Data-Driven Computational Sensing

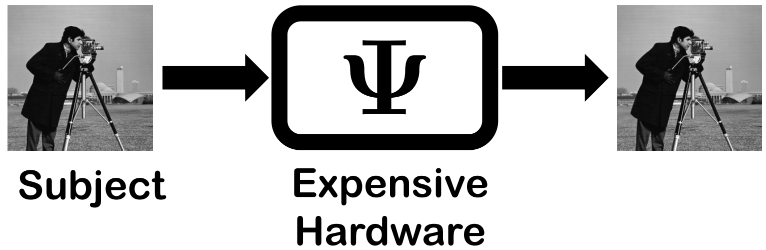
Ali Mousavi

Department of Electrical and Computer Engineering

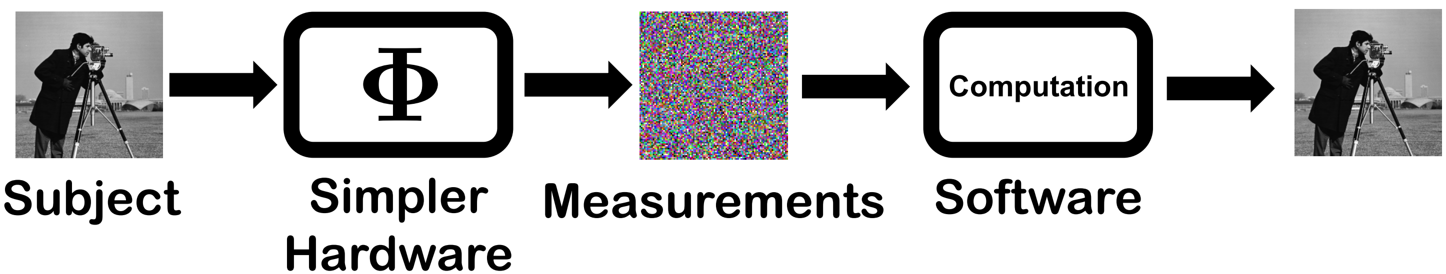
Rice University

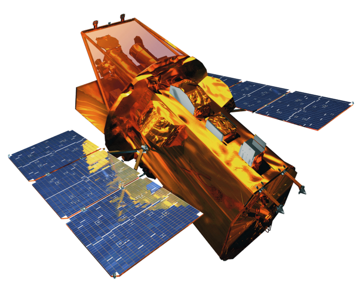
Computational Sensing

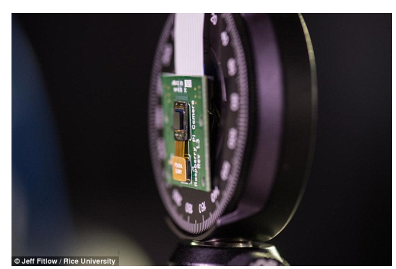
- **Conventional Sensing**

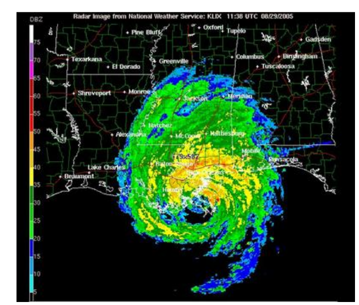


- **Computational Sensing: Reduce costs in acquisition systems by replacing expensive hardware w/ cheap hardware + computation**

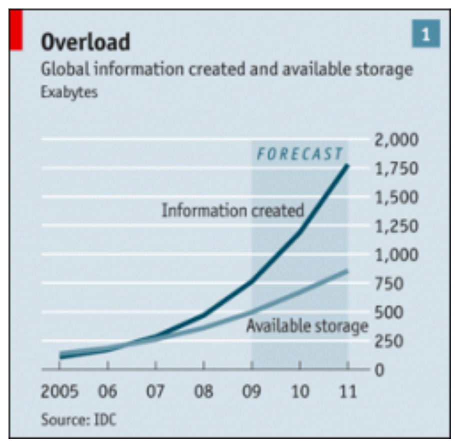


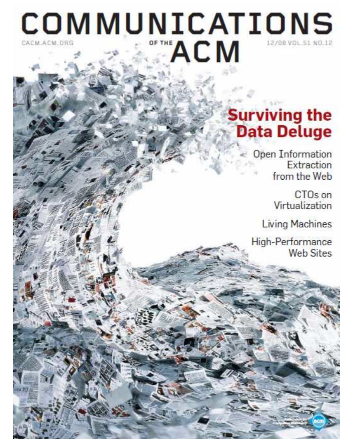




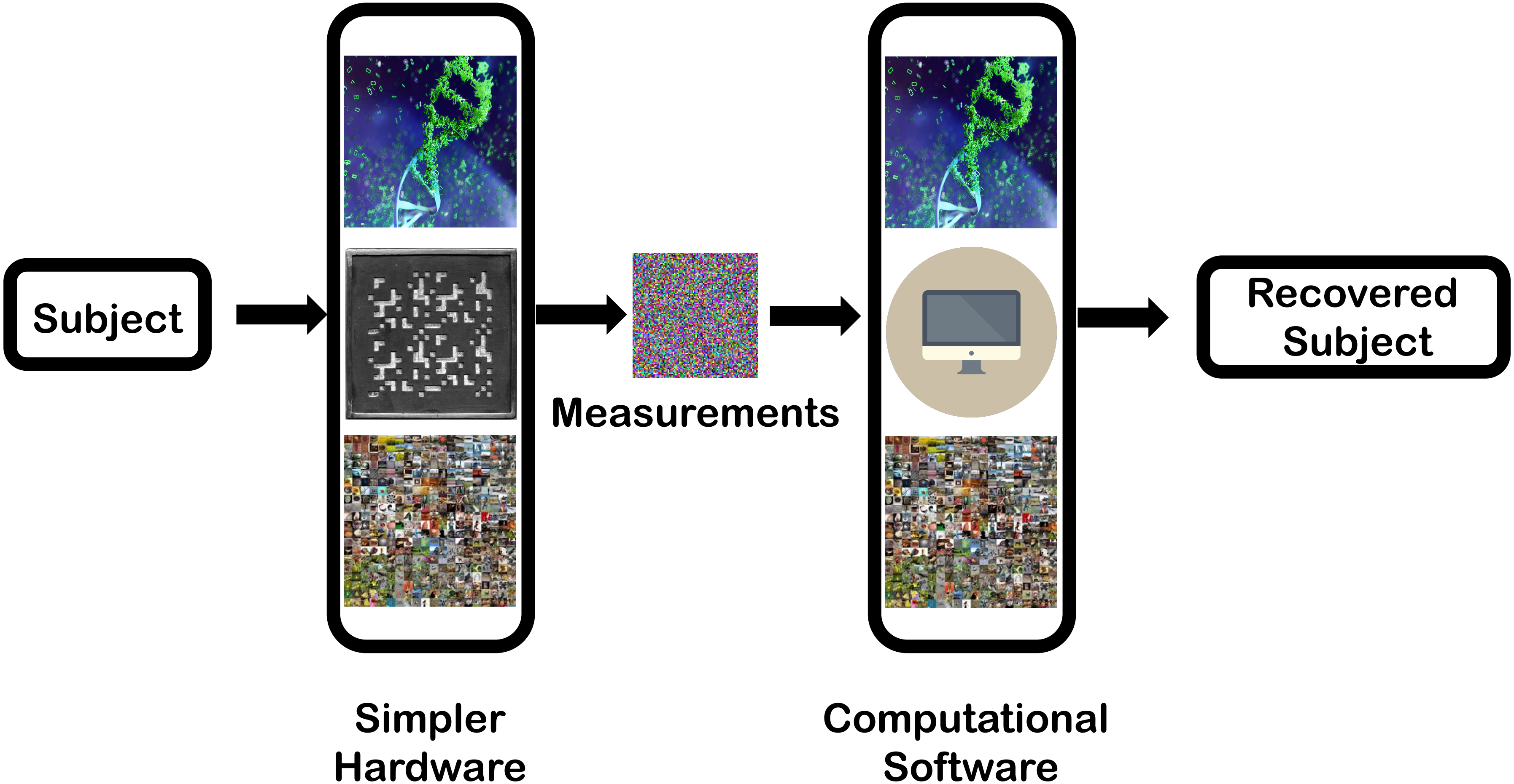


Large Scale Datasets

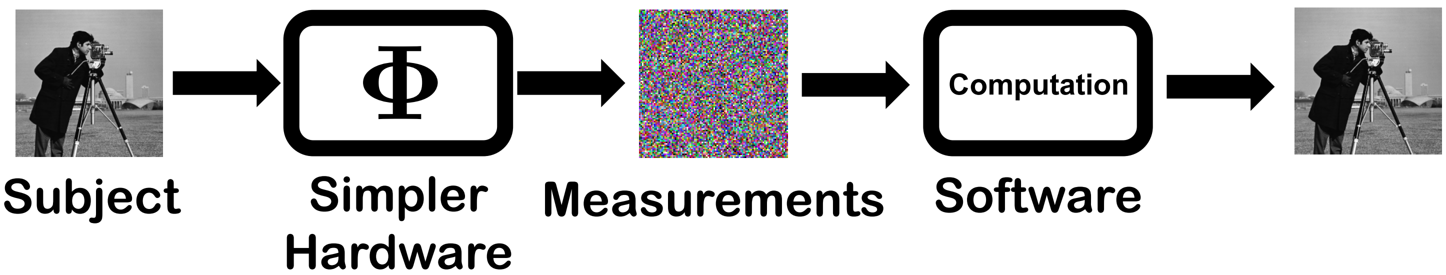




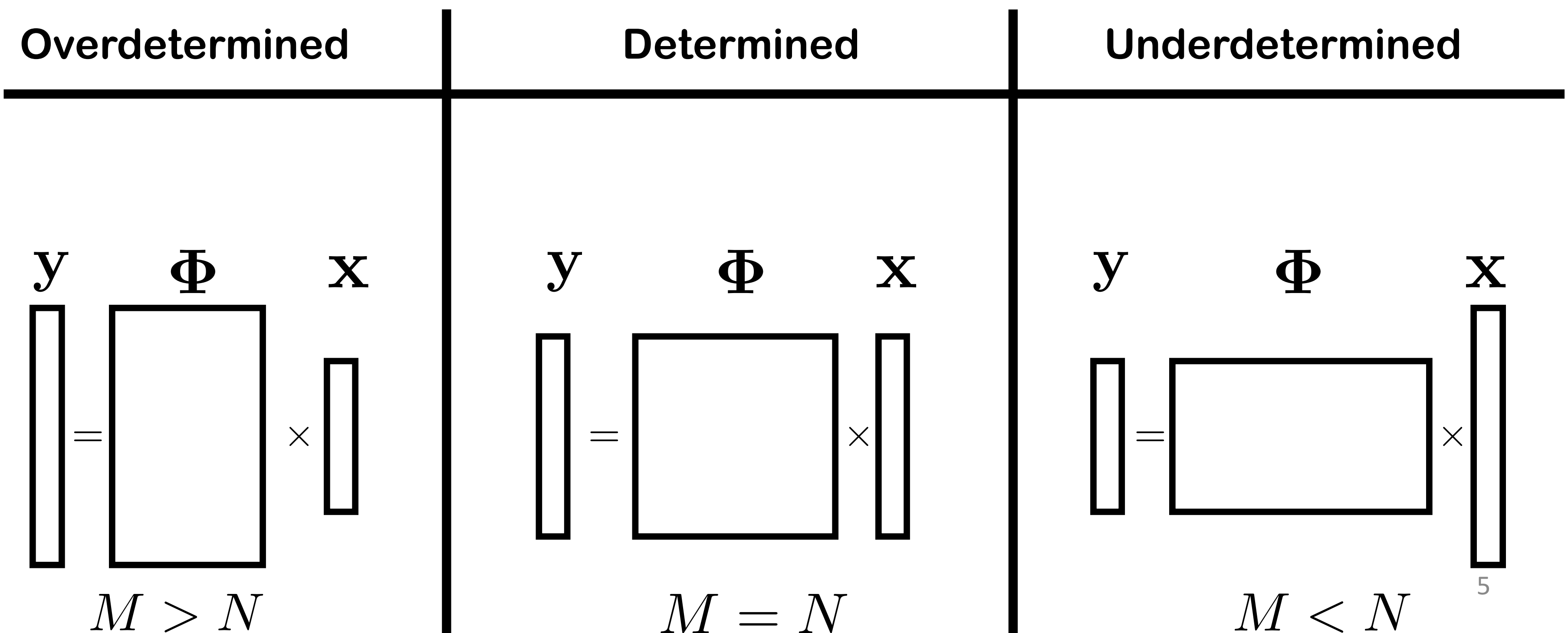
Data-Driven Computational Sensing



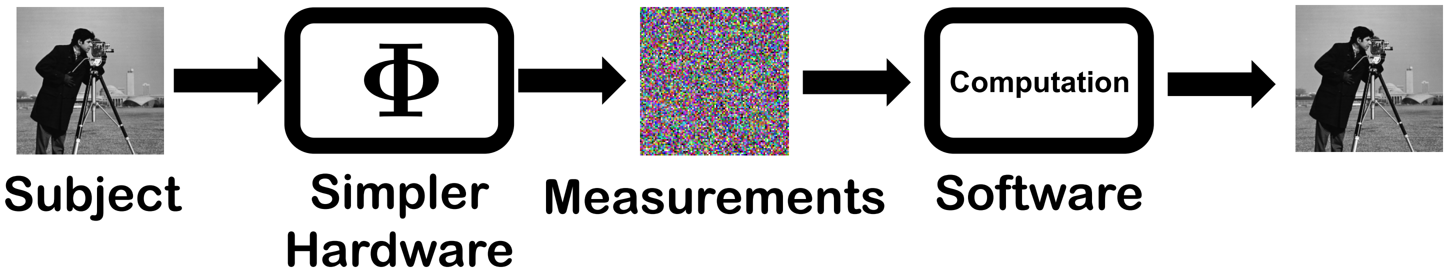
Model



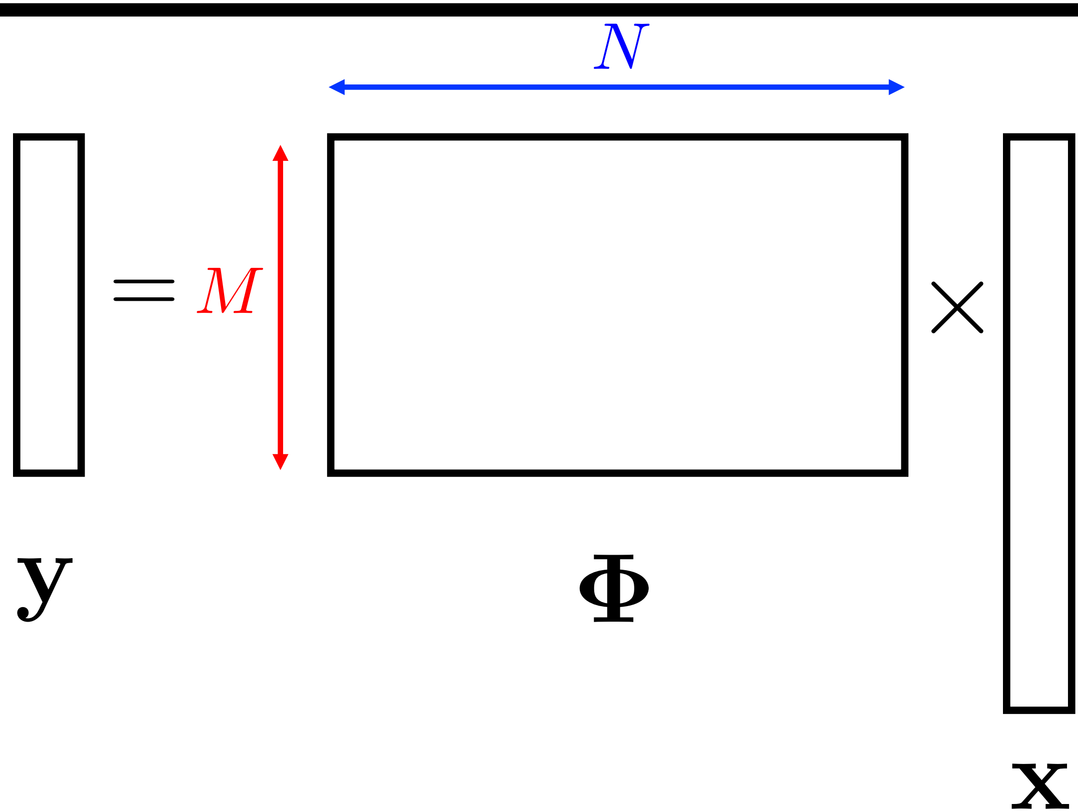
$$\mathbf{x} \in \mathbb{R}^N \xrightarrow{\boldsymbol{\Phi}(.)} \mathbf{y} = \boldsymbol{\Phi}(\mathbf{x}) \in \mathbb{R}^M \xrightarrow{\boldsymbol{\Phi}^{-1}(.)} \hat{\mathbf{x}} \in \mathbb{R}^N$$



