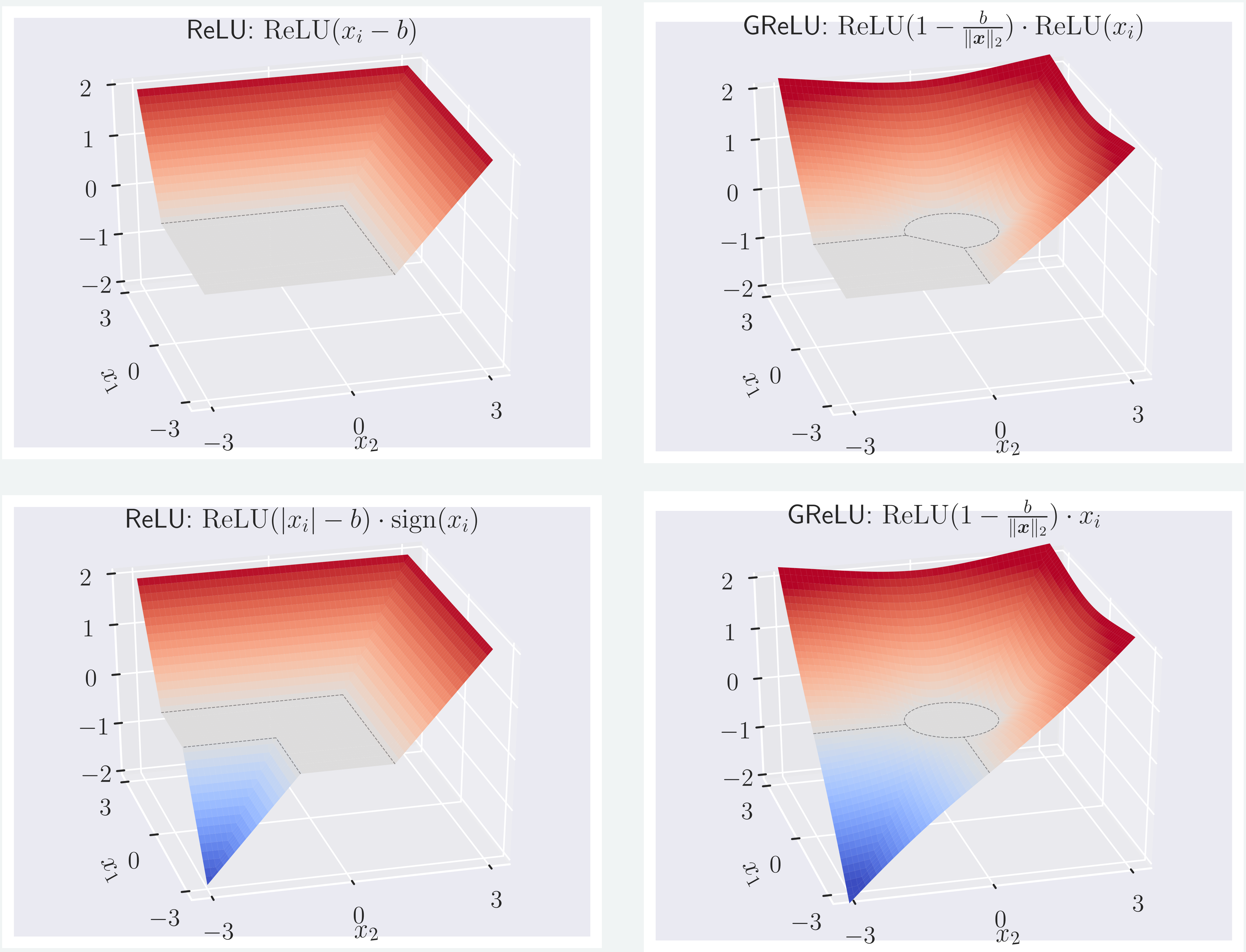
Non-zero coefficients appear in groups

$$\boldsymbol{y} = \sum_{g \in G} A_g \boldsymbol{x}_g$$

Nonlinearity

No longer separable, but more complex.

Sparsity | Group Sparsity



Group sparse priors

Model

Assume a group sparse prior

$$\min_{\boldsymbol{x}} \; \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 \; + \lambda \sum_{g \in S} \|\boldsymbol{x}_g\|_2$$

