DEALing with Image Reconstruction: Deep Attentive Least Squares

Erich Kobler

Insitute for Machine Learning, LIT AI Lab
Department of Virtual Morphology, Clinical Research Institute Medical AI
JKU Linz

Joint work with Mehrsa Pourya, Michael Unser, and Sebastian Neumayer

March 9, 2026

JOHANNES KEPLER
UNIVERSITY LINZ

Inverse Problems in Imaging

Goal:
Estimate unknown image $\mathbf{x}$ given measurement $\mathbf{y}$.

Inverse Problems in Imaging

Goal:
Estimate unknown image $\mathbf{x}$ given measurement $\mathbf{y}$.

Bayesian approach:
Determine quality by posterior

$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x})p(\mathbf{x})$$

Inverse Problems in Imaging

Goal:
Estimate unknown image $\mathbf{x}$ given measurement $\mathbf{y}$.

Bayesian approach:
Determine quality by posterior

$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x})p(\mathbf{x})$$

- Likelihood $p(\mathbf{y}|\mathbf{x})$ models forward measurement process

$$\mathbf{y} = H(\mathbf{x}) + \mathbf{n}$$

  $H\colon \mathcal{X} \to \mathcal{Y}$ forward operator
  $\mathbf{n}$ noise (e.g., Gaussian/Poisson noise, ...)

Inverse Problems in Imaging

Goal:
Estimate unknown image $\mathbf{x}$ given measurement $\mathbf{y}$.

Bayesian approach:
Determine quality by posterior

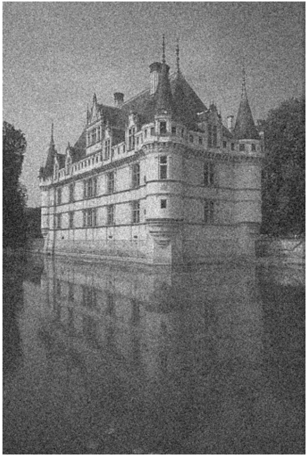
$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x})p(\mathbf{x})$$

- Likelihood $p(\mathbf{y}|\mathbf{x})$ models forward measurement process

$$\mathbf{y} = H(\mathbf{x}) + \mathbf{n}$$

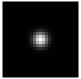
  $H\colon \mathcal{X} \to \mathcal{Y}$ forward operator
  $\mathbf{n}$ noise (e.g., Gaussian/Poisson noise, $\dots$)
- Prior $p(\mathbf{x})$ models a priori knowledge

Image Denoising



$$\xmapsto[\mathbf{n} \sim \mathcal{N}(0, \sigma^2 \mathrm{Id})]{\mathbf{H} = \mathrm{Id}}$$

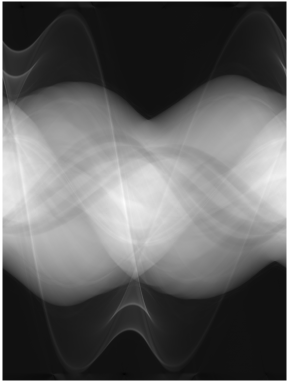
Image Superresolution



$$H\colon \mathbf{x} \mapsto (\mathbf{x} * \mathbf{a})_{\downarrow s}$$

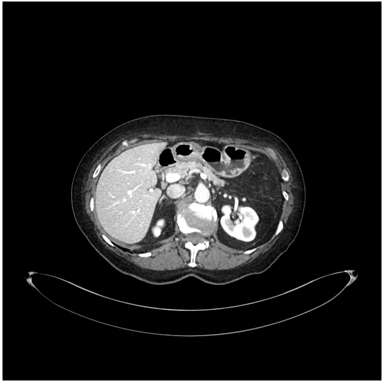
$$\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathrm{Id})$$

Computed Tomography (CT)

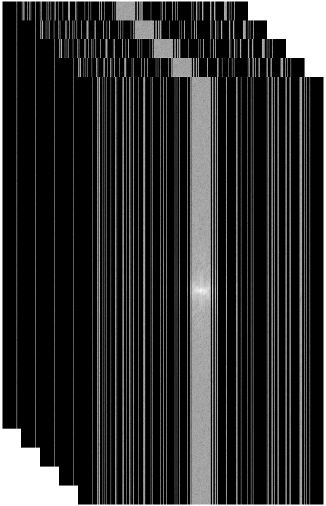


$$H_i : \mathbf{x} \mapsto \int_{\Gamma_i} \mathbf{x}(\xi)\mathrm{d}\xi$$

$$\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \Sigma)$$



Magnetic Resonance Imaging (MRI)



$$\xleftarrow[\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathrm{Id})]{H_i : \mathbf{x} \mapsto \mathbf{m} \odot \mathcal{F}(\mathbf{c} \odot \mathbf{x})}$$

Variational approach:

Find image maximizing posterior

$$\hat{\mathbf{x}} = \operatorname*{argmax}_{\mathbf{x}} p(\mathbf{x}|\mathbf{y})$$

Variational approach:

Find image maximizing posterior

$$\hat{\mathbf{x}} = \operatorname*{argmax}_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}) = \operatorname*{argmin}_{\mathbf{x}} \underbrace{-\log p(\mathbf{y}|\mathbf{x})}_{\text{data fidelity}} \underbrace{-\log p(\mathbf{x})}_{\text{regularizer}}$$

Variational Methods for Inverse Problems

Variational approach:
Find image maximizing posterior

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmax}}\, p(\mathbf{x}|\mathbf{y}) = \underset{\mathbf{x}}{\operatorname{argmin}} \underbrace{- \log p(\mathbf{y}|\mathbf{x})}_{\text{data fidelity}} \underbrace{- \log p(\mathbf{x})}_{\text{regularizer}}$$

Data fidelity:
For linear inverse problem and homoscedastic Gaussian noise, simplifies to

$$D(\mathbf{x}, \mathbf{y}) = - \log p(\mathbf{y}|\mathbf{x}) = \frac{1}{2\sigma^2} \|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2$$

Variational Methods for Inverse Problems

Variational approach:
Find image maximizing posterior

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmax}}\, p(\mathbf{x}|\mathbf{y}) = \underset{\mathbf{x}}{\operatorname{argmin}} \underbrace{-\log p(\mathbf{y}|\mathbf{x})}_{\text{data fidelity}} \underbrace{-\log p(\mathbf{x})}_{\textbf{regularizer}}$$

Data fidelity:
For linear inverse problem and homoscedastic Gaussian noise, simplifies to

$$D(\mathbf{x}, \mathbf{y}) = -\log p(\mathbf{y}|\mathbf{x}) = \frac{1}{2\sigma^2} \|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2$$

Regularizer:
Impose prior knowledge on solution $\mathbf{x}$

$$R(\mathbf{x}) = -\log p(\mathbf{x})$$

Regularizers

- Tikhonov regularization[1]

$$R(\mathbf{x}) = \frac{1}{2}\|\mathbf{Wx}\|_2^2 = \left\langle \mathbf{1}, (\mathbf{Wx})^2 \right\rangle$$

[1] Tikhonov, Arsenin (1977) *Solutions of Ill-Posed Problems*.

[2] Rudin, Osher, Fatemi (1992). *Nonlinear total variation based noise removal algorithms*. Physica D.

[3] Roth, Black (2009). *Fields of experts*. IJCV.

Regularizers

- Tikhonov regularization[1]

$$R(\mathbf{x}) = \frac{1}{2}\left\|\mathbf{W}\mathbf{x}\right\|_2^2 = \left\langle \mathbf{1}, (\mathbf{W}\mathbf{x})^2 \right\rangle$$

- Total variation[2] (TV) regularization

$$R(\mathbf{x}) = \left\|\mathbf{W}\mathbf{x}\right\|_{2,1} = \left\langle \mathbf{1}, \sqrt{(\mathbf{W}_h\mathbf{x})^2 + (\mathbf{W}_v\mathbf{x})^2} \right\rangle$$

[1] Tikhonov, Arsenin (1977) *Solutions of Ill-Posed Problems*.

[2] Rudin, Osher, Fatemi (1992). *Nonlinear total variation based noise removal algorithms*. Physica D.

[3] Roth, Black (2009). *Fields of experts*. IJCV.