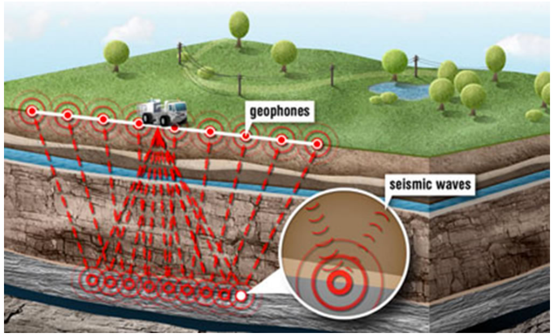
Model



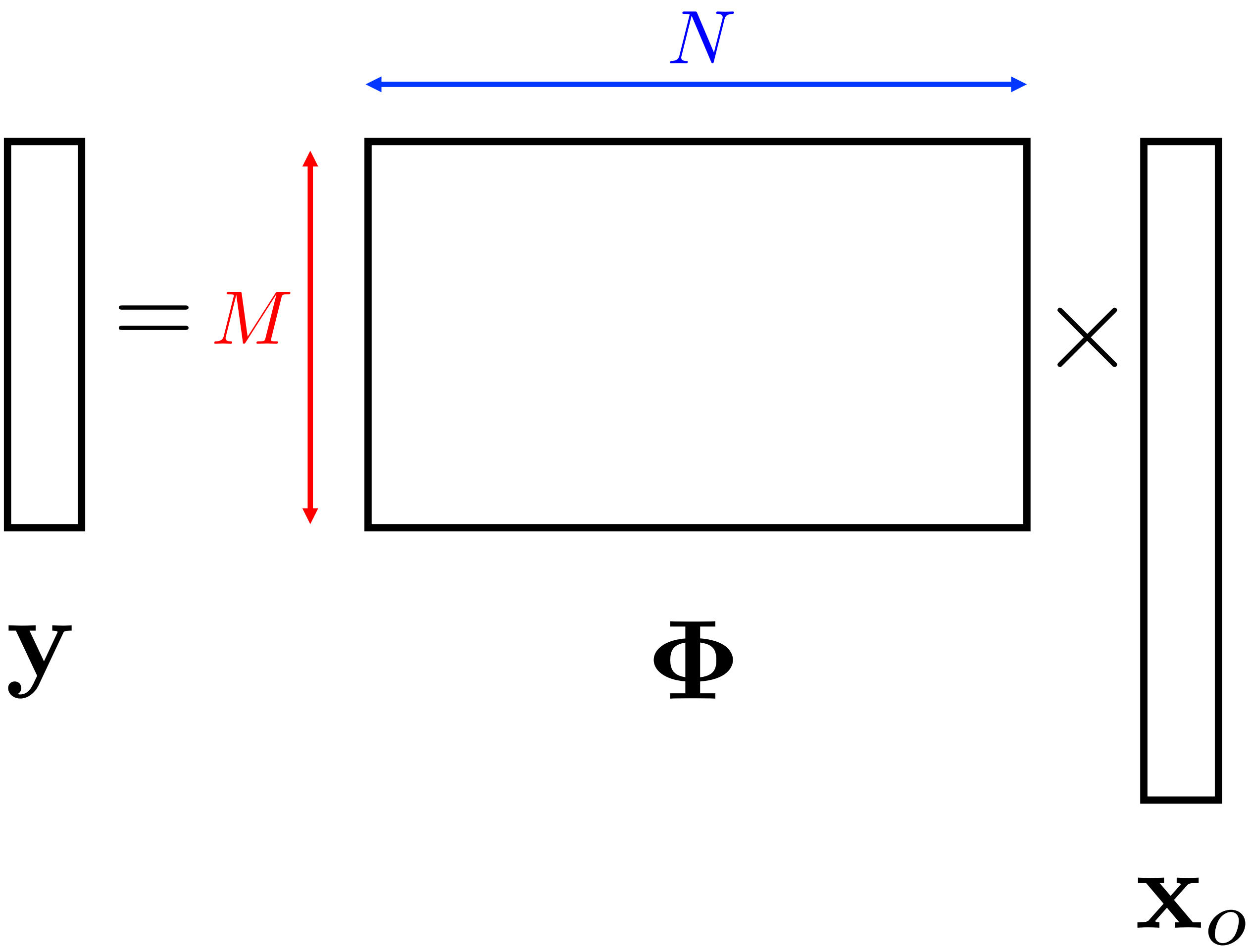
$$\mathbf{x} \in \mathbb{R}^N \xrightarrow{\boldsymbol{\Phi}(.)} \mathbf{y} = \boldsymbol{\Phi}(\mathbf{x}) \in \mathbb{R}^M \xrightarrow{\boldsymbol{\Phi}^{-1}(.)} \hat{\mathbf{x}} \in \mathbb{R}^N$$



Applications

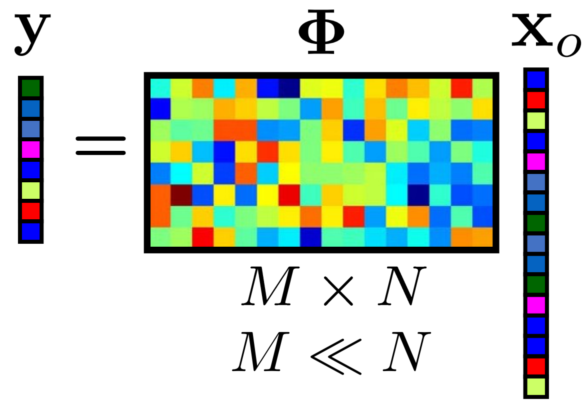


Data-Driven Computational Sensing

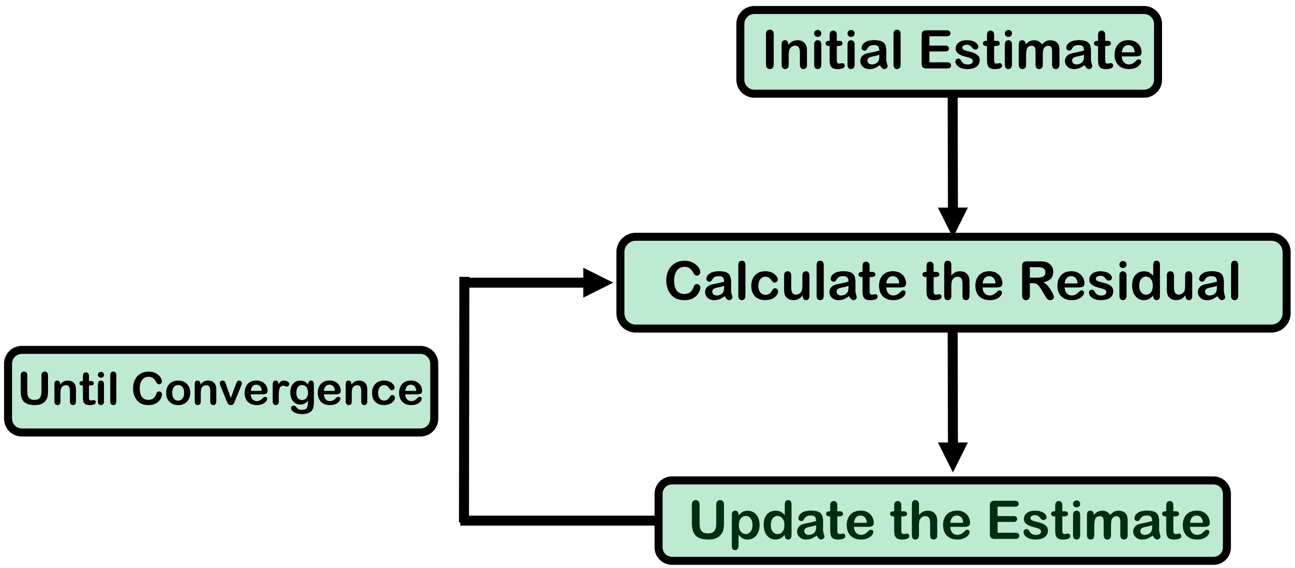


$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_2^2 + \lambda \times f(\mathbf{x})$$

Iterative Algorithms

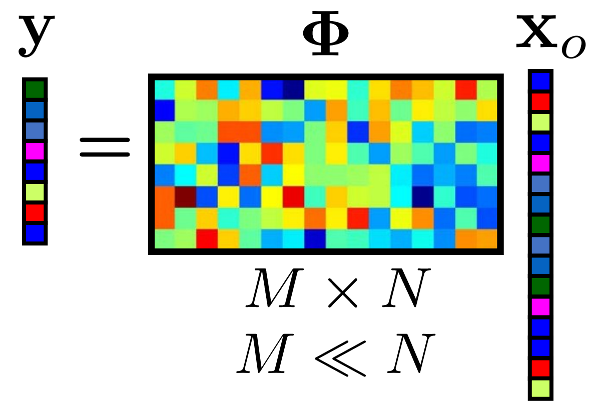


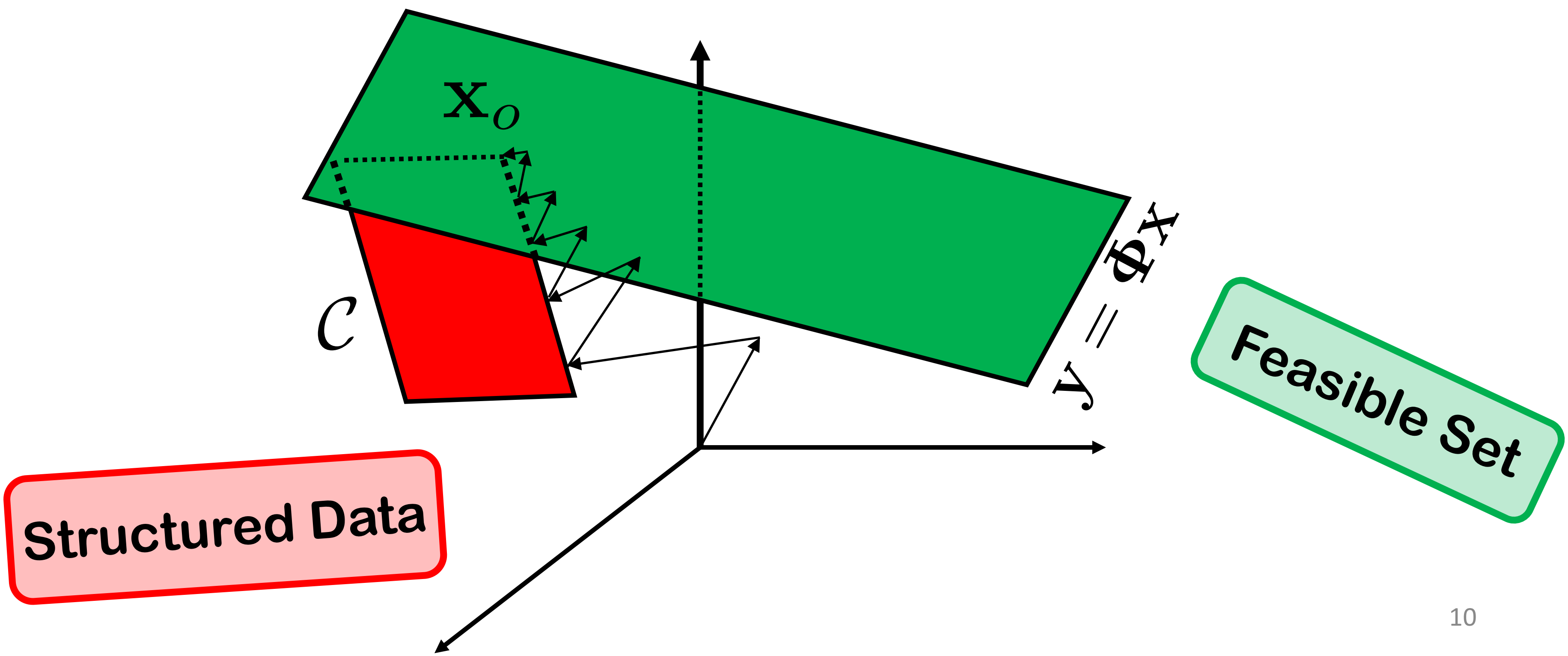
$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{\Phi x}\|_2^2 + \lambda \times f(\mathbf{x})$$



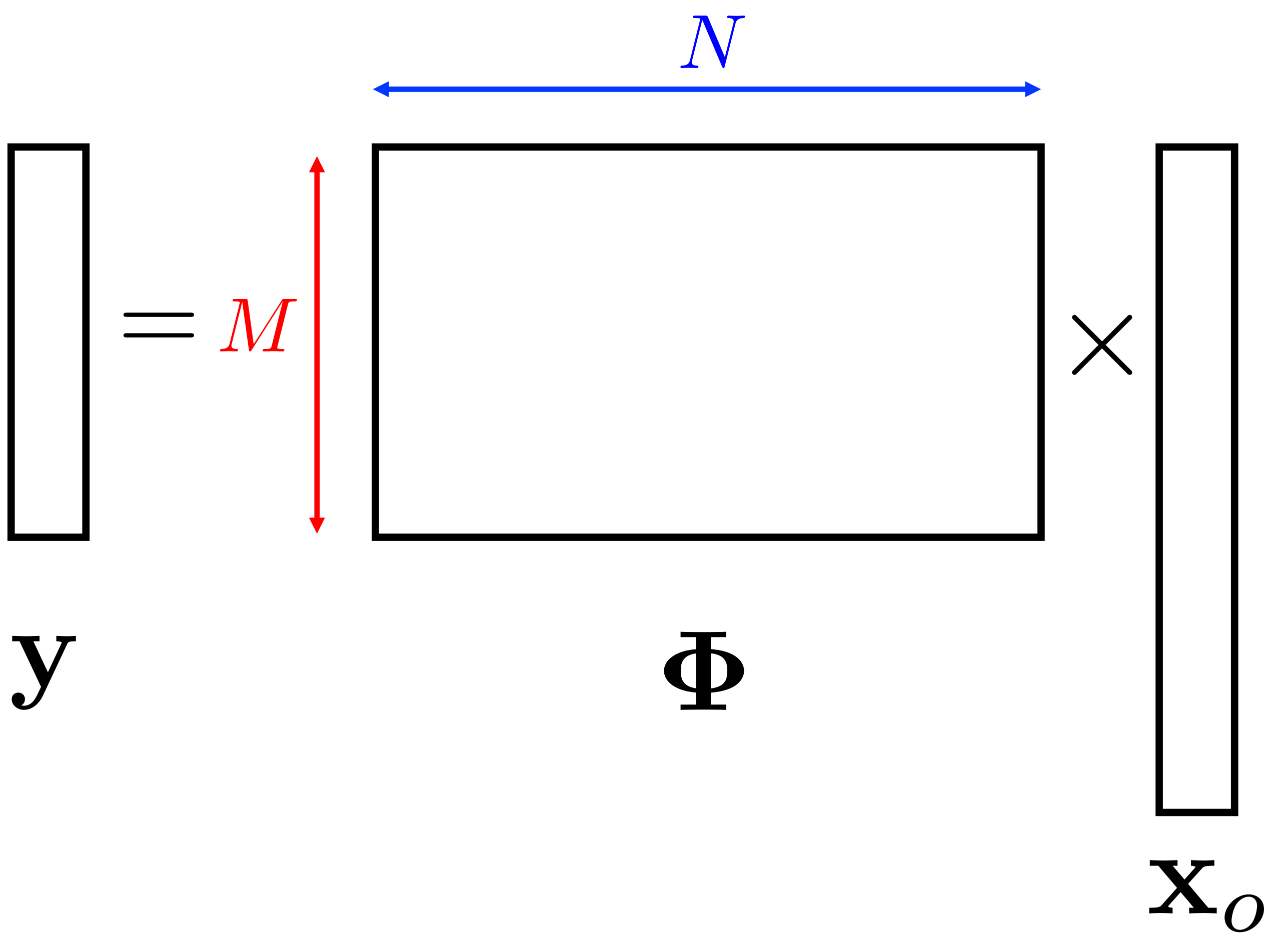
Iterative Algorithms

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{\Phi x}\|_2^2 + \lambda \times f(\mathbf{x})$$



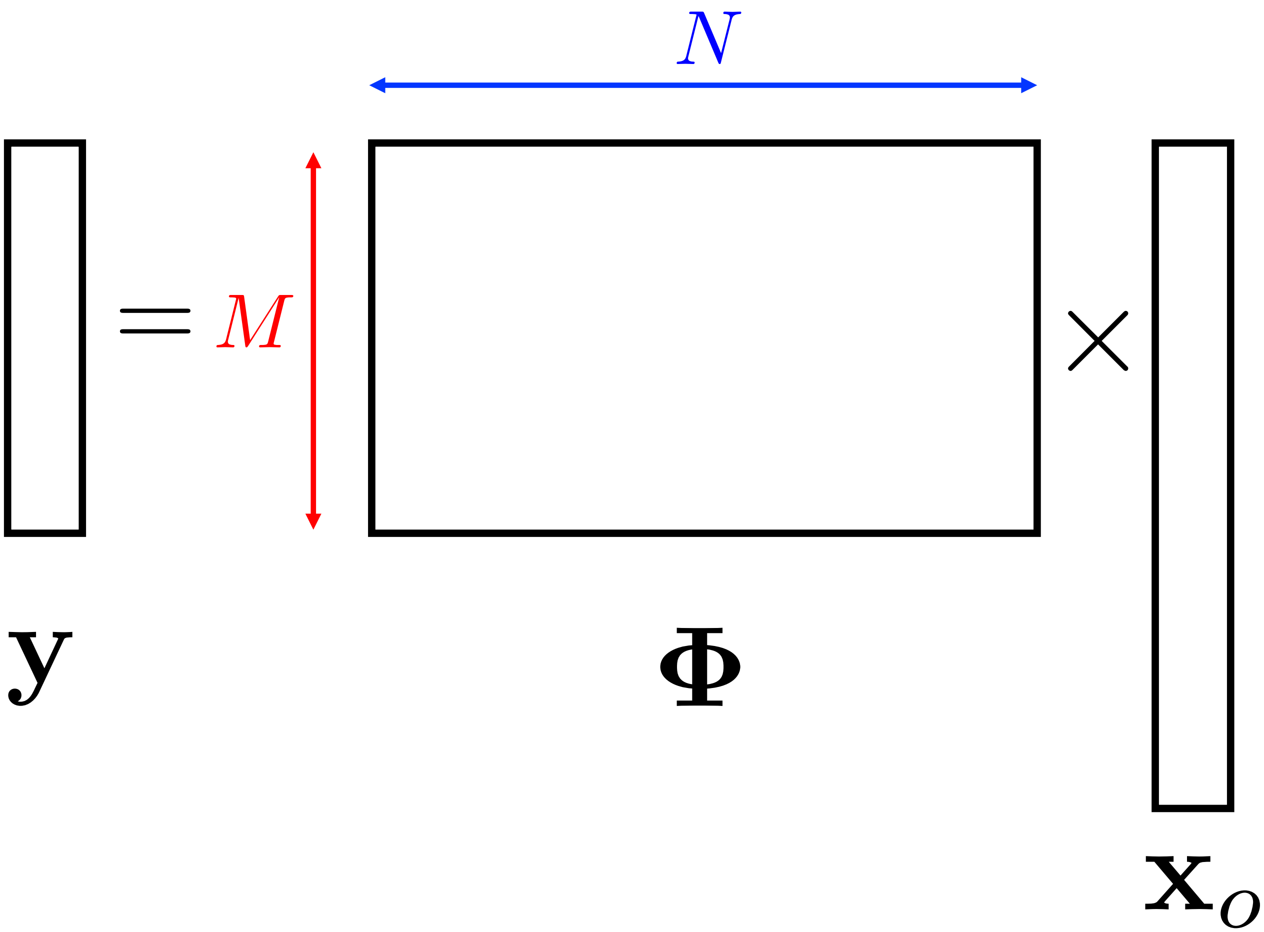


Data-Driven Computational Sensing



$$\min_{\mathbf{x}} \|\mathbf{y} - \boldsymbol{\Phi}\mathbf{x}\|_2^2 + \lambda \times f(\mathbf{x})$$