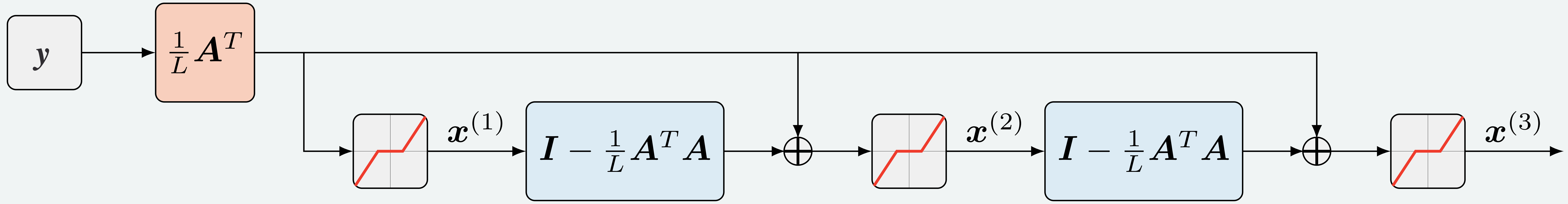
Proximal operator

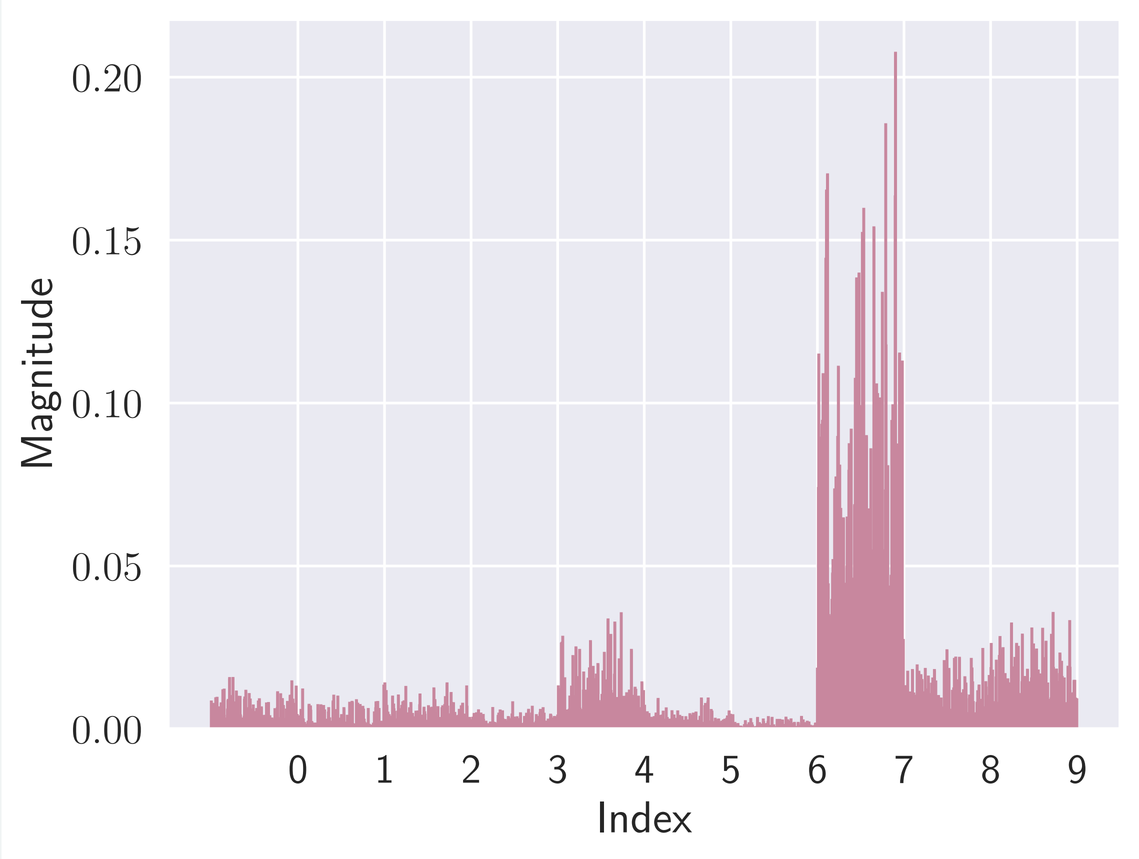
Projection to the solution set via

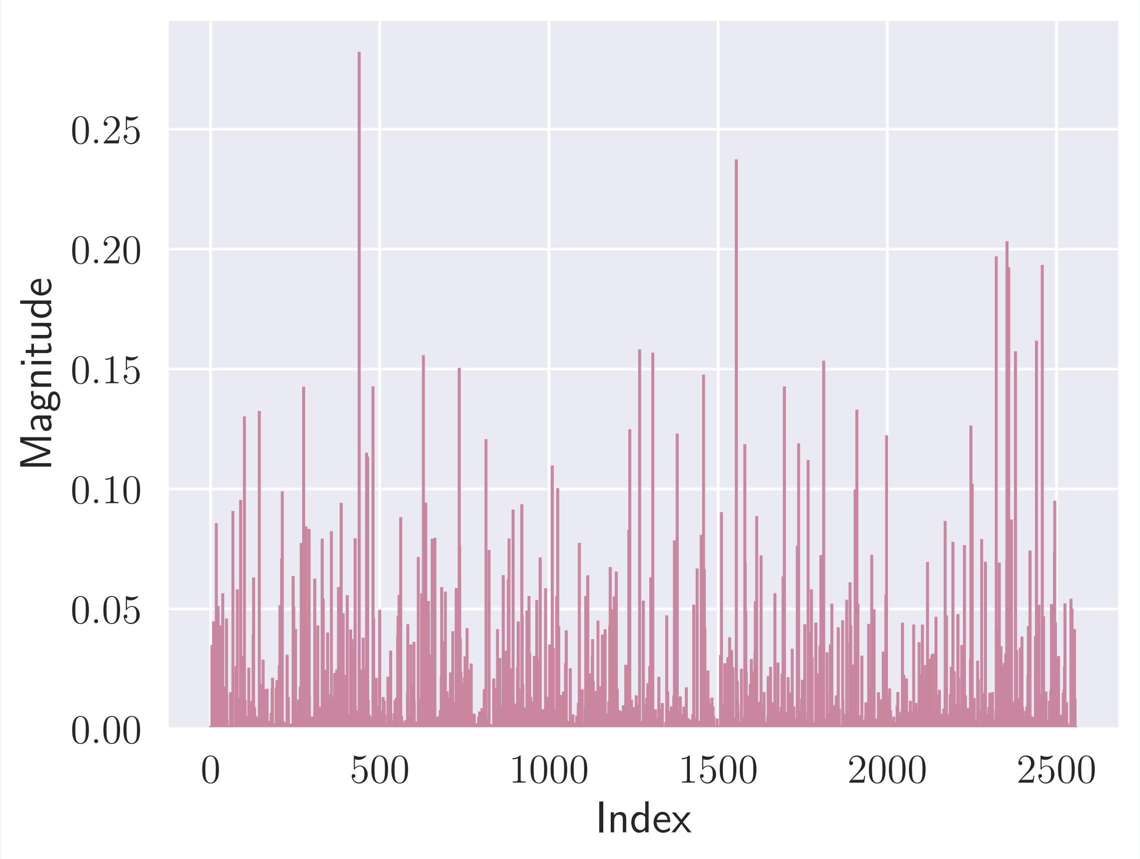
$$\sigma_\lambda(\boldsymbol{x}_g) = \left(1 - \frac{\lambda}{\|\boldsymbol{x}_g\|_2}\right)_+ \boldsymbol{x}_g$$