Regularizers

- Tikhonov regularization[1]

$$R(\mathbf{x}) = \frac{1}{2}\left\|\mathbf{W}\mathbf{x}\right\|_2^2 = \left\langle \mathbf{1}, (\mathbf{W}\mathbf{x})^2 \right\rangle$$

- Total variation[2] (TV) regularization

$$R(\mathbf{x}) = \left\|\mathbf{W}\mathbf{x}\right\|_{2,1} = \left\langle \mathbf{1}, \sqrt{(\mathbf{W}_h\mathbf{x})^2 + (\mathbf{W}_v\mathbf{x})^2} \right\rangle$$

- Fields of experts[3] (FoE)

$$R(\mathbf{x}) = \sum_{c=1}^{N_c} \langle \mathbf{1}, \psi_c(\mathbf{W}_c\mathbf{x}) \rangle$$

[1] Tikhonov, Arsenin (1977) *Solutions of Ill-Posed Problems*.

[2] Rudin, Osher, Fatemi (1992). *Nonlinear total variation based noise removal algorithms*. Physica D.

[3] Roth, Black (2009). *Fields of experts*. IJCV.

Computing Solutions

- Tikhonov regularization $R(\mathbf{x}) = \frac{1}{2}\|\mathbf{Wx}\|_2^2$

$$\hat{\mathbf{x}} = \operatorname*{argmin}_{\mathbf{x}} \left\{ \frac{1}{2\sigma^2}\|\mathbf{Hx}-\mathbf{y}\|_2^2 + \frac{1}{2}\|\mathbf{Wx}\|_2^2 \right\}$$

Computing Solutions

- Tikhonov regularization $R(\mathbf{x}) = \frac{1}{2} \|\mathbf{W}\mathbf{x}\|_2^2$

$$\begin{aligned}\hat{\mathbf{x}} &= \underset{\mathbf{x}}{\operatorname{argmin}} \left\{ \frac{1}{2\sigma^2} \|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2 + \frac{1}{2} \|\mathbf{W}\mathbf{x}\|_2^2 \right\} \\ &= (\mathbf{H}^\top \mathbf{H} + \sigma^2 \mathbf{W}^\top \mathbf{W})^{-1} \mathbf{H}^\top \mathbf{y}\end{aligned}$$

Computing Solutions

- Tikhonov regularization $R(\mathbf{x}) = \frac{1}{2}\|\mathbf{W}\mathbf{x}\|_2^2$

$$\begin{aligned}\hat{\mathbf{x}} &= \operatorname*{argmin}_{\mathbf{x}} \left\{ \frac{1}{2\sigma^2}\|\mathbf{H}\mathbf{x}-\mathbf{y}\|_2^2 + \frac{1}{2}\|\mathbf{W}\mathbf{x}\|_2^2 \right\} \\ &= (\mathbf{H}^\top\mathbf{H} + \sigma^2\mathbf{W}^\top\mathbf{W})^{-1}\mathbf{H}^\top\mathbf{y}\end{aligned}$$

- TV/FoE regularization

$$\hat{\mathbf{x}} = \operatorname*{argmin}_{\mathbf{x}} \left\{ \frac{1}{2\sigma^2}\|\mathbf{H}\mathbf{x}-\mathbf{y}\|_2^2 + \sum_{c=1}^{N_c} \langle \mathbf{1}, \psi_c(\mathbf{W}_c\mathbf{x}) \rangle \right\}$$

Computing Solutions

- Tikhonov regularization $R(\mathbf{x}) = \frac{1}{2}\|\mathbf{W}\mathbf{x}\|_2^2$

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \left\{ \frac{1}{2\sigma^2}\|\mathbf{H}\mathbf{x}-\mathbf{y}\|_2^2 + \frac{1}{2}\|\mathbf{W}\mathbf{x}\|_2^2 \right\}$$
$$= (\mathbf{H}^\top\mathbf{H} + \sigma^2\mathbf{W}^\top\mathbf{W})^{-1}\mathbf{H}^\top\mathbf{y}$$

- TV/FoE regularization

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \left\{ \frac{1}{2\sigma^2}\|\mathbf{H}\mathbf{x}-\mathbf{y}\|_2^2 + \sum_{c=1}^{N_c} \langle \mathbf{1}, \psi_c(\mathbf{W}_c\mathbf{x})\rangle \right\}$$

Iterative solvers such as proximal (sub)gradient descent

$$\mathbf{x}_{k+1} = \mathrm{prox}_{\tau_k D}(\mathbf{x}_k - \tau_k \sum_{c=1}^{N_c} \mathbf{W}_c^\top \nabla\psi_c(\mathbf{W}_c\mathbf{x}))$$

using $\mathrm{prox}_{\tau D}(\mathbf{y}) = \mathrm{argmin}_{\mathbf{x}} \left\{ \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_2^2 + \frac{\tau\lambda}{2}\|\mathbf{H}\mathbf{x}-\mathbf{y}\|_2^2 \right\}$

Computing Solutions

- Tikhonov regularization $R(\mathbf{x}) = \frac{1}{2}\|\mathbf{W}\mathbf{x}\|_2^2$

$$\hat{\mathbf{x}} = \operatorname*{argmin}_{\mathbf{x}} \left\{ \frac{1}{2\sigma^2}\|\mathbf{H}\mathbf{x}-\mathbf{y}\|_2^2 + \frac{1}{2}\|\mathbf{W}\mathbf{x}\|_2^2 \right\}$$
$$= (\mathbf{H}^\top\mathbf{H} + \sigma^2\mathbf{W}^\top\mathbf{W})^{-1}\mathbf{H}^\top\mathbf{y}$$

- TV/FoE regularization

$$\hat{\mathbf{x}} = \operatorname*{argmin}_{\mathbf{x}} \left\{ \frac{1}{2\sigma^2}\|\mathbf{H}\mathbf{x}-\mathbf{y}\|_2^2 + \sum_{c=1}^{N_c} \langle \mathbf{1}, \psi_c(\mathbf{W}_c\mathbf{x})\rangle \right\}$$

Iterative solvers such as proximal (sub)gradient descent

$$\mathbf{x}_{k+1} = \mathrm{prox}_{\tau_k D}\Big(\mathbf{x}_k - \tau_k \sum_{c=1}^{N_c} \mathbf{W}_c{}^\top \nabla\psi_c(\mathbf{W}_c\mathbf{x})\Big)$$

using $\mathrm{prox}_{\tau D}(\mathbf{y}) = \mathrm{argmin}_{\mathbf{x}} \left\{ \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_2^2 + \frac{\tau\lambda}{2}\|\mathbf{H}\mathbf{x}-\mathbf{y}\|_2^2 \right\}$

Deep (implicit) Regularization

- Total Deep Variation[4] regularization

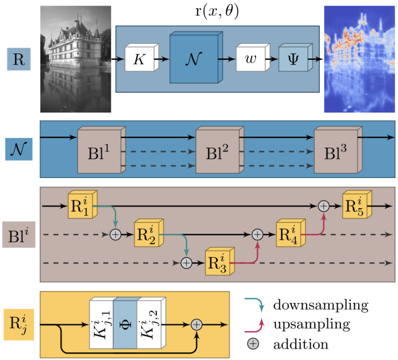
$$R(\mathbf{x}) = \langle \mathbf{1}, \psi(\mathbf{W}_{\theta}(\mathbf{x})) \rangle$$

[4] Kobler et al. (2020) *Total Deep Variation for Linear Inverse Problems*. CVPR.

Deep (implicit) Regularization

- Total Deep Variation[4] regularization

$$R(\mathbf{x}) = \langle \mathbf{1}, \psi(\mathbf{W}_{\theta}(\mathbf{x})) \rangle$$



[4] Kobler et al. (2020) *Total Deep Variation for Linear Inverse Problems*. CVPR.

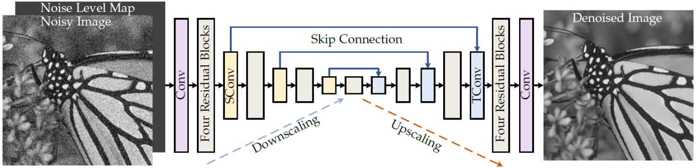
Deep (implicit) Regularization

- Total Deep Variation[5] regularization

$$R(\mathbf{x}) = \langle \mathbf{1}, \phi(\mathbf{W}_\theta(\mathbf{x})) \rangle$$

- DRUNet[6] plug & play regularization

$$\operatorname{prox}_R^\sigma(\mathbf{y}) = \underset{\mathbf{x}}{\operatorname{argmin}} \left\{ \frac{1}{2\sigma^2} \|\mathbf{x} - \mathbf{y}\|_2^2 + R(\mathbf{x}) \right\} \approx \mathbf{N}_\theta(\mathbf{y}, \sigma)$$



[5] Kobler et al. (2020) *Total Deep Variation for Linear Inverse Problems*. CVPR.
[6] Zhang et al. (2021) *Plug-and-Play Image Restoration with Deep Denoiser Prior*. TPAMI.

