Sparse Representation Solutions

From Convex Optimization to Deep Unfolding

Yimian Dai
PCA Lab, NJUST

2024-01-15

Contents

1

Reference

Convex Optimization

[1] Daubechies I, Defrise M, De Mol C. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint[J]. Commun. Pur. Appl. Math., 2004, 57(11): 1413-1457.

[2] Hunter D R, Lange K. A tutorial on MM algorithms[J]. The American Statistician, 2004, 58(1): 30-37.

[3] Landweber L. An iteration formula for Fredholm integral equations of the first kind[J]. American Journal of Mathematics, 1951, 73(3): 615-624.

[4] Beck A, Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems[J]. SIAM Journal on Imaging Sciences, 2009, 2(1): 183-202.

[5] Attouch H, Peypouquet J. The rate of convergence of Nesterov's accelerated forward-backward method is actually faster than 1/k^2[J]. SIAM Journal on Optimization, 2016, 26(3): 1824-1834.

Deep Unfolding

[6] Gregor K, LeCun Y. Learning fast approximations of sparse coding[C]//Proceedings of the 27th international conference on machine learning. 2010: 399-406.

[7] Zhang J, Ghanem B. ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 1828-1837.

2 Preparation

2 Preparation

Soft Thresholding Function[1]

Problem to be solved

$$J(x) = \arg\min_{x} \|x - y\|_2^2 + \lambda \|x\|_1$$

⇩

Resolve for each element

$$\min (x_i - y_i)^2 + \lambda |x_i|$$

⇩

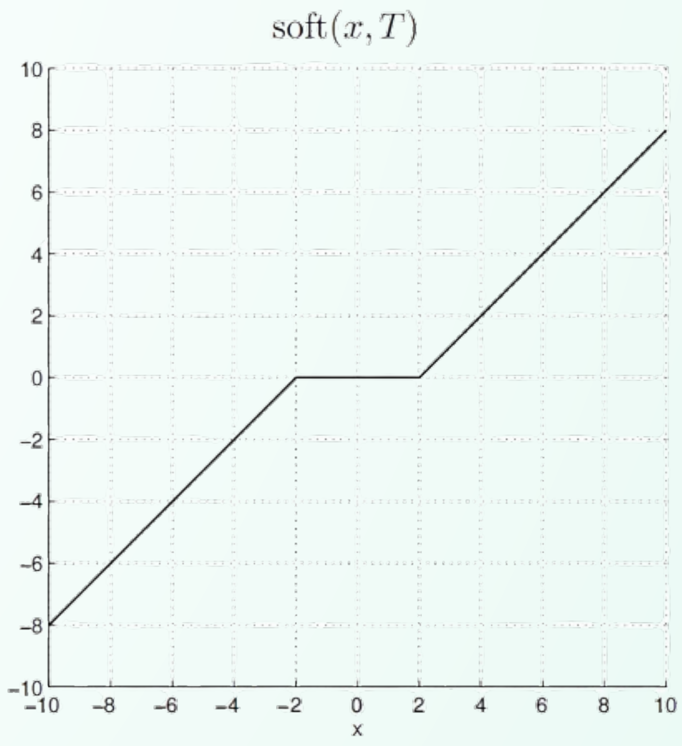
Set the derivative to zero

$$y_i = x_i + \frac{\lambda}{2}\operatorname{sgn}(x_i)$$

⇩

Denote as

$$x_i = soft\left(y_i, \frac{\lambda}{2}\right)$$



[1] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pur. Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, 2004.

2 Preparation

Majorization-Minimization[2] 上界化-最小化

Unconstrained optimization problem

$$\min_x J(x)$$

⇩

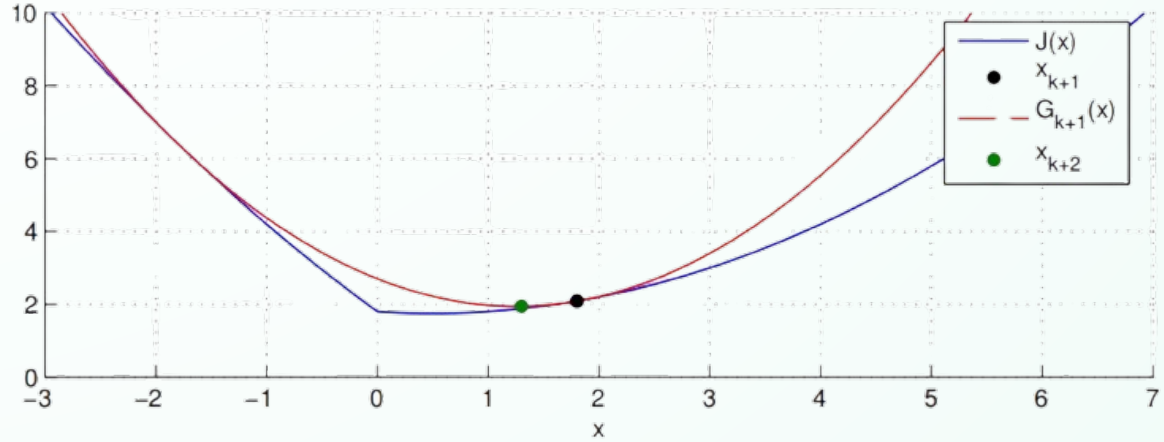
At the $(k+1)$th iteration, construct a globally optimal function $G_k(x)$ that is more tractable for optimization, satisfying:

$$G_k(x) \geq J(x), G_k(x) = J(x_k)$$

⇩

The local minimum $G_k(x)$ is taken as the new iterate x_k. This ensures:

$$J(x_{k+1}) \leq G_{k+1}(x_{k+1}) < G_k(x_k) = J(x_k)$$



[2] David R Hunter & Kenneth Lange (2004) A Tutorial on MM Algorithms, *The American Statistician*, 58:1, 30-37, DOI: 10.1198/0003130042836

2

Preparation

Landweber Iteration[3]

Unconstrained optimization problem

$$\min_{\mathbf{x}} J(\mathbf{x}) = \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_2^2$$

$$\Downarrow$$

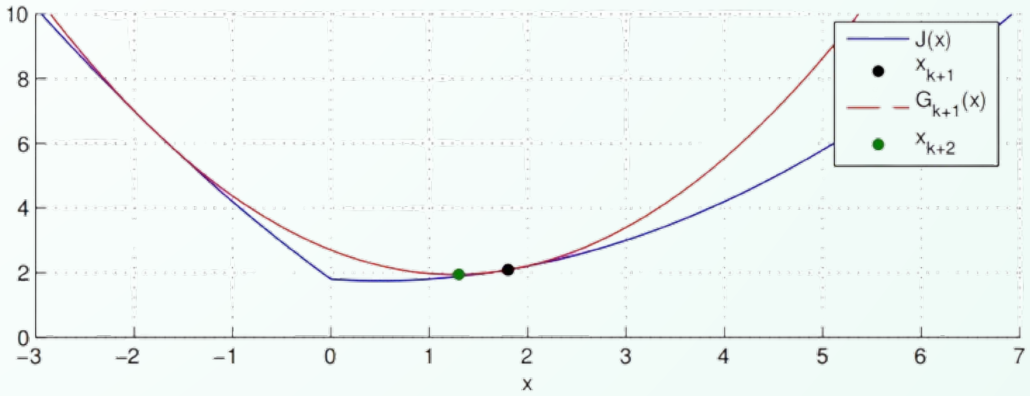
At the $(k+1)$th iteration, construct a globally optimal function $G_k(x)$

$$G_k(\mathbf{x}) = \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_2^2 + (\mathbf{x} - \mathbf{x}_k)^{\mathrm{T}}\left(\alpha\mathbf{I} - \mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}\right)(\mathbf{x} - \mathbf{x}_k)$$

$$\Downarrow$$

The local minimum is obtained by setting the derivative equal to zero

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \frac{1}{\alpha}\mathbf{\Phi}^{\mathrm{T}}(\mathbf{y} - \mathbf{\Phi}\mathbf{x}_k)$$



[3] Landweber, L. (1951). An Iteration Formula for Fredholm Integral Equations of the First Kind. *American Journal of Mathematics*, *73*(3), 615–624.

3 Convex Optimization

3 Convex Optimization

Iterative Soft Thresholding Algorithm[1] (ISTA)

1. **Slow Convergence**
2. Hyperparameter Sensitivity

Problem to be solved

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

At the $(k+1)$th iteration, construct a globally optimal function $G_k(x)$

$$G_k(\mathbf{x}) = \|\mathbf{y} - \mathbf{\Phi x}\|_2^2 + \lambda\|\mathbf{x}\|_1 + (\mathbf{x} - \mathbf{x}_k)^{\mathrm{T}}\left(\alpha\mathbf{I} - \mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}\right)(\mathbf{x} - \mathbf{x}_k)$$
$$= \alpha\left\|\mathbf{x}_k + \frac{1}{\alpha}\mathbf{\Phi}^{\mathrm{T}}(\mathbf{y} - \mathbf{\Phi x}_k) - \mathbf{x}\right\|_2^2 + C + \lambda\|\mathbf{x}\|_1$$

Soft thresholding function

$$\mathbf{x}_{k+1} = \mathrm{soft}\left(\mathbf{x}_k + \frac{1}{\alpha}\mathbf{\Phi}^{\mathrm{T}}(\mathbf{y} - \mathbf{\Phi x}_k), \frac{\lambda}{2\alpha}\right)$$

$$\mathbf{r}_{k+1} = \mathbf{x}_k - \frac{1}{\alpha}\mathbf{\Phi}^{\mathrm{T}}(\mathbf{\Phi x}_k - \mathbf{y})$$
$$\mathbf{x}_{k+1} = \mathrm{soft}\left(\mathbf{r}_{k+1}, \frac{\lambda}{2\alpha}\right)$$

[1] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pur. Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, 2004.