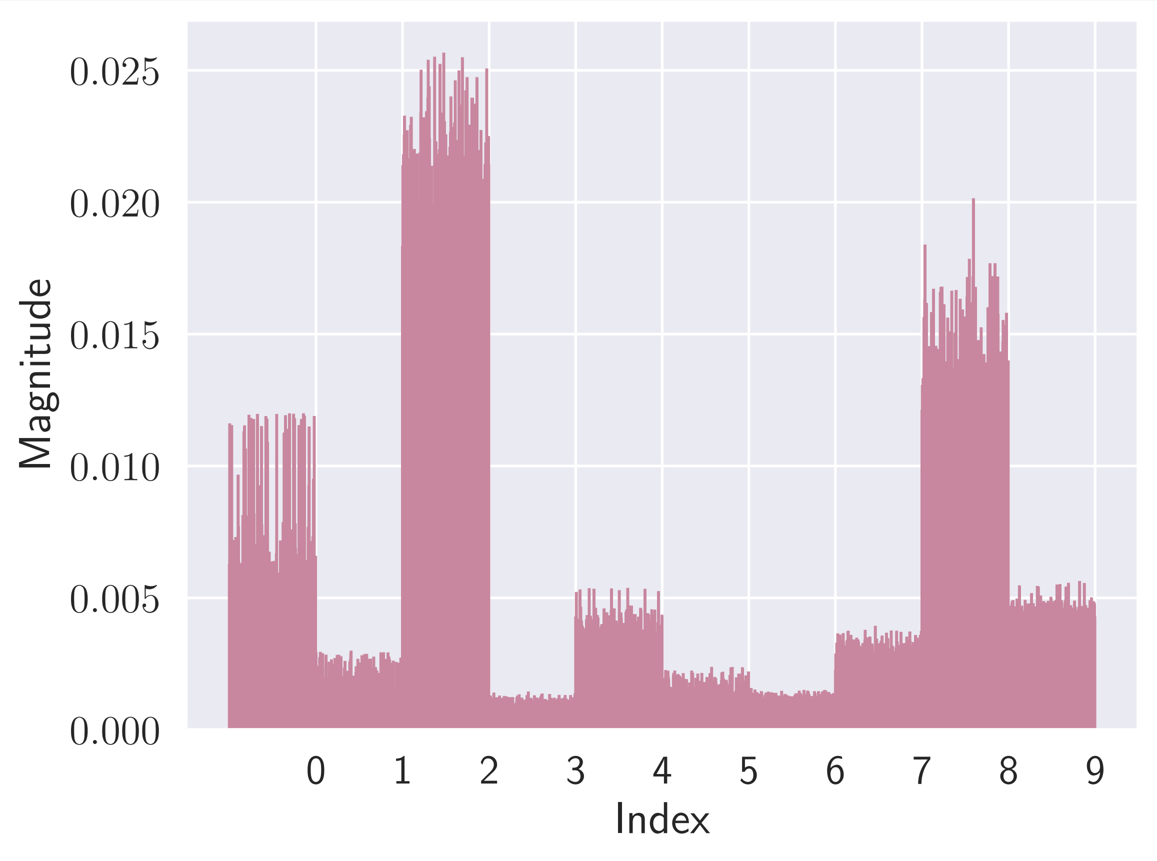
Architecture

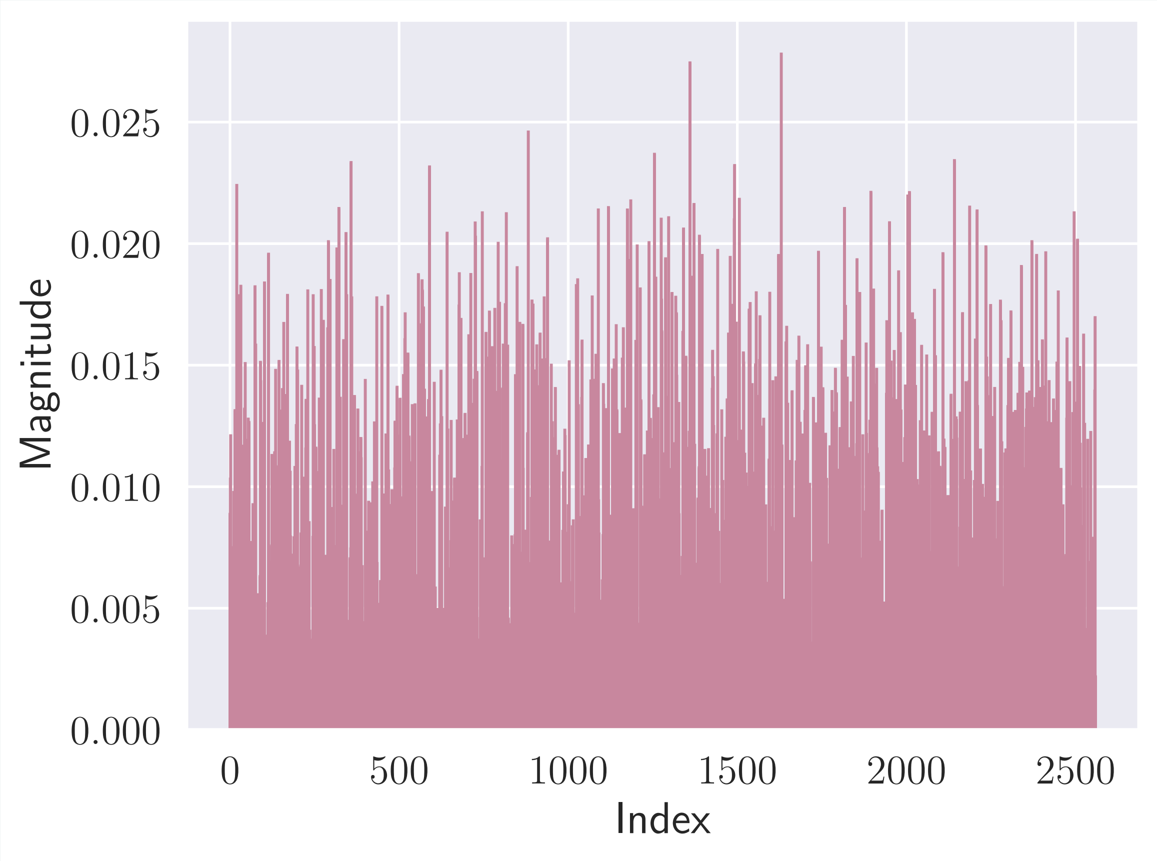
Residual connections to the input

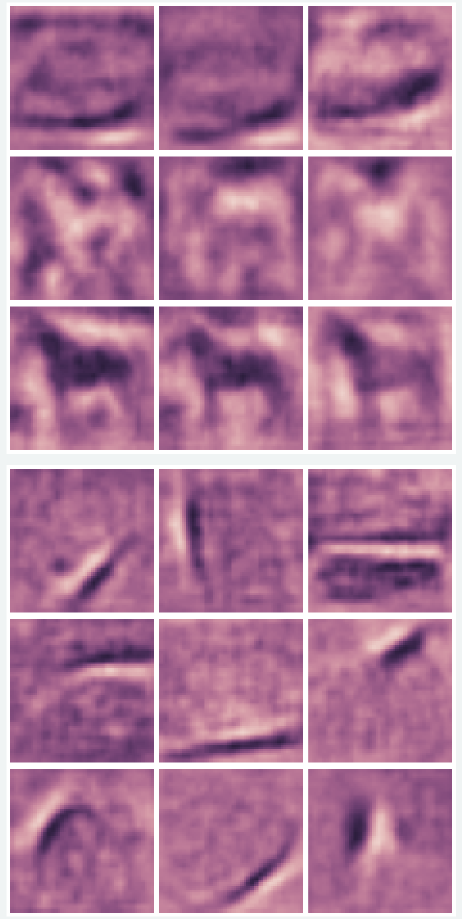










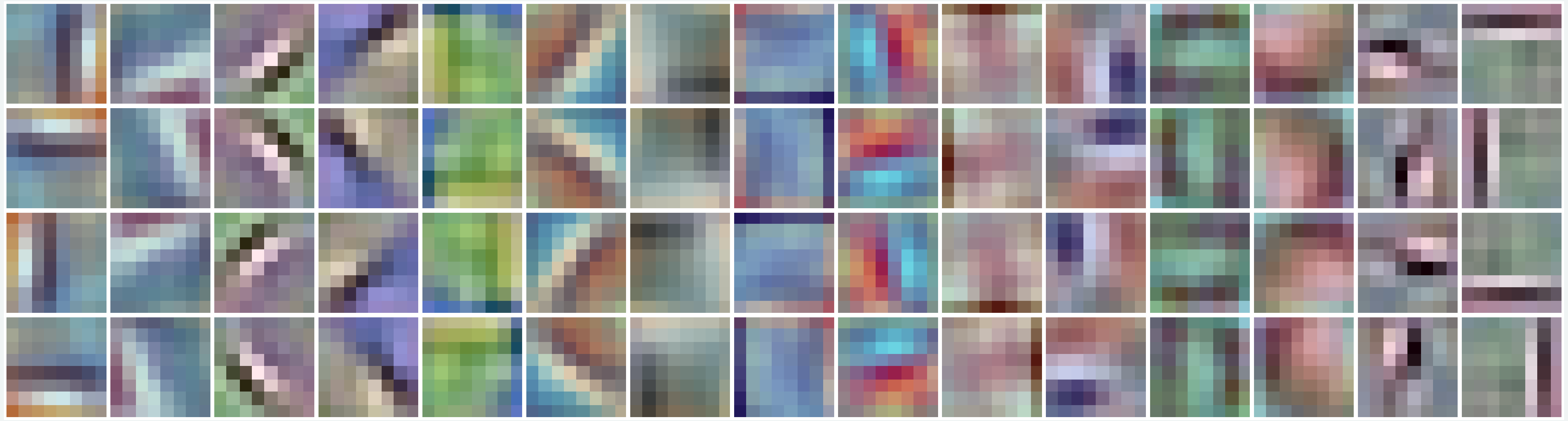


Unrolling equivariances

Equivariance

"Consistent" behavior w.r.t. an operator