Deep (implicit) Regularization

- Total Deep Variation[7] regularization

$$R(\mathbf{x}) = \langle \mathbf{1}, \phi(\mathbf{W}_\theta(\mathbf{x})) \rangle$$

- DRUNet[8] plug & play regularization

$$\operatorname{prox}_R^\sigma(\mathbf{y}) = \operatorname*{argmin}_{\mathbf{x}} \left\{ \frac{1}{2\sigma^2} \|\mathbf{x} - \mathbf{y}\|_2^2 + R(\mathbf{x}) \right\} \approx \mathbf{N}_\theta(\mathbf{y}, \sigma)$$

- Prox-DRUNet[9] gradient step denoiser as proximal operator

$$\mathbf{D}_\sigma = \mathrm{Id} - \nabla g_\theta(\cdot, \sigma) \overset{\nabla g \text{ contractive}}{=} \operatorname{prox}_{\phi_\sigma}$$

[7] Kobler et al. (2020) *Total Deep Variation for Linear Inverse Problems*. CVPR.

[8] Zhang et al. (2021) *Plug-and-Play Image Restoration with Deep Denoiser Prior*. TPAMI.

[9] Hurault et al. (2022) *Proximal Denoiser for Convergent Plug-and-Play Optimization with Nonconvex Regularization*. ICML.

FoE regularization

$$\sum_{c=1}^{N_c} \langle \mathbf{1}, \psi_c(\mathbf{W}_c \mathbf{x}) \rangle$$

[9] Pourya, Kobler, Unser, Neumayer (2025) *DEALing with Image Reconstruction: Deep Attentive Least Squares*. ICML.

Deep Attentive Least Squares (DEAL)

FoE regularization

$$\sum_{c=1}^{N_c} \langle \mathbf{1}, \psi_c(\mathbf{W}_c\mathbf{x}) \rangle$$

$$\Downarrow$$

$$\sum_{c=1}^{N_c} \left\langle \mathbf{m}_c{}^2, (\mathbf{W}_c\mathbf{x})^2 \right\rangle$$

[9] Pourya, Kobler, Unser, Neumayer (2025) *DEALing with Image Reconstruction: Deep Attentive Least Squares*. ICML.

Deep Attentive Least Squares (DEAL)

FoE regularization

$$\sum_{c=1}^{N_c} \langle \mathbf{1}, \psi_c(\mathbf{W}_c\mathbf{x}) \rangle$$

$$\Downarrow$$

$$\sum_{c=1}^{N_c} \left\langle {\mathbf{m}_c}^2, (\mathbf{W}_c\mathbf{x})^2 \right\rangle \quad = \quad \|\mathbf{W}\mathbf{x}\|^2_{\mathbf{m}^2} =: R(\mathbf{x})$$

DEAL regularization

[9] Pourya, Kobler, Unser, Neumayer (2025) *DEALing with Image Reconstruction: Deep Attentive Least Squares*. ICML.

Deep Attentive Least Squares (DEAL)

Role of local weights m

Adapt regularization strength to local image structure.

Deep Attentive Least Squares (DEAL)

Role of local weights m

Adapt regularization strength to local image structure.

How to compute local weights m?

- Ideally, reflect structure of unkown ground truth.
- Idea: use current best estimate $\mathbf{x}_k$ to compute $\mathbf{m}_k = \mathbf{M}(\mathbf{x}_k)$ using CNN

$$\mathbf{M}(\mathbf{x}) = (\phi^{\sigma} \circ \mathbf{W}^2_{\mathrm{mix}} \circ \varphi_2 \circ \mathbf{W}^1_{\mathrm{mix}} \circ \varphi_1 \circ \mathbf{W}_{\mathrm{mask}})(\mathbf{x})$$

Deep Attentive Least Squares (DEAL)

Role of local weights m

Adapt regularization strength to local image structure.

How to compute local weights m?

- Ideally, reflect structure of unkown ground truth.
- Idea: use current best estimate $\mathbf{x}_k$ to compute $\mathbf{m}_k = \mathbf{M}(\mathbf{x}_k)$ using CNN

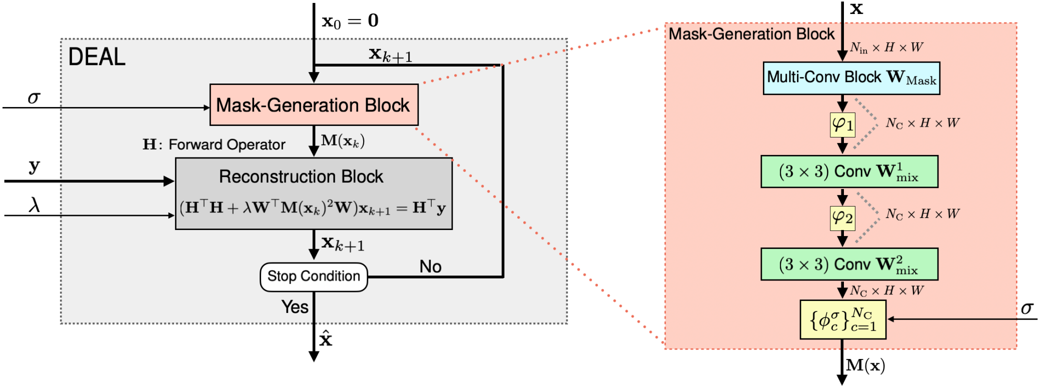
$$\mathbf{M}(\mathbf{x}) = (\phi^\sigma \circ \mathbf{W}^2_{\mathrm{mix}} \circ \varphi_2 \circ \mathbf{W}^1_{\mathrm{mix}} \circ \varphi_1 \circ \mathbf{W}_{\mathrm{mask}})(\mathbf{x})$$

Fixedpoint formulation $\mathbf{x}_{k+1} \in \mathcal{T}(\mathbf{x}_k, \mathbf{y}) \quad \text{with} \quad \mathbf{x}_0 = \mathbf{0}$

$$\mathcal{T}(\mathbf{x}_k, \mathbf{y}) = \operatorname*{argmin}_{\mathbf{x}} \left(\|\mathbf{Hx} - \mathbf{y}\|_2^2 + \lambda \, \|\mathbf{Wx}\|^2_{\mathbf{M}(\mathbf{x}_k)} \right)$$

$$= \left(\mathbf{H}^\top \mathbf{H} + \lambda \mathbf{W}^\top \mathbf{M}(\mathbf{x}_k)^2 \mathbf{W} \right)^{-1} \mathbf{H}^\top \mathbf{y} \iff \mathbf{A}_k \mathbf{x} = \mathbf{b}$$

DEAL Algorithm



DEAL – Training

On image denoising ($\mathbf{H} = \mathrm{Id}$)

$$\mathcal{L}(\theta) = \left\{ \mathbb{E}_{\substack{\mathbf{x}\sim \mathrm{D} \\ \sigma_{\mathrm{n}}\sim\mathcal{U}([0,50]) \\ \mathbf{n}\sim\mathcal{N}(\mathbf{0},\mathrm{Id})}} \left[\|\mathbf{x}_{K_{\mathrm{out}}} - \mathbf{x}\|_2^2 + \frac{\gamma}{N_{\mathrm{C}}} \|\mathbf{M}(\mathbf{x}_{K_{\mathrm{out}}}) - \mathbf{M}(\mathbf{x}_{K_{\mathrm{out}}-1})\|_2^2 \right] + \gamma \mathrm{TV}^2(\theta) \right\}$$

using $K_{\mathrm{out}} \in \{15, \ldots, 60\}$

DEAL – Training

On image denoising ($\mathbf{H} = \mathrm{Id}$)