Regularization Parameter Tuning

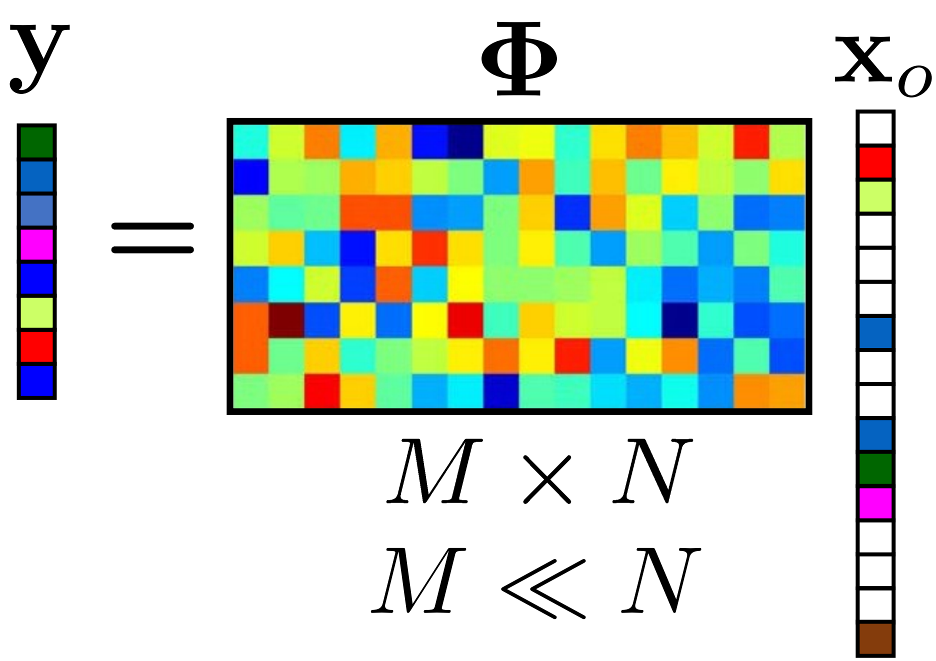


$$\min_{\mathbf{x}} \|\mathbf{y} - \boldsymbol{\Phi}\mathbf{x}\|_2^2 + \lambda \times f(\mathbf{x})$$

Sparse Regression

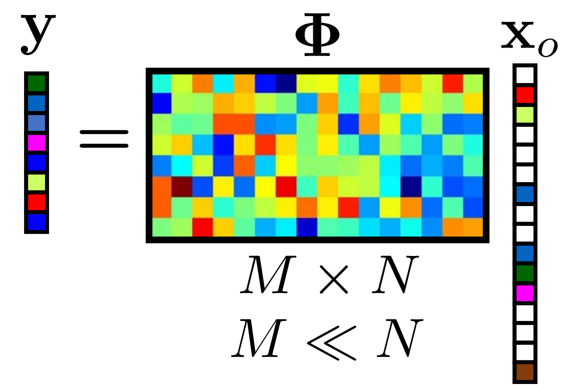
$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_2^2 + \lambda \times f(\mathbf{x})$$

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1$$



Approximate Message Passing

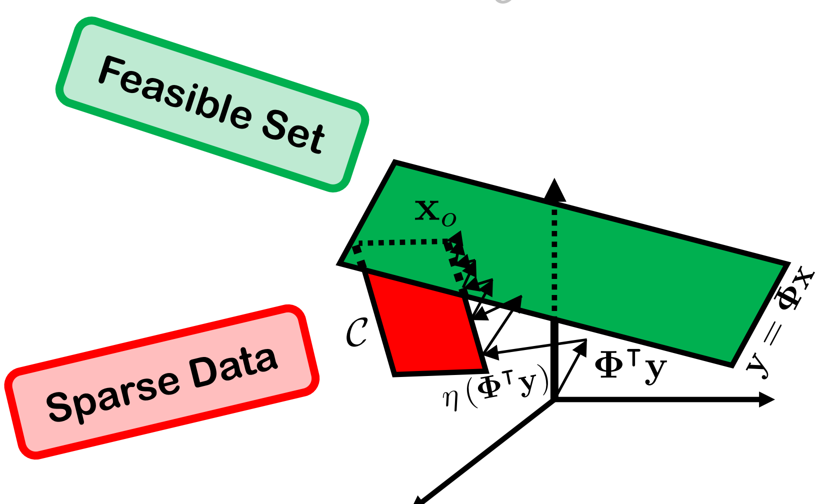
$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{\Phi x}\|_2^2 + \lambda \|\mathbf{x}\|_1$$



- **Approximate Message Passing (AMP)** [Donoho, Maleki, Montanari 2009]

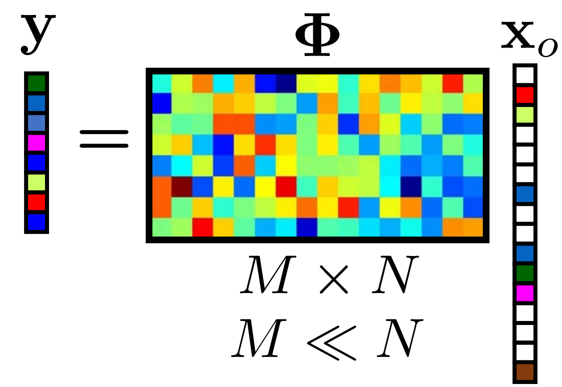
$$\mathbf{x}^{t+1} = \eta(\mathbf{x}^t + \mathbf{\Phi}^\intercal \mathbf{z}^t; \tau^t)$$

$$\mathbf{z}^t = \mathbf{y} - \mathbf{\Phi x}^t + \frac{1}{\delta}\mathbf{z}^{t-1}\left\langle \eta'(\mathbf{x}^{t-1} + \mathbf{\Phi}^\intercal \mathbf{z}^{t-1})\right\rangle$$



Approximate Message Passing

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1$$

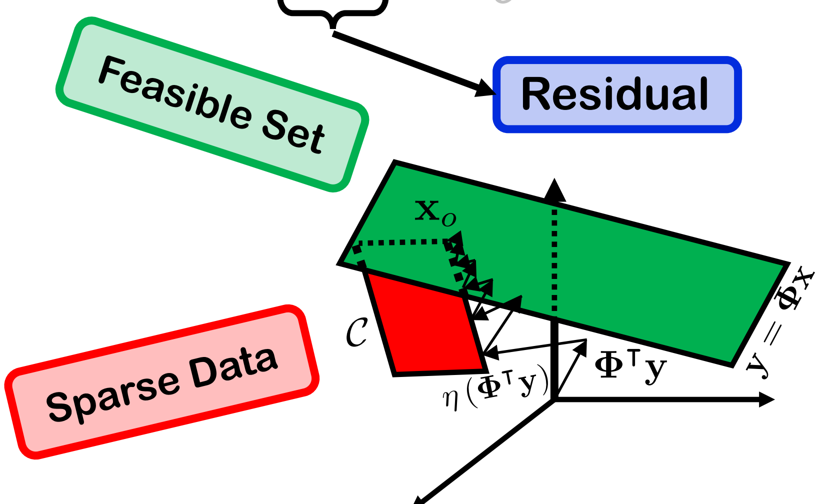


- **Approximate Message Passing (AMP)** [Donoho, Maleki, Montanari 2009]

$$\mathbf{x}^{t+1} = \eta(\mathbf{x}^t + \mathbf{\Phi}^\intercal \mathbf{z}^t; \tau^t)$$

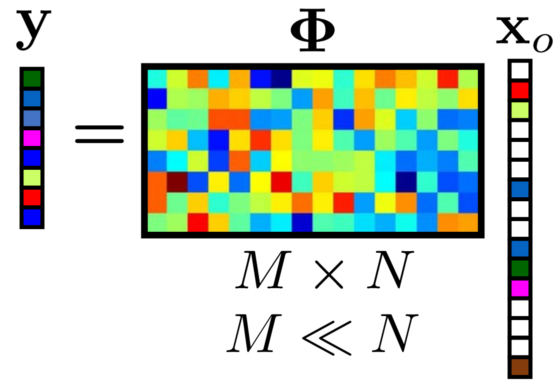
$$\mathbf{z}^t = \mathbf{y} - \mathbf{\Phi}\mathbf{x}^t + \frac{1}{\delta}\mathbf{z}^{t-1}\langle \eta'(\mathbf{x}^{t-1} + \mathbf{\Phi}^\intercal \mathbf{z}^{t-1}) \rangle$$

Residual



Approximate Message Passing

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



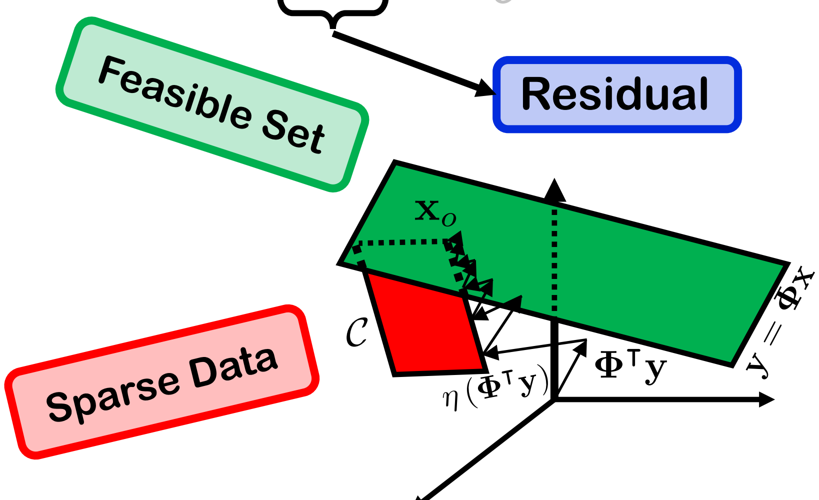
- **Approximate Message Passing (AMP)** [Donoho, Maleki, Montanari 2009]

Gradient Step

$$\mathbf{x}^{t+1} = \eta(\mathbf{x}^t + \overbrace{\boldsymbol{\Phi}^\intercal \mathbf{z}^t}; \tau^t)$$

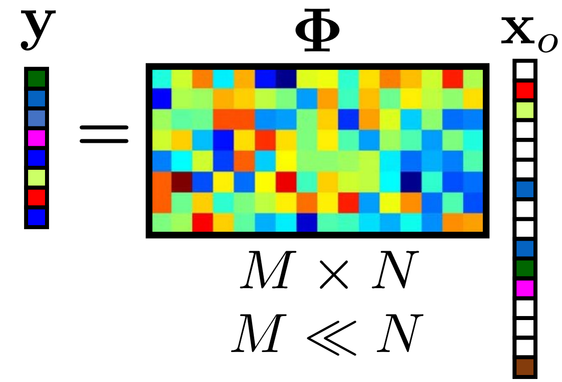
$$\underbrace{\mathbf{z}^t} = \mathbf{y} - \boldsymbol{\Phi}\mathbf{x}^t + \frac{1}{\delta}\mathbf{z}^{t-1}\langle \eta'(\mathbf{x}^{t-1} + \boldsymbol{\Phi}^\intercal \mathbf{z}^{t-1})\rangle$$

Residual



Approximate Message Passing

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{\Phi x}\|_2^2 + \lambda \|\mathbf{x}\|_1$$



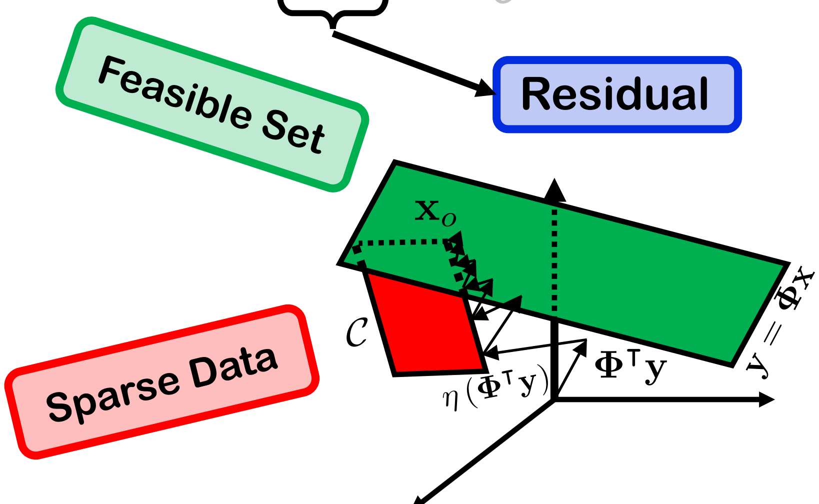
- **Approximate Message Passing (AMP)** [Donoho, Maleki, Montanari 2009]

Projection Operator (η) — Gradient Step ($\mathbf{x}^t + \mathbf{\Phi}^\intercal \mathbf{z}^t$)

$$\mathbf{x}^{t+1} = \eta(\mathbf{x}^t + \mathbf{\Phi}^\intercal \mathbf{z}^t; \tau^t)$$

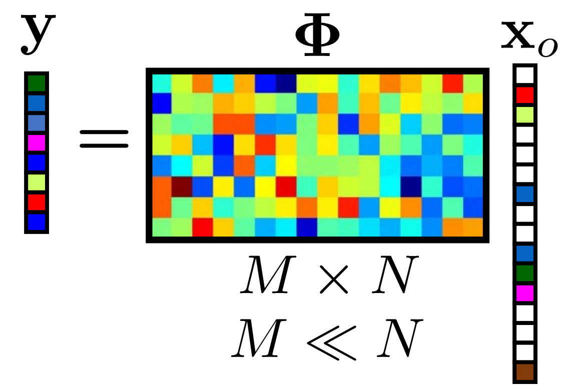
$$\mathbf{z}^t = \mathbf{y} - \mathbf{\Phi x}^t + \frac{1}{\delta}\mathbf{z}^{t-1}\langle \eta'(\mathbf{x}^{t-1} + \mathbf{\Phi}^\intercal \mathbf{z}^{t-1})\rangle$$

Residual ($\mathbf{z}^t$)



Approximate Message Passing

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1$$



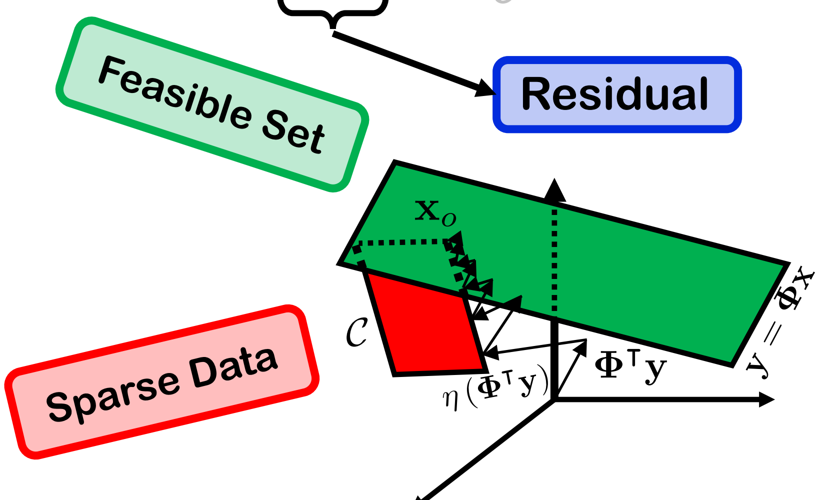
- **Approximate Message Passing (AMP)** [Donoho, Maleki, Montanari 2009]

Projection Operator (η) — **Gradient Step** ($\mathbf{x}^t + \mathbf{\Phi}^\intercal \mathbf{z}^t$)

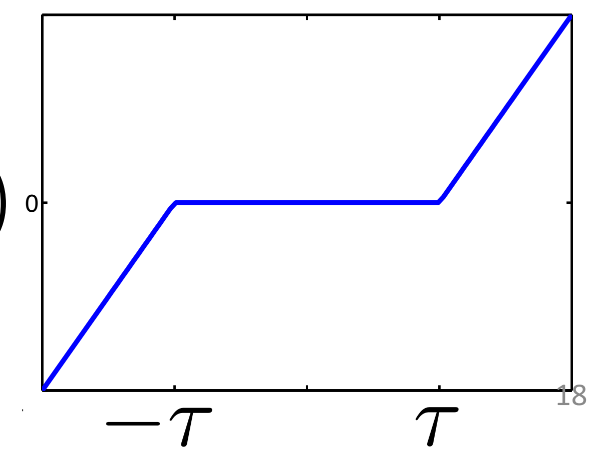
$$\mathbf{x}^{t+1} = \eta(\mathbf{x}^t + \mathbf{\Phi}^\intercal \mathbf{z}^t; \tau^t)$$

$$\mathbf{z}^t = \mathbf{y} - \mathbf{\Phi}\mathbf{x}^t + \frac{1}{\delta}\mathbf{z}^{t-1}\langle \eta'(\mathbf{x}^{t-1} + \mathbf{\Phi}^\intercal \mathbf{z}^{t-1})\rangle$$

Residual ($\mathbf{z}^t$)

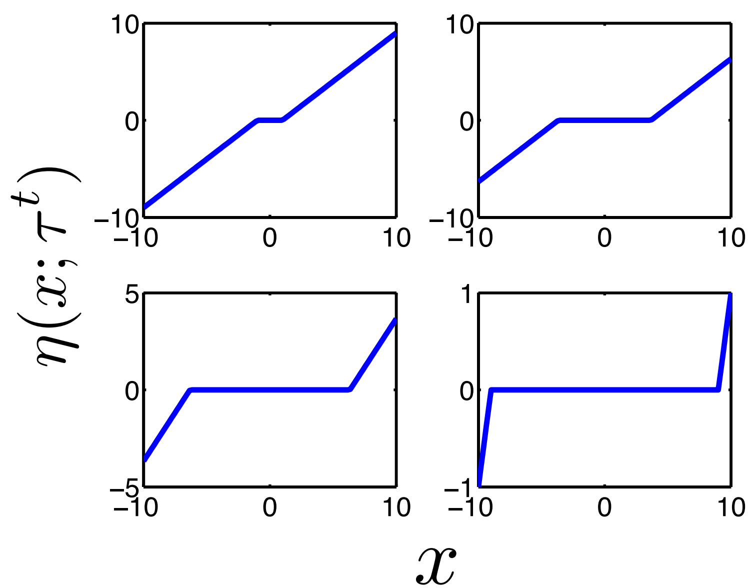


$\eta(x, \tau)$

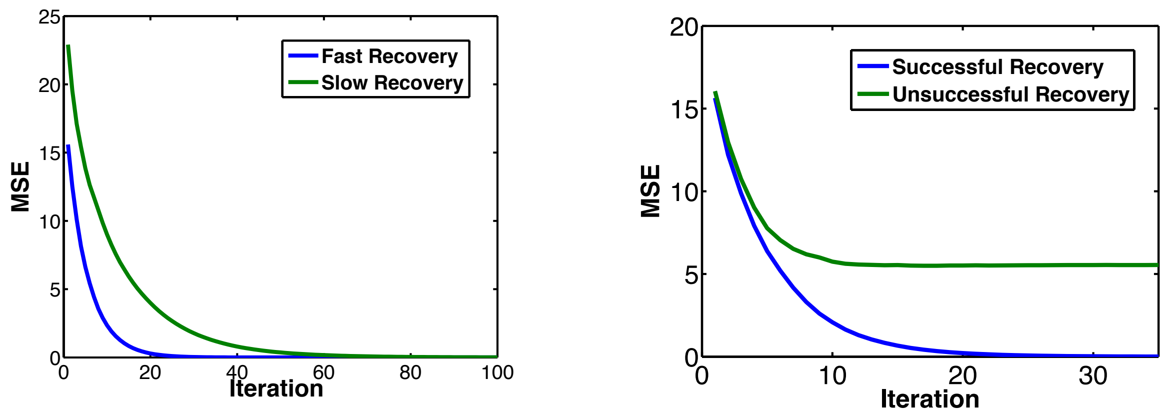


Impact of Tuning

- **Problem:** what is the optimal tuning parameter at every iteration?



