

Equivariant splitting: self-supervised learning from incomplete data



Victor Sechaud^{*1}, Jérémie Scanvic^{*1,2}, Quentin Barthélémy², Patrice Abry¹, Julián Tachella¹

¹CNRS, ENS de Lyon

²Prysm, Lyon, France

2025-10-31

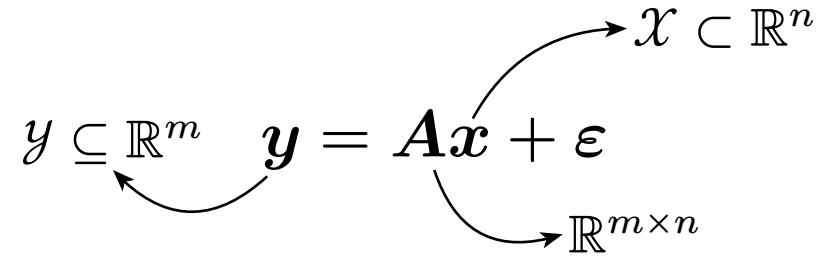
^{*}Equal contribution

$$y = Ax + \epsilon$$

$$y = Ax + \varepsilon \xrightarrow{\quad} \mathcal{X} \subset \mathbb{R}^n$$

$$y = Ax + \epsilon$$

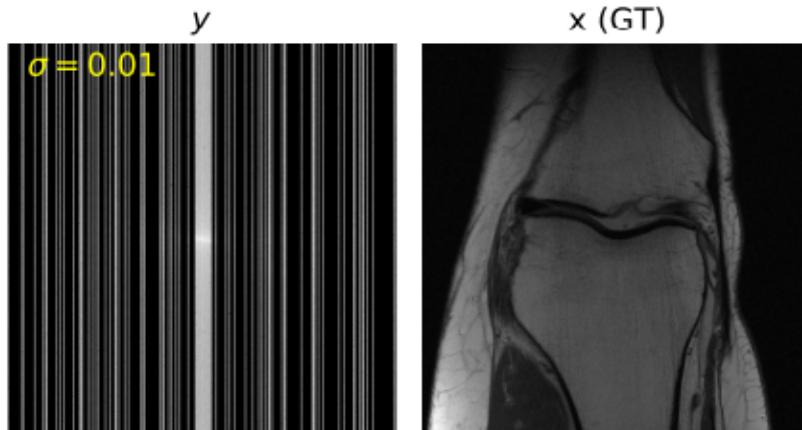
$\xrightarrow{\quad \mathcal{X} \subset \mathbb{R}^n \quad}$
 $\xrightarrow{\quad \mathbb{R}^{m \times n} \quad}$

$$y \in \mathbb{R}^m \quad y = Ax + \epsilon \quad \mathcal{X} \subset \mathbb{R}^n$$
$$\mathbb{R}^{m \times n}$$


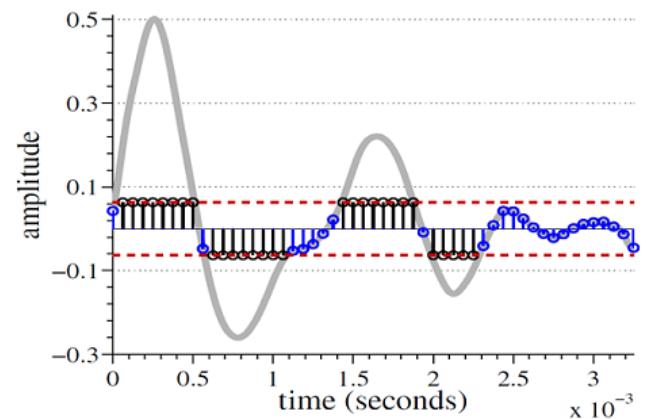
The diagram illustrates the components of a linear model. At the top right is the set $\mathcal{X} \subset \mathbb{R}^n$. Below it is the matrix $\mathbb{R}^{m \times n}$. To the left of the equation is the set $y \subseteq \mathbb{R}^m$. The equation $y = Ax + \epsilon$ is positioned centrally. Curved arrows point from \mathcal{X} to A , from A to $\mathbb{R}^{m \times n}$, and from y to \mathbb{R}^m .

$$\begin{array}{ccc} \mathcal{X} \subset \mathbb{R}^n & & \\ \nearrow & \textbf{\textit{y}} = \mathbf{A}\textbf{\textit{x}} + \boldsymbol{\varepsilon} \longrightarrow \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_n) & \searrow \\ \mathcal{Y} \subseteq \mathbb{R}^m & & \mathbb{R}^{m \times n} \end{array}$$

$$\begin{array}{c} \mathcal{X} \subset \mathbb{R}^n \\ \downarrow \\ \mathcal{Y} \subseteq \mathbb{R}^m \quad \mathbf{y} = \mathbf{A}\mathbf{x} + \boldsymbol{\varepsilon} \xrightarrow{\sim} \mathcal{N}(0, \sigma^2 \mathbf{I}_n) \\ \uparrow \\ \mathbb{R}^{m \times n} \end{array}$$



Linear: MRI, inpainting, ...

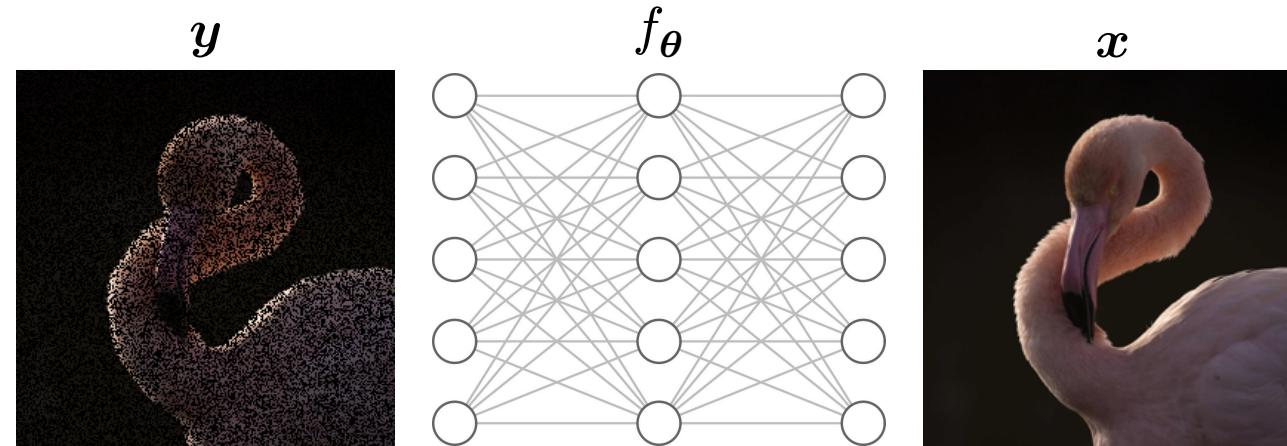


Non-linear: declipping, phase retrieval

- Deep learning is now the state of the art.

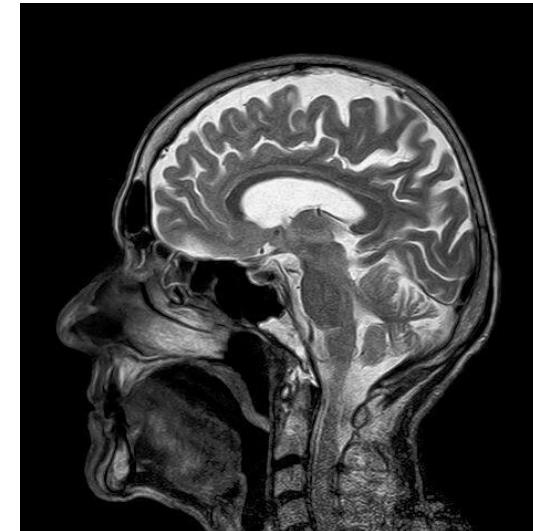
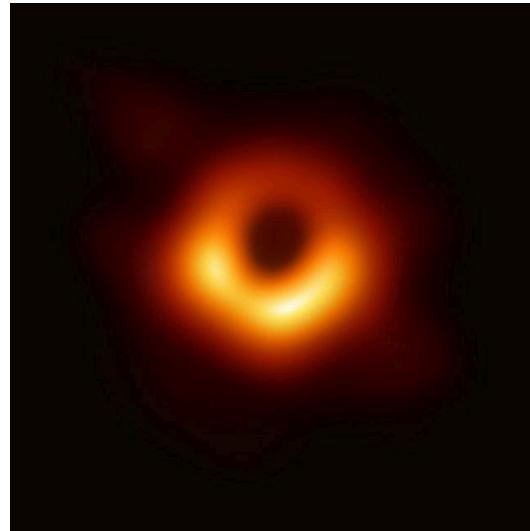
Objective

- **Learn** inverse forward operator of \mathbf{A} , $f_{\theta}(\mathbf{y}, \mathbf{A}) \approx \mathbf{x}$.
- **Using** a dataset $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i \in I}$ + a neural network.



$$\operatorname{argmin}_{\theta} \sum_{i \in I} \|f_{\theta}(\mathbf{y}_i, \mathbf{A}) - \mathbf{x}_i^2\|$$

- Training and testing datasets can be very different.
- We need a large set of $\{x_i\}_{i \in I}$ (ground-truth).



Astronomical and medical imaging.

- **Dataset:** $(\mathbf{x}_i, \mathbf{y}_i)_{i \in I}$ only $(\mathbf{y}_i)_{i \in I}$.
- What can we do with \mathbf{y}_i ?

$$\mathbf{y} = \mathbf{A}\mathbf{x} \Rightarrow \mathbf{y} \approx \mathbf{A}f_{\theta}(\mathbf{y}, \mathbf{A})$$

- **Dataset:** $(\mathbf{x}_i, \mathbf{y}_i)_{i \in I}$ only $(\mathbf{y}_i)_{i \in I}$.
- What can we do with \mathbf{y}_i ?