Fast Iterative Soft Thresholding Algorithm[4] (FISTA)

ISTA

$$\mathbf{x}_{k+1} = \mathcal{P}_{\frac{\lambda}{2\alpha}}(\mathbf{x}_k)$$
$$= \mathrm{soft}\left(\mathbf{x}_k + \frac{1}{\alpha}\mathbf{\Phi}^{\mathrm{T}}(\mathbf{y} - \mathbf{\Phi}\mathbf{x}_k), \frac{\lambda}{2\alpha}\right)$$

Problem to be solved

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

Motivation

An additional point that is smartly chosen[5]

$$\mathbf{x}_{k+1} = \mathcal{P}_{\frac{\lambda}{2\alpha}}(\mathbf{y}_{k+1})$$

where

$$\mathbf{y}_{k+1} = \mathbf{x}_k + \left(\frac{t_k - 1}{t_{k+1}}\right)(\mathbf{x}_k - \mathbf{x}_{k-1})$$

$$t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}$$

FISTA Solution

Convergence Rate: $\mathcal{O}(\frac{1}{k}) \longrightarrow \mathcal{O}(\frac{1}{k^2})$

Hyperparameter Sensitivity

[4] Beck A, Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems[J]. *SIAM Journal on Imaging Sciences*, 2009, 2(1): 183-202.
[5] Attouch H, Peypouquet J. The rate of convergence of Nesterov's accelerated forward-backward method is actually faster than 1/k^2[J]. *SIAM Journal on Optimization*, 2016, 26(3): 1824-1834.

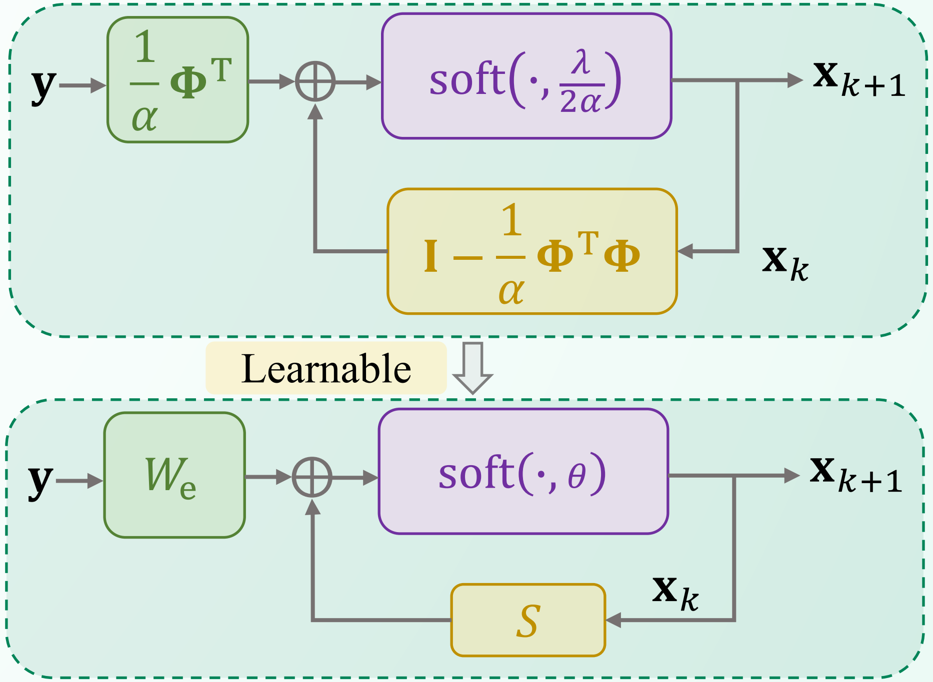
4 Deep Unfolding

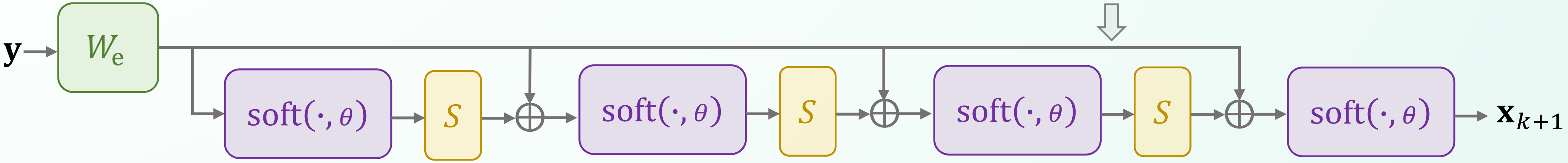
4 Deep Unfolding

Learned Iterative Soft Thresholding Algorithm[6] (LISTA)

ISTA

$$
\begin{aligned}
\mathbf{x}_{k+1} &= \mathcal{P}_{\frac{\lambda}{2\alpha}}(\mathbf{x}_k) \\
&= \mathrm{soft}\left(\mathbf{x}_k + \frac{1}{\alpha}\boldsymbol{\Phi}^{\mathrm{T}}(\mathbf{y} - \boldsymbol{\Phi}\mathbf{x}_k), \frac{\lambda}{2\alpha}\right) \\
&= \mathrm{soft}\left(\mathbf{x}_k - \frac{1}{\alpha}\boldsymbol{\Phi}^{\mathrm{T}}(\boldsymbol{\Phi}\mathbf{x}_k - \mathbf{y}), \frac{\lambda}{2\alpha}\right) \\
&= \mathrm{soft}\left(\frac{1}{\alpha}\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{y} + \left(\mathbf{I} - \frac{1}{\alpha}\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}\right)\mathbf{x}_k, \frac{\lambda}{2\alpha}\right)
\end{aligned}
$$