- **Tuning impact on inferential and computational performance**



Greedy Tuning is Optimal

- **Best Possible error after T iterations** $\min_{\tau^1,\tau^2,\ldots,\tau^T} \frac{\|\mathbf{x}^T - \mathbf{x}_o\|_2^2}{N}$

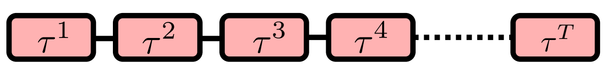
- **Theorem** [**Mousavi**, Maleki, Baraniuk, *Annals of Statistics 2017*]

  $\tau^{*,1},\ldots,\tau^{*,T-1}$ → **Optimal for iteration T**

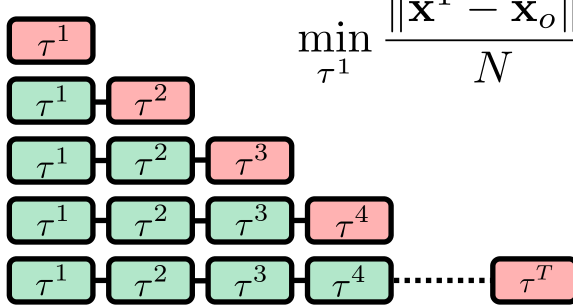
  ⬇

  $\tau^{*,1},\ldots,\tau^{*,t-1}$ → **Optimal for iteration $t < T$**

- **Implication:**
  - **Original Optimization:** $\min_{\tau^1,\tau^2,\ldots,\tau^T} \frac{\|\mathbf{x}^T - \mathbf{x}_o\|_2^2}{N}$



  - **New Optimization:** $\min_{\tau^1} \frac{\|\mathbf{x}^1 - \mathbf{x}_o\|_2^2}{N}, \ldots, \min_{\tau^T} \frac{\|\mathbf{x}^T - \mathbf{x}_o\|_2^2}{N}$



Simplified Optimization

- **If $\tau^1, \ldots, \tau^{t-1}$ are optimally set by $\tau^{*,1}, \ldots, \tau^{*,t-1}$, then we solve**

$$\min_{\tau^t} \frac{\|\mathbf{x}^t - \mathbf{x}_o\|_2^2}{N}$$

- **Lemma** [**Mousavi**, Maleki, Baraniuk, *Annals of Statistics 2017*]

$\frac{\|\mathbf{x}^t - \mathbf{x}_o\|_2^2}{N}$
- **is quasi-convex.**
- **achieves its minimum at a unique and finite $\tau \in \mathbb{R}$**
- **its derivative is zero only at the optimal τ**

Simplified Optimization

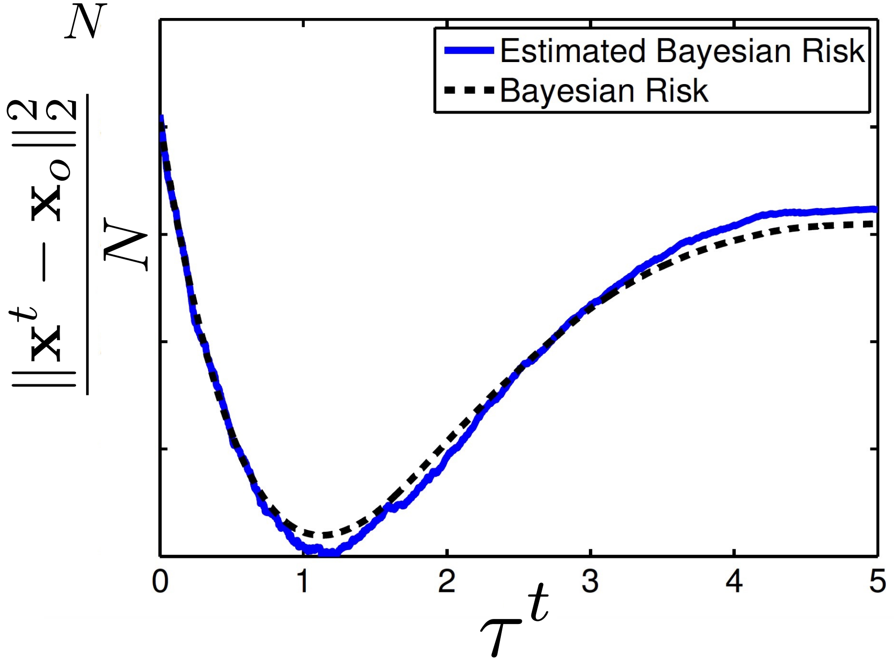
- If $\tau^1, \ldots, \tau^{t-1}$ are optimally set by $\tau^{*,1}, \ldots, \tau^{*,t-1}$, then we solve

$$\min_{\tau^t} \frac{\|\mathbf{x}^t - \mathbf{x}_o\|_2^2}{N}$$

- **Lemma** [**Mousavi**, Maleki, Baraniuk, *Annals of Statistics 2017*]

$\frac{\|\mathbf{x}^t - \mathbf{x}_o\|_2^2}{N}$
- is quasi-convex.
- achieves its minimum at a unique and finite $\tau \in \mathbb{R}$
- its derivative is zero only at the optimal τ

- Estimating $\frac{\|\mathbf{x}^t - \mathbf{x}_o\|_2^2}{N}$ by model selection techniques.



Simplified Optimization

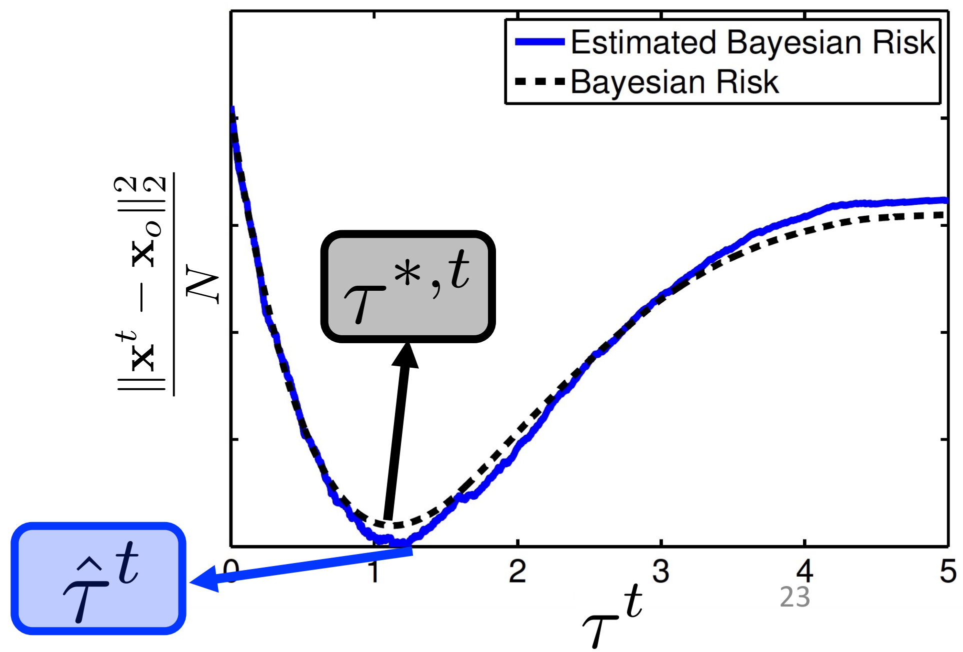
- If $\tau^1, \ldots, \tau^{t-1}$ are optimally set by $\tau^{*,1}, \ldots, \tau^{*,t-1}$, then we solve

$$\min_{\tau^t} \frac{\|\mathbf{x}^t - \mathbf{x}_o\|_2^2}{N}$$

- **Lemma** [**Mousavi**, Maleki, Baraniuk, *Annals of Statistics 2017*]

$\frac{\|\mathbf{x}^t - \mathbf{x}_o\|_2^2}{N}$
  - is quasi-convex.
  - achieves its minimum at a unique and finite $\tau \in \mathbb{R}$
  - its derivative is zero only at the optimal τ

- Estimating $\frac{\|\mathbf{x}^t - \mathbf{x}_o\|_2^2}{N}$ by model selection techniques.



Simplified Optimization

- If $\tau^1, \ldots, \tau^{t-1}$ are optimally set by $\tau^{*,1}, \ldots, \tau^{*,t-1}$, then we solve

$$\min_{\tau^t} \frac{\|\mathbf{x}^t - \mathbf{x}_o\|_2^2}{N}$$

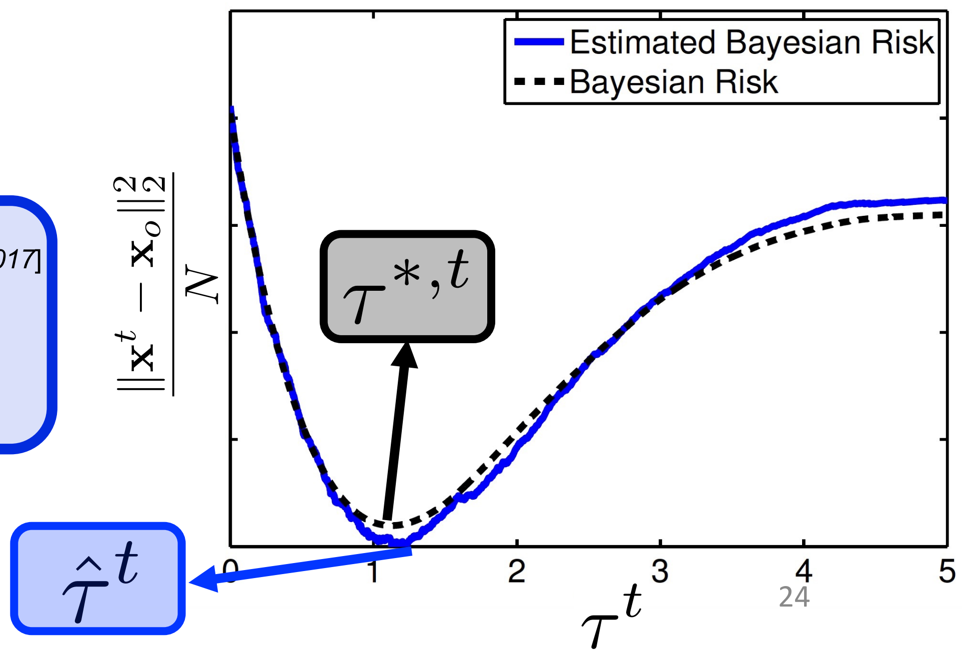
- **Lemma** [**Mousavi**, Maleki, Baraniuk, *Annals of Statistics 2017*]

  $\frac{\|\mathbf{x}^t - \mathbf{x}_o\|_2^2}{N}$
  - is quasi-convex.
  - achieves its minimum at a unique and finite $\tau \in \mathbb{R}$
  - its derivative is zero only at the optimal τ

- Estimating $\frac{\|\mathbf{x}^t - \mathbf{x}_o\|_2^2}{N}$ by model selection techniques.

- **Theorem** [**Mousavi**, Maleki, Baraniuk, *Annals of Statistics 2017*]

  $\hat{\tau}^t \to \tau^{*,t}$ in probability.



Indirect Optimal Tuning

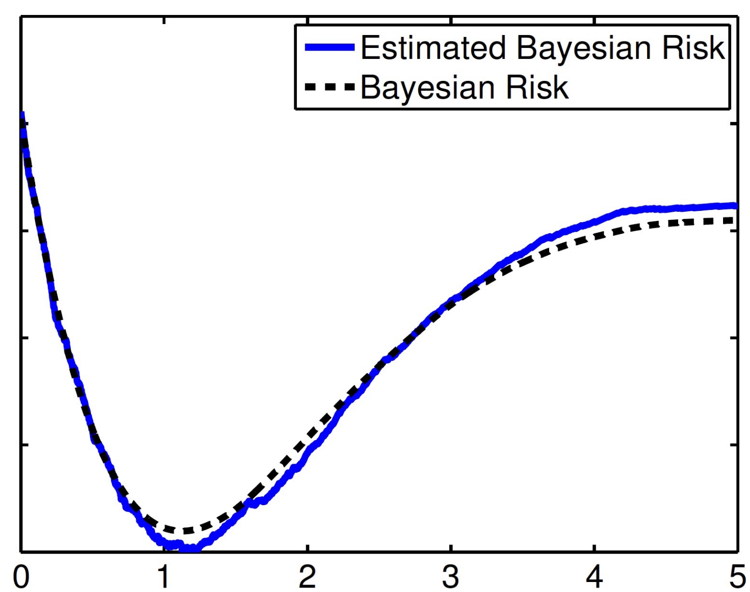
$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1$$

- **Optimal Regularization Parameter** $\longrightarrow \lambda^*$

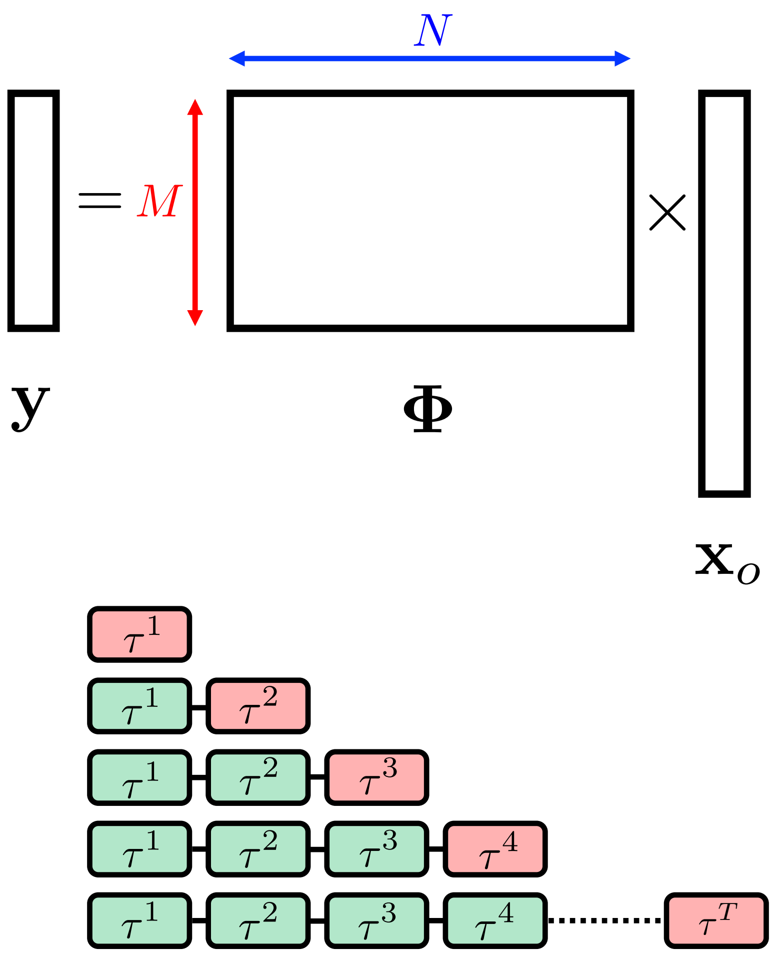
- **Theorem** [**Mousavi**, Maleki, Baraniuk, *Annals of Statistics 2017*]

$$\mathbf{x}^{\tau^*} \longrightarrow \mathbf{x}^{\lambda^*}$$

$$\mathbf{x}^{\hat{\tau}} \longrightarrow \mathbf{x}^{\lambda^*}$$

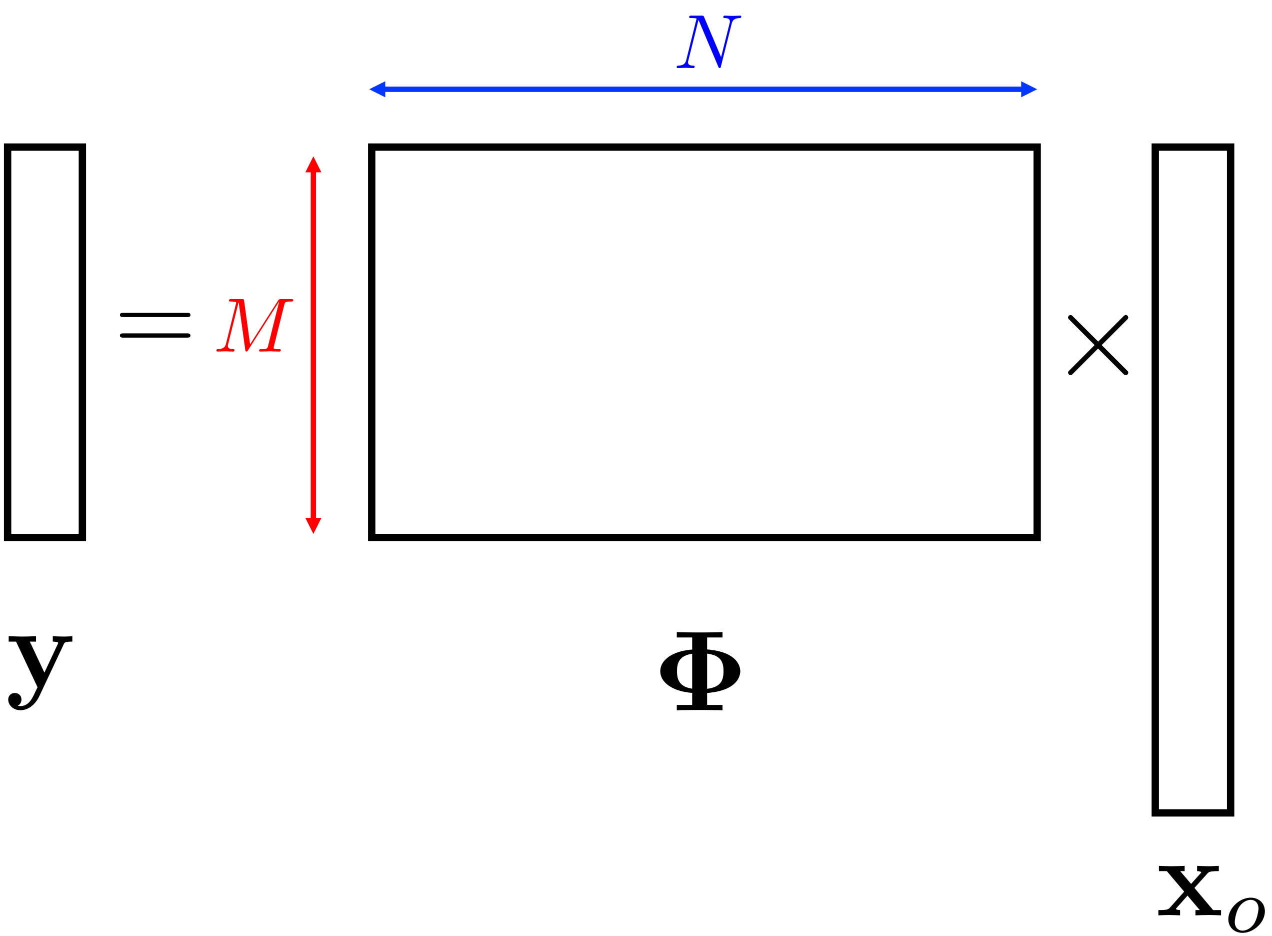


summary so far



$$\min_{\mathbf{x}} \|\mathbf{y} - \boldsymbol{\Phi}\mathbf{x}\|_2^2 + \lambda \times f(\mathbf{x})$$