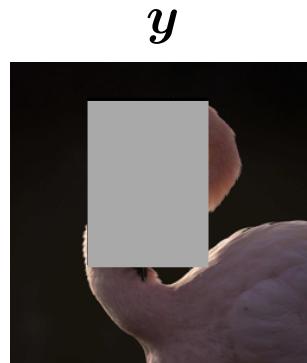
$$\mathbf{y} = \mathbf{A}\mathbf{x} \Rightarrow \mathbf{y} \approx \mathbf{A}f_{\theta}(\mathbf{y}, \mathbf{A})$$

$$\operatorname{argmin}_{\theta} \sum_{i \in I} \|\mathbf{y}_i - \mathbf{A}f_{\theta}(\mathbf{y}_i, \mathbf{A})\|^2$$

- **Dataset:** $(\mathbf{x}_i, \mathbf{y}_i)_{i \in I}$ only $(\mathbf{y}_i)_{i \in I}$.
- What can we do with \mathbf{y}_i ?

$$\mathbf{y} = \mathbf{A}\mathbf{x} \Rightarrow \mathbf{y} \approx \mathbf{A}f_{\theta}(\mathbf{y}, \mathbf{A})$$

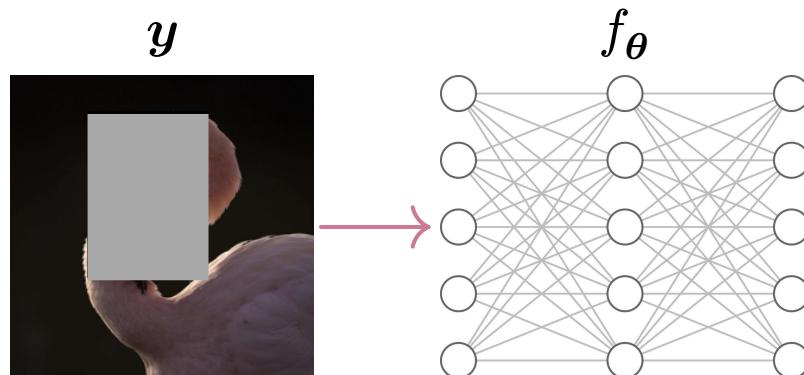
$$\operatorname{argmin}_{\theta} \sum_{i \in I} \|\mathbf{y}_i - \mathbf{A}f_{\theta}(\mathbf{y}_i, \mathbf{A})\|^2$$



- **Dataset:** $(\mathbf{x}_i, \mathbf{y}_i)_{i \in I}$ only $(\mathbf{y}_i)_{i \in I}$.
- What can we do with \mathbf{y}_i ?

$$\mathbf{y} = \mathbf{A}\mathbf{x} \Rightarrow \mathbf{y} \approx \mathbf{A}f_{\theta}(\mathbf{y}, \mathbf{A})$$

$$\operatorname{argmin}_{\theta} \sum_{i \in I} \|\mathbf{y}_i - \mathbf{A}f_{\theta}(\mathbf{y}_i, \mathbf{A})\|^2$$

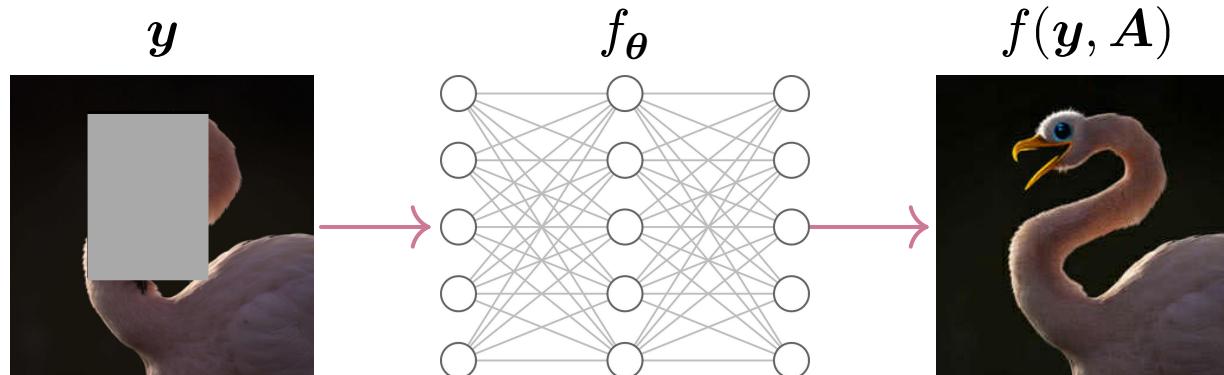


Self-supervised learning

- **Dataset:** $(\mathbf{x}_i, \mathbf{y}_i)_{i \in I}$ only $(\mathbf{y}_i)_{i \in I}$.
- What can we do with \mathbf{y}_i ?

$$\mathbf{y} = \mathbf{A}\mathbf{x} \Rightarrow \mathbf{y} \approx \mathbf{A}f_{\theta}(\mathbf{y}, \mathbf{A})$$

$$\operatorname{argmin}_{\theta} \sum_{i \in I} \|\mathbf{y}_i - \mathbf{A}f_{\theta}(\mathbf{y}_i, \mathbf{A})\|^2$$

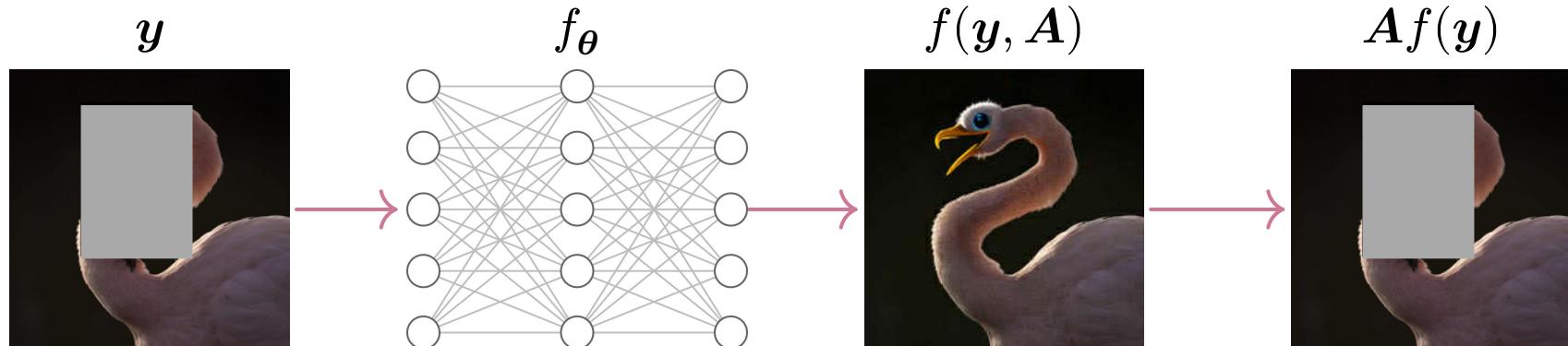


Self-supervised learning

- **Dataset:** $(\mathbf{x}_i, \mathbf{y}_i)_{i \in I}$ only $(\mathbf{y}_i)_{i \in I}$.
- What can we do with \mathbf{y}_i ?

$$\mathbf{y} = \mathbf{A}\mathbf{x} \Rightarrow \mathbf{y} \approx \mathbf{A}f_{\theta}(\mathbf{y}, \mathbf{A})$$

$$\operatorname{argmin}_{\theta} \sum_{i \in I} \|\mathbf{y}_i - \mathbf{A}f_{\theta}(\mathbf{y}_i, \mathbf{A})\|^2$$

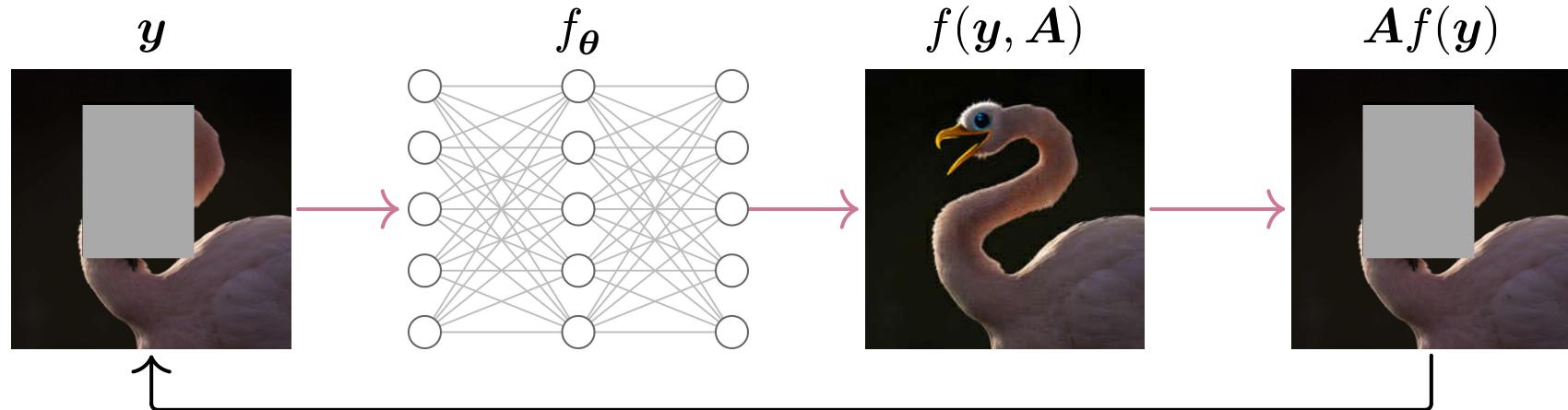


Self-supervised learning

- **Dataset:** $(\mathbf{x}_i, \mathbf{y}_i)_{i \in I}$ only $(\mathbf{y}_i)_{i \in I}$.
- What can we do with \mathbf{y}_i ?

$$\mathbf{y} = \mathbf{A}\mathbf{x} \Rightarrow \mathbf{y} \approx \mathbf{A}f_{\theta}(\mathbf{y}, \mathbf{A})$$

$$\operatorname{argmin}_{\theta} \sum_{i \in I} \|\mathbf{y}_i - \mathbf{A}f_{\theta}(\mathbf{y}_i, \mathbf{A})\|^2$$

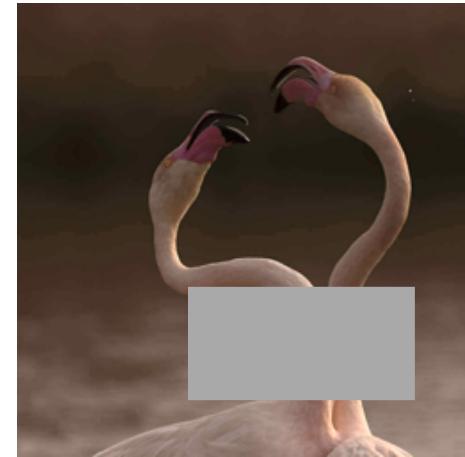
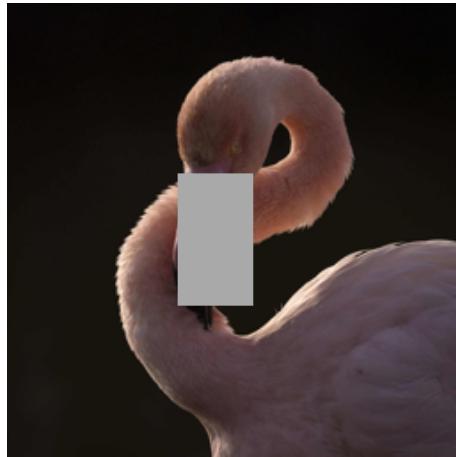


$$\|\mathbf{y} - \mathbf{A}f(\mathbf{y}, \mathbf{A})\|^2 = 0$$

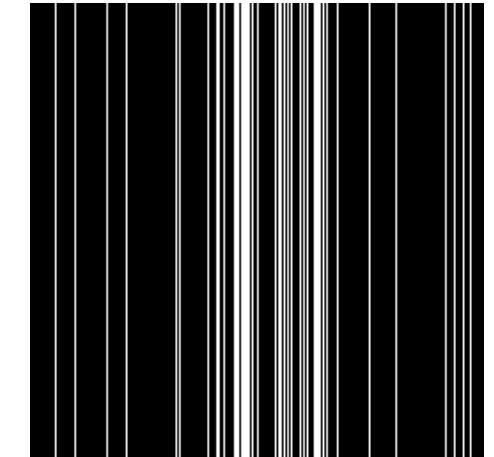
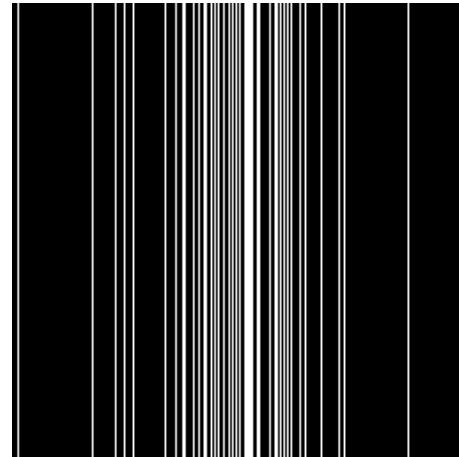
- Leveraging **multiple** operators: $\mathbf{A} \sim p(\mathbf{A})$.
- **Differ** for each measurements $\mathbf{y} \sim p(\mathbf{y} \mid \mathbf{A}\mathbf{x})$.

