Jacobian-Free Unrolling: memory efficient unrolled network for 3D SPECT Reconstruction

Corentin CONSTANZA[1]

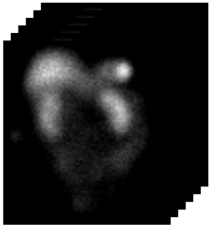
David SARRUT[1] Ane ETXEBESTE[1] Théo KAPRELIAN[1]

[1]INSA-Lyon, Université Claude Bernard Lyon 1, CNRS, Inserm, CREATIS UMR 5220, U1294, F-69373, Lyon, France.

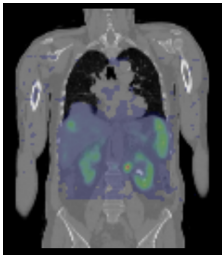
SPECT imaging

- We need accurate SPECT voxel-level dosimetry for RPT [Del Prete et al., 2018; Vergnaud, 2023].
- SPECT forward model:

$$y \sim \mathbf{P}(Ax + s) \tag{1}$$



(a) y: SPECT Projections

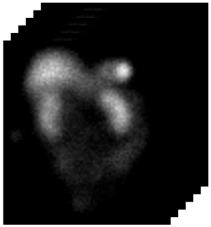


(b) x: SPECT Reconstruction

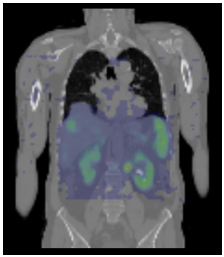
SPECT imaging

- We need accurate SPECT voxel-level dosimetry for RPT [Del Prete et al., 2018; Vergnaud, 2023].
- SPECT forward model:

$$y \sim \mathbf{P}(Ax + s) \tag{1}$$



(a) y: SPECT Projections



(b) x: SPECT Reconstruction

- SPECT degrading effects:
  - Partial volume effect
  - Scattered events
  - Poisson Noise

SPECT Reconstruction

- SPECT optimization problem:

$$x^* = \arg\max_{x>0} \mathcal{L}(x) + \mathcal{R}(x) \tag{2}$$

SPECT Reconstruction

- SPECT optimization problem:

$$x^* = \arg\max_{x>0} \mathcal{L}(x) + \mathcal{R}(x) \tag{2}$$

- Model Based Iterative Reconstruction (MBIR):
  - MLEM/OSEM [Shepp and Vardi, 1982; Hudson and Larkin, 1994]: $x^{[k+1]} = \frac{x^{[k]}}{A^T(\mathbf{1})} A^T \left(\frac{y}{Ax^{[k]}} \right)$

SPECT Reconstruction

- SPECT optimization problem:

$$x^* = \underset{x>0}{\arg\max}\ \mathcal{L}(x) + \mathcal{R}(x) \tag{2}$$

- Model Based Iterative Reconstruction (MBIR):
  - MLEM/OSEM [Shepp and Vardi, 1982; Hudson and Larkin, 1994]: $x^{[k+1]} = \frac{x^{[k]}}{A^T(\mathbf{1})} A^T\left(\frac{y}{Ax^{[k]}}\right)$
  - MAPEM-OSL [Green, 1990; Panin et al., 1998]: $x^{[k+1]} = \frac{x^{[k]}}{A^T(\mathbf{1}) + \nabla\mathcal{R}(x^{[k]})} A^T\left(\frac{y}{Ax^{[k]}}\right)$

SPECT Reconstruction

- SPECT optimization problem:

$$x^* = \underset{x>0}{\arg\max}\ \mathcal{L}(x) + \mathcal{R}(x) \tag{2}$$

- Model Based Iterative Reconstruction (MBIR):
  - MLEM/OSEM [Shepp and Vardi, 1982; Hudson and Larkin, 1994]: $x^{[k+1]} = \frac{x^{[k]}}{A^T(\mathbf{1})} A^T\left(\frac{y}{Ax^{[k]}}\right)$
  - MAPEM-OSL [Green, 1990; Panin et al., 1998]: $x^{[k+1]} = \frac{x^{[k]}}{A^T(\mathbf{1}) + \nabla\mathcal{R}(x^{[k]})} A^T\left(\frac{y}{Ax^{[k]}}\right)$
  - BSREM [Ahn and Fessler, 2003]: $x^{[k+1]} = x^{[k]} + \alpha_k \frac{x^{[k]}}{A^T\mathbf{1}}(\nabla\mathcal{L}(x^{[k]}) - \nabla\mathcal{R}(x^{[k]}))$

SPECT Reconstruction

- SPECT optimization problem:

$$x^* = \arg\max_{x>0} \mathcal{L}(x) + \mathcal{R}(x) \tag{2}$$

- Model Based Iterative Reconstruction (MBIR):
  - MLEM/OSEM [Shepp and Vardi, 1982; Hudson and Larkin, 1994]: $x^{[k+1]} = \frac{x^{[k]}}{A^T(\mathbf{1})} A^T\left(\frac{y}{Ax^{[k]}}\right)$
  - MAPEM-OSL [Green, 1990; Panin et al., 1998]: $x^{[k+1]} = \frac{x^{[k]}}{A^T(\mathbf{1}) + \nabla\mathcal{R}(x^{[k]})} A^T\left(\frac{y}{Ax^{[k]}}\right)$
  - BSREM [Ahn and Fessler, 2003]: $x^{[k+1]} = x^{[k]} + \alpha_k \frac{x^{[k]}}{A^T\mathbf{1}}(\nabla\mathcal{L}(x^{[k]}) - \nabla\mathcal{R}(x^{[k]}))$
  - Classical $\mathcal{R}(x)$: TV [Panin et al., 1998], RDP [Ahn et al., 2015; Kangasmaa et al., 2021], ...