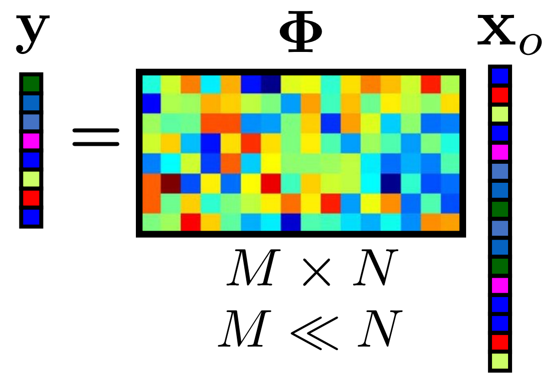
Data-Driven Penalty Selection

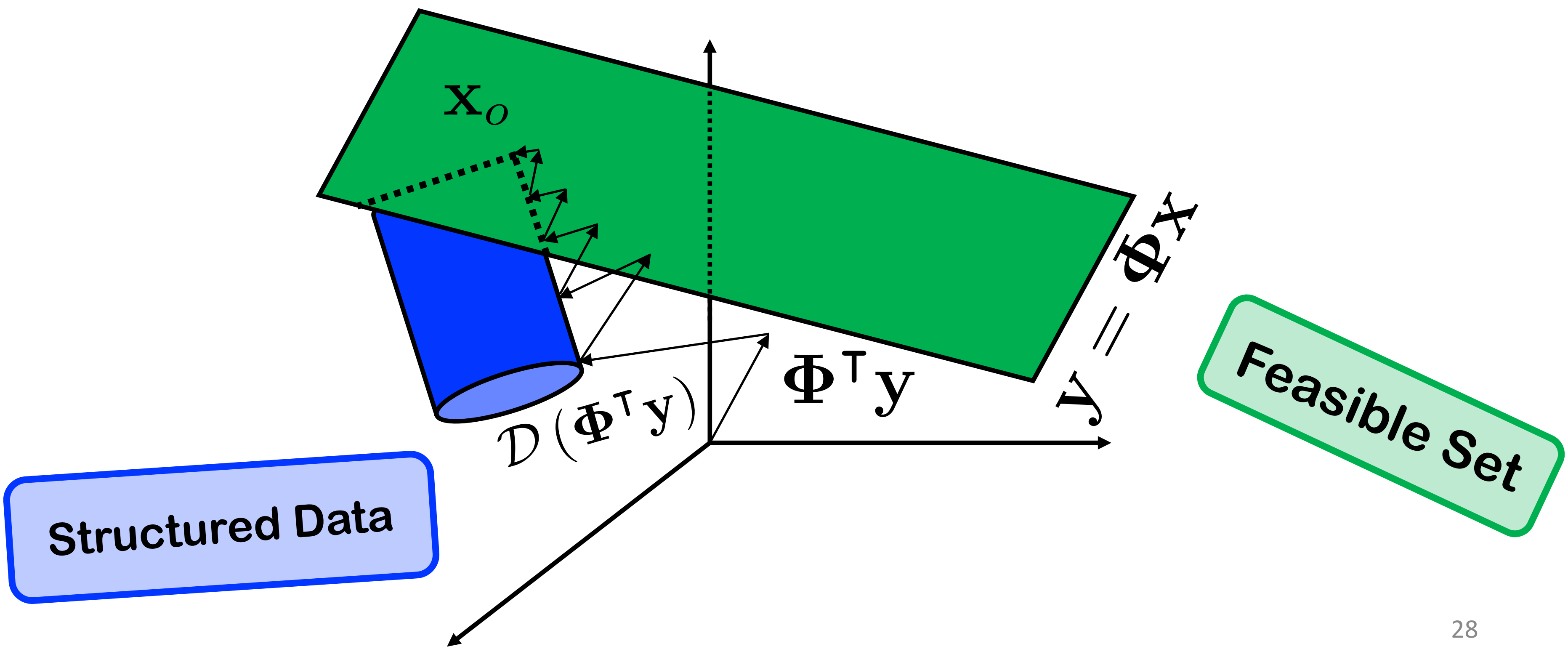


$$\min_{\mathbf{x}} \|\mathbf{y} - \boldsymbol{\Phi}\mathbf{x}\|_2^2 + \lambda \times f(\mathbf{x})$$

Structured Regression

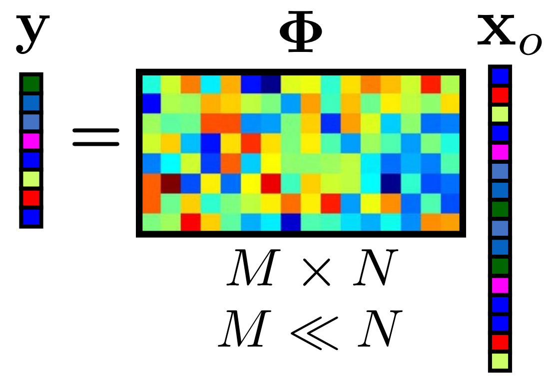
$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{\Phi x}\|_2^2 + \lambda f(\mathbf{x})$$



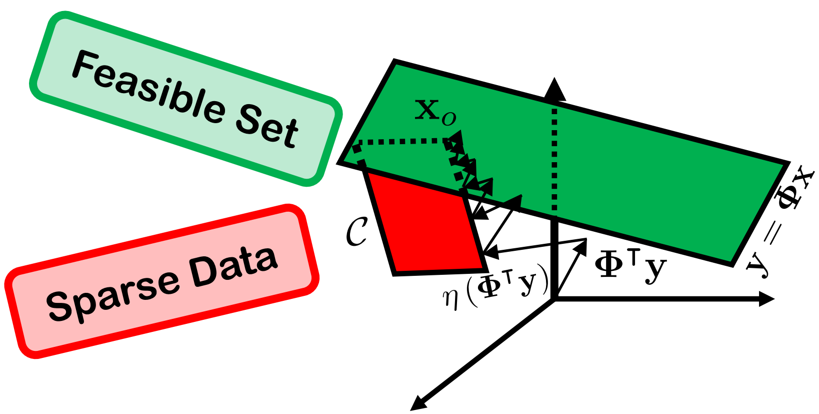


Sparse Regression

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



- **Approximate Message Passing (AMP)** [Donoho, Maleki, Montanari 2009]

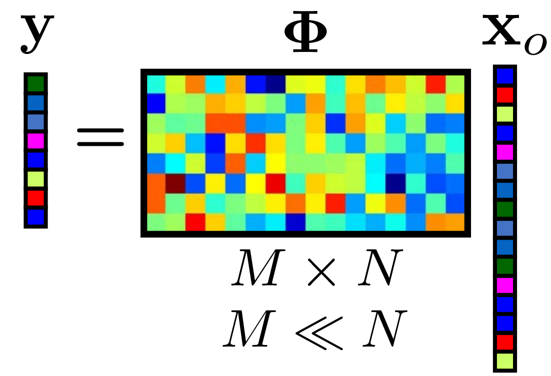


$$\mathbf{x}^{t+1} = \eta(\mathbf{x}^t + \boldsymbol{\Phi}^\intercal \mathbf{z}^t; \tau^t)$$

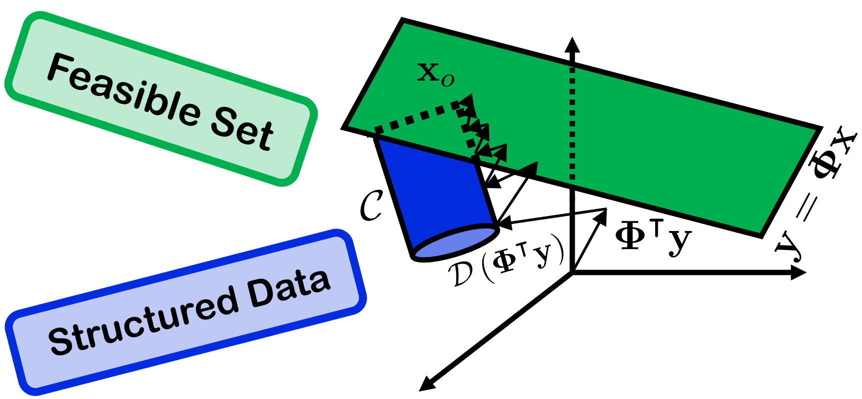
$$\mathbf{x}^t + \boldsymbol{\Phi}^\intercal \mathbf{z}^t = \mathbf{x}_o + \underbrace{\mathbf{v}^t}_{\text{Effective Noise}}$$

Structured Regression

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_2^2 + \lambda f(\mathbf{x})$$

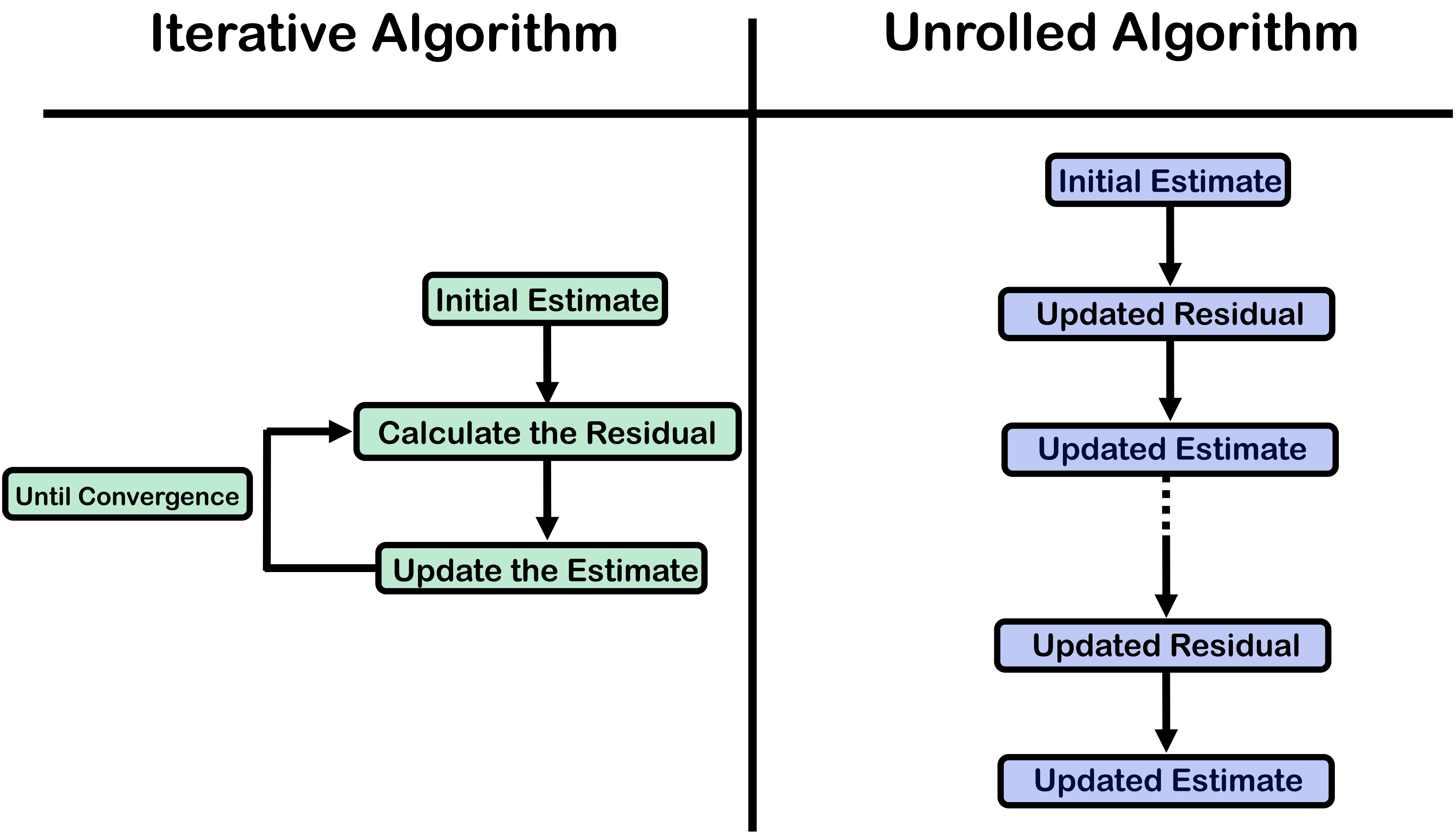


- **Denoising Approximate Message Passing (D-AMP)** [Metzler, Maleki, Baraniuk 2015]



$$\mathbf{x}^{t+1} = \mathcal{D}^t(\mathbf{x}^t + \mathbf{\Phi}^\intercal \mathbf{z}^t)$$

Unrolling Iterative Algorithms



[Gregor and LeCun, 2010]

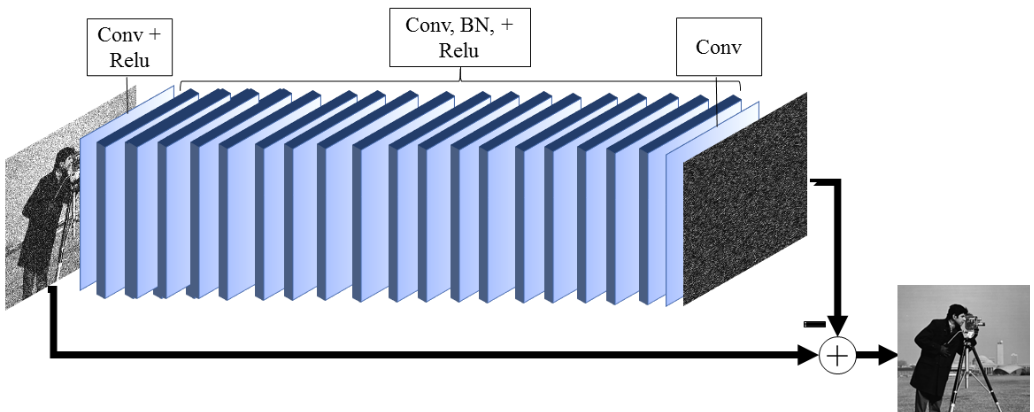
Learned-Denoising-AMP

Learned-Denoising-AMP (LDAMP) [Metzler, **Mousavi**, Baraniuk, *NIPS 2017*]

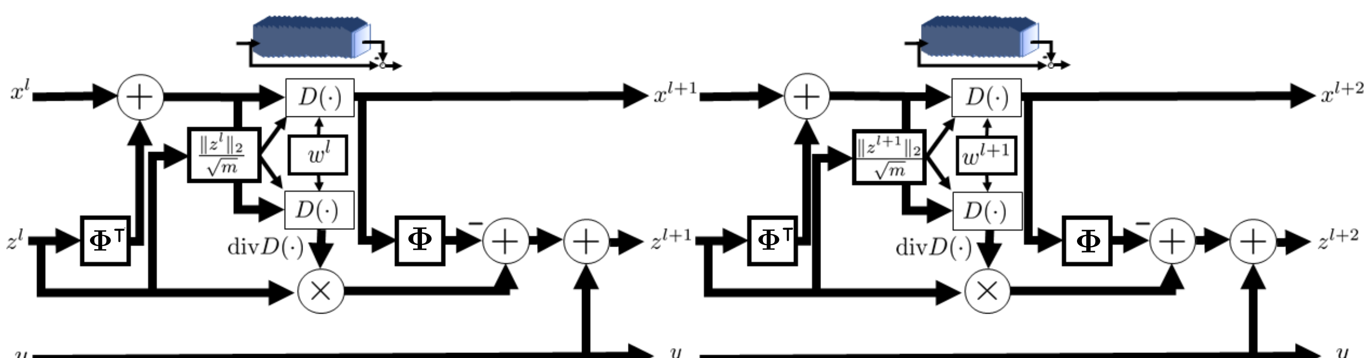
$$\mathbf{x}^{l+1} = \mathcal{D}^l(\mathbf{x}^l + \mathbf{\Phi}^\intercal \mathbf{z}^l)$$

$$\mathbf{z}^l = \mathbf{y} - \mathbf{\Phi}\mathbf{x}^l + \frac{1}{\delta}\mathbf{z}^{l-1}\langle \mathrm{div}\mathcal{D}^l(\mathbf{x}^{l-1} + \mathbf{\Phi}^\intercal \mathbf{z}^{l-1})\rangle$$

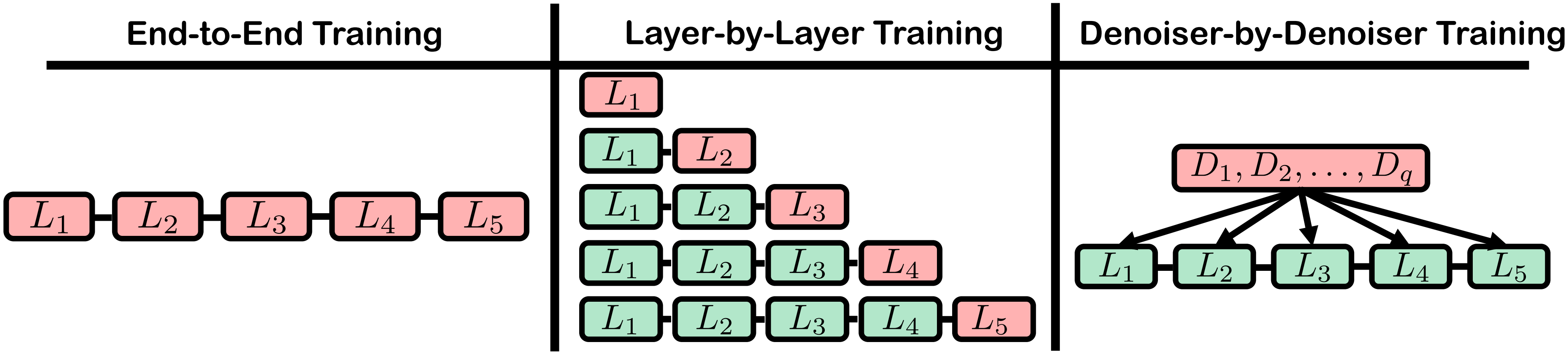
- **We use a 20-layer convolutional network as a denoiser** [Zhang et al. 2017]