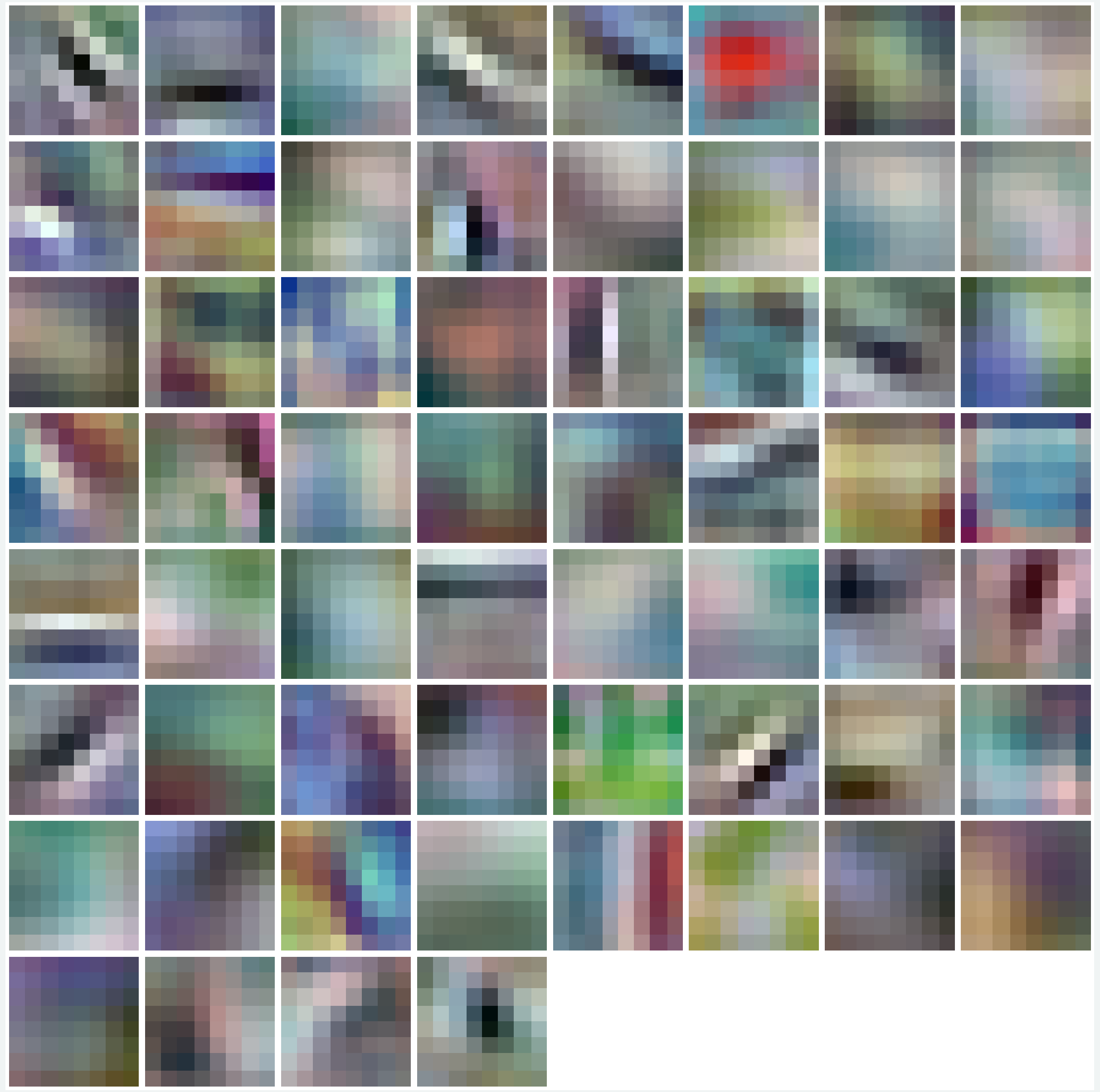
$$f(T(\boldsymbol{x})) = T'(f(\boldsymbol{x}))$$



Symmetry

Model fixed rotations

$$G = \{e, R_\theta, R_\theta^2, \ldots, R_\theta^{k-1}\}$$



Unrolled network

Layer weights

$$\boldsymbol{W}_l = [\boldsymbol{w}_l \quad R_\theta(\boldsymbol{w}_l) \quad \ldots R_\theta^{k-1}(\boldsymbol{w}_l)]$$