SPECT Reconstruction

- SPECT optimization problem:

$$x^* = \underset{x>0}{\arg\max}\ \mathcal{L}(x) + \mathcal{R}(x) \tag{2}$$

- Model Based Iterative Reconstruction (MBIR):
  - MLEM/OSEM [Shepp and Vardi, 1982; Hudson and Larkin, 1994]: $x^{[k+1]} = \frac{x^{[k]}}{A^T(\mathbf{1})} A^T\left(\frac{y}{Ax^{[k]}}\right)$
  - MAPEM-OSL [Green, 1990; Panin et al., 1998]: $x^{[k+1]} = \frac{x^{[k]}}{A^T(\mathbf{1}) + \nabla\mathcal{R}(x^{[k]})} A^T\left(\frac{y}{Ax^{[k]}}\right)$
  - BSREM [Ahn and Fessler, 2003]: $x^{[k+1]} = x^{[k]} + \alpha_k \frac{x^{[k]}}{A^T\mathbf{1}}(\nabla\mathcal{L}(x^{[k]}) - \nabla\mathcal{R}(x^{[k]}))$
  - Classical $\mathcal{R}(x)$: TV [Panin et al., 1998], RDP [Ahn et al., 2015; Kangasmaa et al., 2021], ...
- Deep learning
  - Post/pre-processing methods [Pan et al., 2022; Kaprélian et al., 2024; Leube et al., 2024]
  - **Black-box**, **physics-agnostic**, and **require large training database**

SPECT Reconstruction

- SPECT optimization problem:

$$x^* = \underset{x>0}{\arg\max}\ \mathcal{L}(x) + \mathcal{R}(x) \tag{2}$$

- Model Based Iterative Reconstruction (MBIR):
  - MLEM/OSEM [Shepp and Vardi, 1982; Hudson and Larkin, 1994]: $x^{[k+1]} = \frac{x^{[k]}}{A^T(\mathbf{1})} A^T\left(\frac{y}{Ax^{[k]}}\right)$
  - MAPEM-OSL [Green, 1990; Panin et al., 1998]: $x^{[k+1]} = \frac{x^{[k]}}{A^T(\mathbf{1}) + \nabla\mathcal{R}(x^{[k]})} A^T\left(\frac{y}{Ax^{[k]}}\right)$
  - BSREM [Ahn and Fessler, 2003]: $x^{[k+1]} = x^{[k]} + \alpha_k \frac{x^{[k]}}{A^T\mathbf{1}}(\nabla\mathcal{L}(x^{[k]}) - \nabla\mathcal{R}(x^{[k]}))$
  - Classical $\mathcal{R}(x)$: TV [Panin et al., 1998], RDP [Ahn et al., 2015; Kangasmaa et al., 2021], ...
- Deep learning
  - Post/pre-processing methods [Pan et al., 2022; Kaprélian et al., 2024; Leube et al., 2024]
  - **Black-box**, **physics-agnostic**, and **require large training database**
- Unrolling [Adler and Öktem, 2017; Monga et al., 2021; Li et al., 2023]

Unrolled Algorithm

- **Core idea:** transform a fixed number of optimization iterations into a network architecture, with a learnable component (prior, psf, ...).

Unrolled Algorithm

- **Core idea:** transform a fixed number of optimization iterations into a network architecture, with a learnable component (prior, psf, ...).
- **Bi-level formulation:**

$$\begin{cases} \theta^* \in \arg\min_{\theta} \underbrace{L(\hat{x}_\theta)}_{\text{Loss}} & \textbf{(Outer Problem)} \\ \hat{x}_\theta \in \arg\max_{x} \big(\mathcal{L}(x) + \mathcal{R}_\theta(x)\big) & \textbf{(Inner Problem)} \end{cases} \tag{2}$$

Unrolled Algorithm

- **Core idea:** transform a fixed number of optimization iterations into a network architecture, with a learnable component (prior, psf, ...).
- **Bi-level formulation:**

$$\begin{cases} \theta^* \in \arg\min_{\theta} \underbrace{L(\hat{x}_\theta)}_{\text{Loss}} & \textbf{(Outer Problem)} \\ \hat{x}_\theta \in \arg\max_{x} \big(\mathcal{L}(x) + \mathcal{R}_\theta(x)\big) & \textbf{(Inner Problem)} \end{cases} \tag{2}$$

$$\varphi^K(x^{[0]}, \theta) = \underbrace{\varphi(\cdot, \theta) \circ \cdots \circ \varphi(x^{[0]}, \theta)}_{\text{K times}} \tag{3}$$

Unrolled Algorithm

- **Core idea:** transform a fixed number of optimization iterations into a network architecture, with a learnable component (prior, psf, ...).
- **Bi-level formulation:**

$$\begin{cases} \theta^* \in \arg\min_{\theta} \underbrace{L(\hat{x}_\theta)}_{\text{Loss}} & \textbf{(Outer Problem)} \\ \hat{x}_\theta \in \arg\max_{x} \big(\mathcal{L}(x) + \mathcal{R}_\theta(x)\big) & \textbf{(Inner Problem)} \end{cases} \tag{2}$$

$$\varphi^K(x^{[0]}, \theta) = \underbrace{\varphi(\cdot, \theta) \circ \cdots \circ \varphi(x^{[0]}, \theta)}_{\text{K times}} \tag{3}$$

- **For MAPEM-OSL :**

$$\varphi(x^{[k]}, \theta) = \frac{x^{[k]}}{A^T(\mathbf{1}) + \nabla\mathcal{R}_\theta(x^{[k]})} A^T\left(\frac{y}{Ax^{[k]}}\right) \tag{4}$$

With $\nabla\mathcal{R}_\theta$ a CNN.

Unrolled Algorithm

- **Core idea:** transform a fixed number of optimization iterations into a network architecture, with a learnable component (prior, psf, ...).
- **Bi-level formulation:**

$$\begin{cases} \theta^* \in \arg\min_{\theta} \underbrace{L(\hat{x}_\theta)}_{\text{Loss}} & \textbf{(Outer Problem)} \\ \hat{x}_\theta \in \arg\max_{x} \big(\mathcal{L}(x) + \mathcal{R}_\theta(x)\big) & \textbf{(Inner Problem)} \end{cases} \tag{2}$$

$$\varphi^K(x^{[0]}, \theta) = \underbrace{\varphi(\cdot, \theta) \circ \cdots \circ \varphi(x^{[0]}, \theta)}_{\text{K times}} \tag{3}$$

- **For MAPEM-OSL :**

$$\varphi(x^{[k]}, \theta) = \frac{x^{[k]}}{A^T(\mathbf{1}) + \nabla\mathcal{R}_\theta(x^{[k]})} A^T\left(\frac{y}{Ax^{[k]}}\right) \tag{4}$$

With $\nabla\mathcal{R}_\theta$ a CNN.