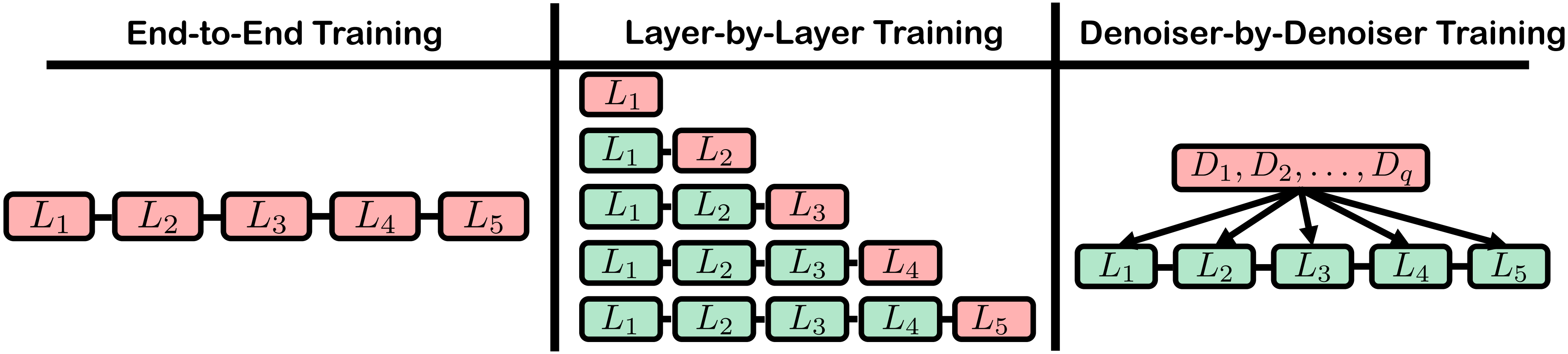
- **Two layers of the LDAMP network**



Training LDAMP and LDIT



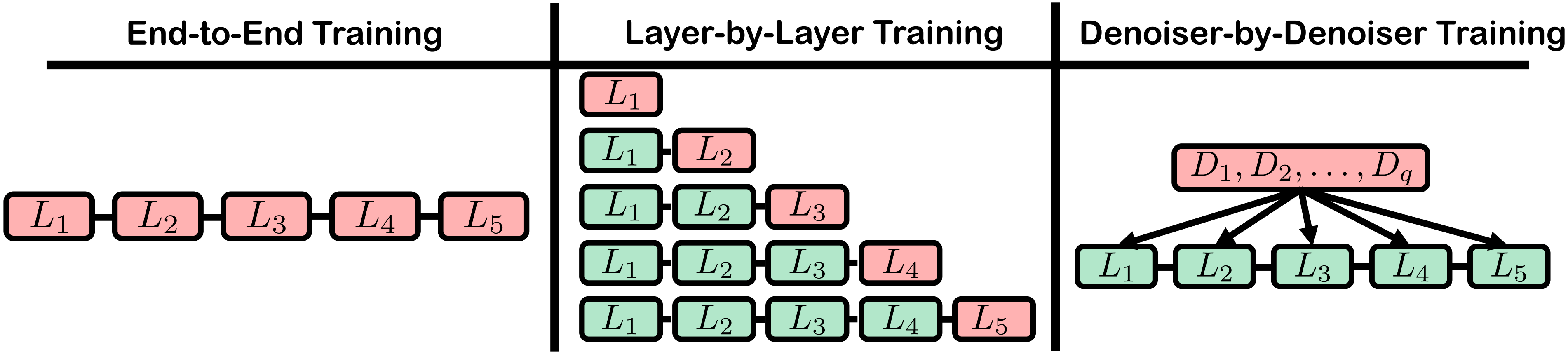
Training LDAMP and LDIT



- **Lemma 1** [Metzler, **Mousavi**, Baraniuk, *NIPS 2017*]
  Layer-by-layer training of LDAMP is MMSE optimal.

- **Lemma 2** [Metzler, **Mousavi**, Baraniuk, *NIPS 2017*]
  Denoiser-by-denoiser training of LDAMP is MMSE optimal.

Training LDAMP and LDIT



- **Lemma 1** [Metzler, **Mousavi**, Baraniuk, *NIPS 2017*]
  Layer-by-layer training of LDAMP is MMSE optimal.

- **Lemma 2** [Metzler, **Mousavi**, Baraniuk, *NIPS 2017*]
  Denoiser-by-denoiser training of LDAMP is MMSE optimal.

- **Noise discretization degrades the performance.**
- **Denoiser-by-denoiser is more generalizable.**

Average PSNR (dB) of one hundred 40x40 images Recovered from i.i.d Gaussian Measurements

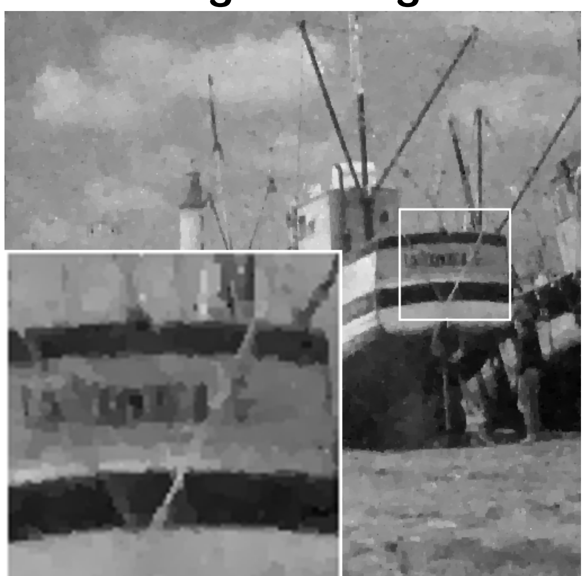
	Training: $\frac{M}{N} = 0.2$ Testing: $\frac{M}{N} = 0.2$		Training: $\frac{M}{N} = 0.2$ Testing: $\frac{M}{N} = 0.05$	
Training Method	LDIT	LDAMP	LDIT	LDAMP
End-to-End	32.1	33.1	8.0	18.7
Layer-by-Layer	26.1	33.1	-2.6	18.7
Denoiser-by-Denoiser	28.0	31.6	22.1	25.9

Compressive Image Recovery

512x512 images, 20x undersampling, noiseless measurements

Original Image

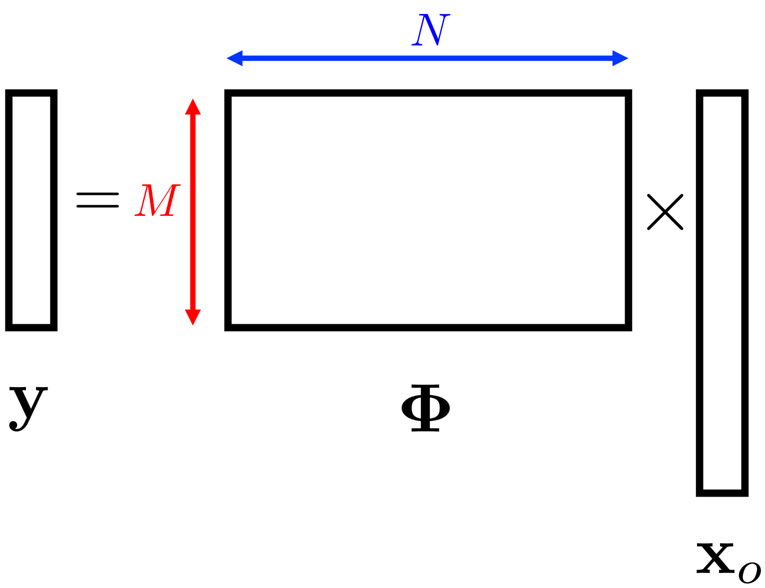
BM3D-AMP (27.2 dB, 75.04 sec)



TVAL3 (26.4 dB, 6.85 sec)

LDAMP (28.1 dB, 1.22 sec)

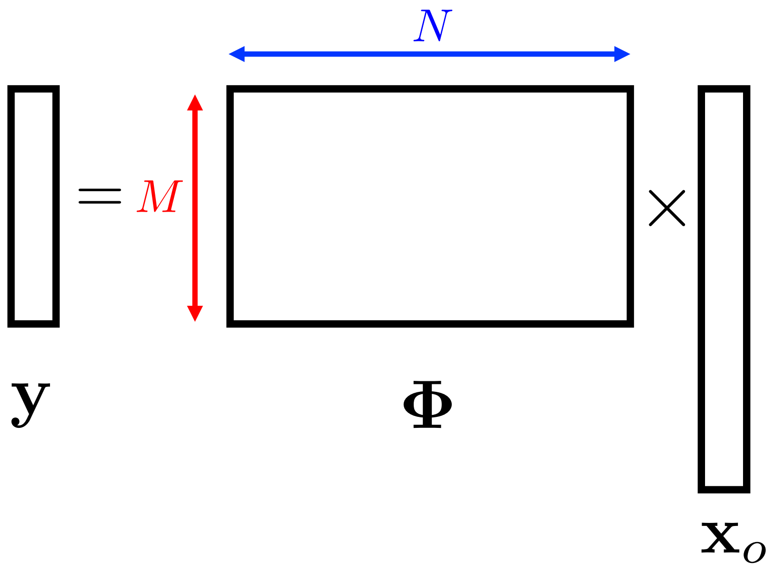
summary so far

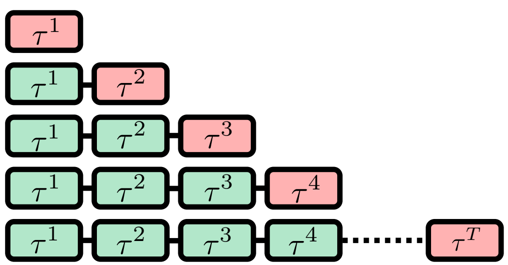


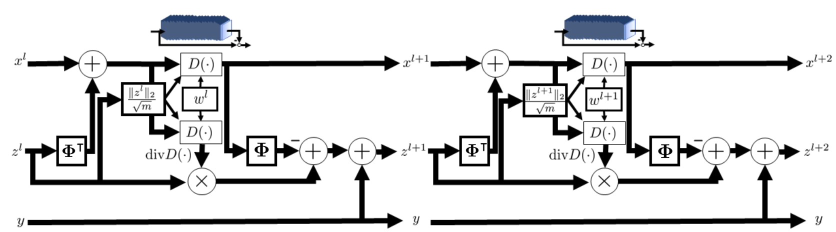
$$\arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_2^2 \text{ subject to } \mathbf{x} \in \mathcal{C}$$

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_2^2 + \lambda \times f(\mathbf{x})$$

summary so far

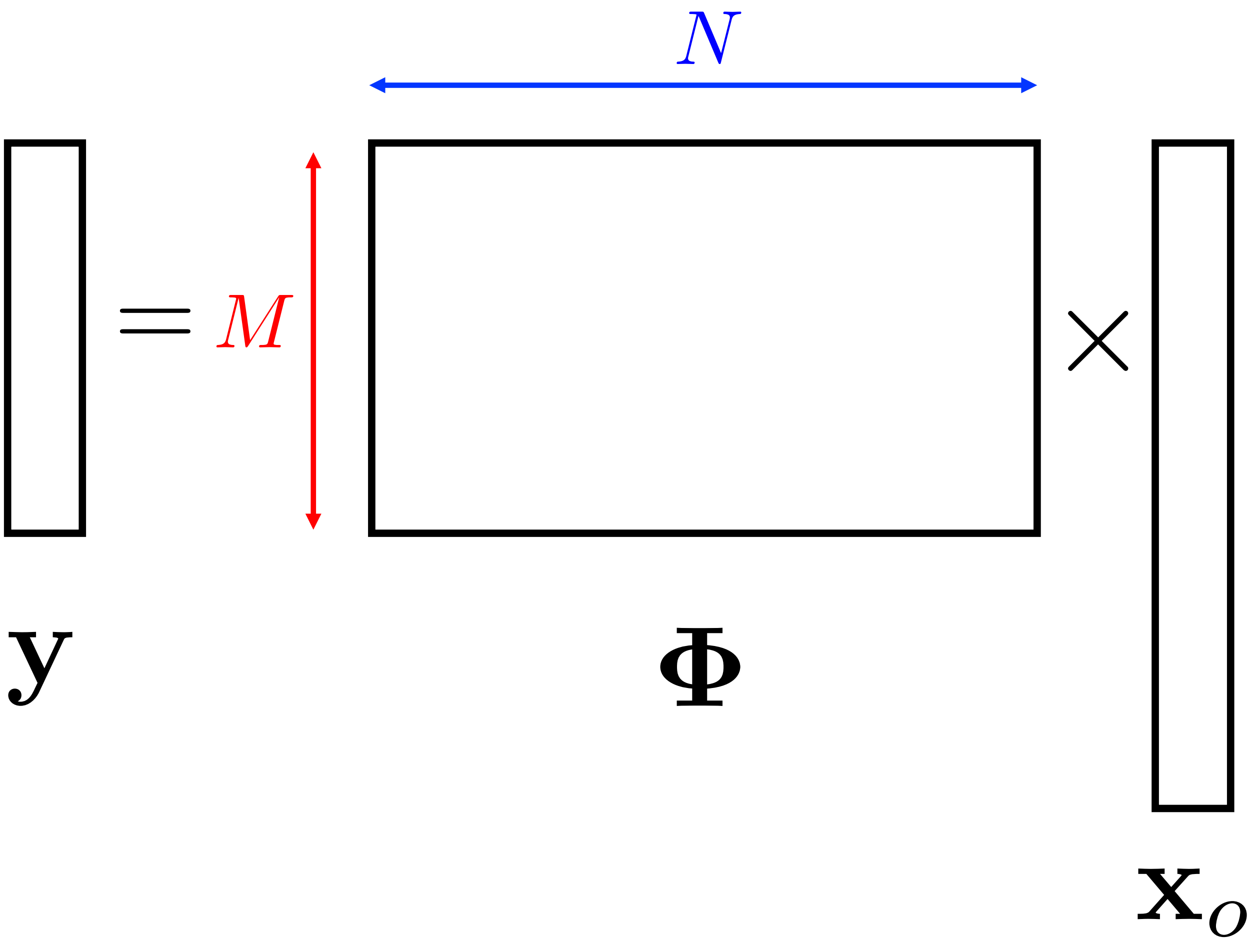






$$\min_{\mathbf{x}} \|\mathbf{y} - \boldsymbol{\Phi}\mathbf{x}\|_2^2 + \lambda \times f(\mathbf{x})$$

Data-Driven Dimensionality Reduction

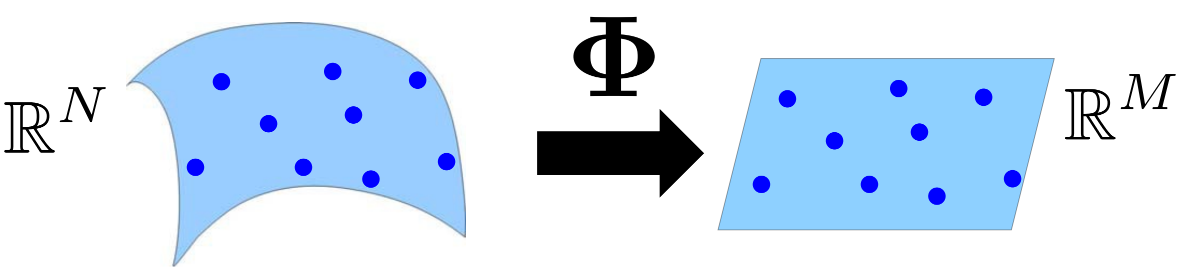


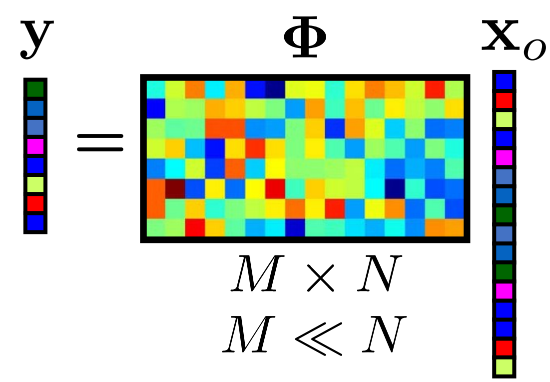
$$\min_{\mathbf{x}} \|\mathbf{y} - \boldsymbol{\Phi}\mathbf{x}\|_2^2 + \lambda \times f(\mathbf{x})$$

Data-Driven Dimensionality Reduction

- Goal: Create a mapping $\mathbf{\Phi}$ from $\mathbb{R}^N$ to $\mathbb{R}^M$ with $M < N$ that preserves the key geometric properties of the data.

- Learning from Data: Given a training set, find "best" $\mathbf{\Phi}$ that preserves its geometry.