Multiple operators

- Leveraging **multiple** operators: $A \sim p(A)$.
- **Differ** for each measurements $y \sim p(y \mid Ax)$.



Inpainting



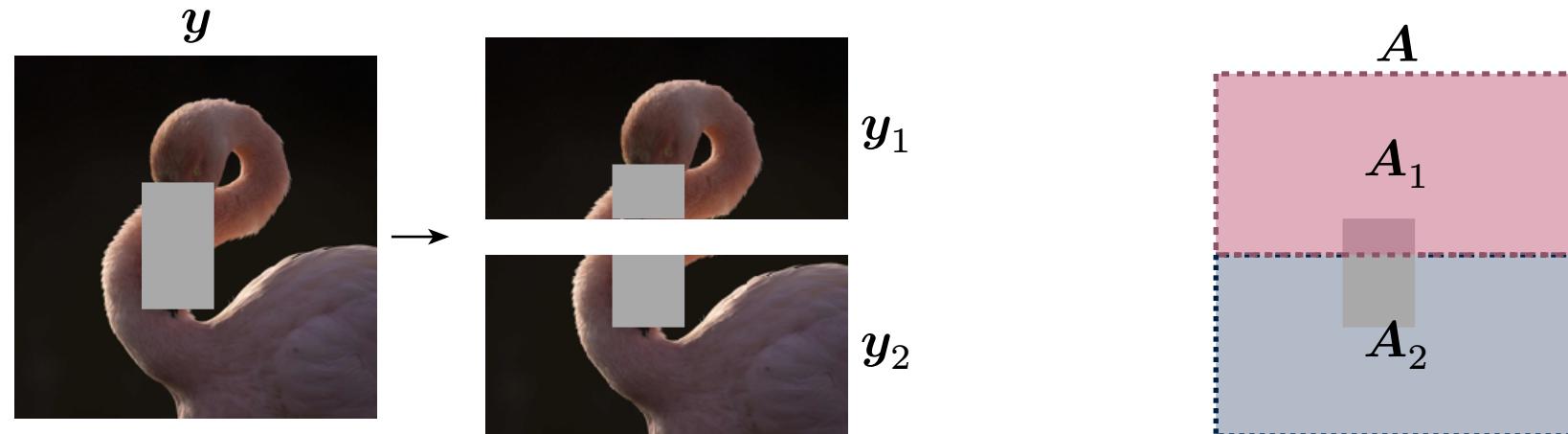
MRI

Measurements splitting

- Leveraging **multiple** operators: $\mathbf{y} \sim p(\mathbf{y} \mid \mathbf{A}\mathbf{x})$ with $\mathbf{A} \sim p(\mathbf{A})$.
- **Split** measurements $\mathbf{y} = [\mathbf{y}_1, \mathbf{y}_2]$ with **corresponding** operators $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2]$.

$$f^* \in \operatorname{argmin}_f \mathbb{E}_{\mathbf{y}, \mathbf{A}} \{\mathcal{L}_{\text{SPLIT}}(\mathbf{y}, \mathbf{A}, f)\}$$

with $\mathcal{L}_{\text{SPLIT}}(\mathbf{y}, \mathbf{A}, f) = \mathbb{E}_{\mathbf{y}_1, \mathbf{A}_1 \mid \mathbf{y}, \mathbf{A}} [\|\mathbf{A}f(\mathbf{y}_1, \mathbf{A}_1) - \mathbf{y}\|^2]$

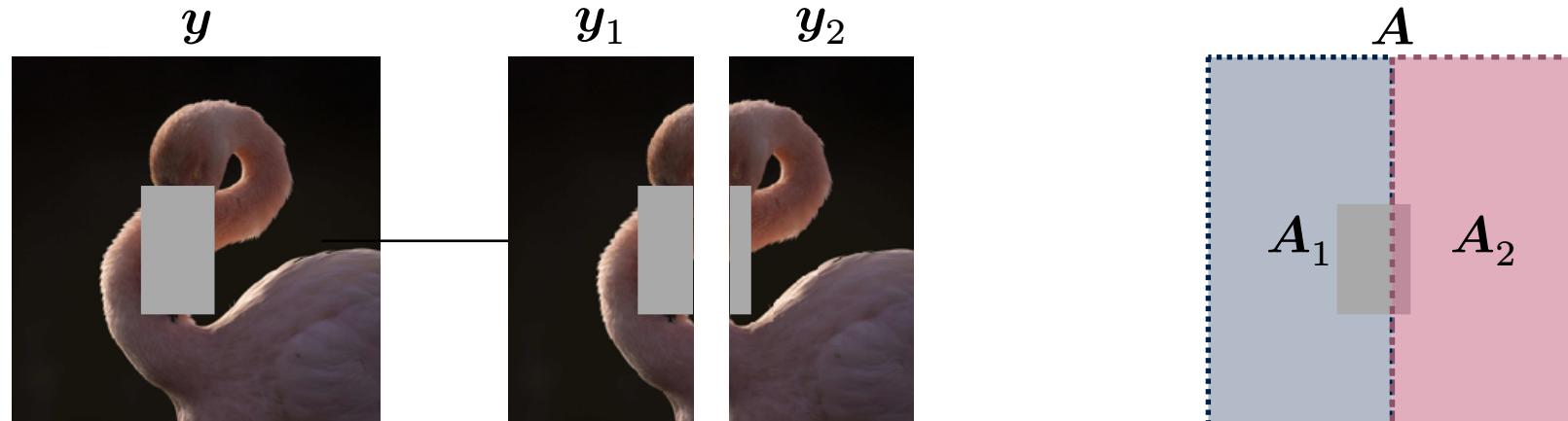


Measurements splitting

- Leveraging **multiple** operators: $\mathbf{y} \sim p(\mathbf{y} \mid \mathbf{A}\mathbf{x})$ with $\mathbf{A} \sim p(\mathbf{A})$.
- **Split** measurements $\mathbf{y} = [\mathbf{y}_1, \mathbf{y}_2]$ with **corresponding** operators $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2]$.

$$f^* \in \operatorname{argmin}_f \mathbb{E}_{\mathbf{y}, \mathbf{A}} \{\mathcal{L}_{\text{SPLIT}}(\mathbf{y}, \mathbf{A}, f)\}$$

with $\mathcal{L}_{\text{SPLIT}}(\mathbf{y}, \mathbf{A}, f) = \mathbb{E}_{\mathbf{y}_1, \mathbf{A}_1 \mid \mathbf{y}, \mathbf{A}} [\|\mathbf{A}f(\mathbf{y}_1, \mathbf{A}_1) - \mathbf{y}\|^2]$



A visual interpretation of Q_{A_1}

Solution with $Q_{A_1} \triangleq \mathbb{E}_{A \mid A_1} \{A^\top A\}$ and v any function in:

$$f^*(y_1, A_1) = Q_{A_1}^\dagger Q_{A_1} \underbrace{\mathbb{E}_{x \mid y_1, A_1} \{x\}}_{\text{MMSE estimator}} + \underbrace{(I - Q_{A_1}^\dagger Q_{A_1})v(y_1)}_{\text{Error}}$$

A visual interpretation of Q_{A_1}

Solution with $Q_{A_1} \triangleq \mathbb{E}_{A \mid A_1} \{A^\top A\}$ and v any function in:

$$f^*(y_1, A_1) = Q_{A_1}^\dagger Q_{A_1} \underbrace{\mathbb{E}_{x \mid y_1, A_1} \{x\}}_{\text{MMSE estimator}} + \underbrace{(I - Q_{A_1}^\dagger Q_{A_1})v(y_1)}_{\text{Error}}$$

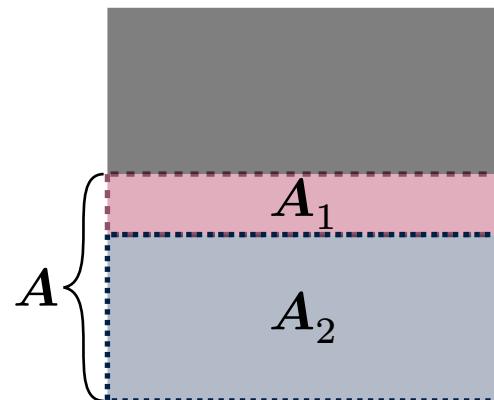
If Q_{A_1} invertible: $f^*(y_1, A_1) = \mathbb{E}_{x \mid y_1, A_1} \{x\}$

A visual interpretation of Q_{A_1}

Solution with $Q_{A_1} \triangleq \mathbb{E}_{A \mid A_1} \{A^\top A\}$ and v any function in:

$$f^*(y_1, A_1) = Q_{A_1}^\dagger Q_{A_1} \underbrace{\mathbb{E}_{x \mid y_1, A_1} \{x\}}_{\text{MMSE estimator}} + \underbrace{(I - Q_{A_1}^\dagger Q_{A_1})v(y_1)}_{\text{Error}}$$