- Advantages
  - Small training dataset
  - Interpretable
  - Physics knowledge

Unrolled Algorithm

- **Core idea:** transform a fixed number of optimization iterations into a network architecture, with a learnable component (prior, psf, ...).
- **Bi-level formulation:**

$$\begin{cases} \theta^* \in \arg\min_{\theta} \underbrace{L(\hat{x}_\theta)}_{\text{Loss}} & \textbf{(Outer Problem)} \\ \hat{x}_\theta \in \arg\max_{x} \big(\mathcal{L}(x) + \mathcal{R}_\theta(x)\big) & \textbf{(Inner Problem)} \end{cases} \tag{2}$$

$$\varphi^K(x^{[0]}, \theta) = \underbrace{\varphi(\cdot, \theta) \circ \cdots \circ \varphi(x^{[0]}, \theta)}_{\text{K times}} \tag{3}$$

- **For MAPEM-OSL :**

$$\varphi(x^{[k]}, \theta) = \frac{x^{[k]}}{A^T(\mathbf{1}) + \nabla\mathcal{R}_\theta(x^{[k]})} A^T\left(\frac{y}{Ax^{[k]}}\right) \tag{4}$$

With $\nabla\mathcal{R}_\theta$ a CNN.

- Advantages
  - Small training dataset
  - Interpretable
  - Physics knowledge
- Drawbacks
  - Memory consumption
  - Computation time

Deep Equilibrium models (DEQ)

- Deep Equilibrium (DEQ) [Bai et al., 2019] is a bi-level framework in the specific case where $\hat{x}_\theta$ is the fixed point of an operator $\varphi(\cdot, \theta)$:

$$\hat{x}_\theta = \varphi(\hat{x}_\theta, \theta) \tag{5}$$

Deep Equilibrium models (DEQ)

- Deep Equilibrium (DEQ) [Bai et al., 2019] is a bi-level framework in the specific case where $\hat{x}_\theta$ is the fixed point of an operator $\varphi(\cdot, \theta)$:

$$\hat{x}_\theta = \varphi(\hat{x}_\theta, \theta) \tag{5}$$

- DEQ also require large amount of memory to train due to Jacobian computation and inversion. To overcome this limitation, [Fung et al., 2022] introduce Jacobian-Free Backpropagation (JFB).

Deep Equilibrium models (DEQ)

- Deep Equilibrium (DEQ) [Bai et al., 2019] is a bi-level framework in the specific case where $\hat{x}_\theta$ is the fixed point of an operator $\varphi(\cdot, \theta)$:

$$\hat{x}_\theta = \varphi(\hat{x}_\theta, \theta) \tag{5}$$

- DEQ also require large amount of memory to train due to Jacobian computation and inversion. To overcome this limitation, [Fung et al., 2022] introduce Jacobian-Free Backpropagation (JFB).
- **Link between JFB and Unrolled network** [Davy et al., 2025]: To solve the **(Outer Problem)** of our bi-level problem, we use gradient decent algorithm :

$$\theta^{[m+1]} = \theta^{[m]} - \tau g(\theta^{[m]}) \tag{6}$$

Deep Equilibrium models (DEQ)

- Deep Equilibrium (DEQ) [Bai et al., 2019] is a bi-level framework in the specific case where $\hat{x}_\theta$ is the fixed point of an operator $\varphi(\cdot, \theta)$:

$$\hat{x}_\theta = \varphi(\hat{x}_\theta, \theta) \tag{5}$$

- DEQ also require large amount of memory to train due to Jacobian computation and inversion. To overcome this limitation, [Fung et al., 2022] introduce Jacobian-Free Backpropagation (JFB).
- **Link between JFB and Unrolled network** [Davy et al., 2025]: To solve the **(Outer Problem)** of our bi-level problem, we use gradient decent algorithm :

$$\theta^{[m+1]} = \theta^{[m]} - \tau g(\theta^{[m]}) \tag{6}$$

$$\begin{aligned} g^{\mathrm{JFB}}(\theta) &= \partial_x L(\hat{x}_\theta)\, \partial_\theta \varphi(\hat{x}_\theta, \bullet)(\theta) \\ &= \partial_x L(\varphi(\hat{x}_\theta, \theta))\, \partial_\theta \varphi(\hat{x}_\theta, \bullet)(\theta) \end{aligned} \tag{7}$$

Deep Equilibrium models (DEQ)

- Deep Equilibrium (DEQ) [Bai et al., 2019] is a bi-level framework in the specific case where $\hat{x}_\theta$ is the fixed point of an operator $\varphi(\cdot, \theta)$:

$$\hat{x}_\theta = \varphi(\hat{x}_\theta, \theta) \tag{5}$$

- DEQ also require large amount of memory to train due to Jacobian computation and inversion. To overcome this limitation, [Fung et al., 2022] introduce Jacobian-Free Backpropagation (JFB).
- **Link between JFB and Unrolled network** [Davy et al., 2025]: To solve the **(Outer Problem)** of our bi-level problem, we use gradient decent algorithm :

$$\theta^{[m+1]} = \theta^{[m]} - \tau g(\theta^{[m]}) \tag{6}$$

$$\begin{aligned} g^{\mathrm{JFB}}(\theta) &= \partial_x L(\hat{x}_\theta)\, \partial_\theta \varphi(\hat{x}_\theta, \bullet)(\theta) \\ &= \partial_x L(\varphi(\hat{x}_\theta, \theta))\, \partial_\theta \varphi(\hat{x}_\theta, \bullet)(\theta) \end{aligned} \tag{7}$$

- Suppose that $\tilde{x}_{K-1} = \varphi^{K-1}(x^{[0]}, \theta) \approx \hat{x}_\theta$, then

$$g^{\mathrm{JFB}}(\theta) \simeq \partial_x L(\varphi(\tilde{x}_{K-1}, \theta))\, \partial_\theta \varphi(\tilde{x}_{K-1}, \bullet)(\theta) \tag{8}$$

$\Rightarrow$ **Equivalent to backpropagate through the last iteration only**

Jacobian-Free Unrolling

	Classical Unrolling	**Jacobian-Free Unrolling (JFU)**
Max itration (K)	$\sim 6-8$	∞
Parameters	$\theta = \{\theta_1, \dots, \theta_k\}$	Forced to be tight
Memory	Scale with K	Constant
Training Time	Scale with K	$\approx 2\times$ faster for the same K
Backpropagation	All iterations	Only the last iteration
$g(\theta) =$	$\partial_x L(\varphi^K(x^{[0]}, \theta))\, \partial_\theta \varphi^K(x^{[0]}, \bullet)(\theta)$	$\partial_x L(\varphi(\tilde{x}_{K-1}, \theta))\, \partial_\theta \varphi(\tilde{x}_{K-1}, \bullet)(\theta)$