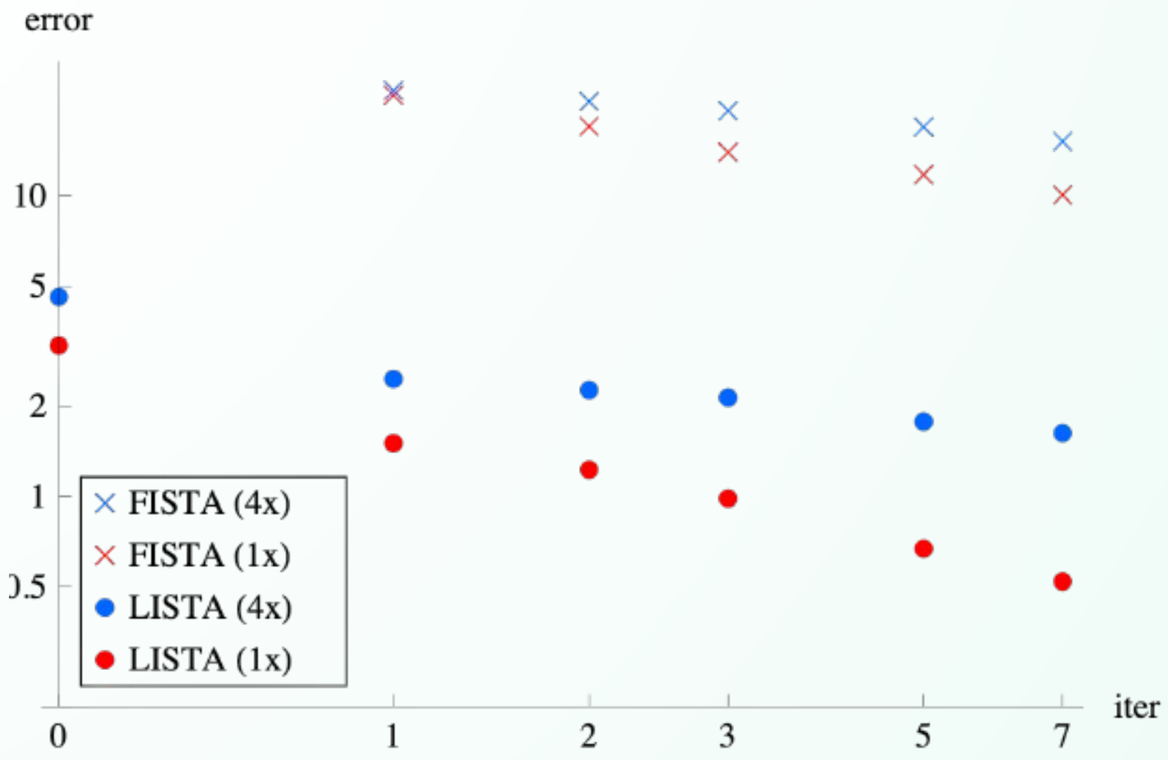




[6] Gregor K, **LeCun Y**. Learning fast approximations of sparse coding[C]//Proceedings of the 27th international conference on international conference on machine learning. 2010: 399-406.

4 Deep Unfolding

Learned Iterative Soft Thresholding Algorithm[6] (LISTA)