If Q_{A_1} invertible: $f^*(y_1, A_1) = \mathbb{E}_{x \mid y_1, A_1} \{x\}$

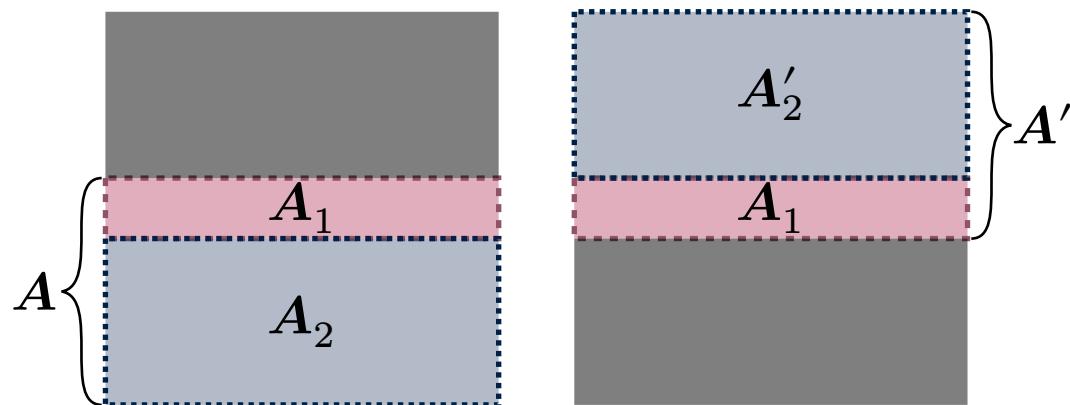


A visual interpretation of Q_{A_1}

Solution with $Q_{A_1} \triangleq \mathbb{E}_{A \mid A_1} \{A^\top A\}$ and v any function in:

$$f^*(y_1, A_1) = Q_{A_1}^\dagger Q_{A_1} \underbrace{\mathbb{E}_{x \mid y_1, A_1} \{x\}}_{\text{MMSE estimator}} + \underbrace{(I - Q_{A_1}^\dagger Q_{A_1})v(y_1)}_{\text{Error}}$$

If Q_{A_1} invertible: $f^*(y_1, A_1) = \mathbb{E}_{x \mid y_1, A_1} \{x\}$

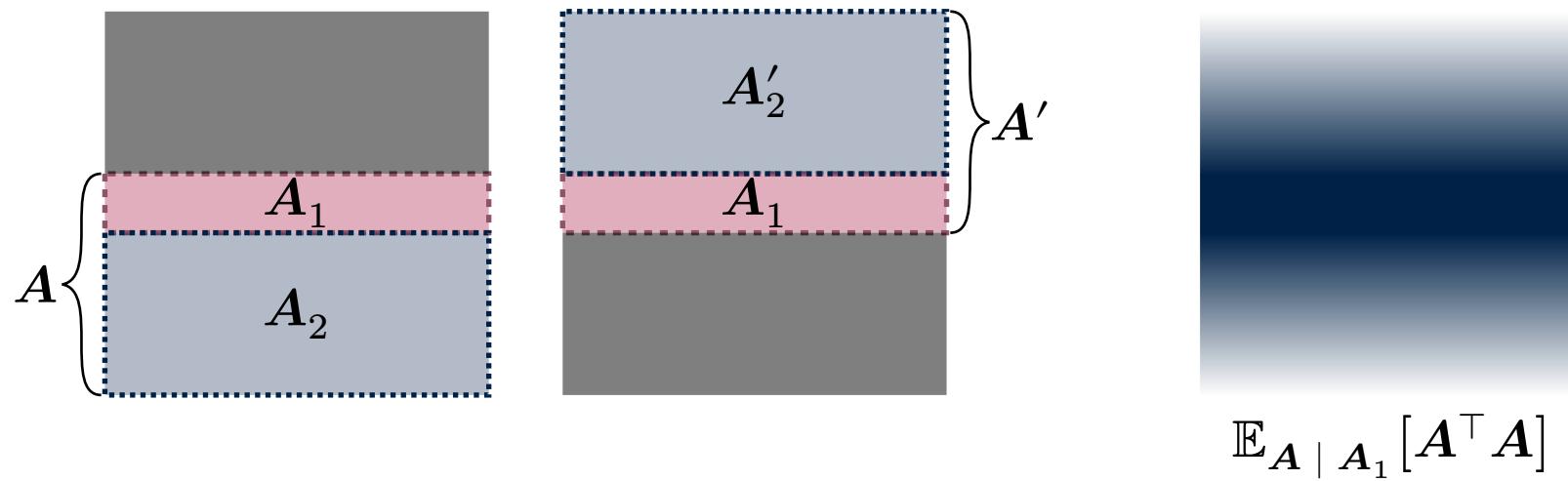


A visual interpretation of Q_{A_1}

Solution with $Q_{A_1} \triangleq \mathbb{E}_{A \mid A_1} \{A^\top A\}$ and v any function in:

$$f^*(y_1, A_1) = Q_{A_1}^\dagger Q_{A_1} \underbrace{\mathbb{E}_{x \mid y_1, A_1} \{x\}}_{\text{MMSE estimator}} + \underbrace{(I - Q_{A_1}^\dagger Q_{A_1})v(y_1)}_{\text{Error}}$$

If Q_{A_1} invertible: $f^*(y_1, A_1) = \mathbb{E}_{x \mid y_1, A_1} \{x\}$

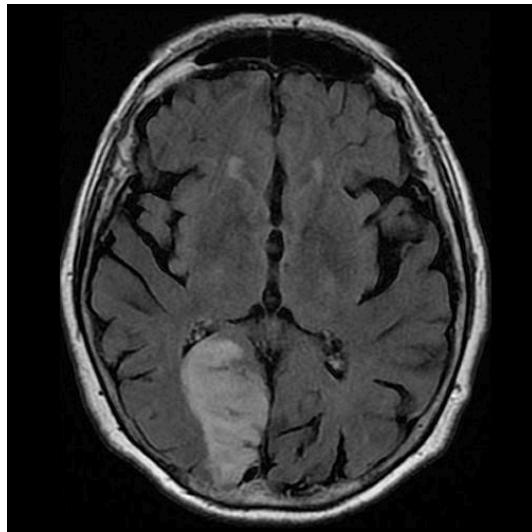


$$\mathbb{E}_{A \mid A_1} [A^\top A]$$

- **A priori:** Invariance to certain transformations: $\forall g \in \mathcal{G} : p(T_g x) = p(x)$.

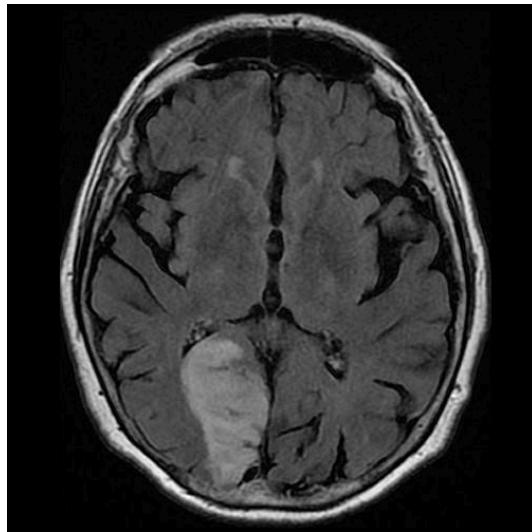
- **A priori:** Invariance to certain transformations: $\forall g \in \mathcal{G} : p(T_g x) = p(x)$.

$x \in \mathcal{X}$

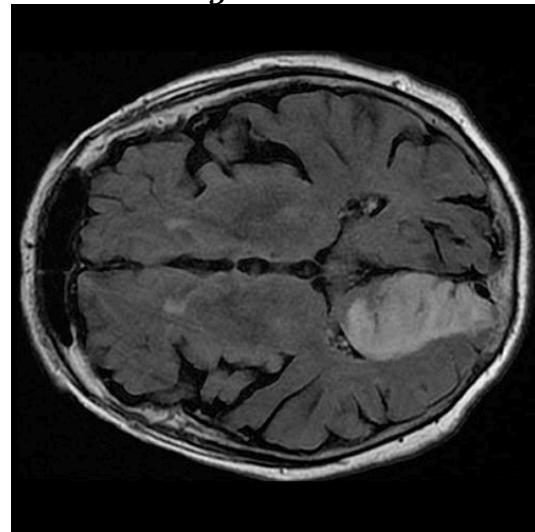


- **A priori:** Invariance to certain transformations: $\forall g \in \mathcal{G} : p(T_g x) = p(x)$.

$x \in \mathcal{X}$

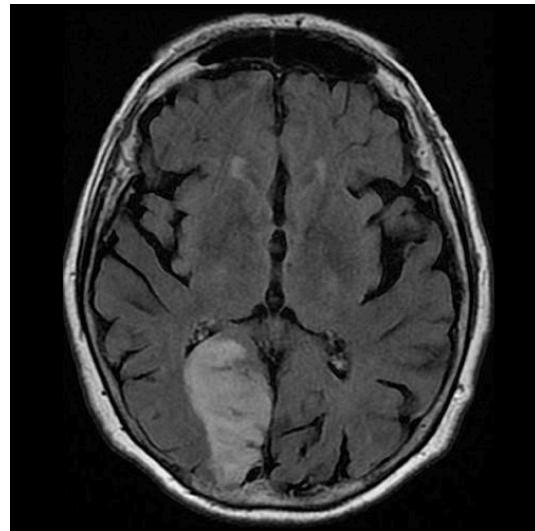


$T_g x \in \mathcal{X}$

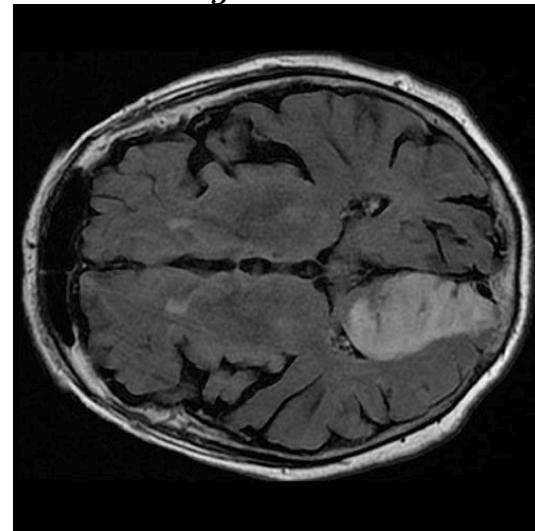


- **A priori:** Invariance to certain transformations: $\forall g \in \mathcal{G} : p(T_g x) = p(x)$.

$$x \in \mathcal{X}$$



$$T_g x \in \mathcal{X}$$



$$y = Ax = AT_g T_g^{-1} x = A_g x'$$

- **Assumptions:**
 - **Single** operator A .
 - **Noiseless** measurements: $y = Ax$.
-

- **Assumptions:**
 - **Single** operator A .
 - **Noiseless** measurements: $\mathbf{y} = \mathbf{A}\mathbf{x}$.
- **Method:**
 - **Split** measurements $\mathbf{y} = [\mathbf{y}_1, \mathbf{y}_2]$ for **different** operators $\mathbf{AT}_g = [\mathbf{A}_1, \mathbf{A}_2]$.

$$f^* \in \operatorname{argmin}_f \mathbb{E}_{\mathbf{y}, f} \{ \mathcal{L}_{\text{ES}}(\mathbf{y}, \mathbf{A}, f) \}$$

$$\begin{aligned} \mathcal{L}_{\text{ES}}(\mathbf{y}, \mathbf{A}, f) &\triangleq \mathbb{E}_g \{ \mathcal{L}_{\text{SPLIT}}(\mathbf{y}, \mathbf{AT}_g, f) \} \\ &= \mathbb{E}_g \mathbb{E}_{\mathbf{y}_1, \mathbf{A}_1 \mid \mathbf{y}, \mathbf{AT}_g} \left\{ \| \mathbf{AT}_g f(\mathbf{y}_1, \mathbf{A}_1) - \mathbf{y} \|^2 \right\} \end{aligned}$$