Learning equivariances

Discover symmetries

Model finite group actions

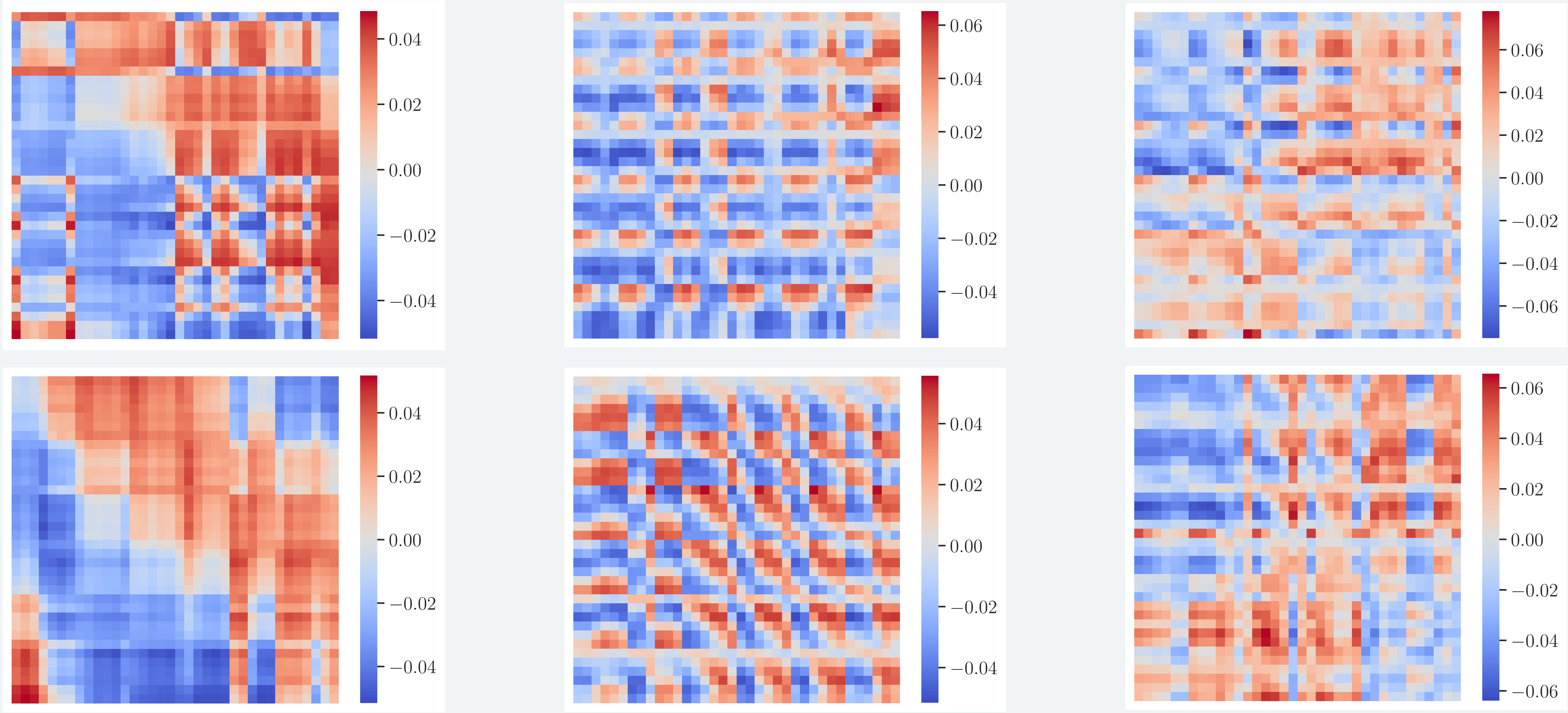
$$G = \{e, g, g^2, \ldots, g^{k-1}\}$$

"Lifting" filters

To learn any group we need to "lift" (flatten) filters

$$(\cdot)_f : \mathbb{R}^{p \times q} \to \mathbb{R}^{p \cdot q}$$

$$\boldsymbol{X} \mapsto [\boldsymbol{X}_1; \boldsymbol{X}_2; \ldots; \boldsymbol{X}_q]$$



(Semi-) Skew-symmetric **Toeplitz** **Block/multi-scale**

The Bayesian perspective: From optimization to uncertainty

Optimization-informed deep learning

Linear generative model

$$y = Ax$$

Inverse problem

$$x^* = (A^T A)^{-1} A^T y$$

In deep learning A is overcomplete

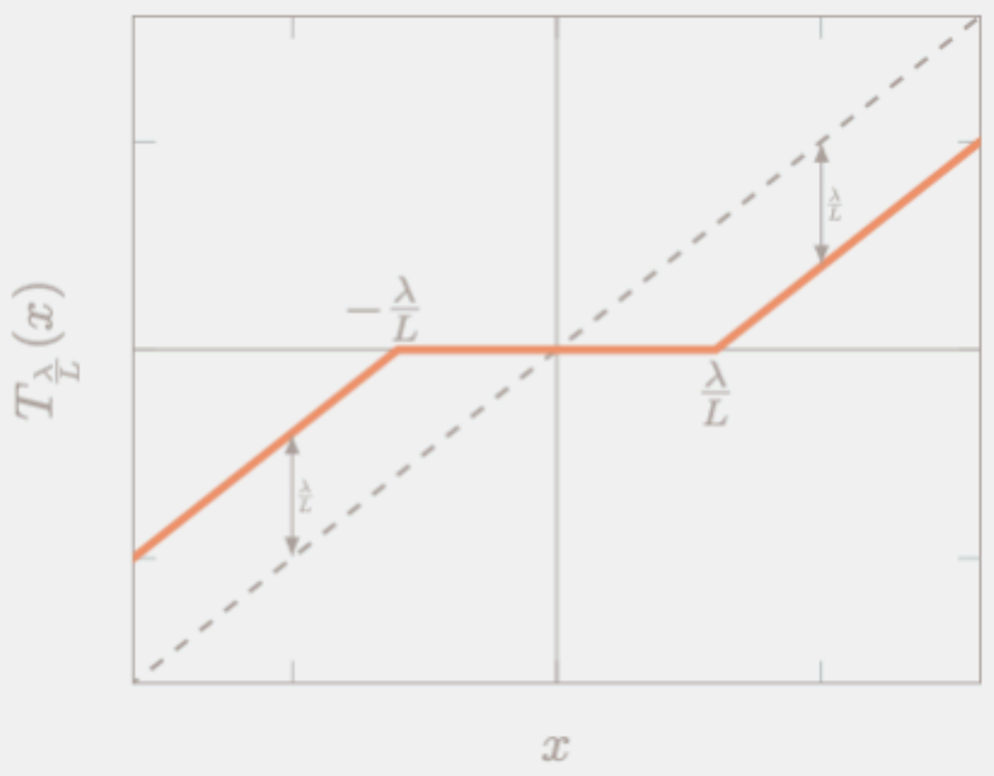
$$y = Ax \quad \text{s.t.} \quad x \text{ sparse}$$

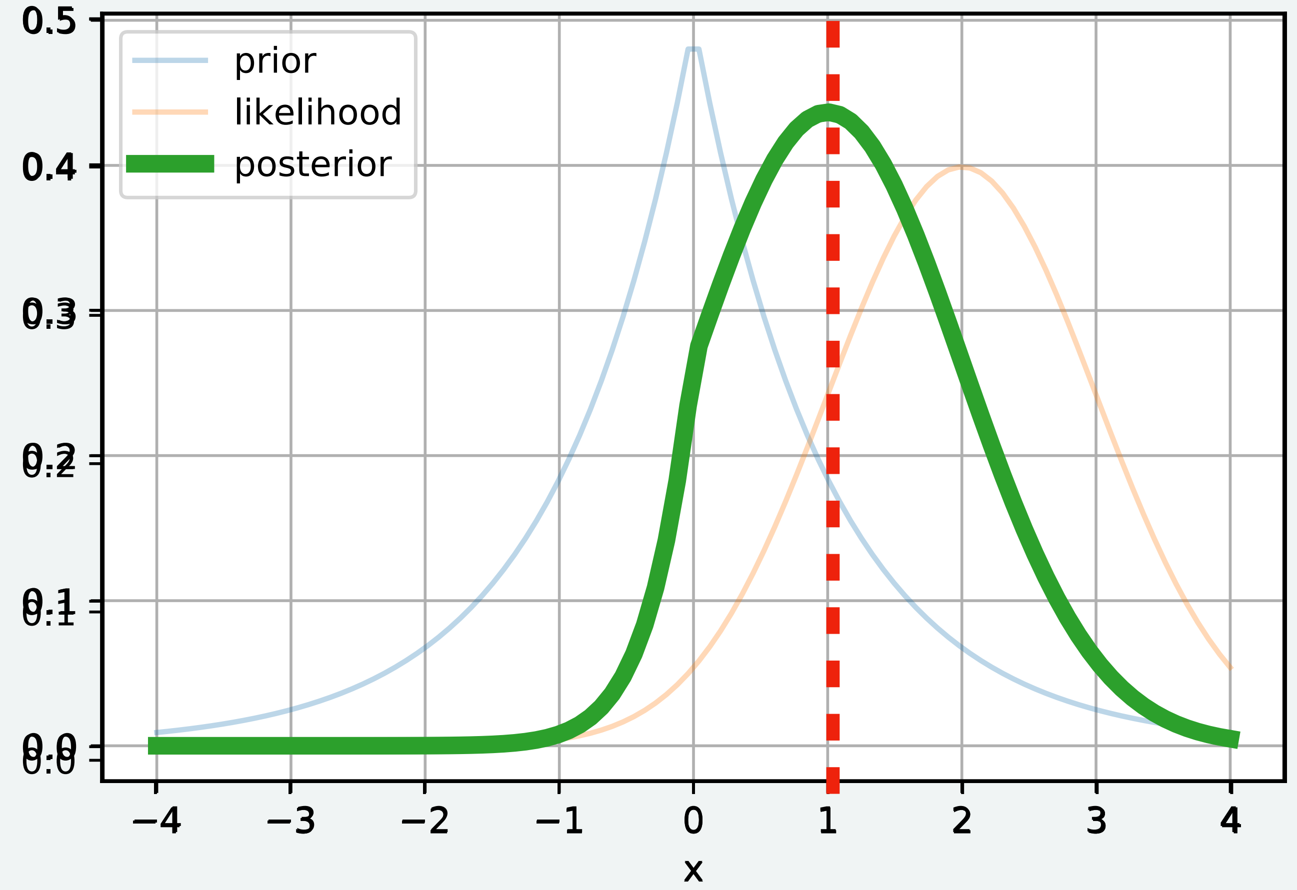
$$\min_x ||y - Ax||_2^2 + \lambda ||x||_1$$