4 Deep Unfolding

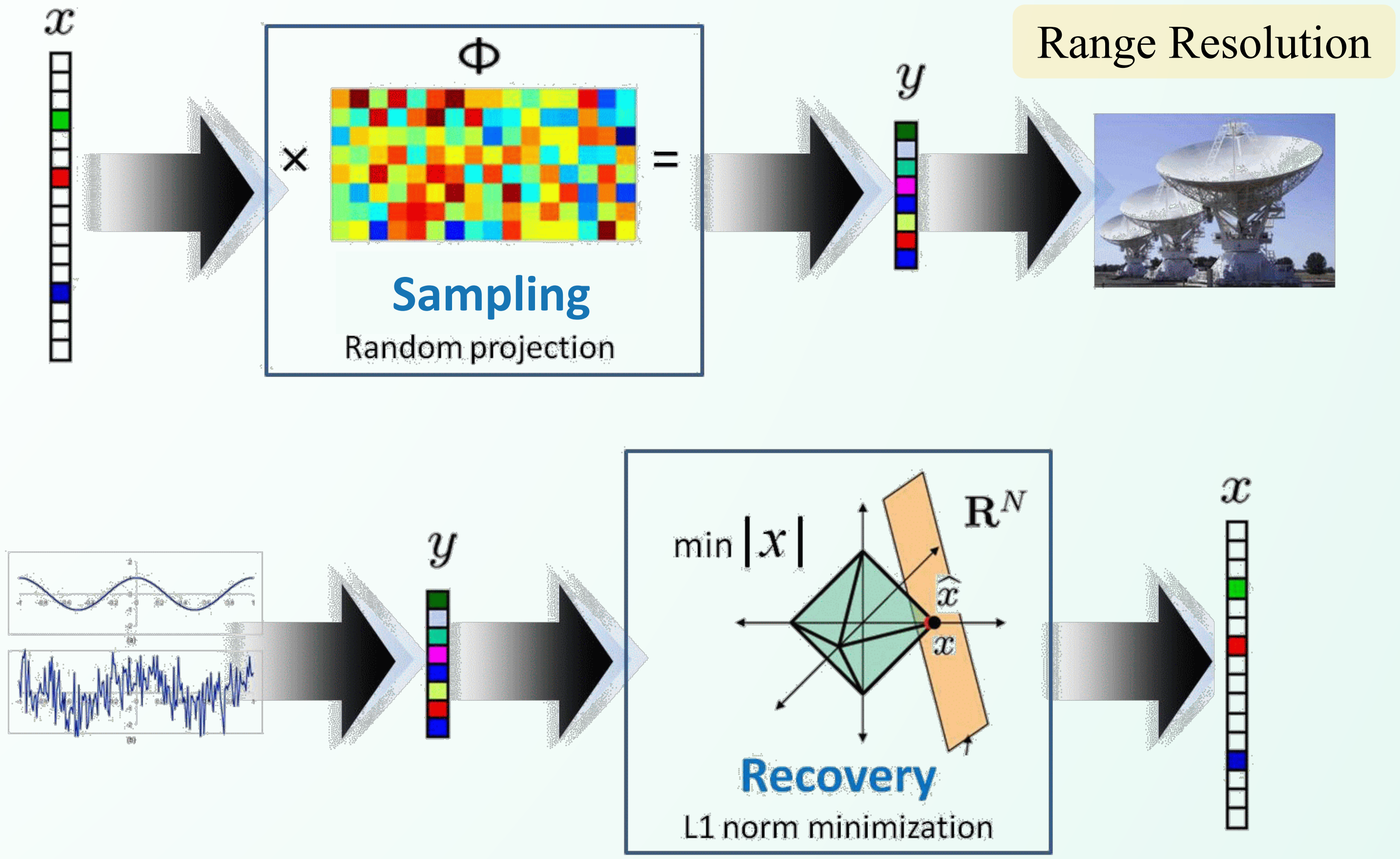
ISTA-Net[7]

Sparse Coding

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y}-\mathbf{\Phi}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1$$

Compressive Sensing Reconstruction

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y}-\mathbf{\Phi}\mathbf{x}\|_2^2 + \lambda\|\mathbf{\Psi}\mathbf{x}\|_1$$



[7] Zhang J, Ghanem B. ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 1828-1837.

Compression

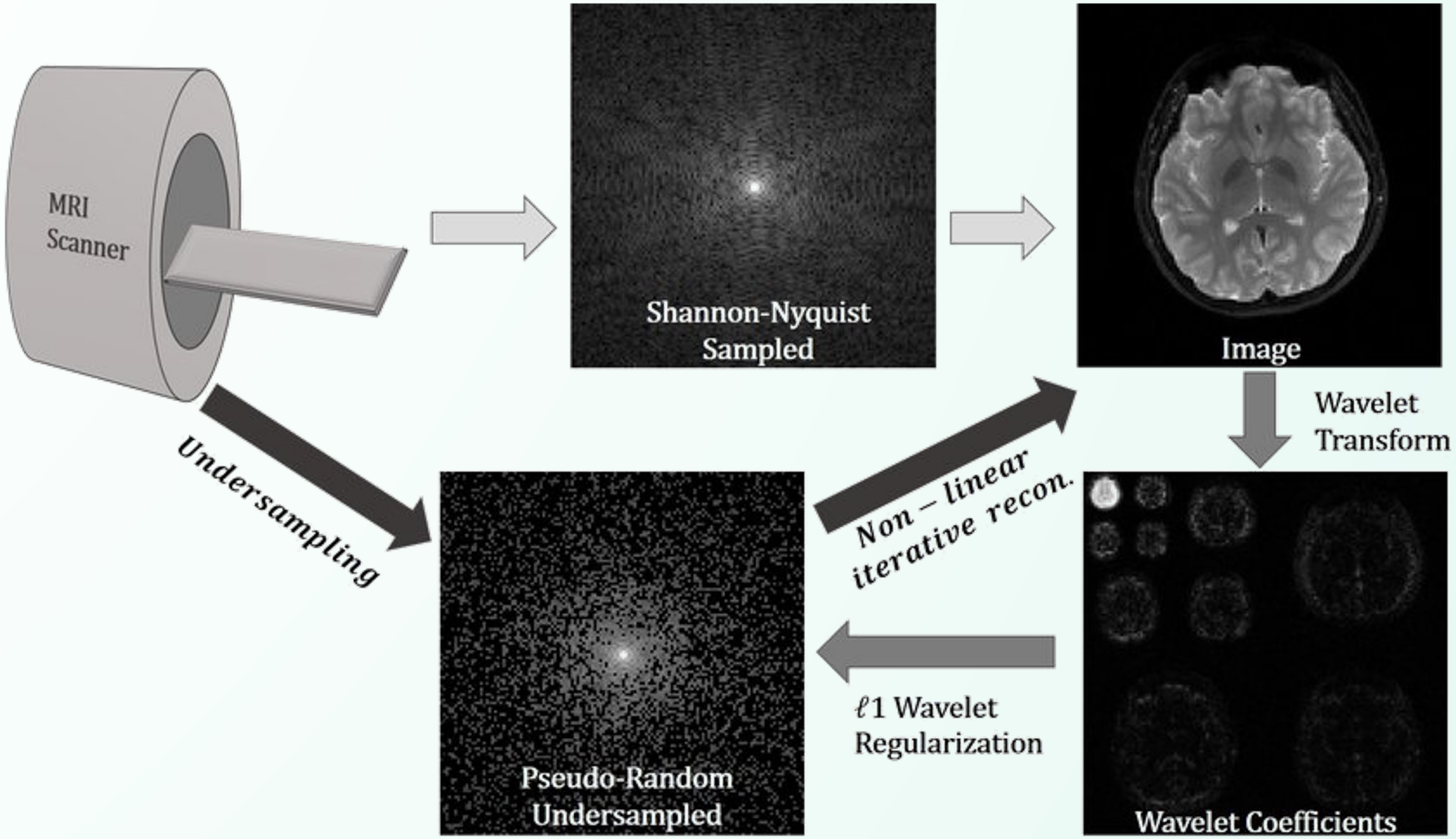
ISTA-Net[7]