A visual interpretation of Q_{A_1}

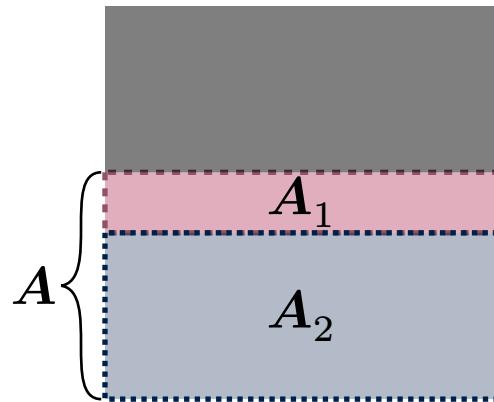
Solution with $Q_{A_1} \triangleq \mathbb{E}_{g \mid A_1} \left\{ (AT_g)^\top AT_g \right\}$:

$$f^*(y_1, A_1) = Q_{A_1}^\dagger Q_{A_1} \underbrace{\mathbb{E}_{x \mid y_1, A_1} \{x\}}_{\text{MMSE estimator}} + \underbrace{\left(I - Q_{A_1}^\dagger Q_{A_1} \right)}_{\text{Error}} v(y_1)$$

A visual interpretation of Q_{A_1}

Solution with $Q_{A_1} \triangleq \mathbb{E}_{g \mid A_1} \left\{ (AT_g)^\top AT_g \right\}$:

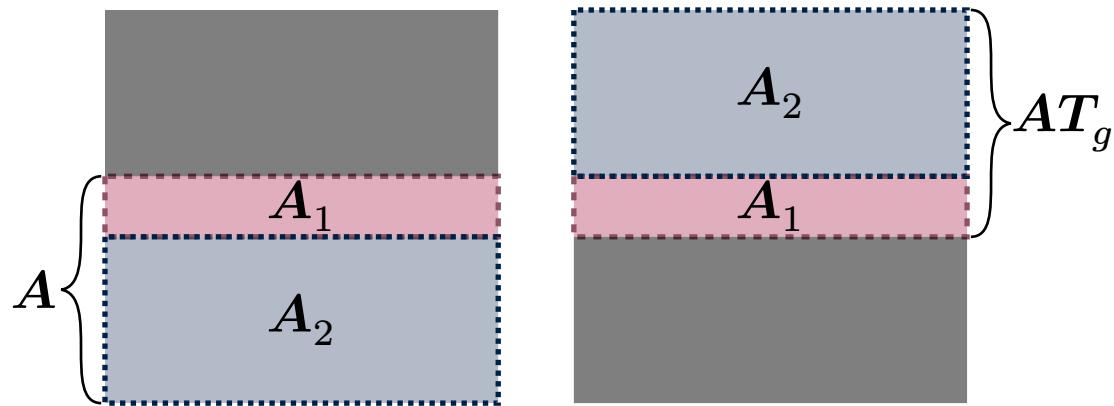
$$f^*(y_1, A_1) = Q_{A_1}^\dagger Q_{A_1} \underbrace{\mathbb{E}_{x \mid y_1, A_1} \{x\}}_{\text{MMSE estimator}} + \underbrace{(I - Q_{A_1}^\dagger Q_{A_1})v(y_1)}_{\text{Error}}$$



A visual interpretation of Q_{A_1}

Solution with $Q_{A_1} \triangleq \mathbb{E}_{g \mid A_1} \left\{ (AT_g)^\top AT_g \right\}$:

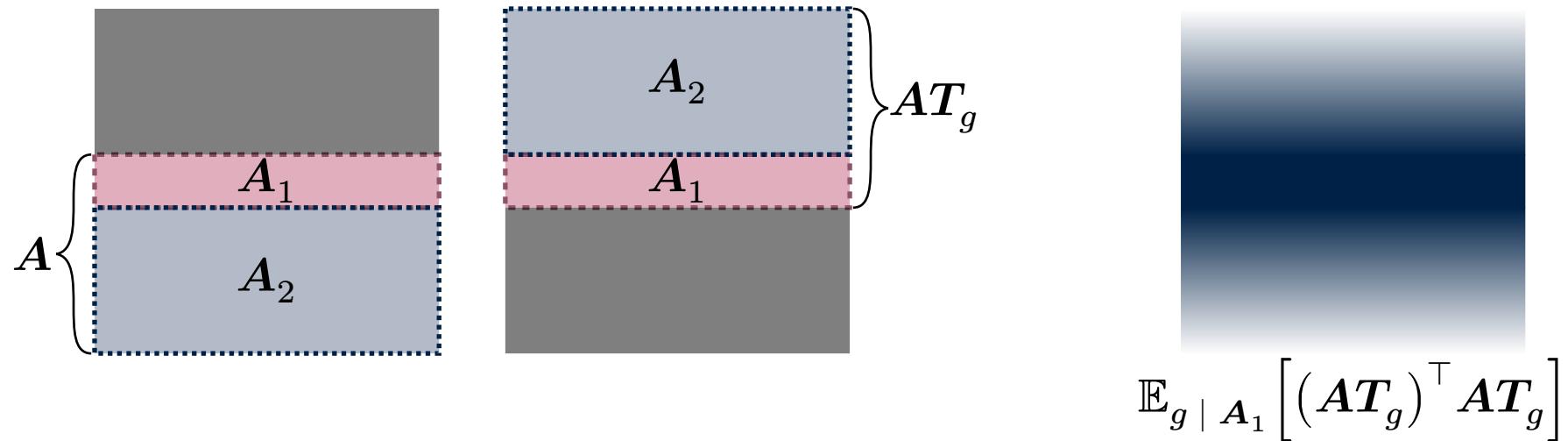
$$f^*(y_1, A_1) = Q_{A_1}^\dagger Q_{A_1} \underbrace{\mathbb{E}_{x \mid y_1, A_1} \{x\}}_{\text{MMSE estimator}} + \underbrace{(I - Q_{A_1}^\dagger Q_{A_1}) v(y_1)}_{\text{Error}}$$



A visual interpretation of Q_{A_1}

Solution with $Q_{A_1} \triangleq \mathbb{E}_{g \mid A_1} \left\{ (AT_g)^\top AT_g \right\}$:

$$f^*(y_1, A_1) = Q_{A_1}^\dagger Q_{A_1} \underbrace{\mathbb{E}_{x \mid y_1, A_1} \{x\}}_{\text{MMSE estimator}} + \underbrace{(I - Q_{A_1}^\dagger Q_{A_1}) v(y_1)}_{\text{Error}}$$



$$f^* \in \operatorname{argmin}_f \mathbb{E}_{\mathbf{y}} \{ \mathcal{L}_{\text{ES}}(\mathbf{y}, A, f) \}$$

- **At test time:** several splits $A_1 \longrightarrow$ **Average** the reconstructions.

-

-

$$f^* \in \operatorname{argmin}_f \mathbb{E}_{\mathbf{y}} \{\mathcal{L}_{\text{ES}}(\mathbf{y}, A, f)\}$$

- **At test time:** several splits $A_1 \rightarrow$ **Average** the reconstructions.
- If the average $\overline{Q}_A \triangleq \mathbb{E}_{A_1 \mid A} \{Q_{A_1}\}$ is **invertible**:

$$\begin{aligned}\overline{f}(\mathbf{y}, A) &\triangleq \mathbb{E}_{\mathbf{y}_1, A_1 \mid \mathbf{y}, A} \left\{ \overline{Q}_A^{-1} Q_{A_1} f^*(\mathbf{y}_1, A_1) \right\} \\ &= \underbrace{\mathbb{E}_{\mathbf{y}_1, A_1 \mid \mathbf{y}, A} \left\{ \overline{Q}_A^{-1} Q_{A_1} \mathbb{E}_{\mathbf{x} \mid \mathbf{y}_1, A_1} \{\mathbf{x}\} \right\}}_{\text{Convex combination of MMSE estimators}}.\end{aligned}$$

-

$$f^* \in \operatorname{argmin}_f \mathbb{E}_{\mathbf{y}} \{ \mathcal{L}_{\text{ES}}(\mathbf{y}, A, f) \}$$

- **At test time:** several splits $A_1 \rightarrow$ **Average** the reconstructions.
- If the average $\overline{\mathbf{Q}}_A \triangleq \mathbb{E}_{A_1 \mid A} \{ \mathbf{Q}_{A_1} \}$ is **invertible**:

$$\begin{aligned} \overline{f}(\mathbf{y}, A) &\triangleq \mathbb{E}_{\mathbf{y}_1, A_1 \mid \mathbf{y}, A} \left\{ \overline{\mathbf{Q}}_A^{-1} \mathbf{Q}_{A_1} f^*(\mathbf{y}_1, A_1) \right\} \\ &= \underbrace{\mathbb{E}_{\mathbf{y}_1, A_1 \mid \mathbf{y}, A} \left\{ \overline{\mathbf{Q}}_A^{-1} \mathbf{Q}_{A_1} \mathbb{E}_{\mathbf{x} \mid \mathbf{y}_1, A_1} \{ \mathbf{x} \} \right\}}_{\text{Convex combination of MMSE estimators}}. \end{aligned}$$

- If $\overline{\mathbf{Q}}_A$ cannot be computed efficiently \rightarrow **non-weighted average**:

$$\overline{f}(\mathbf{y}, A) = \frac{1}{J} \sum_{j=1}^J f^*\left(\mathbf{y}_1^{(j)}, A_i^{(j)}\right) \quad \text{with} \quad \left(\mathbf{y}_1^{(j)}, A_i^{(j)}\right) \sim p(\mathbf{y}_1, A_1 \mid \mathbf{y}, AT_g).$$

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \boldsymbol{\varepsilon}$$

- Divide the \mathcal{L}_{ES} into 2 terms:

$$\mathcal{L}_{\text{ES}}(\mathbf{y}, \mathbf{A}, f) = \mathbb{E}_g \mathbb{E}_{\mathbf{y}_1, \mathbf{A}_1 \mid \mathbf{y}, \mathbf{A} T_g} \left\{ \underbrace{\|\mathbf{A}_1 f(\mathbf{y}_1, \mathbf{A}_1) - \mathbf{y}_1\|^2}_{\text{Measurement consistency (MC)}} + \underbrace{\|\mathbf{A}_2 f(\mathbf{y}_1, \mathbf{A}_1) - \mathbf{y}_2\|^2}_{\text{prediction accuracy}} \right\}.$$

•

•

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \boldsymbol{\varepsilon}$$

- Divide the \mathcal{L}_{ES} into 2 terms:

$$\mathcal{L}_{\text{ES}}(\mathbf{y}, \mathbf{A}, f) = \mathbb{E}_g \mathbb{E}_{\mathbf{y}_1, \mathbf{A}_1 | \mathbf{y}, \mathbf{A} T_g} \left\{ \underbrace{\|\mathbf{A}_1 f(\mathbf{y}_1, \mathbf{A}_1) - \mathbf{y}_1\|^2}_{\text{Measurement consistency (MC)}} + \underbrace{\|\mathbf{A}_2 f(\mathbf{y}_1, \mathbf{A}_1) - \mathbf{y}_2\|^2}_{\text{prediction accuracy}} \right\}.$$