Jacobian-Free Unrolling

	Classical Unrolling	**Jacobian-Free Unrolling (JFU)**
Max itration (K)	$\sim 6-8$	∞
Parameters	$\theta = \{\theta_1, \dots, \theta_k\}$	Forced to be tight
Memory	Scale with K	Constant
Training Time	Scale with K	$\approx 2\times$ faster for the same K
Backpropagation	All iterations	Only the last iteration
$g(\theta) =$	$\partial_x L(\varphi^K(x^{[0]}, \theta))\, \partial_\theta \varphi^K(x^{[0]}, \bullet)(\theta)$	$\partial_x L(\varphi(\tilde{x}_{K-1}, \theta))\, \partial_\theta \varphi(\tilde{x}_{K-1}, \bullet)(\theta)$

- Theoretically, $\tilde{x}_{K-1} \simeq \hat{x}_\theta$. This assumption does not hold in SPECT. Can we still use JFU ?

Jacobian-Free Unrolling

	Classical Unrolling	**Jacobian-Free Unrolling (JFU)**
Max itration (K)	$\sim 6-8$	∞
Parameters	$\theta = \{\theta_1, \dots, \theta_k\}$	Forced to be tight
Memory	Scale with K	Constant
Training Time	Scale with K	$\approx 2\times$ faster for the same K
Backpropagation	All iterations	Only the last iteration
$g(\theta) =$	$\partial_x L(\varphi^K(x^{[0]}, \theta))\, \partial_\theta \varphi^K(x^{[0]}, \bullet)(\theta)$	$\partial_x L(\varphi(\tilde{x}_{K-1}, \theta))\, \partial_\theta \varphi(\tilde{x}_{K-1}, \bullet)(\theta)$

- Theoretically, $\tilde{x}_{K-1} \simeq \hat{x}_\theta$. This assumption does not hold in SPECT. Can we still use JFU ?

  $\Rightarrow$ The JFB update still provides a descent direction for poor estimate of $\hat{x}_\theta$
  [corollary 0.1 [Fung et al., 2022]]

Jacobian-Free Unrolling

	Classical Unrolling	**Jacobian-Free Unrolling (JFU)**
Max itration (K)	$\sim 6-8$	∞
Parameters	$\theta = \{\theta_1, \dots, \theta_k\}$	Forced to be tight
Memory	Scale with K	Constant
Training Time	Scale with K	$\approx 2\times$ faster for the same K
Backpropagation	All iterations	Only the last iteration
$g(\theta) =$	$\partial_x L(\varphi^K(x^{[0]}, \theta))\, \partial_\theta \varphi^K(x^{[0]}, \bullet)(\theta)$	$\partial_x L(\varphi(\tilde{x}_{K-1}, \theta))\, \partial_\theta \varphi(\tilde{x}_{K-1}, \bullet)(\theta)$

- Theoretically, $\tilde{x}_{K-1} \simeq \hat{x}_\theta$. This assumption does not hold in SPECT. Can we still use JFU ?

  $\Rightarrow$ The JFB update still provides a descent direction for poor estimate of $\hat{x}_\theta$
  [corollary 0.1 [Fung et al., 2022]]

- In our experiments:

	Classical Unrolling	**Jacobian-Free Unrolling (JFU)**
K	6	$6 \rightarrow 18$
VRAM	11.5 GiB	4.96 GiB
Parameters	$\sim$ 840000	$\sim$ 140000
Time/epochs	$\sim$ 25 min	$\sim$ 12 min $\rightarrow \sim$ 37 min

Dataset Generation

No ground truth available: synthetic dataset [Kaprélian et al., 2024]

Dataset Generation

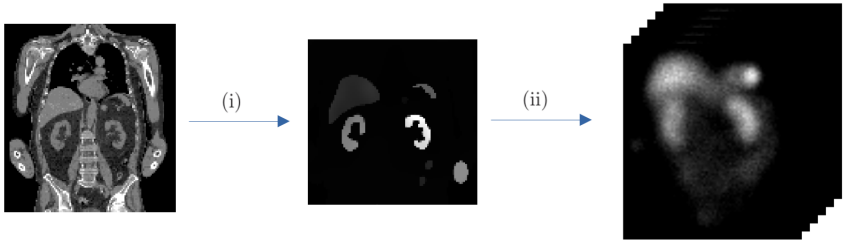
No ground truth available: synthetic dataset [Kaprélian et al., 2024]

- (i) **Source**: random activity sources from segmented CT[Wasserthal et al., 2023]
- (ii) **Acquisition simulation**: attenuation + noise + PSF

Dataset Generation

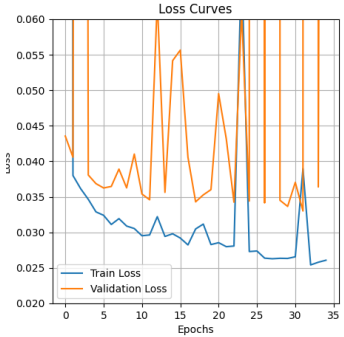
No ground truth available: synthetic dataset [Kaprélian et al., 2024]

- (i) **Source**: random activity sources from segmented CT[Wasserthal et al., 2023]
- (ii) **Acquisition simulation**: attenuation + noise + PSF

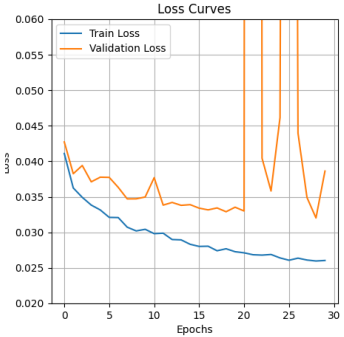


16 patients CT
10 Training/3 validation/3 test $\Rightarrow$ 150/45/45 simulated acquisitions

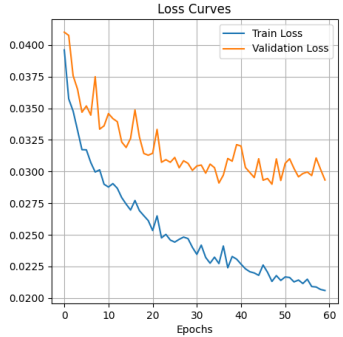
Training convergence and stability



(a) U-MAPEM (6it)