Data Likelihood Prior

Iterative soft thresholding

$$x^{(l+1)} \leftarrow T_{\frac{\lambda}{L}}(x^{(l)} + \frac{1}{L} A^T (y - Ax^{(l)}))$$



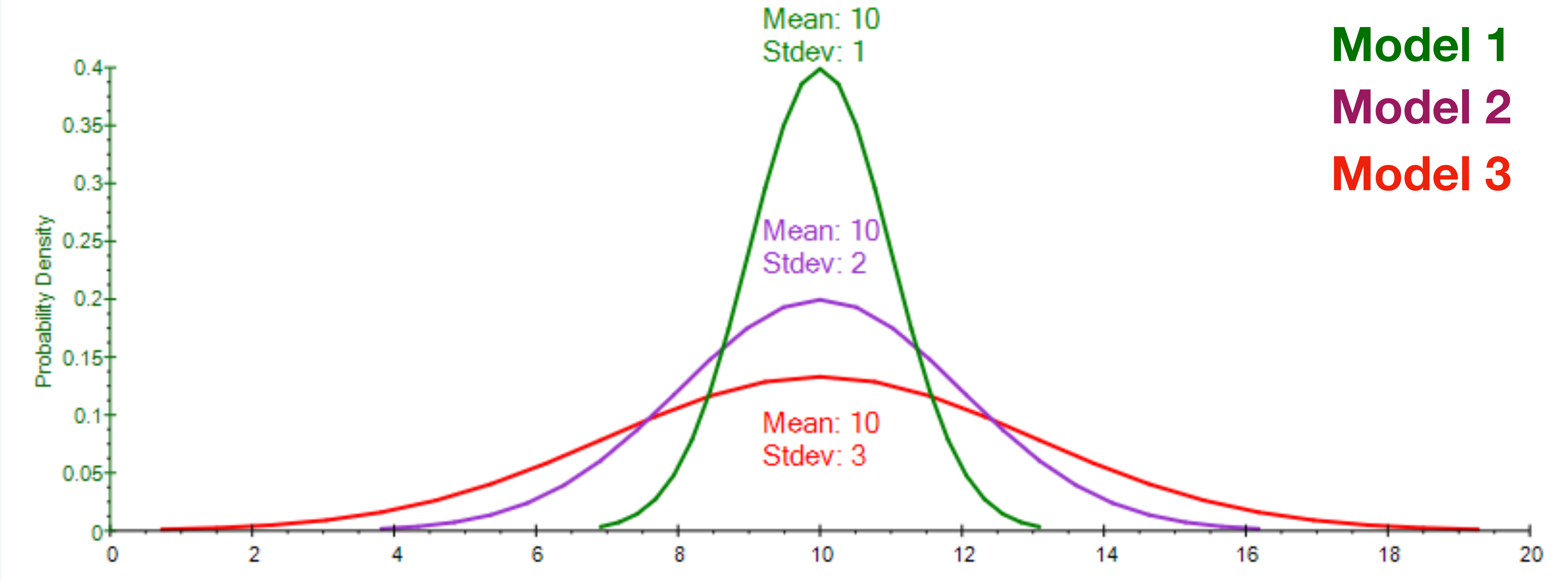


Optimization finds the most likely point

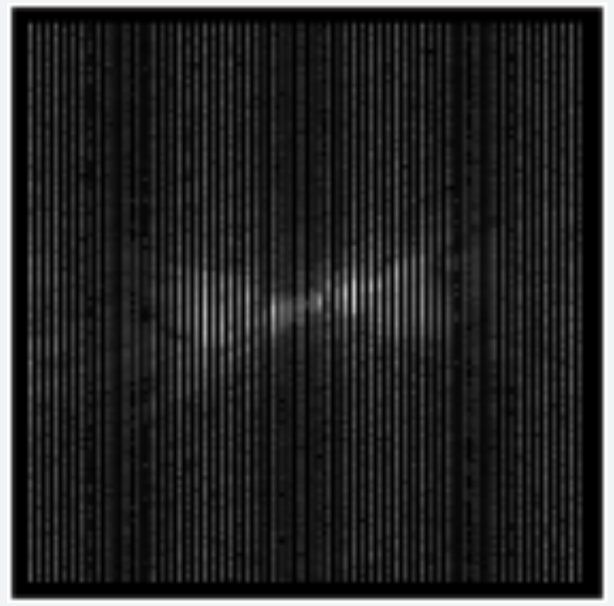
But what if we recover the whole distribution instead?

Why care about Bayesian uncertainty?