Sparse Coding

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y}-\mathbf{\Phi}\mathbf{x}\|_2^2+\lambda\|\mathbf{x}\|_1$$

Compressive Sensing Reconstruction

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y}-\mathbf{\Phi}\mathbf{x}\|_2^2+\lambda\|\mathbf{\Psi}\mathbf{x}\|_1$$



[7] Zhang J, Ghanem B. ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 1828-1837.

4

Deep Unfolding

ISTA-Net

Compressive Sensing Reconstruction

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y}-\mathbf{\Phi}\mathbf{x}\|_2^2+\lambda\|\mathbf{\Psi}\mathbf{x}\|_1$$

$$\mathbf{r}^{(k)}=\mathbf{x}^{(k-1)}-\rho\mathbf{\Phi}^{\mathrm{T}}\left(\mathbf{\Phi}\mathbf{x}^{(k-1)}-\mathbf{y}\right)$$

$$\mathbf{x}^{(k)}=\arg\min_{\mathbf{x}} \tfrac{1}{2}\left\|\mathbf{y}-\mathbf{r}^{(k)}\right\|_2^2+\lambda\|\mathbf{\Psi}\mathbf{x}\|_1$$

$$\mathrm{prox}_{\lambda\phi}(\mathbf{r})=\arg\min_{\mathbf{x}} \tfrac{1}{2}\|\mathbf{y}-\mathbf{r}\|_2^2+\lambda\phi(\mathbf{x})$$

When $\phi(\mathbf{x})=\|\mathbf{x}\|_1$ ⇨ $\mathrm{prox}_{\lambda\phi}(\mathbf{r})=\mathrm{soft}(\mathbf{r},\lambda)$

When $\phi(\mathbf{x})=\|\mathbf{W}\mathbf{x}\|_1$ ⇨ $\mathrm{prox}_{\lambda\phi}(\mathbf{r})=\mathbf{W}^{\mathrm{T}}\mathrm{soft}(\mathbf{W}\mathbf{r},\lambda)$

Wavelet Transform Matrix ⇨

1. Linear
2. Orthogonal

4

Deep Unfolding

ISTA-Net

Linear Transform $\boldsymbol{\Psi}$

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y}-\boldsymbol{\Phi}\mathbf{x}\|_2^2 + \lambda\|\boldsymbol{\Psi}\mathbf{x}\|_1$$

Nonlinear Transform $\mathcal{F}$

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y}-\boldsymbol{\Phi}\mathbf{x}\|_2^2 + \lambda\|\mathcal{F}(\mathbf{x})\|_1$$

$$\mathcal{F}(\mathbf{x}) = \mathbf{B}\mathrm{ReLU}(\mathbf{A}\mathbf{x})$$

$$\mathbf{r}^{(k)} = \mathbf{x}^{(k-1)} - \rho^{(k)}\boldsymbol{\Phi}^{\mathrm{T}}\left(\boldsymbol{\Phi}\mathbf{x}^{(k-1)} - \mathbf{y}\right)$$

$$\mathbf{x}^{(k)} = \min_{\mathbf{x}} \frac{1}{2}\left\|\mathbf{x}-\mathbf{r}^{(k)}\right\|_2^2 + \lambda\|\mathcal{F}(\mathbf{x})\|_1$$

4

Deep Unfolding

ISTA-Net

$$\mathbf{x}^{(k)} = \min_{\mathbf{x}} \frac{1}{2}\left\|\mathbf{x} - \mathbf{r}^{(k)}\right\|_2^2 + \lambda\|\mathcal{F}(\mathbf{x})\|_1$$