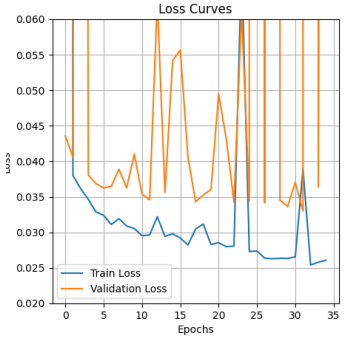
(b) JFU-MAPEM (6it)



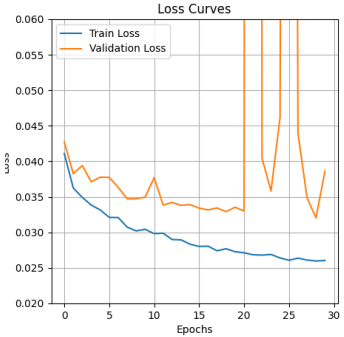
(c) JFU-MAPEM (18it)

Figure: Training and validation loss for ours models

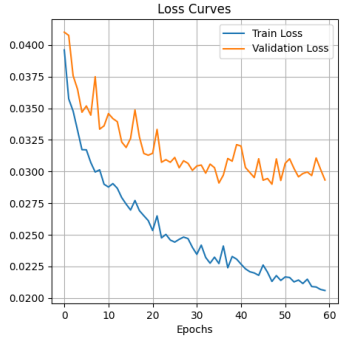
Training convergence and stability



(a) U-MAPEM (6it)



(b) JFU-MAPEM (6it)



(c) JFU-MAPEM (18it)

Figure: Training and validation loss for ours models

- JFU helps to **stabilize the training**, especially with more iterations.

Evaluation on simulated dataset - Global metrics

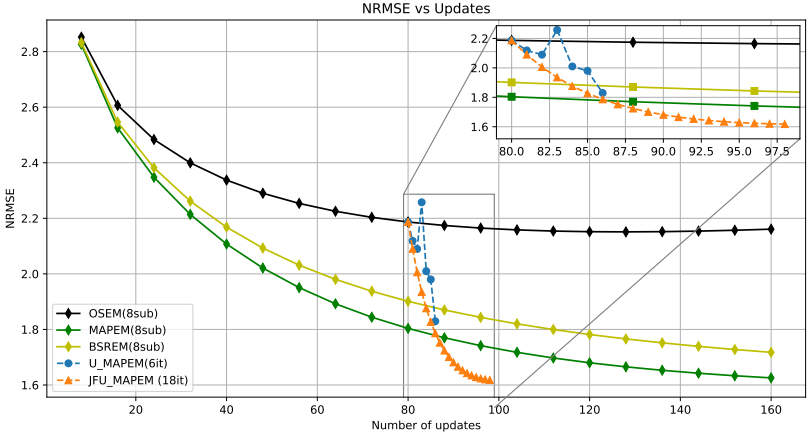
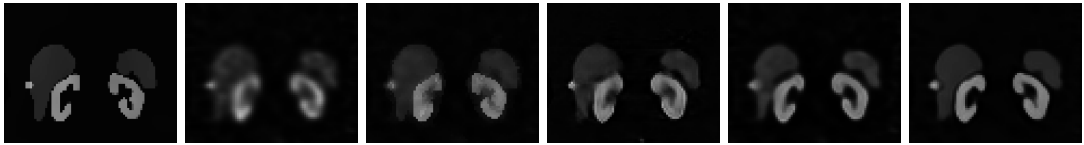


Figure: Evolution of the NRMSE as a function of the number of inner problem updates. Averaged on 45 simulated patients

Evaluation on simulated dataset - Global metrics

	OSEM(10i,8s)	MAPEM(20i,8s)	PVCNet-sino [Kaprélian et al., 2024]	U-MAPEM(6i)	JFU-MAPEM(18i)
PSNR↑	39.51	42.16	41.23	41.03	42.12
NRMSE↓	2.19	1.63	1.79	1.83	1.62
NMAE↓	0.37	0.27	0.25	0.24	0.24

Table: Mean of global metrics (PSNR, NRMSE, NMAE) obtained on the 45 simulated acquisitions. Best values are shown in green and second best in blue.



(a) source (b) OSEM (c) MAPEM (d) PVCNet-Sino (e) U-MAPEM (f) JFU-MAPEM

Evaluation on simulated dataset - Local metrics

- ROI-level metric : Recovery Coefficient (RC), Volume Activity Accuracy (VAA) [Leube et al., 2024]

$$\mathrm{VAA}(ROI^{\mathrm{ref}}, ROI) = \frac{1}{\#\mathrm{voxels}} \#\{i \text{ such that } \frac{|ROI_i^{\mathrm{ref}} - ROI_i|}{|ROI_i^{\mathrm{ref}}|} \leq 0.05\} \times 100\% \tag{9}$$

	VAA↑				
organ	OSEM	MAPEM	PVCNet-sino	U-MAPEM	JFU-MAPEM(18it)
liver	0.18	0.30	0.25	$0.26^{**}_{\diamond\diamond}$	$0.53^{**}_{\diamond\diamond}$
spleen	0.11	0.19	0.20	$0.16^{**}_{\diamond\diamond}$	$0.37^{**}_{\diamond\diamond}$
bladder	0.09	0.12	0.06	$0.07^{*}_{\diamond\diamond}$	$0.11_{\diamond\diamond}$
kidney	0.09	0.12	0.18	0.21^{**}	$0.29^{**}_{\diamond}$
lesions	0.02	0.02	0.07	$0.04^{**}_{\diamond}$	$0.05^{**}_{\diamond}$

Table: Mean of local metrics VAA obtained for different ROIs on 45 simulated patients. Best values are shown in green and second best in blue. We conducted paired t-tests to assess the significance of the differences between our methods and the reference methods. Significance w.r.t RM is denote with * ($p < 5 \times 10^{-3}$), ** ($p < 5 \times 10^{-4}$) and significance w.r.t MAPEM is denote with $\diamond$ ($p < 5 \times 10^{-3}$), $\diamond\diamond$ ($p < 5 \times 10^{-4}$).