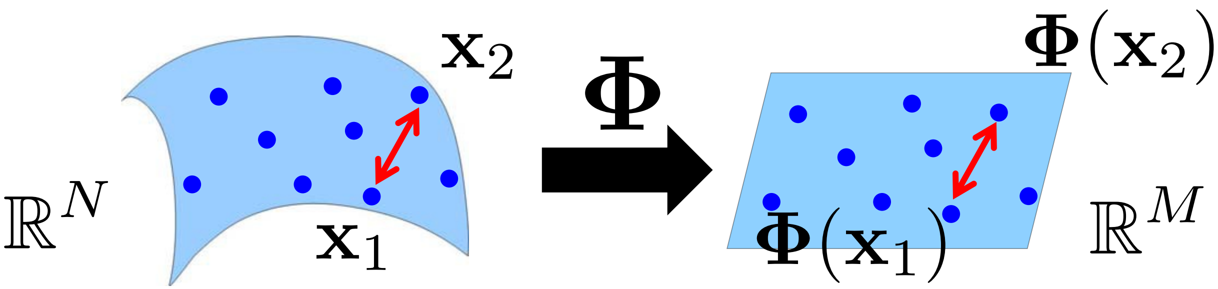
Near-Isometric Embedding

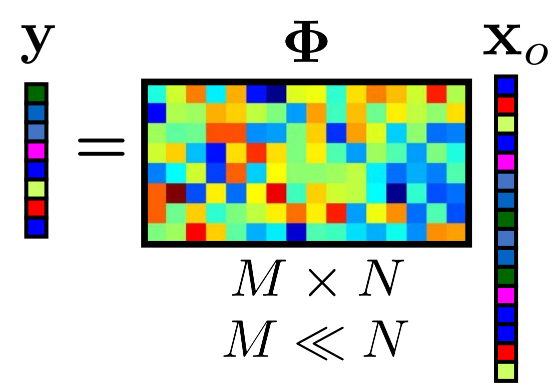
- **Goal: Learn "best" $\boldsymbol{\Phi}$ from data that preserves geometry.**

- **Design $\boldsymbol{\Phi}(.)$ to preserve inter-point distances.**

$$(1-\epsilon)\|\mathbf{x}_1 - \mathbf{x}_2\|_2 \leq \|\boldsymbol{\Phi}(\mathbf{x}_1) - \boldsymbol{\Phi}(\mathbf{x}_2)\|_2 \leq (1+\epsilon)\|\mathbf{x}_1 - \mathbf{x}_2\|_2$$

- **Applications:**
  - **Computational Sensing**
  - **Machine Learning**
  - **Approximate Nearest Neighbors**





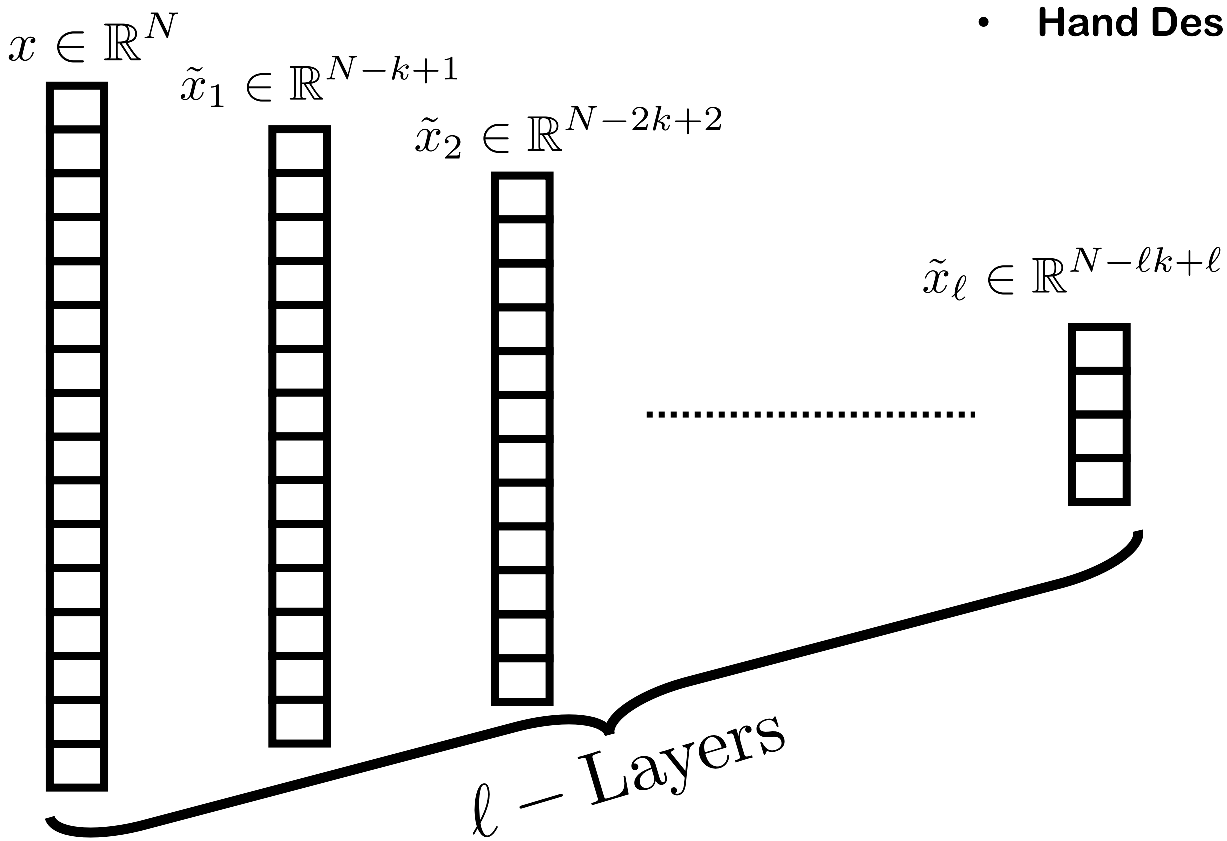
Dim. Reduction with Convolution

- **Required Number of Layers:** $\ell \simeq \left\lfloor \frac{N-M}{k} \right\rfloor$
- **If $M \ll N$, we need many layers.**
  - **Problem:** Vanishing Gradient
  - **Solution:** Large Filter Size, Skip Connections, Pooling

Large Filter Size: Inefficient

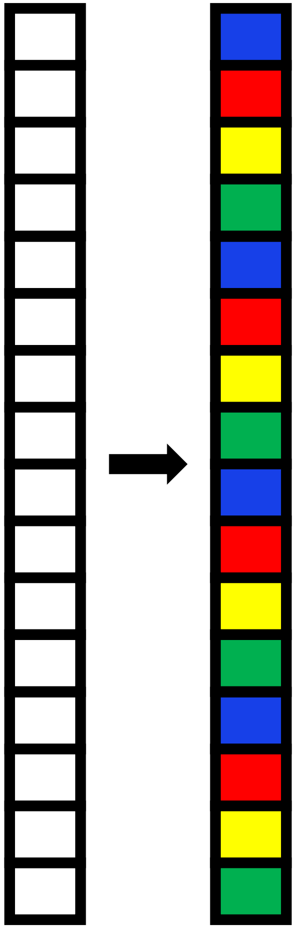
Skip Connections: Unequal Layer Size

Pooling:
- Loss of Information
- Hand Designed



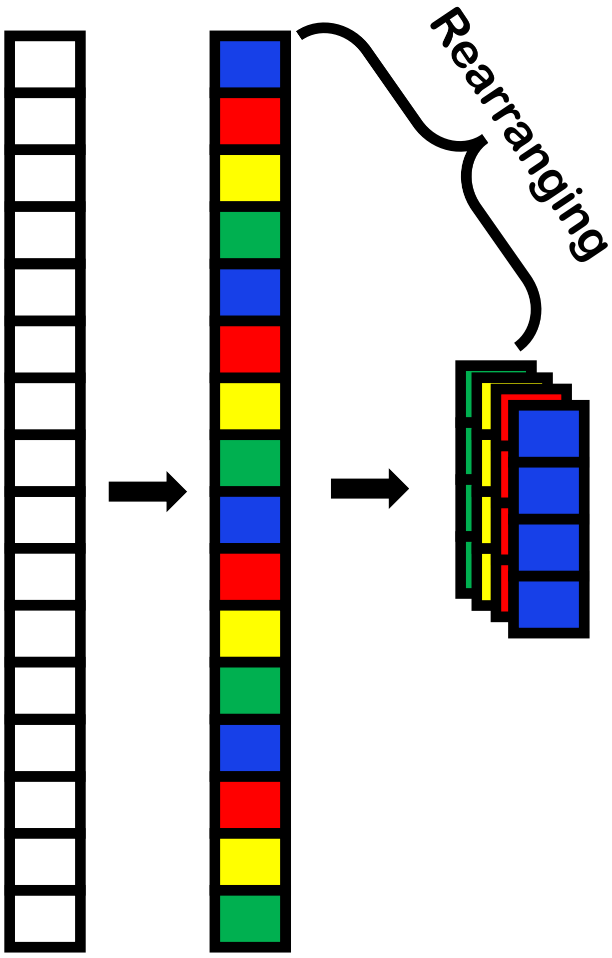
Convolutional Encoder

$\mathbf{x} \in \mathbb{R}^N$

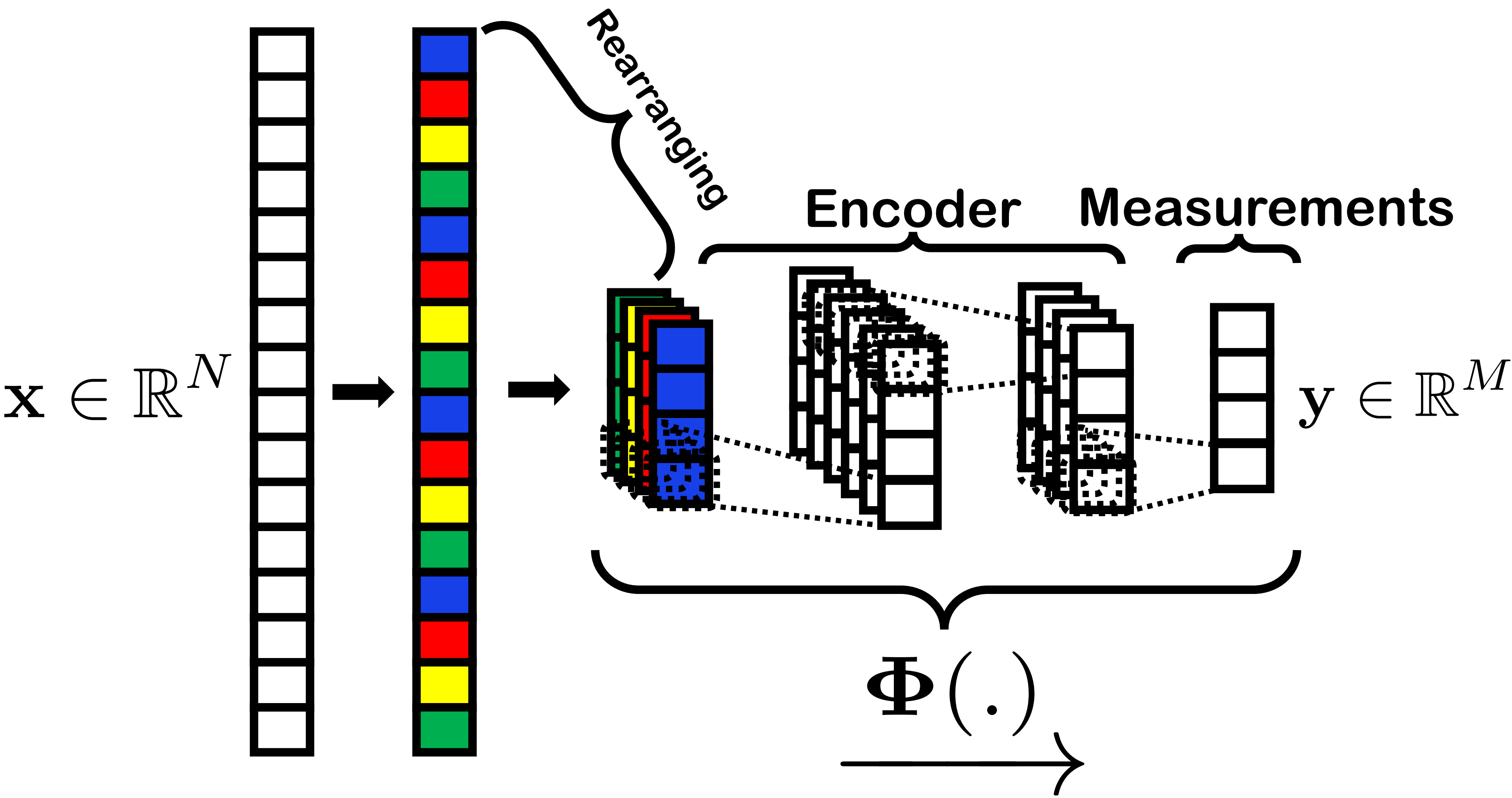


Convolutional Encoder

$\mathbf{x} \in \mathbb{R}^N$



Convolutional Encoder



[**Mousavi**, Dasarathy, Baraniuk, *Allerton 2017*]

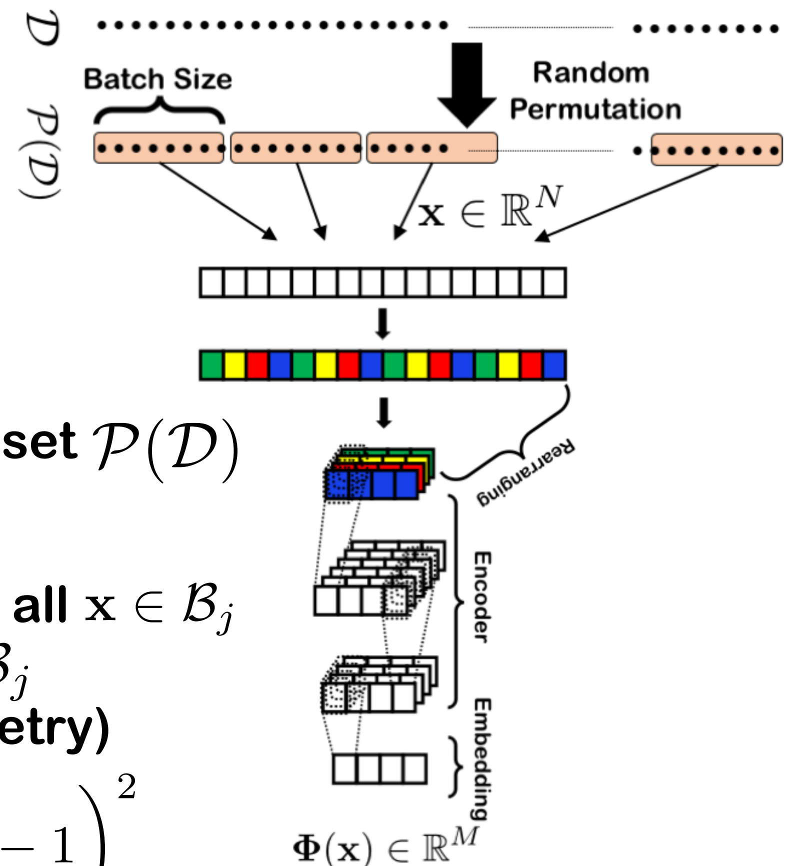
Near-Isometric Embedding with DeepCodec

- **Inputs:**
  - Training set: $\mathcal{D}$
  - Number of epochs: n_{epochs}
  - Network parameters: Ω_e

- **For** $i = 1$ **to** n_{epochs}
  - Generate a randomly permuted dataset $\mathcal{P}(\mathcal{D})$
  - For every $\mathcal{B}_j \in \mathcal{P}(\mathcal{D})$
    - ➢ Compute embedding $\boldsymbol{\Phi}(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{B}_j$
    - ➢ Compute the loss function for $\mathcal{B}_j$ (maximum deviation from isometry)

$$\mathcal{L}_{\mathcal{B}_j} = \max_{l,k} \left(\frac{\|\boldsymbol{\Phi}(\mathbf{x}_l) - \boldsymbol{\Phi}(\mathbf{x}_k)\|_2}{\|\mathbf{x}_l - \mathbf{x}_k\|_2} - 1 \right)^2$$

  - Compute the aggregated loss function $\mathcal{L}(\Omega_e) = \operatorname{avg}_j(\mathcal{L}_{\mathcal{B}_j})$
  - Use an optimizer and $\mathcal{L}(\Omega_e)$ to update Ω_e

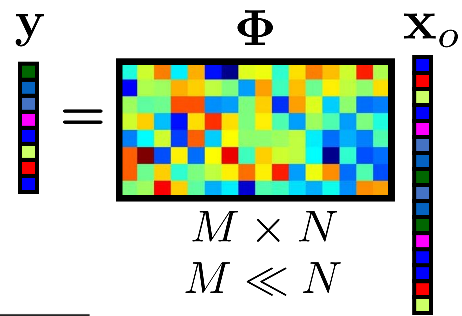


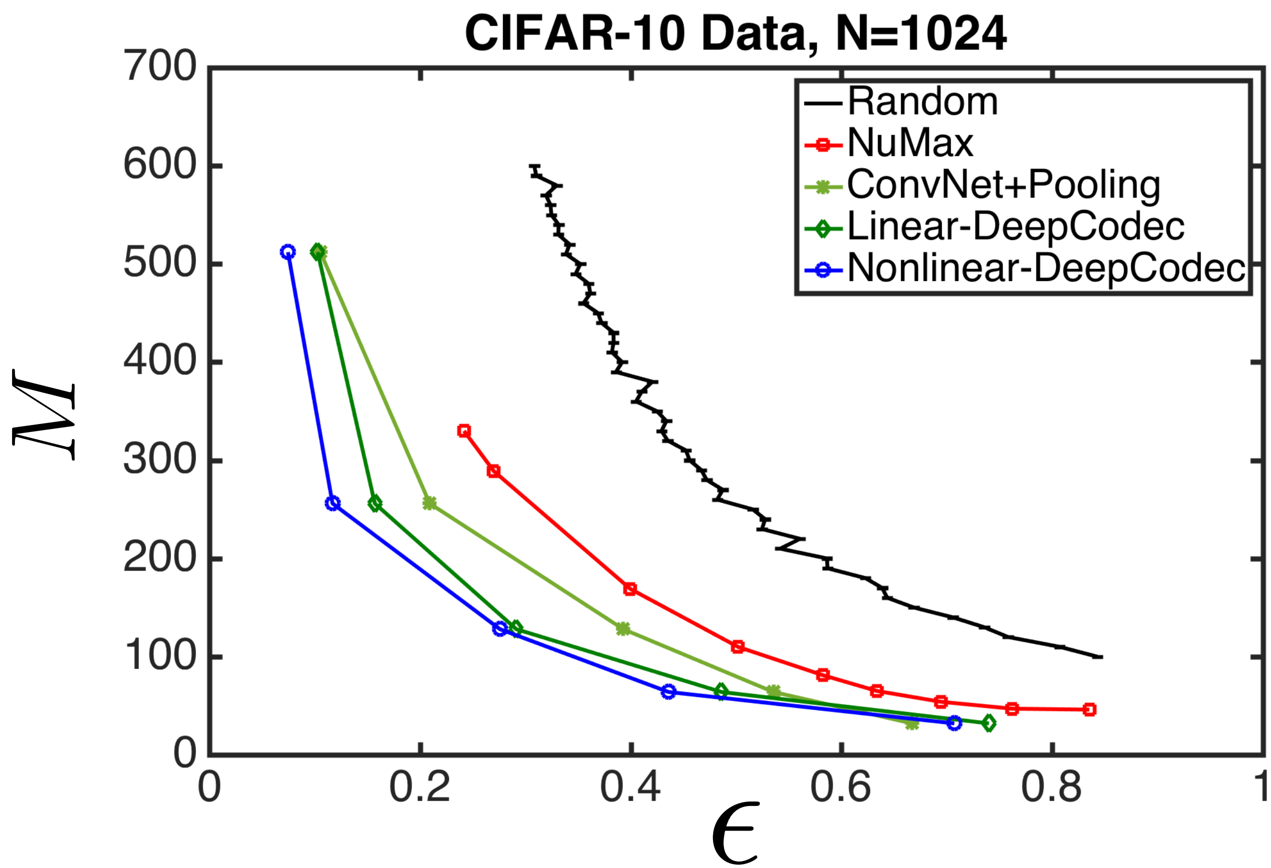
[**Mousavi**, Dasarathy, Baraniuk, *Allerton 2017*]