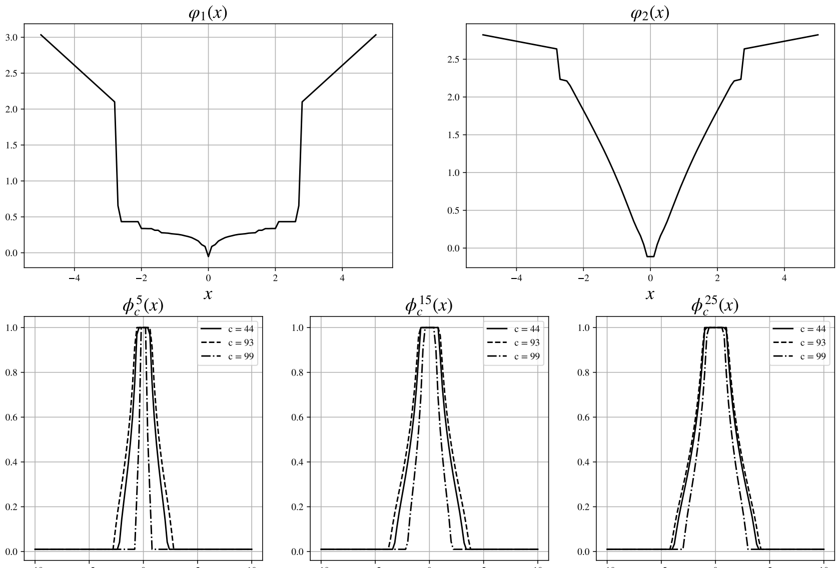
$$\mathcal{L}(\theta) = \left\{ \mathbb{E}_{\substack{\mathbf{x}\sim D \\ \sigma_{\mathrm{n}}\sim\mathcal{U}([0,50]) \\ \mathbf{n}\sim\mathcal{N}(\mathbf{0},\mathrm{Id})}} \left[\|\mathbf{x}_{K_{\mathrm{out}}} - \mathbf{x}\|_2^2 + \frac{\gamma}{N_{\mathrm{C}}} \|\mathbf{M}(\mathbf{x}_{K_{\mathrm{out}}}) - \mathbf{M}(\mathbf{x}_{K_{\mathrm{out}}-1})\|_2^2 \right] + \gamma \mathrm{TV}^2(\theta) \right\}$$

using $K_{\mathrm{out}} \in \{15, \ldots, 60\}$

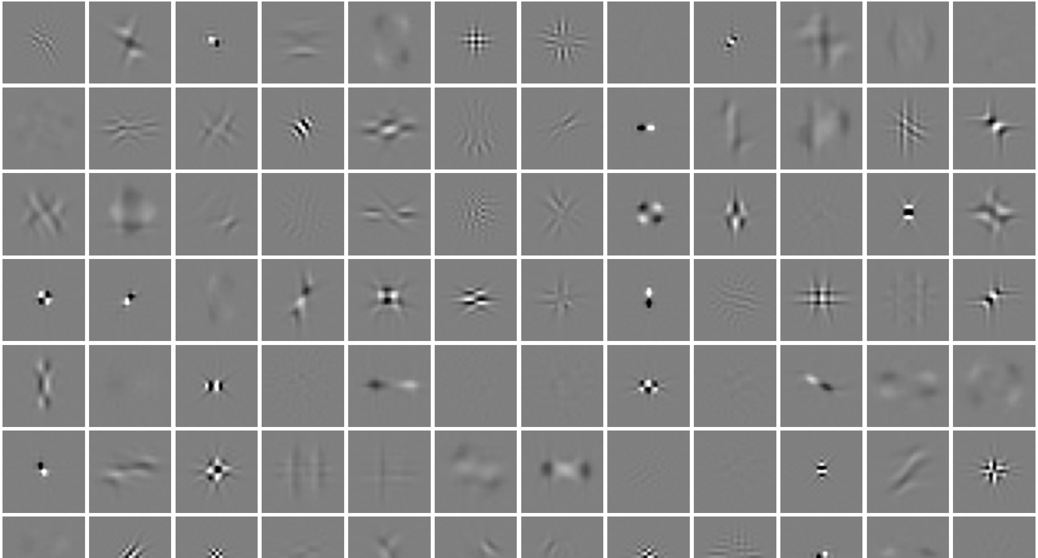
Backpropagation though CG by implicit differentiation

$$\mathbf{A}_k \partial_\theta \mathbf{x}_{k+1} = -\lambda \partial_\theta \left(\mathbf{W}^\top \mathbf{M}(\mathbf{x}_k) \mathbf{W} \right) \mathbf{x}_{k+1}$$

Learned Functions

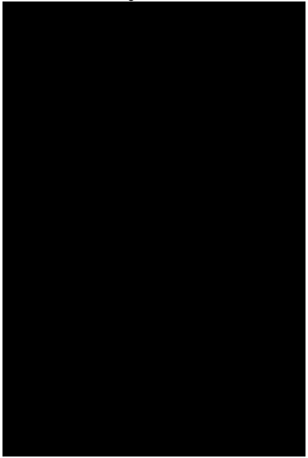


Learned Kernels

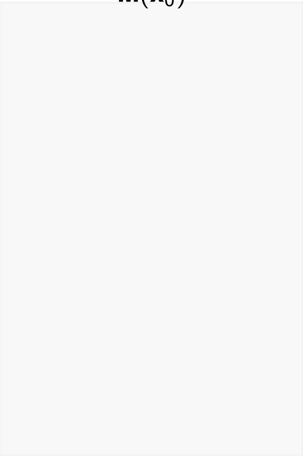


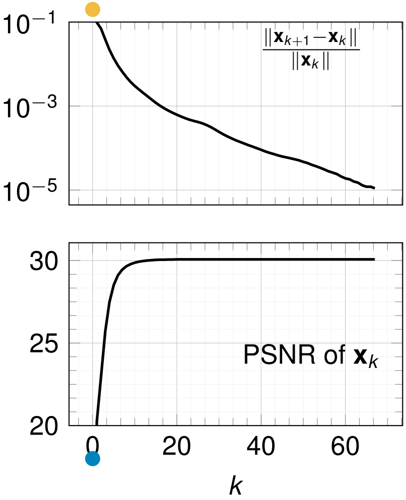
Results Image Denoising

$\mathbf{x}_0$ 6.12



$\overline{\mathbf{M}}(\mathbf{x}_0)$

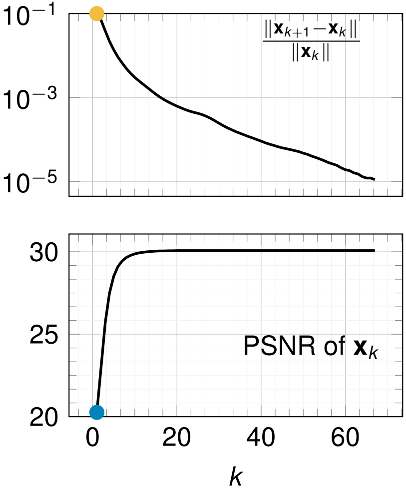




Results Image Denoising

$\mathbf{x}_1$ 20.26

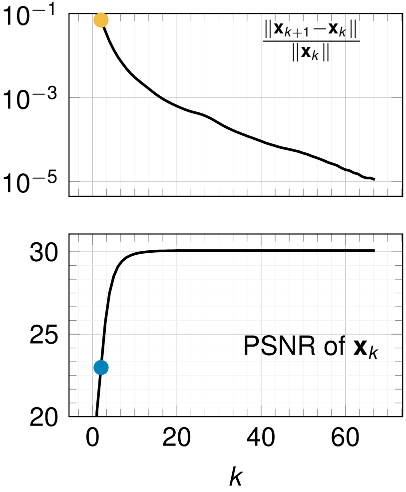
$\overline{\mathbf{M}}(\mathbf{x}_1)$



Results Image Denoising

$\mathbf{x}_2$ 22.98

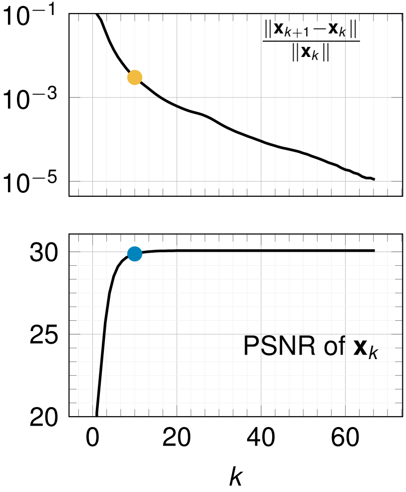
$\overline{\mathbf{M}}(\mathbf{x}_2)$



Results Image Denoising

$\mathbf{x}_{10}$ 29.87

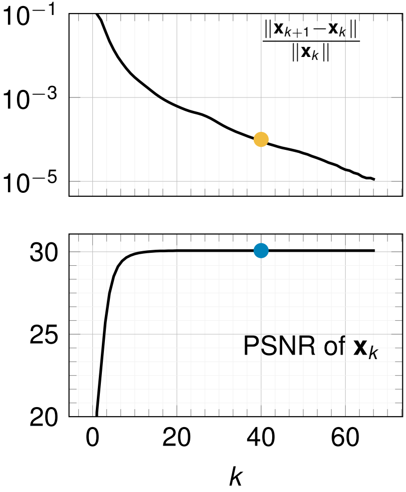
$\overline{\mathbf{M}}(\mathbf{x}_{10})$