Experimental results - RC on the IEC phantom

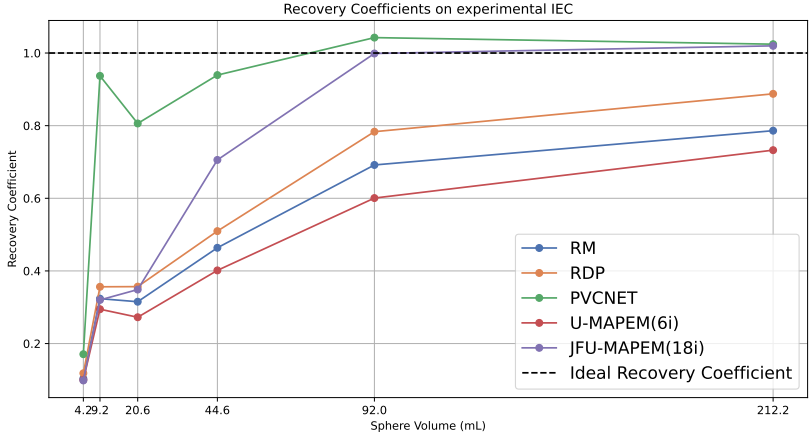
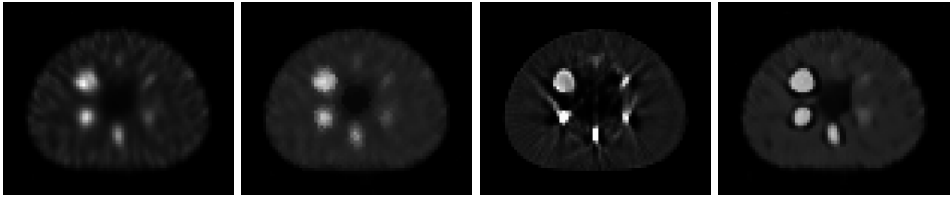


Figure: Recovery Coefficients (RCs) obtained with the 6-sphere IEC phantom experimental acquisition (^{177}Lu, Siemens Intevo Bold).

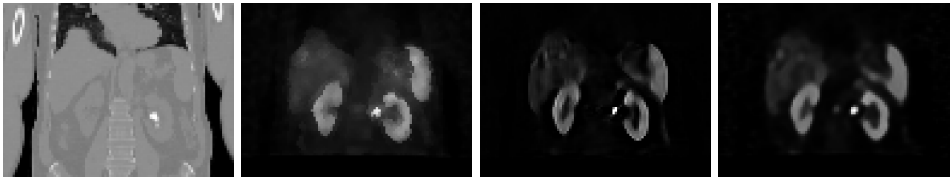
Experimental results

- JFU-MAPEM introduce less artefact than PVCNet-Sino



(a) OSEM (b) MAPEM (c) PVCNet-Sino (d) JFU-MAPEM

- Real clinical data (qualitative)



(a) Attmap (b) MAPEM (c) PVCNet-Sino (d) JFU-MAPEM

Conclusions and perspective

- Key take home message
  - There is a clear connection between DEQ and Unrolled methods. [Davy et al., 2025]
  - With JFU we are **no longer limited by VRAM**, only by computation time !
  - JFU **improves training stability** even though the unrolled iterative scheme has no convergence guaranties.
  - JFU have better quantitative performance.

Conclusions and perspective

- Key take home message
  - There is a clear connection between DEQ and Unrolled methods. [Davy et al., 2025]
  - With JFU we are **no longer limited by VRAM**, only by computation time !
  - JFU **improves training stability** even though the unrolled iterative scheme has no convergence guaranties.
  - JFU have better quantitative performance.

- Perspective:
  - Application to new SPECT geometries (Veriton/StarGuide)
  - Investigate subset acceleration.
  - Leveraging larger neural networks.
  - Impact on voxel based dosimetry.

Thank you!
Do you have any questions?

Bibliography I

Jonas Adler and Ozan Öktem. Solving ill-posed inverse problems using iterative deep neural networks. *Inverse Problems*, 33(12):124007, 2017.

Sangtae Ahn and Jeffrey A Fessler. Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms. *IEEE transactions on medical imaging*, 22(5):613–626, 2003.

Sangtae Ahn, Steven G Ross, Evren Asma, Jun Miao, Xiao Jin, Lishui Cheng, Scott D Wollenweber, and Ravindra M Manjeshwar. Quantitative comparison of osem and penalized likelihood image reconstruction using relative difference penalties for clinical pet. *Physics in Medicine & Biology*, 60(15):5733, 2015.

Shaojie Bai, J Zico Kolter, and Vladlen Koltun. Deep equilibrium models. *Advances in neural information processing systems*, 32, 2019.

Leo Davy, Luis M Briceno-Arias, and Nelly Pustelnik. Restarted contractive operators to learn at equilibrium. *arXiv preprint arXiv:2506.13239*, 2025.

M Del Prete, F Arsenault, N Saighi, LO Bouchard, A Beaulieu, FA Buteau, and JM Beauregard. Personalized lu-177-octreotate peptide receptor radionuclide therapy of neuroendocrine tumours: preliminary dosimetry, efficacy and safety results from the p-prrt trial. In *EUROPEAN JOURNAL OF NUCLEAR MEDICINE AND MOLECULAR IMAGING*, volume 45, pages S60–S61. SPRINGER 233 SPRING ST, NEW YORK, NY 10013 USA, 2018.

Samy Wu Fung, Howard Heaton, Qiuwei Li, Daniel McKenzie, Stanley Osher, and Wotao Yin. Jfb: Jacobian-free backpropagation for implicit networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 6648–6656, 2022.

Peter J Green. Bayesian reconstructions from emission tomography data using a modified em algorithm. *IEEE transactions on medical imaging*, 9(1):84–93, 1990.

H Malcolm Hudson and Richard S Larkin. Accelerated image reconstruction using ordered subsets of projection data. *IEEE transactions on medical imaging*, 13 (4):601–609, 1994.

Tuija S Kangasmaa, Chris Constable, and Antti O Sohlberg. Quantitative bone spect/ct reconstruction utilizing anatomical information. *EJNMMI physics*, 8(1):2, 2021.

Théo Kaprélian, Ane Etxebeste, and David Sarrut. Partial volume correction on 177 lu-spect sinogram with deep learning trained on synthetic data. In *2024 IEEE Nuclear Science Symposium (NSS), Medical Imaging Conference (MIC) and Room Temperature Semiconductor Detector Conference (RTSD)*, pages 1–2. IEEE, 2024.

Julian Leube, Johan Gustafsson, Michael Lassmann, Maikol Salas-Ramirez, and Johannes Tran-Gia. A deep-learning–based partial-volume correction method for quantitative 177lu spect/ct imaging. *Journal of Nuclear Medicine*, 65(6):980–987, 2024.