- Replace MC by a self-supervised denoising loss:

$$\mathbb{E}_g \mathbb{E}_{\mathbf{y}_1, \mathbf{A}_1, \boldsymbol{\omega} | \mathbf{y}, \mathbf{A} T_g} \left\{ \underbrace{\left\| \mathbf{A}_1 f(\mathbf{y}_1 + \alpha \boldsymbol{\omega}, \mathbf{A}_1) - \left(\mathbf{y}_1 - \frac{\boldsymbol{\omega}}{\alpha} \right) \right\|^2}_{\text{R2R (denoising) loss}} + \|\mathbf{A}_2 f(\mathbf{y}_1 + \alpha \boldsymbol{\omega}, \mathbf{A}_1) - \mathbf{y}_2\|^2 \right\}.$$

•

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \boldsymbol{\varepsilon}$$

- **Divide** the \mathcal{L}_{ES} into 2 terms:

$$\mathcal{L}_{\text{ES}}(\mathbf{y}, \mathbf{A}, f) = \mathbb{E}_g \mathbb{E}_{\mathbf{y}_1, \mathbf{A}_1 | \mathbf{y}, \mathbf{A} T_g} \left\{ \underbrace{\|\mathbf{A}_1 f(\mathbf{y}_1, \mathbf{A}_1) - \mathbf{y}_1\|^2}_{\text{Measurement consistency (MC)}} + \underbrace{\|\mathbf{A}_2 f(\mathbf{y}_1, \mathbf{A}_1) - \mathbf{y}_2\|^2}_{\text{prediction accuracy}} \right\}.$$

- **Replace** MC by a self-supervised **denoising** loss:

$$\mathbb{E}_g \mathbb{E}_{\mathbf{y}_1, \mathbf{A}_1, \boldsymbol{\omega} | \mathbf{y}, \mathbf{A} T_g} \left\{ \underbrace{\left\| \mathbf{A}_1 f(\mathbf{y}_1 + \alpha \boldsymbol{\omega}, \mathbf{A}_1) - \left(\mathbf{y}_1 - \frac{\boldsymbol{\omega}}{\alpha} \right) \right\|^2}_{\text{R2R (denoising) loss}} + \|\mathbf{A}_2 f(\mathbf{y}_1 + \alpha \boldsymbol{\omega}, \mathbf{A}_1) - \mathbf{y}_2\|^2 \right\}.$$

- $\mathbf{Q}_{\mathbf{A}_1}$ invertible $\Rightarrow f^*(\mathbf{y}_1, \mathbf{A}_1) = \mathbb{E}_{\mathbf{x} | \mathbf{y}_1, \mathbf{A}_1} \{ \mathbf{x} \}.$

Computing the training loss efficiently

- In many methods: $f(\mathbf{y}, \mathbf{A}) = \varphi_{\boldsymbol{\theta}}(\mathbf{A}^\top \mathbf{y})$ or $f(\mathbf{y}, \mathbf{A}) = \varphi_{\boldsymbol{\theta}}(\mathbf{A}^\dagger \mathbf{y})$
- Image-to-image equivariant network: $\varphi(\mathbf{T}_g \mathbf{x}) = \mathbf{T}_g \varphi(\mathbf{x})$

Computing the training loss efficiently

- In many methods: $f(\mathbf{y}, \mathbf{A}) = \varphi_{\boldsymbol{\theta}}(\mathbf{A}^\top \mathbf{y})$ or $f(\mathbf{y}, \mathbf{A}) = \varphi_{\boldsymbol{\theta}}(\mathbf{A}^\dagger \mathbf{y})$
- Image-to-image equivariant network: $\varphi(\mathbf{T}_g \mathbf{x}) = \mathbf{T}_g \varphi(\mathbf{x})$

$$\mathcal{L}_{\text{ES}}(\mathbf{y}, \mathbf{A}, f) = \mathbb{E}_g \mathbb{E}_{\mathbf{y}_1, \mathbf{A}_1 \mid \mathbf{y}, \mathbf{A} \mathbf{T}_g} \left\{ \| \mathbf{A} \mathbf{T}_g f(\mathbf{y}_1, \mathbf{A}_1) - \mathbf{y} \|^2 \right\}$$

Computing the training loss efficiently

- In many methods: $f(\mathbf{y}, \mathbf{A}) = \varphi_{\boldsymbol{\theta}}(\mathbf{A}^\top \mathbf{y})$ or $f(\mathbf{y}, \mathbf{A}) = \varphi_{\boldsymbol{\theta}}(\mathbf{A}^\dagger \mathbf{y})$
- Image-to-image equivariant network: $\varphi(\mathbf{T}_g \mathbf{x}) = \mathbf{T}_g \varphi(\mathbf{x})$

$$\begin{aligned}\mathcal{L}_{\text{ES}}(\mathbf{y}, \mathbf{A}, f) &= \mathbb{E}_g \mathbb{E}_{\mathbf{y}_1, \mathbf{A}_1 \mid \mathbf{y}, \mathbf{A} \mathbf{T}_g} \left\{ \| \mathbf{A} \mathbf{T}_g f(\mathbf{y}_1, \mathbf{A}_1) - \mathbf{y} \|^2 \right\} \\ &= \mathbb{E}_g \mathbb{E}_{\mathbf{M} \mid \mathbf{y}, g} \left\{ \| \mathbf{A} \mathbf{T}_g f(\mathbf{M} \mathbf{y}, \mathbf{M} \mathbf{A} \mathbf{T}_g) - \mathbf{y} \|^2 \right\}\end{aligned}$$

Computing the training loss efficiently

- In many methods: $f(\mathbf{y}, \mathbf{A}) = \varphi_{\boldsymbol{\theta}}(\mathbf{A}^\top \mathbf{y})$ or $f(\mathbf{y}, \mathbf{A}) = \varphi_{\boldsymbol{\theta}}(\mathbf{A}^\dagger \mathbf{y})$
- Image-to-image equivariant network: $\varphi(\mathbf{T}_g \mathbf{x}) = \mathbf{T}_g \varphi(\mathbf{x})$

$$\begin{aligned}\mathcal{L}_{\text{ES}}(\mathbf{y}, \mathbf{A}, f) &= \mathbb{E}_g \mathbb{E}_{\mathbf{y}_1, \mathbf{A}_1 \mid \mathbf{y}, \mathbf{A} \mathbf{T}_g} \left\{ \| \mathbf{A} \mathbf{T}_g f(\mathbf{y}_1, \mathbf{A}_1) - \mathbf{y} \|^2 \right\} \\ &= \mathbb{E}_g \mathbb{E}_{\mathbf{M} \mid \mathbf{y}, g} \left\{ \| \mathbf{A} \mathbf{T}_g f(\mathbf{M} \mathbf{y}, \mathbf{M} \mathbf{A} \mathbf{T}_g) - \mathbf{y} \|^2 \right\} \\ &= \mathbb{E}_g \mathbb{E}_{\mathbf{M} \mid \mathbf{y}, g} \left\{ \| \mathbf{A} \mathbf{T}_g \varphi_{\boldsymbol{\theta}}(\mathbf{T}_g^{-1} \mathbf{A}^\top \mathbf{M}^\top \mathbf{M} \mathbf{y}) - \mathbf{y} \|^2 \right\}\end{aligned}$$

Computing the training loss efficiently

- In many methods: $f(\mathbf{y}, \mathbf{A}) = \varphi_{\boldsymbol{\theta}}(\mathbf{A}^\top \mathbf{y})$ or $f(\mathbf{y}, \mathbf{A}) = \varphi_{\boldsymbol{\theta}}(\mathbf{A}^\dagger \mathbf{y})$
- Image-to-image equivariant network: $\varphi(\mathbf{T}_g \mathbf{x}) = \mathbf{T}_g \varphi(\mathbf{x})$

$$\begin{aligned}\mathcal{L}_{\text{ES}}(\mathbf{y}, \mathbf{A}, f) &= \mathbb{E}_g \mathbb{E}_{\mathbf{y}_1, \mathbf{A}_1 \mid \mathbf{y}, \mathbf{A} \mathbf{T}_g} \left\{ \| \mathbf{A} \mathbf{T}_g f(\mathbf{y}_1, \mathbf{A}_1) - \mathbf{y} \|^2 \right\} \\ &= \mathbb{E}_g \mathbb{E}_{\mathbf{M} \mid \mathbf{y}, g} \left\{ \| \mathbf{A} \mathbf{T}_g f(\mathbf{M} \mathbf{y}, \mathbf{M} \mathbf{A} \mathbf{T}_g) - \mathbf{y} \|^2 \right\} \\ &= \mathbb{E}_g \mathbb{E}_{\mathbf{M} \mid \mathbf{y}, g} \left\{ \| \mathbf{A} \mathbf{T}_g \varphi_{\boldsymbol{\theta}}(\mathbf{T}_g^{-1} \mathbf{A}^\top \mathbf{M}^\top \mathbf{M} \mathbf{y}) - \mathbf{y} \|^2 \right\} \\ &= \mathbb{E}_g \mathbb{E}_{\mathbf{M} \mid \mathbf{y}, g} \left\{ \| \mathbf{A} \mathbf{T}_g \cancel{\mathbf{T}_g^{-1}} \varphi_{\boldsymbol{\theta}}(\mathbf{A}^\top \mathbf{M}^\top \mathbf{M} \mathbf{y}) - \mathbf{y} \|^2 \right\} \text{ if } \varphi \text{ is equivariant!}\end{aligned}$$

Computing the training loss efficiently

- In many methods: $f(\mathbf{y}, \mathbf{A}) = \varphi_{\boldsymbol{\theta}}(\mathbf{A}^\top \mathbf{y})$ or $f(\mathbf{y}, \mathbf{A}) = \varphi_{\boldsymbol{\theta}}(\mathbf{A}^\dagger \mathbf{y})$
- Image-to-image equivariant network: $\varphi(\mathbf{T}_g \mathbf{x}) = \mathbf{T}_g \varphi(\mathbf{x})$

$$\begin{aligned}\mathcal{L}_{\text{ES}}(\mathbf{y}, \mathbf{A}, f) &= \mathbb{E}_g \mathbb{E}_{\mathbf{y}_1, \mathbf{A}_1 \mid \mathbf{y}, \mathbf{A} \mathbf{T}_g} \left\{ \| \mathbf{A} \mathbf{T}_g f(\mathbf{y}_1, \mathbf{A}_1) - \mathbf{y} \|^2 \right\} \\ &= \mathbb{E}_g \mathbb{E}_{\mathbf{M} \mid \mathbf{y}, g} \left\{ \| \mathbf{A} \mathbf{T}_g f(\mathbf{M} \mathbf{y}, \mathbf{M} \mathbf{A} \mathbf{T}_g) - \mathbf{y} \|^2 \right\} \\ &= \mathbb{E}_g \mathbb{E}_{\mathbf{M} \mid \mathbf{y}, g} \left\{ \| \mathbf{A} \mathbf{T}_g \varphi_{\boldsymbol{\theta}}(\mathbf{T}_g^{-1} \mathbf{A}^\top \mathbf{M}^\top \mathbf{M} \mathbf{y}) - \mathbf{y} \|^2 \right\} \\ &= \mathbb{E}_g \mathbb{E}_{\mathbf{M} \mid \mathbf{y}, g} \left\{ \| \mathbf{A} \cancel{\mathbf{T}_g} \cancel{\mathbf{T}_g^{-1}} \varphi_{\boldsymbol{\theta}}(\mathbf{A}^\top \mathbf{M}^\top \mathbf{M} \mathbf{y}) - \mathbf{y} \|^2 \right\} \text{ if } \varphi \text{ is equivariant!} \\ &= \mathbb{E}_{\mathbf{M} \mid \mathbf{y}} \left\{ \| \mathbf{A} f(\mathbf{M} \mathbf{y}, \mathbf{M} \mathbf{A}) - \mathbf{y} \|^2 \right\}\end{aligned}$$