Results Image Denoising

$\mathbf{x}_{40}$ 30.07

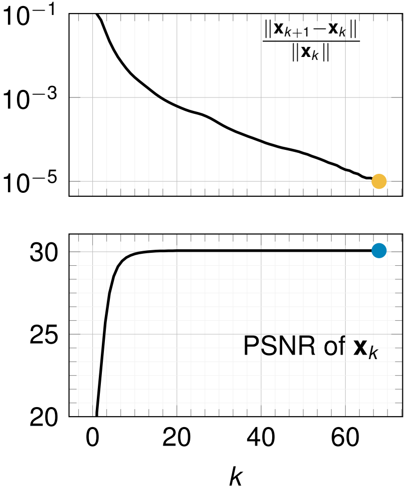
$\overline{\mathbf{M}}(\mathbf{x}_{40})$



Results Image Denoising

$\mathbf{x}_{68}$ 30.07

$\overline{\mathbf{M}}(\mathbf{x}_{68})$



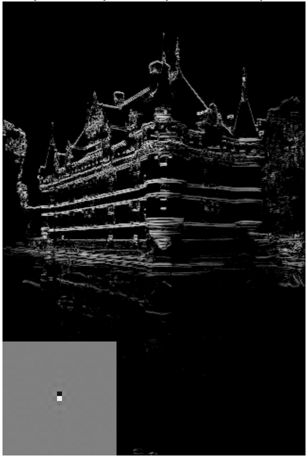
Denoising of BSD68 and CBSD68 images

	Gray			Color		
σ_{n}	5	15	25	5	15	25
BM3D	37.54	31.13	28.61	40.19	33.52	30.71
WCRR	37.65	31.20	28.68	–	–	–
SAFI	37.90	31.56	29.05	–	–	–
DEAL (Ours)	37.85	31.61	29.16	40.33	<u>33.95</u>	<u>31.31</u>
ProxDRUNet	<u>37.97</u>	31.70	29.18	<u>40.40</u>	33.91	31.14
DnCNN	–	<u>31.72</u>	<u>29.23</u>	-	33.90	31.24
DRUNet	**38.09**	**31.94**	**29.48**	**40.59**	**34.30**	**31.69**

DEAL – Effect of Local Weights

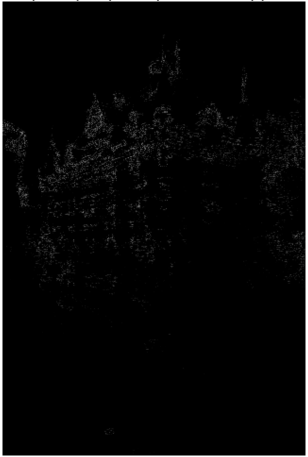
Filter channel $c = 93$

$(\mathbf{W}_c \mathbf{x}_k)^2 = (\mathbf{W}_c * \mathbf{x}_k)^2$



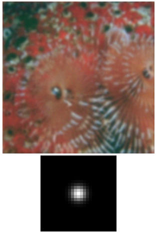
$\mathbf{M}_c(\mathbf{x}_k)$

$(\mathbf{M}_c(\mathbf{x}_k) \odot (\mathbf{W}_c * \mathbf{x}_k))^2$



DEAL - Single-Image Superresolution

$$\mathcal{T}(\mathbf{x}_k, \mathbf{y}) = \operatorname*{argmin}_{\mathbf{x}} \left(\|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{W}\mathbf{x}\|^2_{\mathbf{M}(\mathbf{x}_k)} \right)$$



$$H : \mathbf{x} \mapsto (\mathbf{x} * \mathbf{a})_{\downarrow s}$$

$$\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathrm{Id})$$

Results Single-Image Superresolution

SISR results on CBSD68 dataset

	Category	$\#\theta$	PSNR (s = 2)				PSNR (s=3)			
			<u>bicubic</u>	<u>4 different kernels</u>			<u>bicubic</u>	<u>4 different kernels</u>		
Noise σ_{n}			0	2.55	7.65	12.75	0	2.55	7.65	12.75
DEAL	Explicit Reg	<u>0.85</u>	<u>29.91</u>	**27.99**	<u>26.58</u>	25.75	<u>26.83</u>	**26.20**	<u>25.27</u>	24.59
Prox-PnP	Conv. PnP	32.64	-	<u>27.93</u>	**26.61**	<u>25.79</u>	-	<u>26.13</u>	**25.29**	**24.67**
IRCNN	PnP	**0.19**	29.84	26.97	25.86	25.45	26.74	25.60	25.72	24.38
DPIR	PnP	32.64	29.63	27.79	<u>26.58</u>	**25.83**	26.70	26.05	<u>25.27</u>	<u>24.66</u>
DiffPIR [4]	Diff. Model	93.56	29.73	27.84	26.48	25.63	-	-	-	-
SwinIR [3]	End to End	11.75	**30.88**	24.56	22.84	20.73	**27.76**	22.41	21.24	19.53

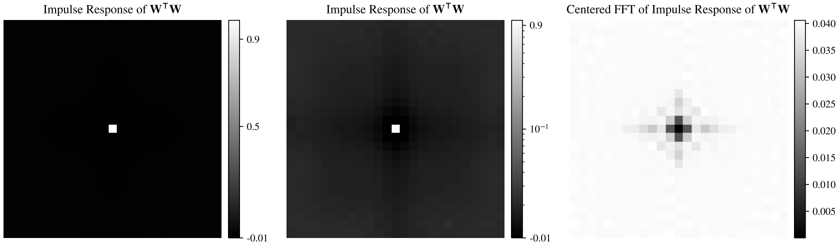
Theorem: Fixedpoint Existence
Assume that $\ker(\mathbf{H}) \cap \ker(\mathbf{W}) = \{\mathbf{0}\}$ and $\mathbf{M}^2(\mathbf{x}) \succeq \epsilon_M \mathrm{Id}$. Then, $\mathcal{T}(\cdot, \mathbf{y})\colon \mathbb{R}^d \to B_r(\mathbf{0})$ maps into a ball around $\mathbf{0}$ with radius $r = \|\mathbf{Hy}\|_2 / \lambda_\epsilon$, where $\lambda_\epsilon = \lambda_{\min}(\mathbf{A}_k)$. If $\mathbf{M}^2\colon B_r(\mathbf{0}) \to [\epsilon, 1]^{N_{\mathrm{out}}HW}$ is Lipschitz continuous with constant L, then $\mathcal{T}(\cdot, \mathbf{y})$ admits a fixed point and

$$\|\mathcal{T}(\mathbf{x}_1, \mathbf{y}) - \mathcal{T}(\mathbf{x}_2, \mathbf{y})\|_2 \leq \frac{L\,\|\mathbf{Hy}\|_2}{\lambda_\epsilon^2} \|\mathbf{x}_1 - \mathbf{x}_2\|_2.$$

DEAL – Analysis

Theorem: Fixedpoint Existence
Assume that $\ker(\mathbf{H}) \cap \ker(\mathbf{W}) = \{\mathbf{0}\}$ and $\mathbf{M}^2(\mathbf{x}) \succeq \epsilon_M \mathrm{Id}$. Then, $\mathcal{T}(\cdot, \mathbf{y})\colon \mathbb{R}^d \to B_r(\mathbf{0})$ maps into a ball around $\mathbf{0}$ with radius $r = \|\mathbf{Hy}\|_2 / \lambda_\epsilon$, where $\lambda_\epsilon = \lambda_{\min}(\mathbf{A}_k)$. If $\mathbf{M}^2\colon B_r(\mathbf{0}) \to [\epsilon, 1]^{N_{\text{out}}HW}$ is Lipschitz continuous with constant L, then $\mathcal{T}(\cdot, \mathbf{y})$ admits a fixed point and

$$\|\mathcal{T}(\mathbf{x}_1, \mathbf{y}) - \mathcal{T}(\mathbf{x}_2, \mathbf{y})\|_2 \leq \frac{L\,\|\mathbf{Hy}\|_2}{\lambda_\epsilon^2} \|\mathbf{x}_1 - \mathbf{x}_2\|_2.$$