It can distinguish between good & bad models

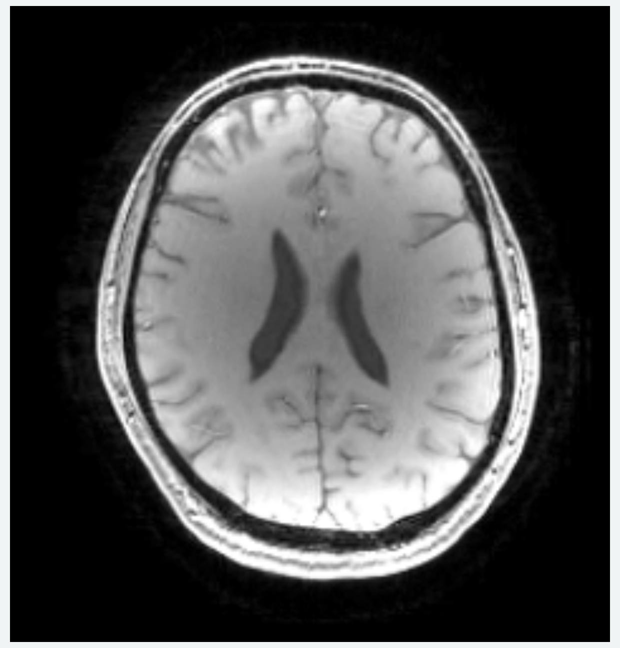


It can tell you how much to trust your model

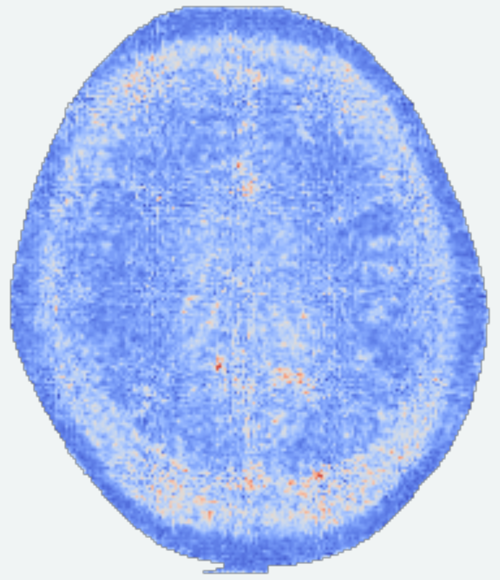


Scanner Data

MODEL →



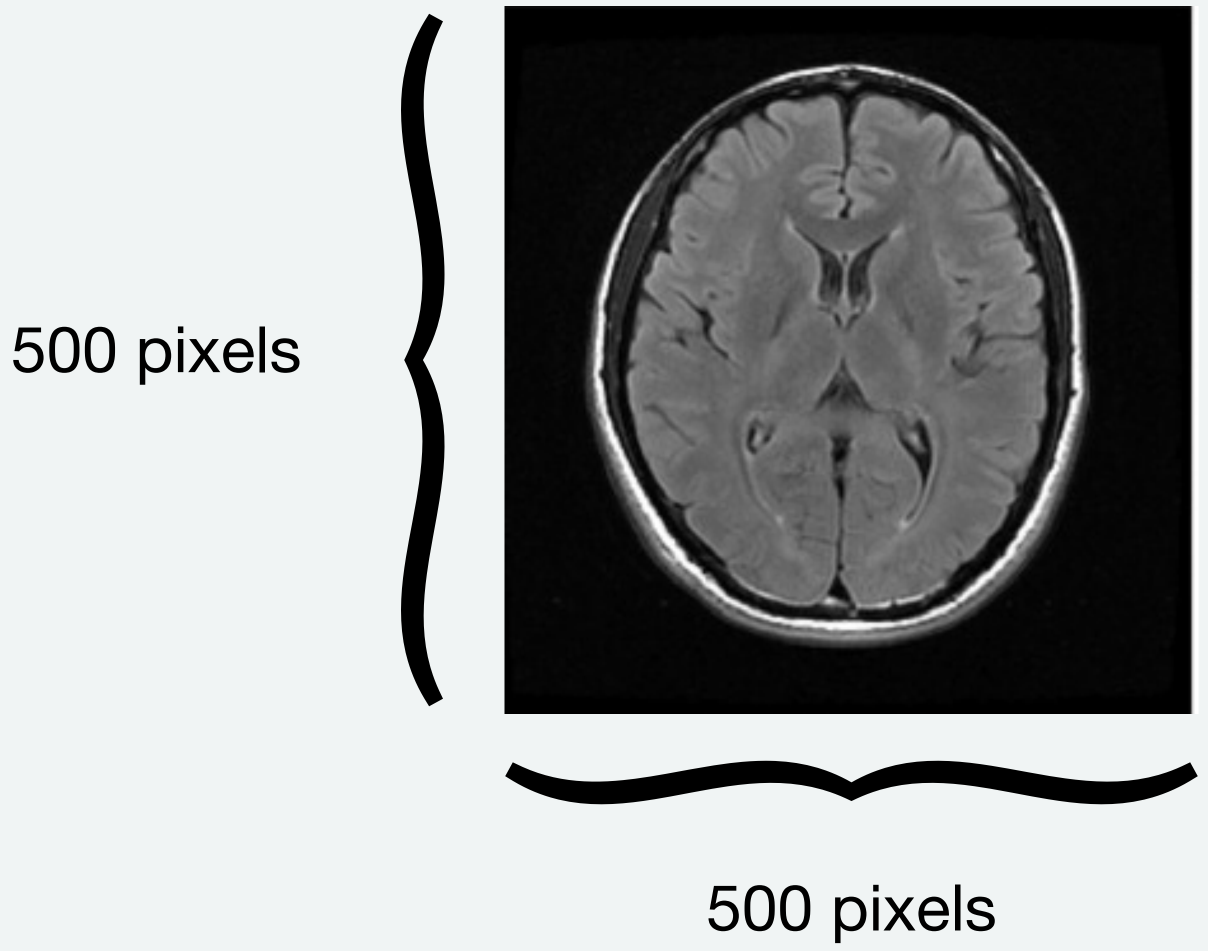
Reconstructed Image



Variance Map

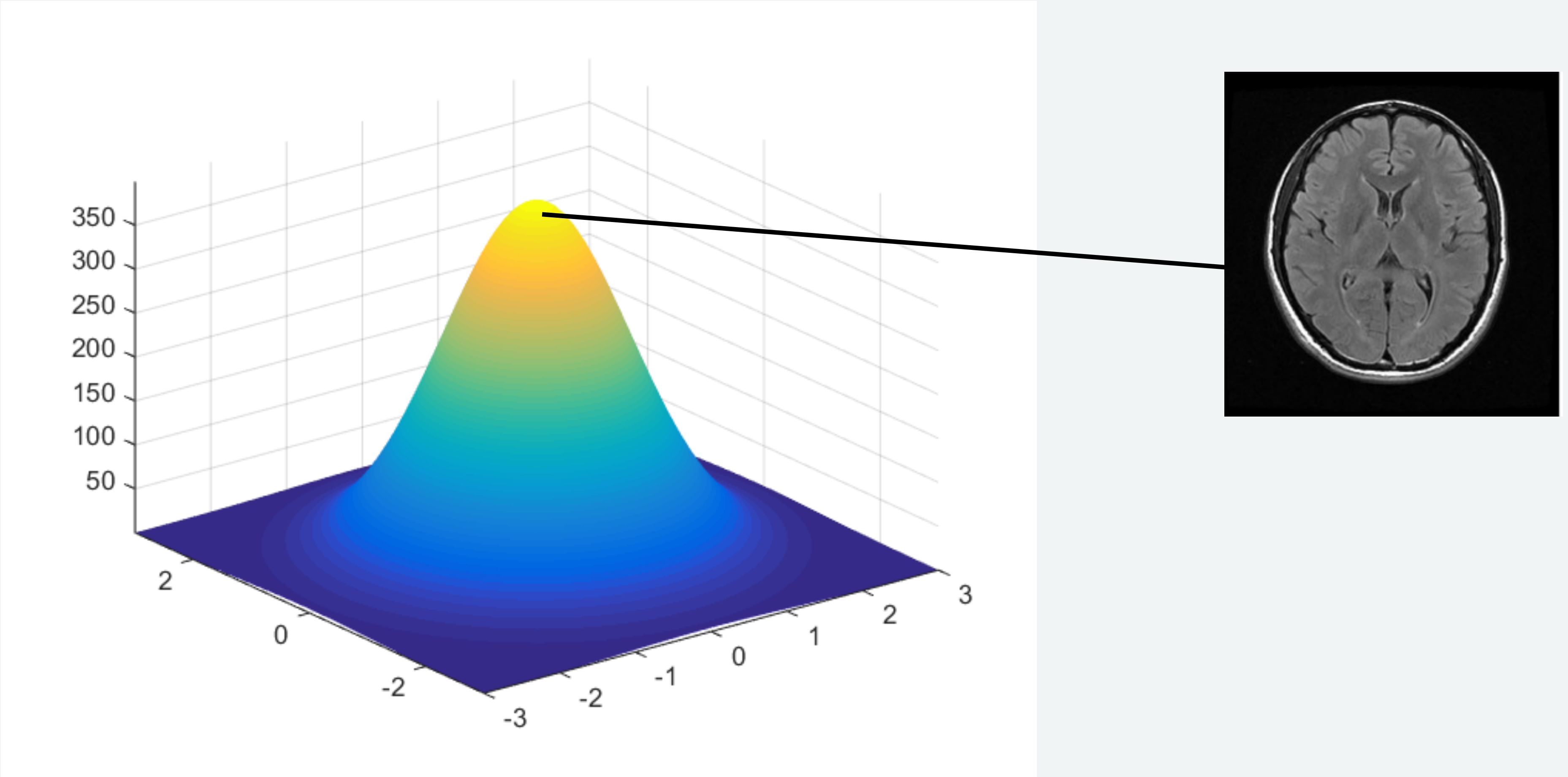
But…uncertainty is expensive (computationally)!