Computing the training loss efficiently

- In many methods: $f(\mathbf{y}, \mathbf{A}) = \varphi_{\boldsymbol{\theta}}(\mathbf{A}^\top \mathbf{y})$ or $f(\mathbf{y}, \mathbf{A}) = \varphi_{\boldsymbol{\theta}}(\mathbf{A}^\dagger \mathbf{y})$
- Image-to-image equivariant network: $\varphi(\mathbf{T}_g \mathbf{x}) = \mathbf{T}_g \varphi(\mathbf{x})$

$$\begin{aligned}
 \mathcal{L}_{\text{ES}}(\mathbf{y}, \mathbf{A}, f) &= \mathbb{E}_g \mathbb{E}_{\mathbf{y}_1, \mathbf{A}_1 \mid \mathbf{y}, \mathbf{A} \mathbf{T}_g} \left\{ \| \mathbf{A} \mathbf{T}_g f(\mathbf{y}_1, \mathbf{A}_1) - \mathbf{y} \|^2 \right\} \\
 &= \mathbb{E}_g \mathbb{E}_{\mathbf{M} \mid \mathbf{y}, g} \left\{ \| \mathbf{A} \mathbf{T}_g f(\mathbf{M} \mathbf{y}, \mathbf{M} \mathbf{A} \mathbf{T}_g) - \mathbf{y} \|^2 \right\} \\
 &= \mathbb{E}_g \mathbb{E}_{\mathbf{M} \mid \mathbf{y}, g} \left\{ \| \mathbf{A} \mathbf{T}_g \varphi_{\boldsymbol{\theta}}(\mathbf{T}_g^{-1} \mathbf{A}^\top \mathbf{M}^\top \mathbf{M} \mathbf{y}) - \mathbf{y} \|^2 \right\} \\
 &= \mathbb{E}_g \mathbb{E}_{\mathbf{M} \mid \mathbf{y}, g} \left\{ \| \mathbf{A} \mathbf{T}_g \cancel{\mathbf{T}_g^{-1}} \varphi_{\boldsymbol{\theta}}(\mathbf{A}^\top \mathbf{M}^\top \mathbf{M} \mathbf{y}) - \mathbf{y} \|^2 \right\} \text{ if } \varphi \text{ is equivariant!} \\
 &= \mathbb{E}_{\mathbf{M} \mid \mathbf{y}} \left\{ \| \mathbf{A} f(\mathbf{M} \mathbf{y}, \mathbf{M} \mathbf{A}) - \mathbf{y} \|^2 \right\} \\
 &= \mathcal{L}_{\text{SPLIT}}(\mathbf{y}, \mathbf{A}, f)
 \end{aligned}$$

Definition

We say that the reconstruction function $f(\mathbf{y}, \mathbf{A})$ is an equivariant reconstructor if

$$f(\mathbf{y}, \mathbf{A}\mathbf{T}_g) = \mathbf{T}_g^{-1} f(\mathbf{y}, \mathbf{A}), \quad \forall \mathbf{y} \in \mathbb{R}^m, \quad \forall g \in \mathcal{G}, \quad \forall \mathbf{A} \in \mathbb{R}^{m \times n}.$$

Definition

We say that the reconstruction function $f(\mathbf{y}, \mathbf{A})$ is an equivariant reconstructor if

$$f(\mathbf{y}, \mathbf{A}\mathbf{T}_g) = \mathbf{T}_g^{-1} f(\mathbf{y}, \mathbf{A}), \quad \forall \mathbf{y} \in \mathbb{R}^m, \quad \forall g \in \mathcal{G}, \quad \forall \mathbf{A} \in \mathbb{R}^{m \times n}.$$

- **Artifact removal and unrolled network.** $f(\mathbf{y}, \mathbf{A}) = \mathbf{x}_L$ with

$$\mathbf{x}_{k+1} = \varphi\left(\mathbf{x}_k - \gamma \nabla_{\mathbf{x}_k} d(\mathbf{A}\mathbf{x}_k, \mathbf{y})\right).$$

-

-

Definition

We say that the reconstruction function $f(\mathbf{y}, \mathbf{A})$ is an equivariant reconstructor if

$$f(\mathbf{y}, \mathbf{A}\mathbf{T}_g) = \mathbf{T}_g^{-1} f(\mathbf{y}, \mathbf{A}), \quad \forall \mathbf{y} \in \mathbb{R}^m, \quad \forall g \in \mathcal{G}, \quad \forall \mathbf{A} \in \mathbb{R}^{m \times n}.$$

- **Artifact removal and unrolled network.** $f(\mathbf{y}, \mathbf{A}) = \mathbf{x}_L$ with
$$\mathbf{x}_{k+1} = \varphi\left(\mathbf{x}_k - \gamma \nabla_{\mathbf{x}_k} d(\mathbf{A}\mathbf{x}_k, \mathbf{y})\right).$$
- **Reynolds averaging.**

$$f(\mathbf{y}, \mathbf{A}) = \frac{1}{|G|} \sum_{g \in \mathcal{G}} \mathbf{T}_g r(\mathbf{y}, \mathbf{A}\mathbf{T}_g).$$

Definition

We say that the reconstruction function $f(\mathbf{y}, \mathbf{A})$ is an equivariant reconstructor if

$$f(\mathbf{y}, \mathbf{A}\mathbf{T}_g) = \mathbf{T}_g^{-1} f(\mathbf{y}, \mathbf{A}), \quad \forall \mathbf{y} \in \mathbb{R}^m, \quad \forall g \in \mathcal{G}, \quad \forall \mathbf{A} \in \mathbb{R}^{m \times n}.$$

- **Artifact removal and unrolled network.** $f(\mathbf{y}, \mathbf{A}) = \mathbf{x}_L$ with

$$\mathbf{x}_{k+1} = \varphi\left(\mathbf{x}_k - \gamma \nabla_{\mathbf{x}_k} d(\mathbf{A}\mathbf{x}_k, \mathbf{y})\right).$$

- **Reynolds averaging.**

$$f(\mathbf{y}, \mathbf{A}) = \frac{1}{|\mathcal{G}|} \sum_{g \in \mathcal{G}} \mathbf{T}_g r(\mathbf{y}, \mathbf{A}\mathbf{T}_g).$$

- **Maximum a posteriori (MAP).**

$$f(\mathbf{y}, \mathbf{A}) = \operatorname{argmax}_{\mathbf{x} \in \mathbb{R}^n} \{p(\mathbf{x} \mid \mathbf{y}, \mathbf{A})\}.$$

.

Definition

We say that the reconstruction function $f(\mathbf{y}, \mathbf{A})$ is an equivariant reconstructor if

$$f(\mathbf{y}, \mathbf{A}\mathbf{T}_g) = \mathbf{T}_g^{-1} f(\mathbf{y}, \mathbf{A}), \quad \forall \mathbf{y} \in \mathbb{R}^m, \quad \forall g \in \mathcal{G}, \quad \forall \mathbf{A} \in \mathbb{R}^{m \times n}.$$

- **Artifact removal and unrolled network.** $f(\mathbf{y}, \mathbf{A}) = \mathbf{x}_L$ with

$$\mathbf{x}_{k+1} = \varphi\left(\mathbf{x}_k - \gamma \nabla_{\mathbf{x}_k} d(\mathbf{A}\mathbf{x}_k, \mathbf{y})\right).$$

- **Reynolds averaging.**

$$f(\mathbf{y}, \mathbf{A}) = \frac{1}{|\mathcal{G}|} \sum_{g \in \mathcal{G}} \mathbf{T}_g r(\mathbf{y}, \mathbf{A}\mathbf{T}_g).$$

- **Maximum a posteriori (MAP).**

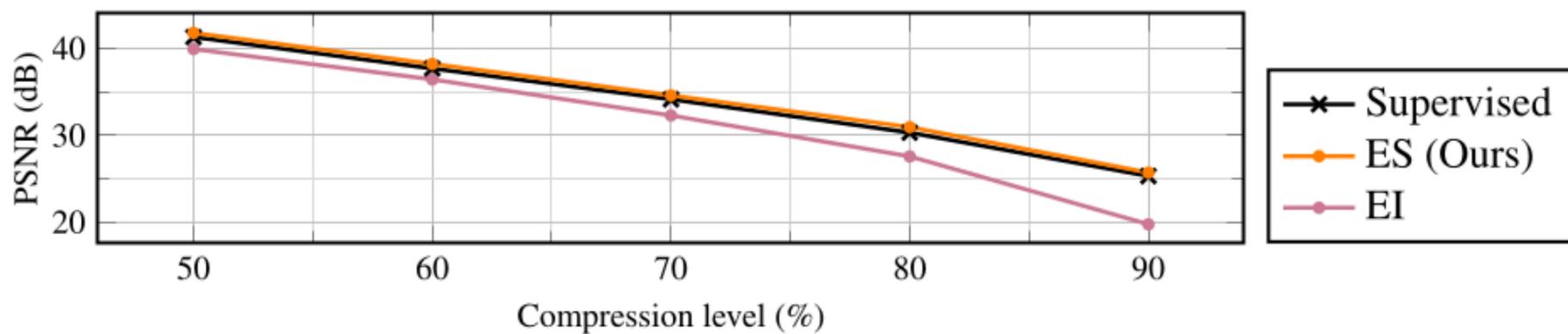
$$f(\mathbf{y}, \mathbf{A}) = \underset{\mathbf{x} \in \mathbb{R}^n}{\operatorname{argmax}} \{p(\mathbf{x} \mid \mathbf{y}, \mathbf{A})\}.$$

- **Minimum mean squared error (MMSE).**

$$f(\mathbf{y}, \mathbf{A}) = \mathbb{E}[\mathbf{x} \mid \mathbf{y}, \mathbf{A}].$$

Problem	Equivariant architecture	Dataset	Transformation
Compressive sensing	Unrolled equivariant UNet	MNIST 20000 images	Translations
Image inpainting	Equivariant UNet	DIV2K 800 images	Translations
MRI	Unrolled UNet + reynold averaging	FastMRI 900 images	Rotation + Flip