⇩

https://jianzhang.tech/papers/CVPR_Supplementary.pdf

⇩

$$\left\|\mathcal{F}(\mathbf{x}) - \mathcal{F}\left(\mathbf{r}^{(k)}\right)\right\|_2^2 \approx \alpha\left\|\mathbf{x} - \mathbf{r}^{(k)}\right\|_2^2$$

⇩

$$\mathbf{x}^{(k)} = \min_{\mathbf{x}} \frac{1}{2}\left\|\mathcal{F}(\mathbf{x}) - \mathcal{F}\left(\mathbf{r}^{(k)}\right)\right\|_2^2 + \theta\|\mathcal{F}(\mathbf{x})\|_1$$

⇩

$$\mathcal{F}\left(\mathbf{x}^{(k)}\right) = \mathrm{soft}\left(\mathcal{F}\left(\mathbf{r}^{(k)}\right), \theta\right)$$

$$\phi(\mathbf{x}) = \|\mathbf{W}\mathbf{x}\|_1 \;\Rightarrow\; \mathrm{prox}_{\lambda\phi}\left(\mathbf{r}^{(k)}\right) = \mathbf{W}^{\mathrm{T}}\mathrm{soft}\left(\mathbf{W}\mathbf{r}^{(k)}, \lambda\right)$$

⇩

$$\phi(\mathbf{x}) = \|\mathcal{F}(\mathbf{x})\|_1 \;\Rightarrow\; \mathrm{prox}_{\lambda\phi}\left(\mathbf{r}^{(k)}\right) = \tilde{\mathcal{F}}\left(\mathrm{soft}\left(\mathcal{F}\left(\mathbf{r}^{(k)}\right), \theta\right)\right)$$

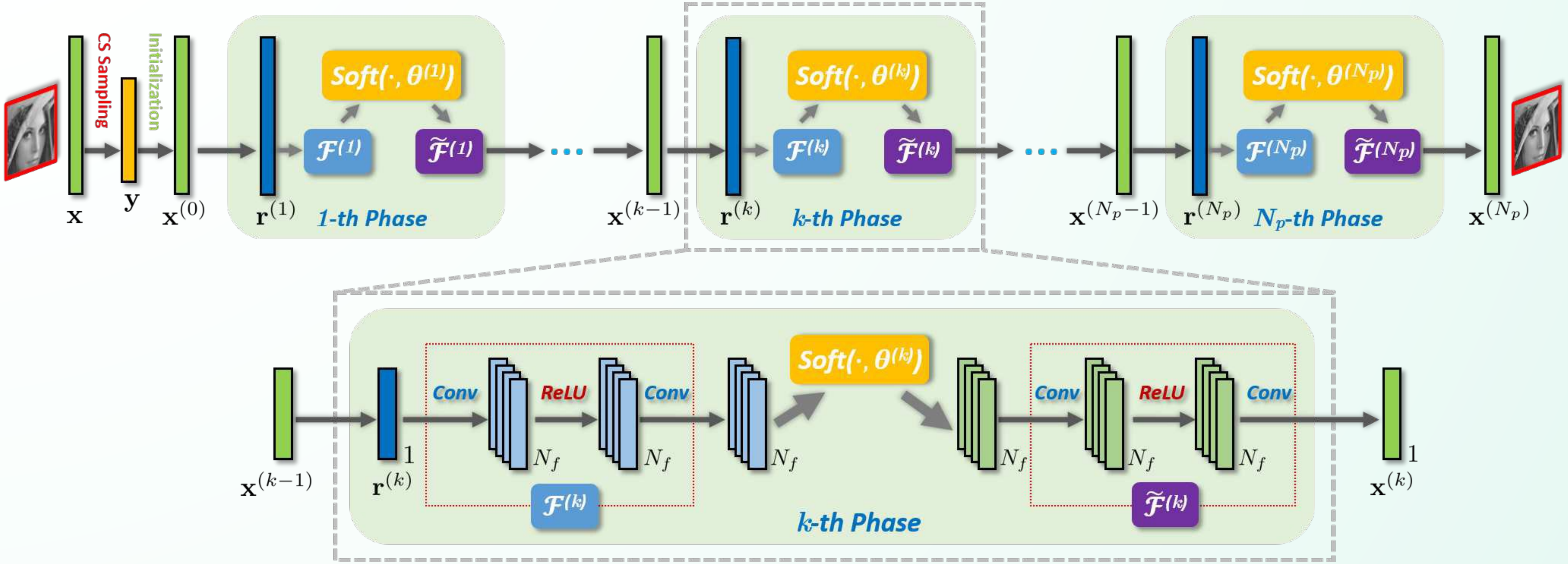
where

$$\widetilde{\mathcal{F}} \circ \mathcal{F} = \mathcal{I}$$

4 Deep Unfolding

$$\mathbf{r}^{(k)} = \mathbf{x}^{(k-1)} - \rho^{(k)}\boldsymbol{\Phi}^{\mathrm{T}}\left(\boldsymbol{\Phi}\mathbf{x}^{(k-1)} - \mathbf{y}\right)$$

$$\mathrm{prox}_{\lambda\phi}(\mathbf{r}^{(k)}) = \tilde{\mathcal{F}}^{(k)}\left(\mathrm{soft}(\mathcal{F}^{(k)}(\mathbf{r}^{(k)}), \theta^{(k)})\right)$$