Theorem: Unique Fixedpoint
Assume that $\ker(\mathbf{H}) \cap \ker(\mathbf{W}) = \{\mathbf{0}\}$ and $\mathbf{M}^2(\mathbf{x}) \succeq \epsilon_M \mathrm{Id}$. If $\mathcal{T}(\cdot, \mathbf{y})\colon \mathbb{R}^d \to \mathbb{R}^d$ is contractive in the sense that if $\|\mathcal{T}(\mathbf{x}_1, \mathbf{y}) - \mathcal{T}(\mathbf{x}_2, \mathbf{y})\|_2 \leq q\|\mathbf{x}_1 - \mathbf{x}_2\|_2$ with $q < 1$ implies that the iterations converge to a unique fixed point $\hat{\mathbf{x}}$ and that

$$\|\mathbf{x}_k - \hat{\mathbf{x}}\|_2 \leq q^{k-1}\|\mathbf{x}_1 - \mathbf{x}_0\|_2.$$

In particular, we have exponential convergence.

Theorem: Unique Fixedpoint
Assume that $\ker(\mathbf{H}) \cap \ker(\mathbf{W}) = \{\mathbf{0}\}$ and $\mathbf{M}^2(\mathbf{x}) \succeq \epsilon_M \mathrm{Id}$. If $\mathcal{T}(\cdot, \mathbf{y}) \colon \mathbb{R}^d \to \mathbb{R}^d$ is contractive in the sense that if $\|\mathcal{T}(\mathbf{x}_1, \mathbf{y}) - \mathcal{T}(\mathbf{x}_2, \mathbf{y})\|_2 \leq q\|\mathbf{x}_1 - \mathbf{x}_2\|_2$ with $q < 1$ implies that the iterations converge to a unique fixed point $\hat{\mathbf{x}}$ and that

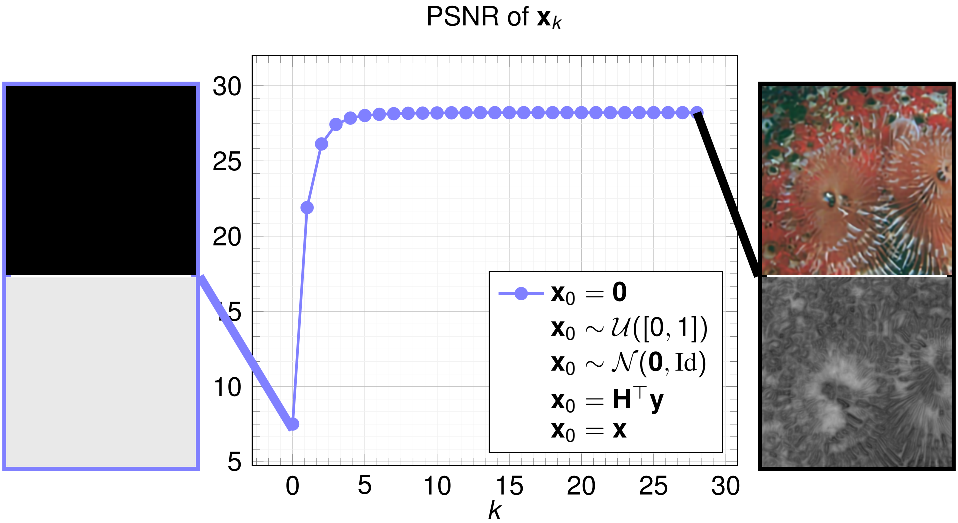
$$\|\mathbf{x}_k - \hat{\mathbf{x}}\|_2 \leq q^{k-1}\|\mathbf{x}_1 - \mathbf{x}_0\|_2.$$

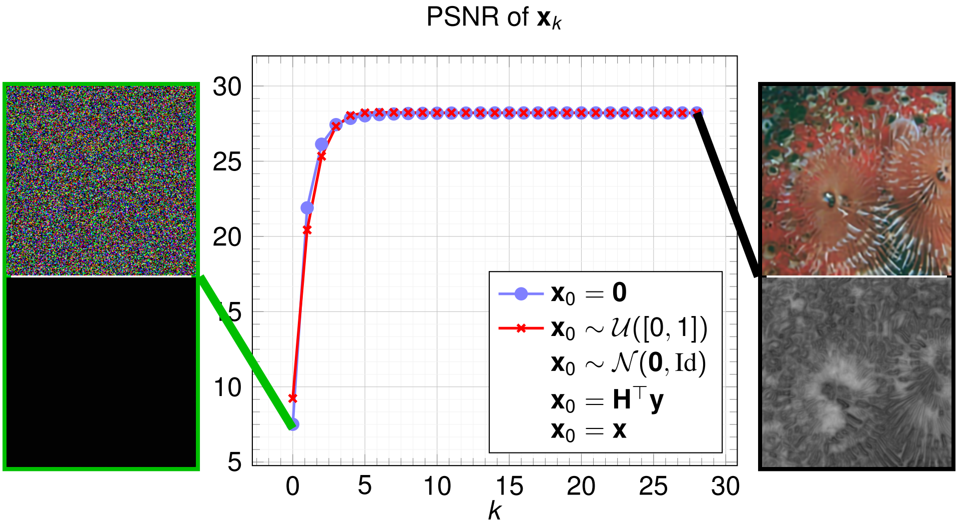
In particular, we have exponential convergence.

Moreover, if $\hat{\mathbf{x}} = \mathcal{T}(\hat{\mathbf{x}}, \mathbf{y}_1)$ and $\hat{\mathbf{z}} = \mathcal{T}(\hat{\mathbf{z}}, \mathbf{y}_2)$, then it holds that

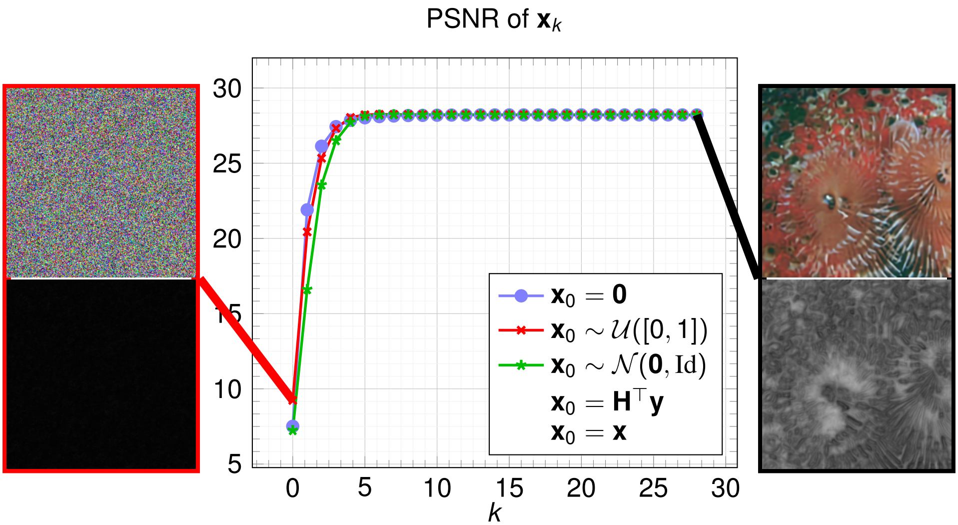
$$\|\hat{\mathbf{x}} - \hat{\mathbf{z}}\| \leq \frac{1}{1-q} \frac{\|\mathbf{H}\|_2}{\lambda_\epsilon} \|\mathbf{y}_1 - \mathbf{y}_2\|_2.$$