Bibliography II

Zongyu Li, Yuni K Dewaraja, and Jeffrey A Fessler. Training end-to-end unrolled iterative neural networks for spect image reconstruction. *IEEE transactions on radiation and plasma medical sciences*, 7(4):410–420, 2023.

Vishal Monga, Yuelong Li, and Yonina C Eldar. Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. *IEEE Signal Processing Magazine*, 38(2):18–44, 2021.

Boyang Pan, Na Qi, Qingyuan Meng, Jiachen Wang, Siyue Peng, Chengxiao Qi, Nan-Jie Gong, and Jun Zhao. Ultra high speed spect bone imaging enabled by a deep learning enhancement method: a proof of concept. *EJNMMI physics*, 9(1):43, 2022.

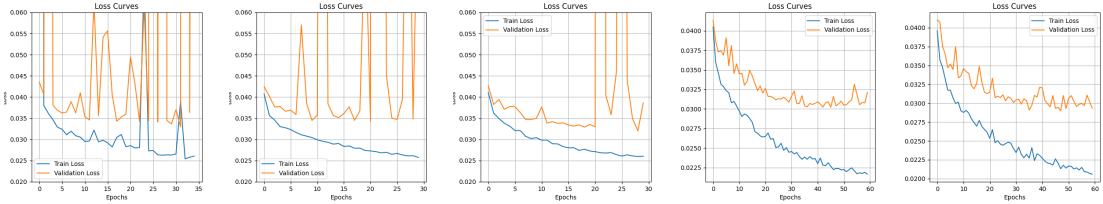
VY Panin, GL Zeng, and GT Gullberg. Total variation regulated em algorithm. In *1998 IEEE Nuclear Science Symposium Conference Record. 1998 IEEE Nuclear Science Symposium and Medical Imaging Conference (Cat. No. 98CH36255)*, volume 3, pages 1562–1566. IEEE, 1998.

Lawrence A Shepp and Yehuda Vardi. Maximum likelihood reconstruction for emission tomography. *IEEE transactions on medical imaging*, 1(2):113–122, 1982.

Laure Vergnaud. *Dosimetry for 177Lu and 90Y radionuclide therapies through imaging and Monte Carlo simulations*. PhD thesis, INSA de Lyon, 2023.

Jakob Wasserthal, Hanns-Christian Breit, Manfred T Meyer, Maurice Pradella, Daniel Hinck, Alexander W Sauter, Tobias Heye, Daniel T Boll, Joshy Cyriac, Shan Yang, et al. Totalsegmentator: robust segmentation of 104 anatomic structures in ct images. *Radiology: Artificial Intelligence*, 5(5):e230024, 2023.

Backup - MAPEM extended Results I



(a) U-MAPEM (6it) (b) U-MAPEM (tied weights) (c) JFU-MAPEM (6it) (d) U-MAPEM (12it) (e) JFU-MAPEM (18it)

Figure: Training and validation loss for ours models

	OSEM(10i,8s)	MAPEM(20i,8s)	PVCNet-sino [Kaprélian et al., 2024]	U-MAPEM(6i)	U-MAPEM(18i)	U-MAPEM(18i)	JFU-MAPEM(18i)
PSNR↑	39.51	42.16	41.23	41.03	41.65	42.00	42.12
NRMSE↓	2.19	1.63	1.79	1.83	1.70	1.64	1.62
NMAE↓	0.37	0.27	0.25	0.34	0.30	0.24	0.24

Table: Mean of global metrics (PSNR, NRMSE, NMAE) obtained on the 45 simulated acquisitions. Best values are shown in green and second best in blue.

Backup - MAPEM extended Results II

	RC				
organ	OSEM	MAPEM	PVCNet-sino	U-MAPEM	JFU-MAPEM(18it)
liver	0.91	0.96	0.89	$0.97^{**}_{\diamond\diamond}$	$0.94^{**}_{\diamond\diamond}$
pancreas	0.54	0.52	0.56	$0.41^{**}_{\diamond\diamond}$	$0.50^{**}_{\diamond\diamond}$
spleen	0.84	0.87	0.91	$0.95^{**}_{\diamond\diamond}$	$0.90^{**}_{\diamond\diamond}$
bladder	0.80	0.82	0.72	$0.89^{**}_{\diamond}$	$0.85_{\diamond\diamond}$
kidney	0.76	0.82	0.88	$0.8^{**}_{\diamond\diamond}$	0.86^{**}
lesions	0.46	0.42	0.64	$0.52^{**}_{\diamond\diamond}$	$0.54^{**}_{\diamond\diamond}$

Table: Mean of RC obtained for different ROIs on 45 simulated patients. Best values are shown in green and second best in blue. We conducted paired t-tests to assess the significance of the differences between our methods and the reference methods. Significance w.r.t RM is denote with * ($p < 5 \times 10^{-3}$), ** ($p < 5 \times 10^{-4}$) and significance w.r.t MAPEM is denote with $\diamond$ ($p < 5 \times 10^{-3}$), $\diamond\diamond$ ($p < 5 \times 10^{-4}$).

Backup - BSREM Results I

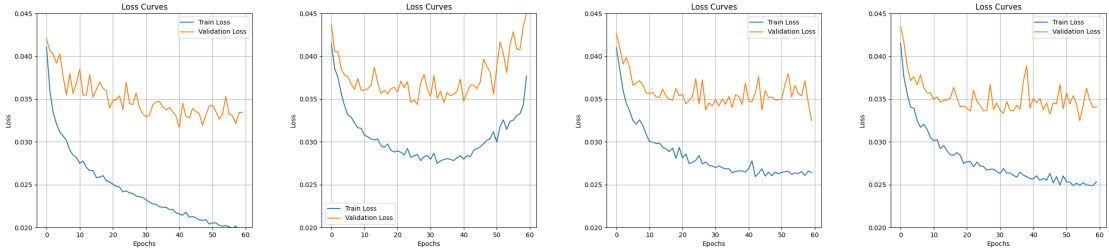
	OSEM(10i,8s)	BSREM(20i,8s)	PVCNet-sino	U-BSREM(6i)	JFU-BSREM(6i)	JFU-BSREM(12i)	JFU-BSREM(18i)
PSNR	39.51	41.66	41.23	41.99	41.34	41.67	41.82
NRMSE	2.19	1.72	1.79	1.64	1.77	1.70	1.68
NMAE	0.37	0.28	0.25	0.25	0.30	0.28	0.25

TABLE III: Mean of global metrics (PSNR, NRMSE, NMAE) for the 7 compared reconstruction methods, obtained on the 45 simulated patients. Best values are shown in green and second best in blue. We conducted paired t-tests to assess the significance of the differences between our methods and the two reference methods. Across all tests and method combinations, we obtained p-values smaller than 1×10^{-5}.