Result: near-isometry mapping

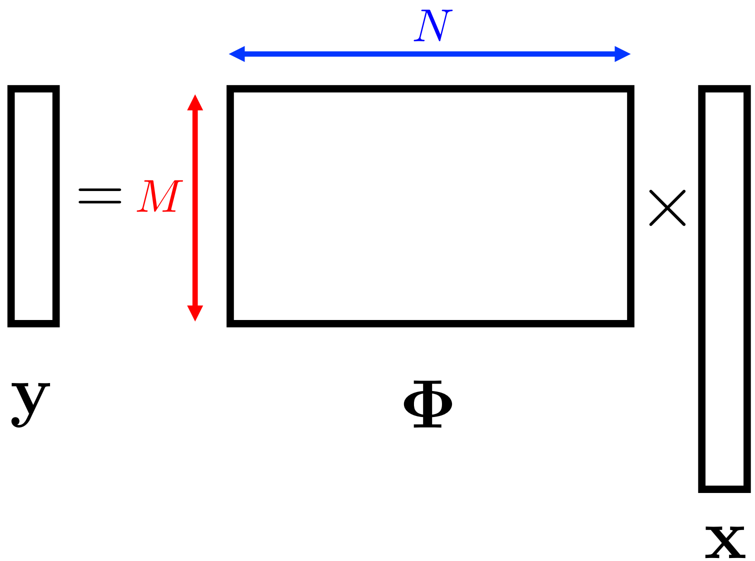
$$(1-\epsilon)\|\mathbf{x}_1-\mathbf{x}_2\|_2 \leq \|\mathbf{\Phi}(\mathbf{x}_1)-\mathbf{\Phi}(\mathbf{x}_2)\|_2 \leq (1+\epsilon)\|\mathbf{x}_1-\mathbf{x}_2\|_2$$

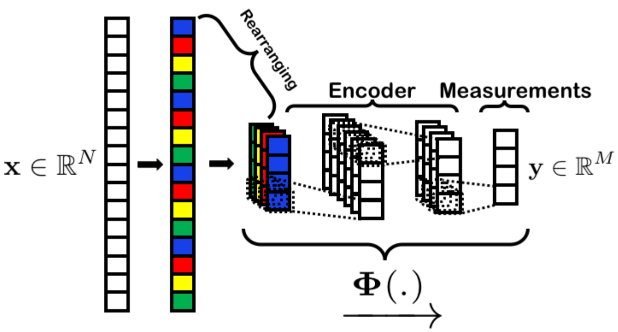
$$\mathbf{\Phi}(\mathbf{x}) \in \mathbb{R}^M$$

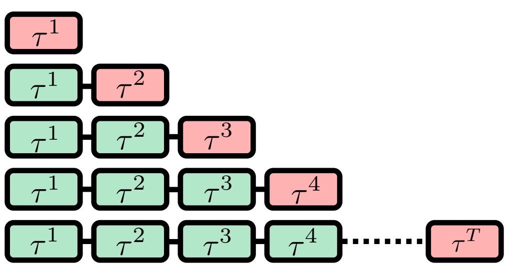


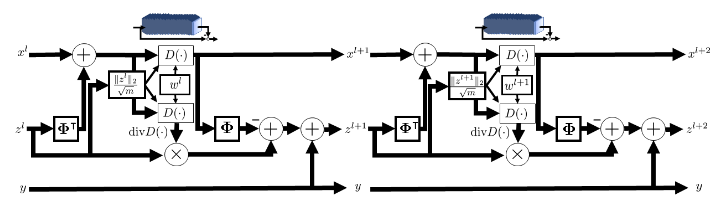


summary









$$\min_{\mathbf{x}} \|\mathbf{y} - \boldsymbol{\Phi}\mathbf{x}\|_2^2 + \lambda \times f(\mathbf{x})$$

The Road Ahead

- Still early days for coupling model and data for inference problems.

- Two general trend:

  - Advancing data-driven approaches
    - Medical Imaging
    - Generative Modeling

  - Theoretical foundation for data-driven approaches
    - Necessity and sufficiency for deep learning

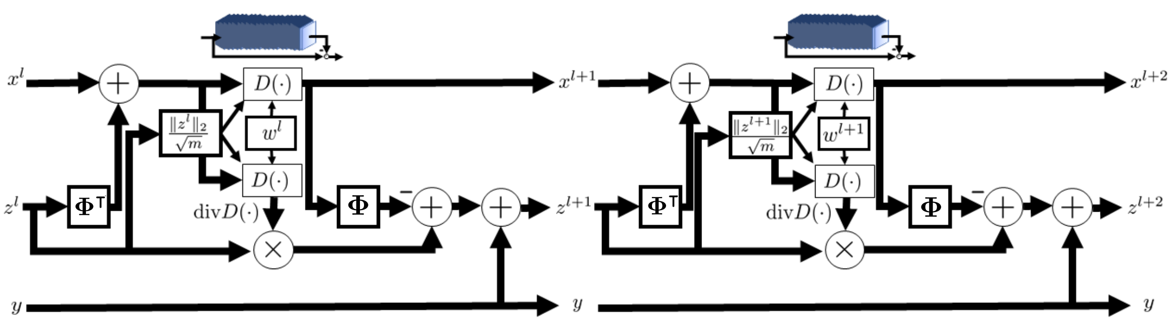
- Sensing for intelligent systems (Machine Sensing):

  - Depth sensing
  - Range acquisition

The Road Ahead: Model vs. Data

- **Still early days for coupling model and data for inference problems.**

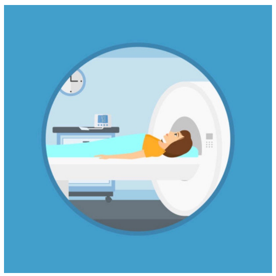
Extensively Model-Based **Extensively Data-Driven**



The Road Ahead: Data Science and AI for Imaging

- Today's Medical Imaging

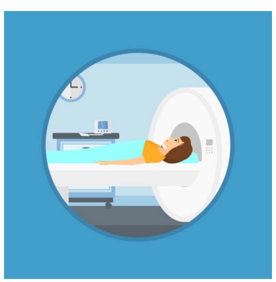
45 Minutes



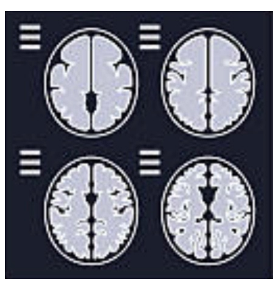
2 Hours

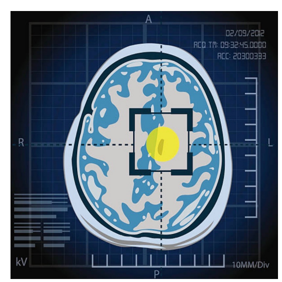
- Future of Medical Imaging

Few Minutes



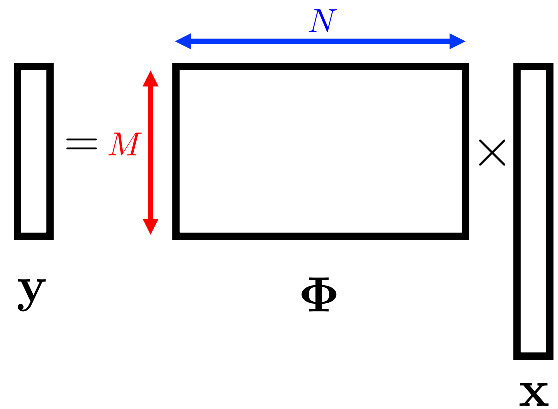
Few Minutes





The Road Ahead: Generative Modeling

- **Generative Modeling as an inverse problem.**



- **Generating a movie by artificial intelligence**

Generate simple objects

Generate phrase-based objects

Generate scenes from sentences

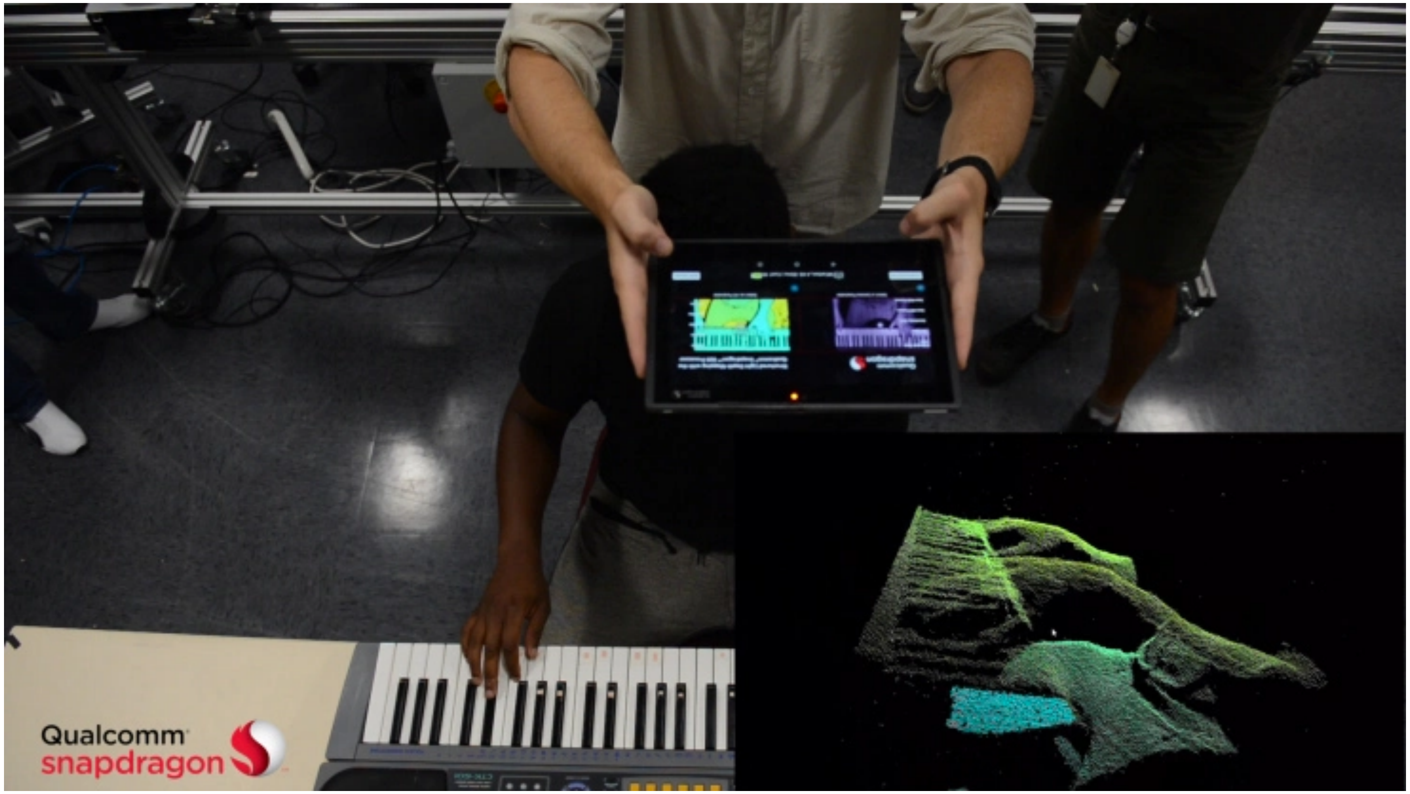
Adding Dynamics between the scenes

The Road Ahead: Theoretical foundation for data-driven approaches

- **Necessity and sufficiency for deep learning**

The Road Ahead: Machine Sensing

- **Sensing for intelligent systems (Machine Sensing):**
  - **Depth sensing**
  - **Range acquisition**



More Information

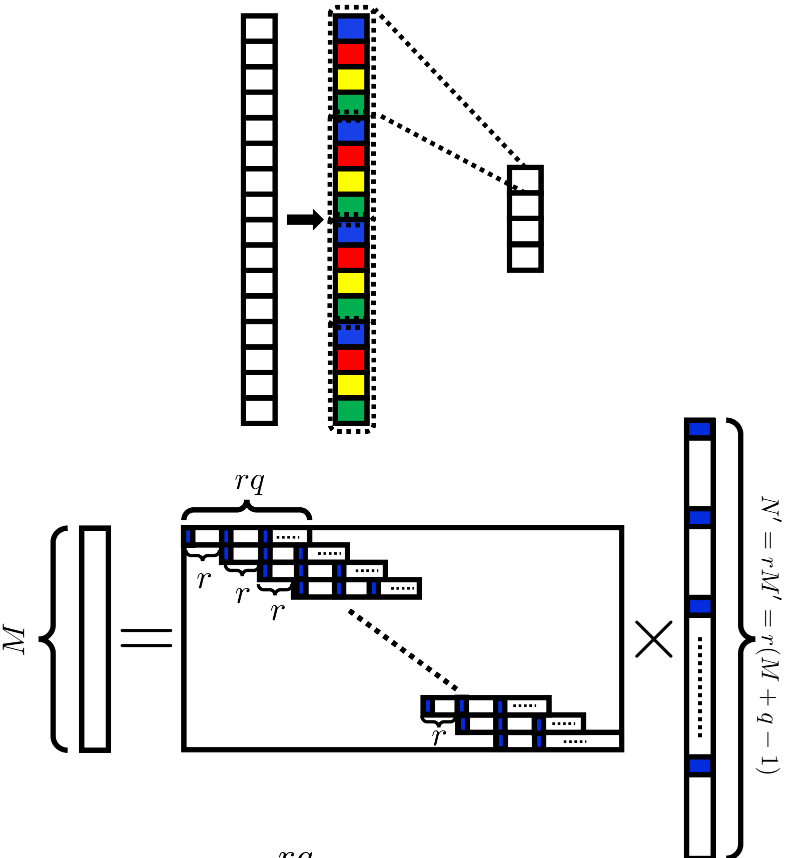
http://alim.blogs.rice.edu

ali.mousavi@rice.edu

Backup Slides

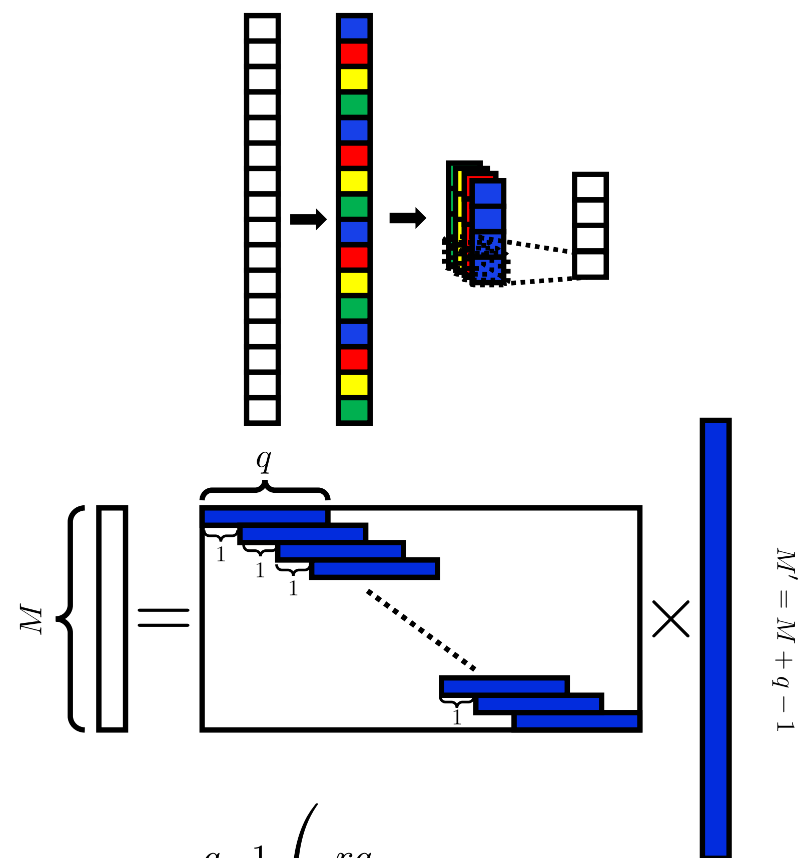
Why Rearrangement in the Encoder?

Strided Convolution



$$(\mathbf{x}_{\text{conv}})_j = \sum_{i=1}^{rq} (w_{\text{conv}})_i (\mathbf{x}_{\text{in}})_{(j-1)r+i}$$

Rearranged Convolution



$$(\mathbf{x}_{\text{conv}})_j = \sum_{z=0}^{q-1} \left(\sum_{\substack{i=1 \\ i \overset{q}{\equiv} z}}^{rq} (w_{\text{conv}})_i (\mathbf{x}_{\text{in}})_{(j-1)r+i} \right)$$

Effective Noise in AMP

- **AMP Iterations:**

$$\mathbf{x}^{t+1} = \eta(\mathbf{x}^t + \mathbf{\Phi}^\intercal \mathbf{z}^t; \tau^t)$$

$$\mathbf{z}^t = \mathbf{y} - \mathbf{\Phi}\mathbf{x}^t + \frac{1}{\delta}\mathbf{z}^{t-1}\langle \eta'(\mathbf{x}^{t-1} + \mathbf{\Phi}^\intercal \mathbf{z}^{t-1})\rangle$$

- **Effective noise at every iteration:**

$$\mathbf{x}^t + \mathbf{\Phi}^\intercal \mathbf{z}^t = \mathbf{x}_o + \mathbf{v}^t$$

- $\mathbf{v}^t$ **has a Gaussian distribution.**

$$(\sigma^t)^2 \triangleq \mathrm{Var}(\mathbf{v}^t)$$

- $\|\mathbf{x}^t - \mathbf{x}_o\|_2^2/N$ **is accurately predicted by** σ^t

$$\frac{\|\mathbf{x}^t - \mathbf{x}_o\|_2^2}{N} = \mathbb{E}_{X,W}[(\eta(X + \sigma^t W; \tau^t) - X)^2]$$

$$W \sim N(0,1)$$

$$X \sim P_{\mathbf{x}_o}$$

→ **State Evolution**

[Donoho et al. 2009, 2011]

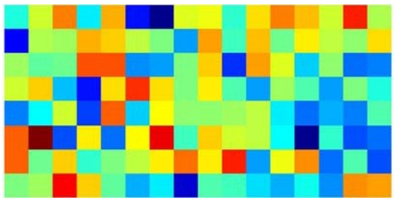
[Bayati and Montanari 2011]

Previous Works on Near-Isometry

[J-L, 84],[Indyk et al., 99],[Hegde et al., 15]

- **Johnson-Lindenstrauss Lemma: Considering a point set $Q \in \mathbb{R}^N$, there exists a lipchitz map that achieves near-isometry with constant δ given:** $M = O\left(\frac{\log(|Q|)}{\delta^2}\right)$

Ψ



$M \times N$

- **Example:**
  - **Random Embedding**
    - **Poor constants**
    - **Oblivious to data structure**
  - **Designed Embedding: Nuclear norm minimization with Max-norm constraints (NuMax)**

$$v_k = \frac{x_i - x_j}{\|x_i - x_j\|_2}$$

$$1 - \delta \leq \|\Psi v_i\|_2^2 \leq 1 + \delta$$

$$i = 1, 2, \ldots, S$$

$$P = \Psi^T \Psi$$

$$\mathcal{A} : P \mapsto \{v_i^T P v_i\}_{i=1}^S$$

$$\begin{aligned} &\text{minimize} \quad \|P\|_* \\ &\|\mathcal{A}(P) - 1\|_\infty \leq \delta \\ &P \succeq 0, \quad P = P^T \end{aligned}$$