DEAL – Stability

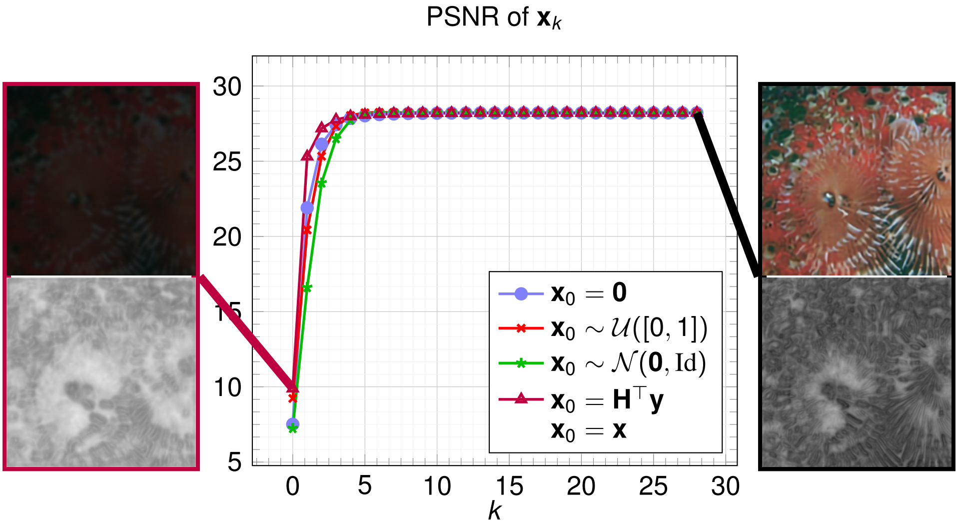



PSNR of $\mathbf{x}_k$
30
25
20
15
10
5
0
5
10
15
20
25
30
k
$\mathbf{x}_0 = \mathbf{0}$
$\mathbf{x}_0 \sim \mathcal{U}([0,1])$
$\mathbf{x}_0 \sim \mathcal{N}(\mathbf{0}, \mathrm{Id})$
$\mathbf{x}_0 = \mathbf{H}^\top \mathbf{y}$
$\mathbf{x}_0 = \mathbf{x}$

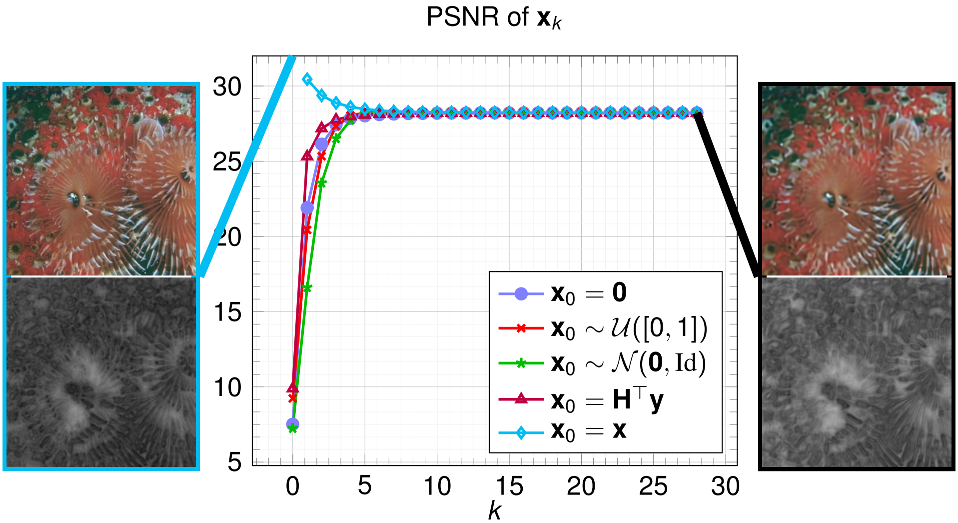
DEAL – Stability



DEAL – Stability

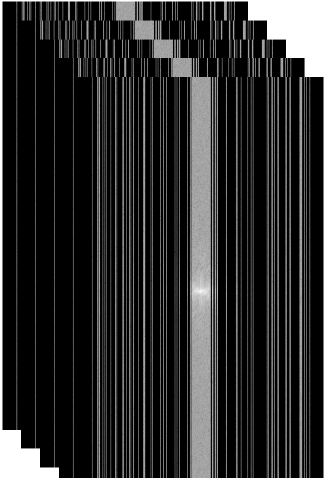


DEAL – Stability



DEAL - Magentic Resonance Imaging

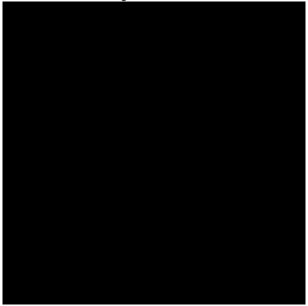
$$\mathcal{T}(\mathbf{x}_k, \mathbf{y}) = \operatorname*{argmin}_{\mathbf{x}} \left(\|\mathbf{Hx} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{Wx}\|^2_{\mathbf{M}(\mathbf{x}_k)} \right)$$



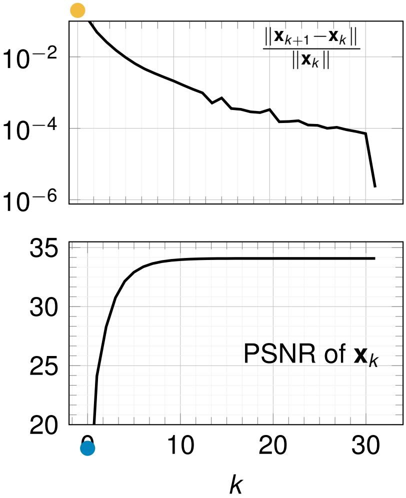
$$H_i : \mathbf{x} \mapsto \mathbf{m} \odot \mathcal{F}(\mathbf{c} \odot \mathbf{x})$$