	VAA					RC				
organ	OSEM	BSREM	PVCNet-sino	U-BSREM	JFU-BSREM(18it)	OSEM	BSREM	PVCNet-sino	U-BSREM	JFU-BSREM(18it)
liver	0.18	0.32	0.25	$0.50^{**}_{\diamond\diamond}$	$0.51^{**}_{\diamond\diamond}$	0.91	0.93	0.89	$0.94^{**}_{\diamond\diamond}$	$0.95^{**}_{\diamond\diamond}$
pancreas	0.02	0.01	0.03	0.01	0.01	0.52	0.66	0.56	$0.50^{**}_{\diamond\diamond}$	$0.59^{**}_{\diamond\diamond}$
spleen	0.11	0.23	0.20	$0.30^{**}_{\diamond\diamond}$	$0.30^{**}_{\diamond\diamond}$	0.84	0.90	0.91	$0.90^{**}_{\diamond\diamond}$	$0.91^{**}_{\diamond\diamond}$
bladder	0.09	0.14	0.06	$0.10^{*}_{\diamond\diamond}$	$0.11_{\diamond\diamond}$	0.80	0.84	0.72	$0.78^{**}_{\diamond}$	$0.79_{\diamond\diamond}$
kidney	0.09	0.16	0.18	0.21^{**}	$0.29^{**}_{\diamond}$	0.76	0.84	0.88	$0.86^{**}_{\diamond\diamond}$	0.88^{**}
lesions	0.02	0.02	0.07	$0.05^{**}_{\diamond}$	$0.05^{**}_{\diamond}$	0.46	0.44	0.64	$0.58^{**}_{\diamond\diamond}$	$0.52^{**}_{\diamond\diamond}$

TABLE IV: Mean of local metrics (VAA and RC) obtained for different ROIs on 45 simulated patients. Best values are shown in green and second best in blue. We conducted paired t-tests to assess the significance of the differences between our methods and the reference methods. Significance w.r.t RM is denote with * ($p < 5 \times 10^{-3}$), ** ($p < 5 \times 10^{-4}$) and significance w.r.t RDP is denote with $\diamond$ ($p < 5 \times 10^{-3}$), $\diamond\diamond$ ($p < 5 \times 10^{-4}$).

Backup - BSREM Results II



(a) U-BSREM (6it) (b) JFU-BSREM (6it) (c) U-BSREM (12it) (d) JFU-BSREM (18it)

Figure: Training and validation loss for U/JFU-BSREM

Backup - Jacobian-Free Backpropagation (JFB) I

$$\begin{cases} \theta^* \in \arg\min_{\theta} L(\hat{x}_\theta) & \textbf{(Outer Problem)} \\ \hat{x}_\theta \in \arg\max_{x} \big(\mathcal{L}(x) + \mathcal{R}_\theta(x)\big) & \textbf{(Inner Problem)} \end{cases} \tag{10}$$

$$\theta^{[m+1]} = \theta^{[m]} - \tau g(\theta^{[m]}) \tag{11}$$

Using the chain rule,

$$\begin{aligned} g(\theta) &= \partial_\theta L\,(\hat{x}_\theta) && (12) \\ &= \partial_x L(\hat{x}_\theta)^T \partial_\theta \hat{x}_\theta && (13) \end{aligned}$$

- The gradient $\partial_x L$ can be computed explicitly in most cases
- The main challenge is the computation of $\partial_\theta \hat{x}_\theta$
- Several approaches exist to address this problem:
  - Deep Equilibrium Networks (DEQ) [Bai et al., 2019] (Jacobian calculation)
  - Unrolled algorithms [Monga et al., 2021] (end to end autodiff)

Backup - Jacobian-Free Backpropagation (JFB) II

$$\hat{x}_\theta = \varphi(\hat{x}_\theta, \theta) \tag{14}$$

Deriving with respect to θ, we get :

$$\partial_\theta \hat{x}_\theta = [\partial_x \varphi(\bullet, \theta)(\hat{x}_\theta)] \partial_\theta \hat{x}_\theta \; + \; \partial_\theta \varphi(\hat{x}_\theta, \bullet)(\theta) \tag{15}$$

$$[\mathrm{Id} - \partial_x \varphi(\bullet, \theta)(\hat{x}_\theta)] \partial_\theta \hat{x}_\theta \; = \; \partial_\theta \varphi(\hat{x}_\theta, \bullet)(\theta) \tag{16}$$

So we have,

$$g^{\mathrm{DEQ}}(\theta) = \partial_x L(\hat{x}_\theta) [\mathrm{Id} - \partial_x \varphi(\bullet, \theta)(\hat{x}_\theta)]^{-1} \; \partial_\theta \varphi(\hat{x}_\theta, \bullet)(\theta) \tag{17}$$

Backup - Jacobian-Free Backpropagation (JFB) III

3 main steps to solve the **Outer problem** :

- Finding $\hat{x}_\theta$ such that $\hat{x}_\theta = \varphi(\hat{x}_\theta, \theta)$ *(any solver can be used)*
- Computing the Jacobian $\partial_x \varphi(\bullet, \theta)(\hat{x}_\theta)$ *(auto-differentiation)*
- Inverting $J_\theta = [\mathrm{Id} - \partial_x \varphi(\bullet, \theta)(\hat{x}_\theta)]$ *(Neumann inversion)*

$$\mathrm{I}_P = \sum_{p=0}^{P} [\partial_x \varphi(\cdot, \theta)(\hat{x}_\theta)]^p \qquad \text{(Neumann Inversion)}$$

To avoid the Jacobian calculation/inversion, we can use Jacobian-Free Backpropagation (JFB) [Fung et al., 2022].

$$g^{JFB}(\theta) = \partial_x L(\hat{x}_\theta)\ \partial_\theta \varphi(\hat{x}_\theta, \bullet)(\theta) \tag{18}$$

Interpretations :

- A case where $\mathrm{J}_\theta^{-1} = \mathrm{Id}$
- Neumann inversion truncated at order 0, leading to $\mathrm{J}_\theta^{-1} = \mathrm{Id}$

Advantages: Faster and fixed-memory approch