4

Deep Unfolding

$$\widetilde{\mathcal{F}} \circ \mathcal{F} = \mathcal{I}$$

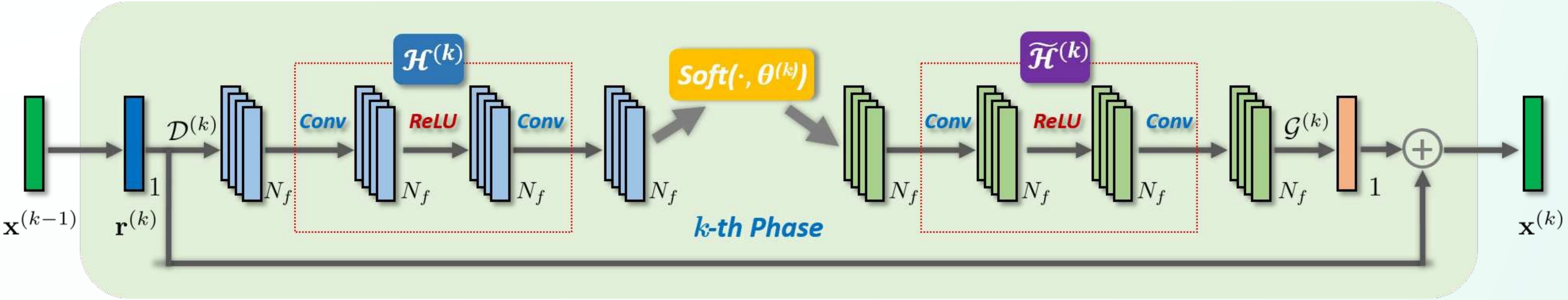
$$\mathcal{L}_{total}(\boldsymbol{\Theta}) = \mathcal{L}_{discrepancy} + \gamma \mathcal{L}_{constraint}$$

$$\text{with: } \begin{cases} \mathcal{L}_{discrepancy} = \frac{1}{N_b N} \sum_{i=1}^{N_b} \|\mathbf{x}_i^{(N_p)} - \mathbf{x}_i\|_2^2 \\ \mathcal{L}_{constraint} = \frac{1}{N_b N} \sum_{i=1}^{N_b} \sum_{k=1}^{N_p} \|\widetilde{\mathcal{F}}^{(k)}(\mathcal{F}^{(k)}(\mathbf{x}_i)) - \mathbf{x}_i\|_2^2 \end{cases}$$

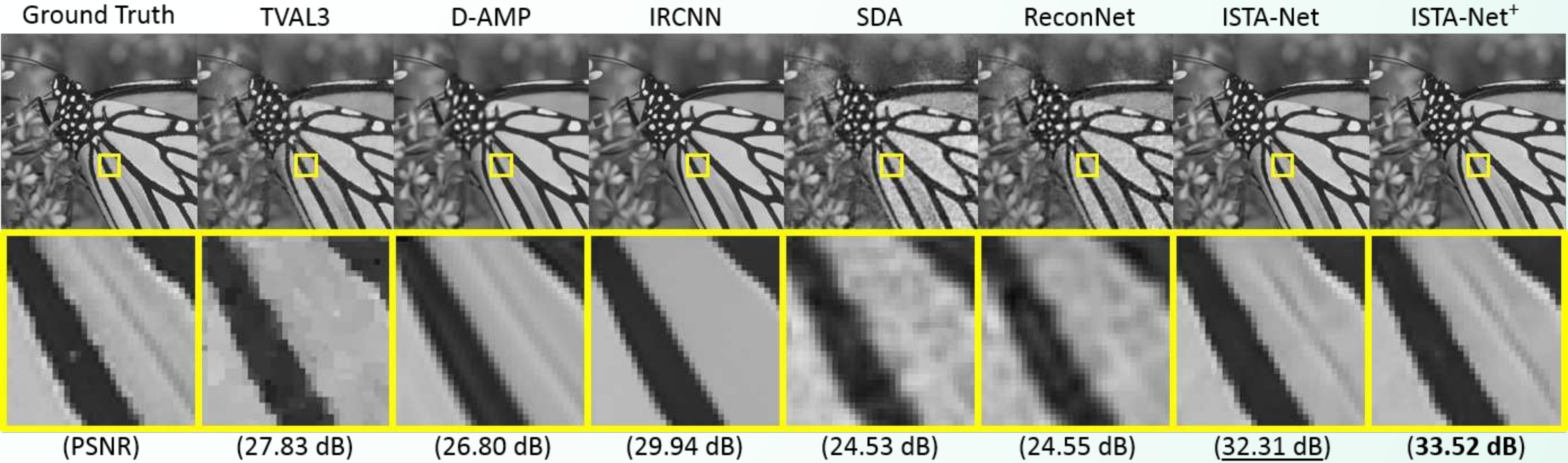
4

Deep Unfolding

ISTA-Net+



4 Deep Unfolding



THANKS