$$\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathrm{Id})$$

$\mathbf{x}_0$ 10.56

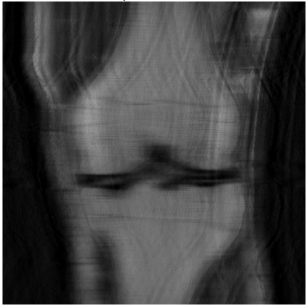


$\overline{\mathbf{M}}(\mathbf{x}_0)$

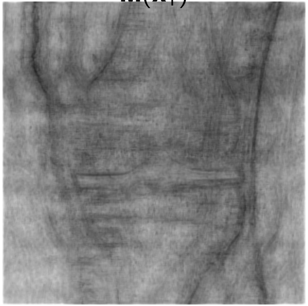


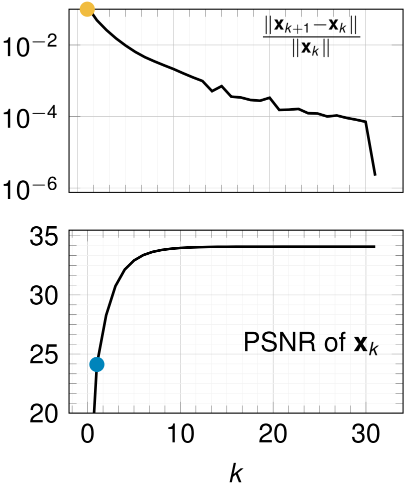
MRI Reconstruction

$\mathbf{x}_1$ 24.12



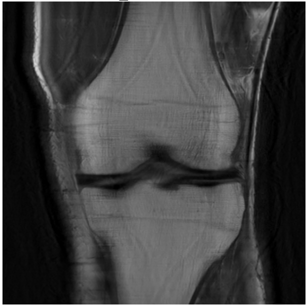
$\overline{\mathbf{M}}(\mathbf{x}_1)$



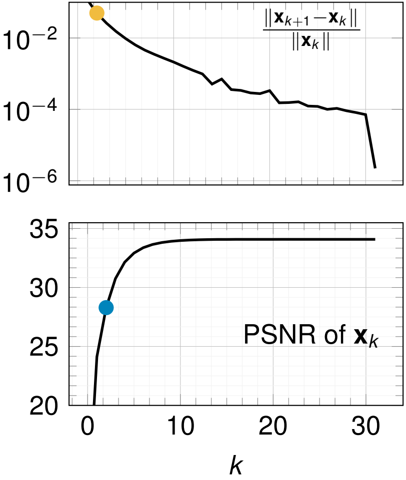


MRI Reconstruction

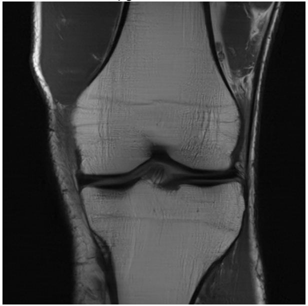
$\mathbf{x}_2$ 28.29



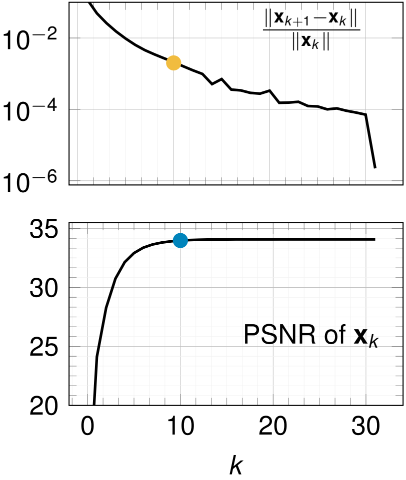
$\overline{\mathbf{M}}(\mathbf{x}_2)$



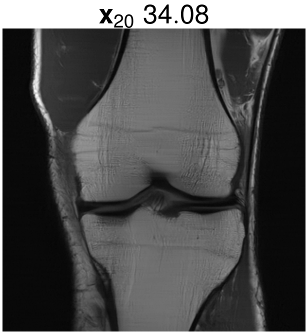
$\mathbf{x}_{10}$ 33.98

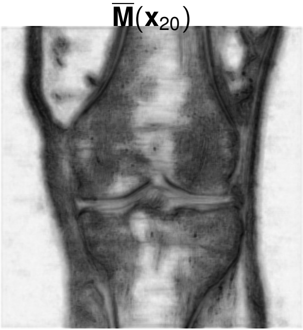


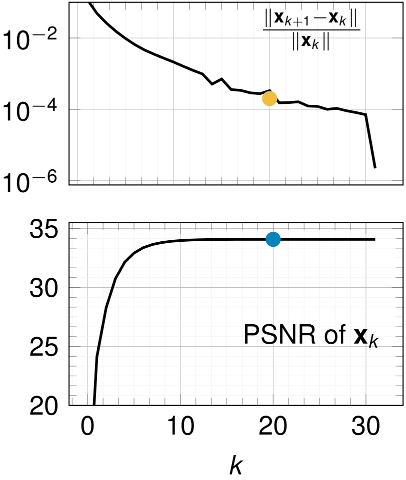
$\overline{\mathbf{M}}(\mathbf{x}_{10})$



MRI Reconstruction

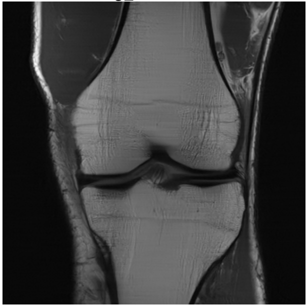




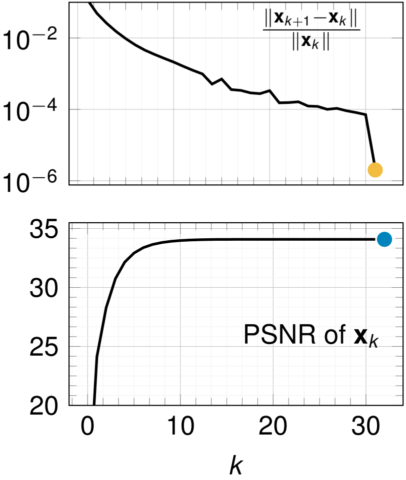


MRI Reconstruction

$\mathbf{x}_{32}$ 34.08



$\overline{\mathbf{M}}(\mathbf{x}_{32})$

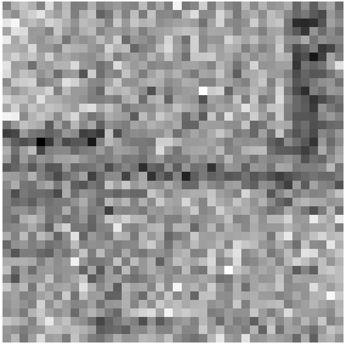


PSNR values for the MRI experiment

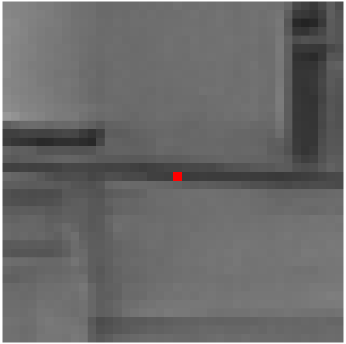
	4-fold single coil		8-fold multi-coil	
	PD	PDFS	PD	PDFS
Zero-fill ($\mathbf{H}^\top \mathbf{y}$)	27.40	29.68	23.80	27.19
TV	32.44	32.67	32.77	33.38
CRR	33.99	33.75	34.29	34.50
WCRR	35.78	34.63	35.57	35.16
SAFI	<u>36.43</u>	<u>34.92</u>	<u>36.06</u>	**35.36**
DEAL (Ours)	**36.59**	<u>34.92</u>	**36.21**	<u>35.32</u>
ProxDRUNet	36.20	**35.05**	35.82	35.12
PnP-DnCNN	35.24	34.63	35.11	35.14

DEAL – Interpretation

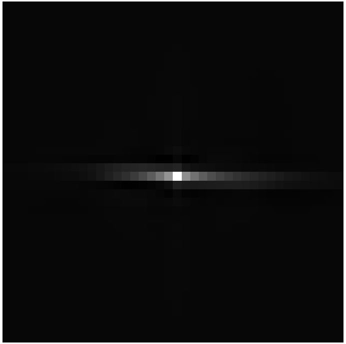
$\mathbf{y}$



$\mathbf{x}_k = \mathbf{A}_{k-1}^{-1}\mathbf{y}$

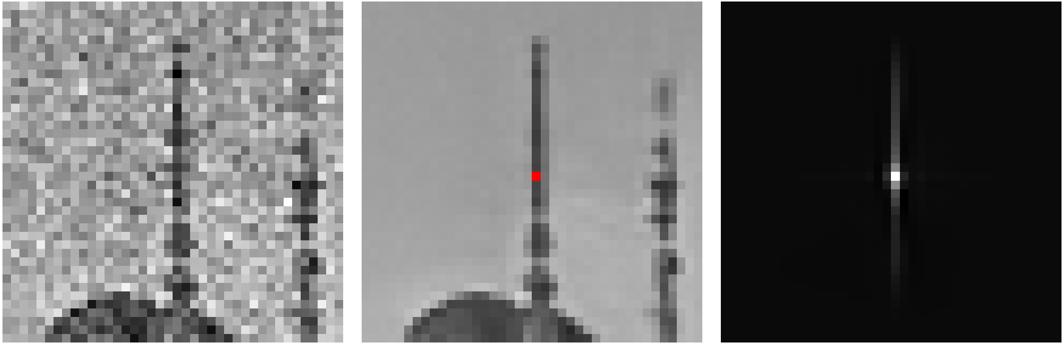


$\mathbf{A}_{k-1}^{-1}\mathbf{e}_n$



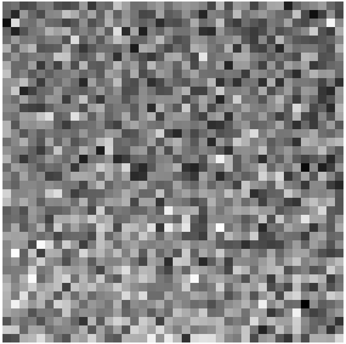
DEAL – Interpretation

$\mathbf{y}$ $\qquad$ $\mathbf{x}_k = \mathbf{A}_{k-1}^{-1}\mathbf{y}$ $\qquad$ $\mathbf{A}_{k-1}^{-1}\mathbf{e}_n$

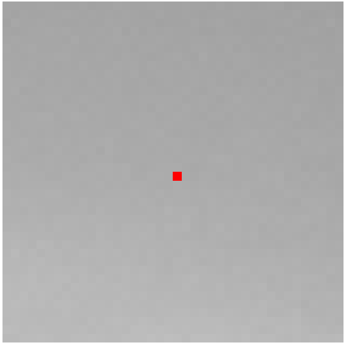


DEAL – Interpretation

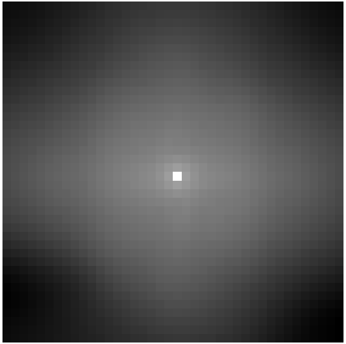
$\mathbf{y}$



$\mathbf{x}_k = \mathbf{A}_{k-1}^{-1}\mathbf{y}$



$\mathbf{A}_{k-1}^{-1}\mathbf{e}_n$



Thank you for your attention!