The optimization approach finds a <u>single point</u> (e.g. an image)



D = 250,000 pixels

The Bayesian approach finds a <u>distribution</u> over possible images



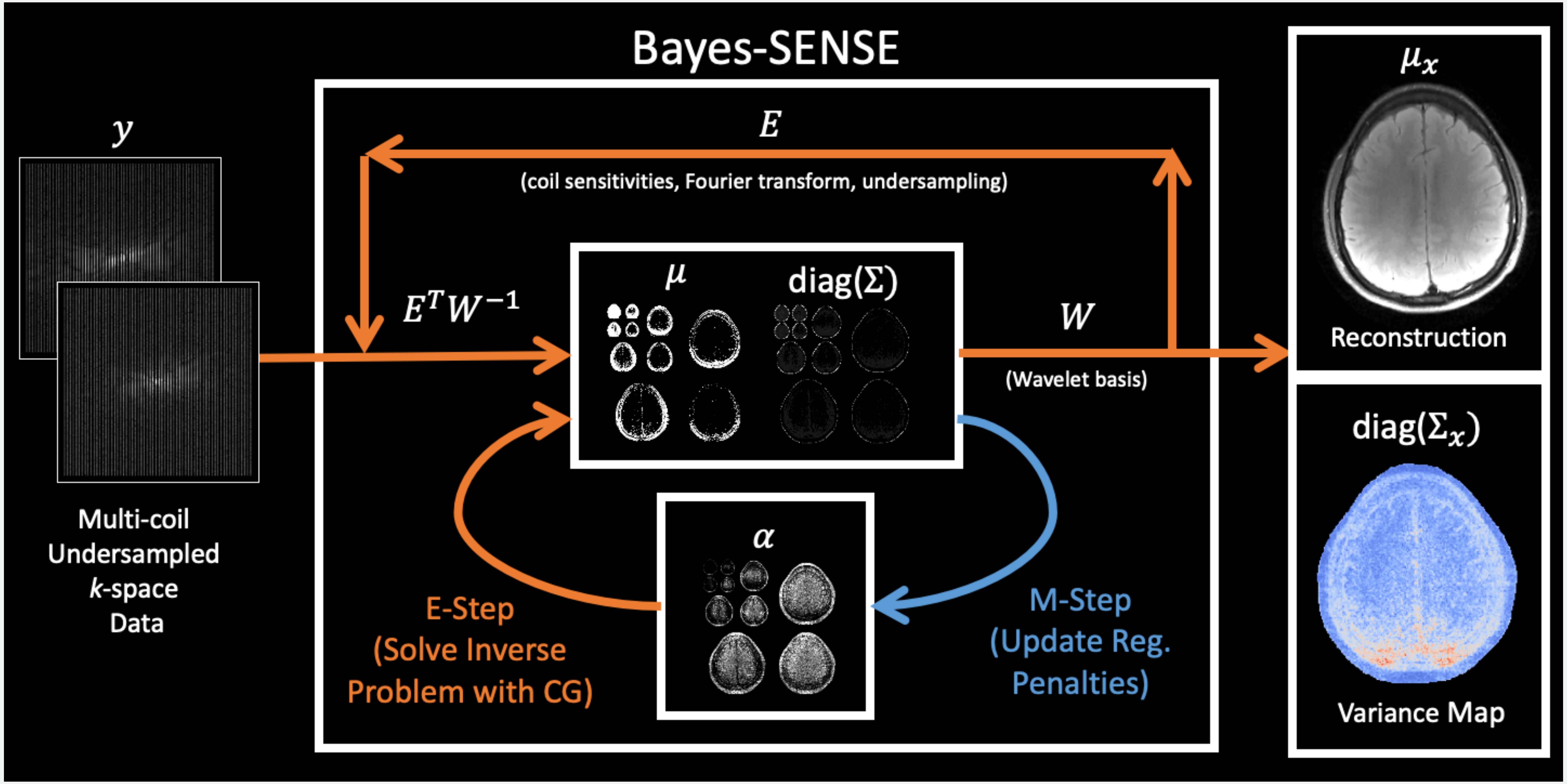
Multivariate Normal $\mathcal{N}(\mu, \Sigma)$

$\mu \in \mathbb{R}^D$ **and** $\Sigma \in \mathbb{R}^{D \times D}$

***D* x *D* = 62,500,000,000 values**

Bayesian Unrolling

A scalable framework for uncertainty quantification



Collaborators and publications

Publications

E. Theodosis and D. Ba, "Learning silhouettes with group sparse autoencoders", in *International Conference on Acoustics, Speech, and Signal Processing*, 2023

A. Tasissa, **E. Theodosis**, B. Tolooshams, and D. Ba, "Discriminative reconstruction via simultaneous dense and sparse coding", *Under review*, 2022

E. Theodosis, B. Tolooshams, P. Tankala, A. Tasissa, and D. Ba, "On the convergence of group sparse autoencoders", in *arXiv*, 2020

A. Lin, B. Tolooshams, Y. Atchadé, "Bayesian unrolling: Scalable, inverse-free maximum likelihood estimation of latent Gaussian models", *Under review*, 2023

A. Lin, A. Song, B. Bilgic, and D. Ba, "Covariance-free sparse Bayesian learning", in *IEEE Transactions on Signal Processing*, 2022

Prof. Demba Ba
demba-ba.org/

Bahareh Tolooshams, PhD
btolooshams.github.io/

Prof. Abiy Tasissa
sites.tufts.edu/atasissa/

Andrew Song, PhD
andrewhsong.wordpress.com/

Prof. Berkin Bilgic
martinos.org/~berkin/

Prof. Yves Atchadé
math.bu.edu/people/atchade/

THANK YOU

Questions?