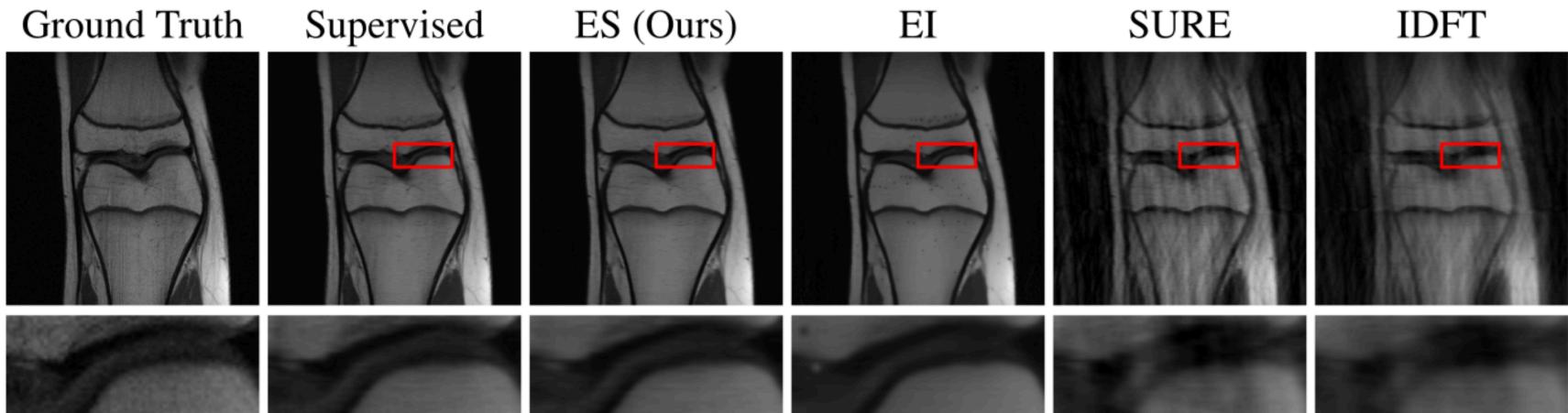
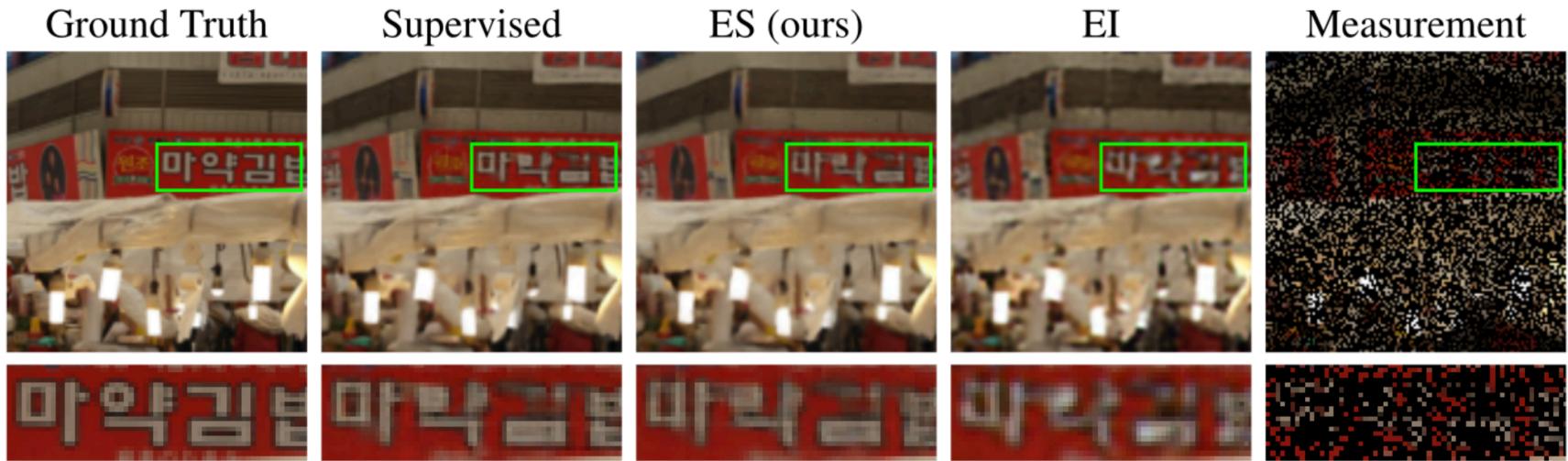
Problem	Equivariant architecture	Dataset	Transformation
Compressive sensing	Unrolled equivariant UNet	MNIST 20000 images	Translations
Image inpainting	Equivariant UNet	DIV2K 800 images	Translations
MRI	Unrolled UNet + reynold averaging	FastMRI 900 images	Rotation + Flip



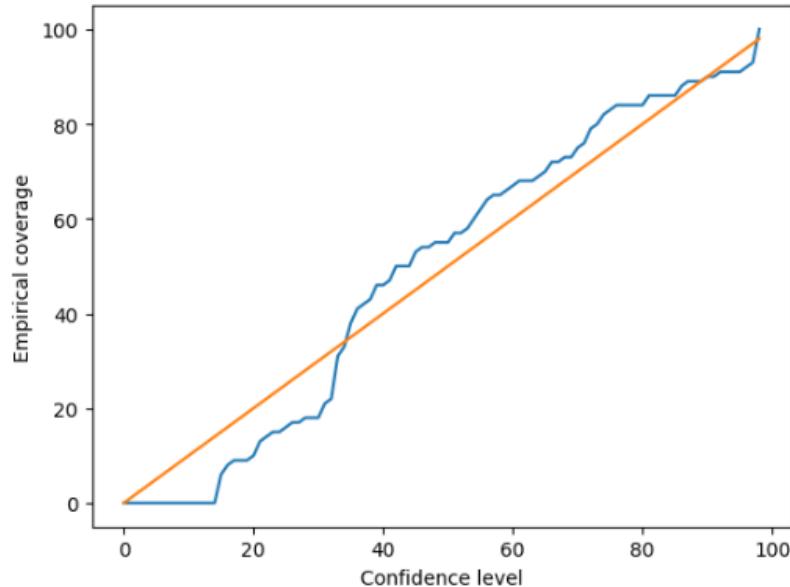
- **Inpainting:** image size: 128×128 , A : keep 30% of the pixels, no noise.
- **MRI:** image size: 265×256 , acceleration: $\times 8$, noise: $\sigma = 0.05$.

Training loss	Eq. arch.	PSNR ↑	
		Inpainting	MRI
Supervised	✓	28.46	28.74
	✗	28.62	28.48
Splitting (Ours)	✓	27.45	28.54
	✗	27.20	28.18
EI	✓	25.89	27.88
	✗	26.33	-
Incomplete image/ IDFT	-	8.22	23.62

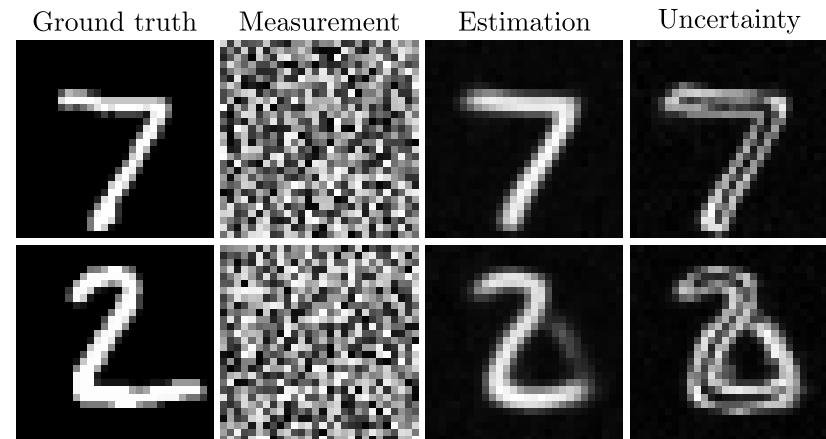
Images



- Collaborative project with Marcelo Pereyra: **Uncertainty Quantification.**



Coverage plot

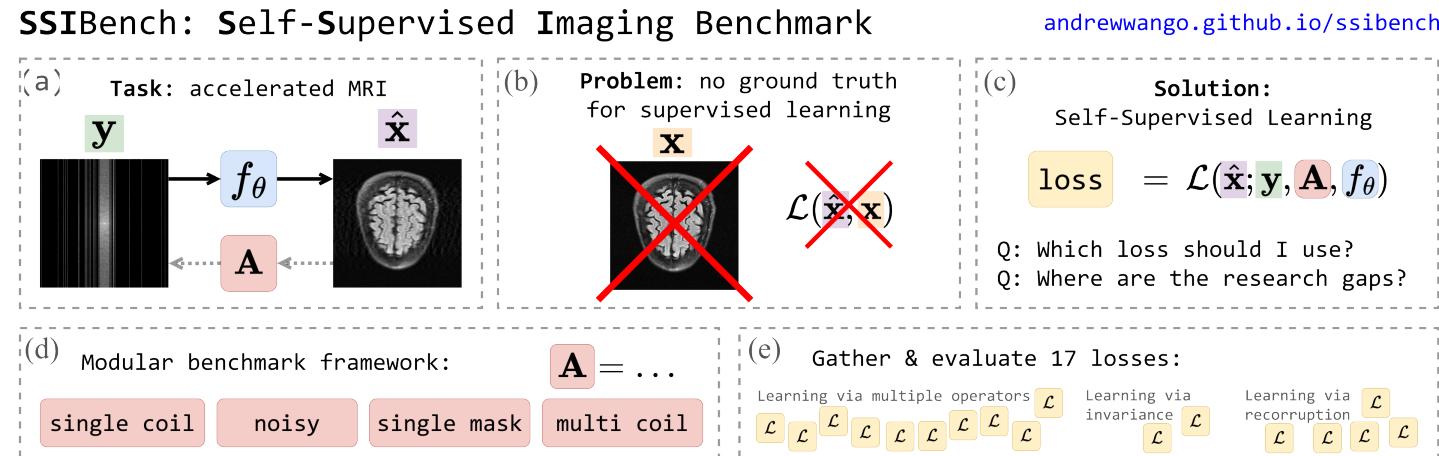


Predicted error

-Code: <https://github.com/vsechaud/Equivariant-Splitting>



-MRI:



-Photographs of flamingos: Émile Sechaud

Appendix

Complementaries results

Method	PSNR ↑	SSIM ↑
Supervised	28.46 ± 2.97	0.8982 ± 0.0411
ES (ours)	27.45 ± 2.86	0.8737 ± 0.0461
EI	25.89 ± 2.65	0.8332 ± 0.0521
Incomplete image	8.22 ± 2.47	0.0973 ± 0.0542

Method	PSNR ↑	SSIM ↑
Supervised	28.74 ± 2.81	0.6445 ± 0.1094
ES (ours)	28.54 ± 2.75	0.6195 ± 0.1188
EI	27.88 ± 2.64	0.5731 ± 0.1299
SURE	24.45 ± 1.86	0.5479 ± 0.0740
IDFT	23.62 ± 1.90	0.5052 ± 0.0900

Training loss	Eq. arch.	Image inpainting	
		PSNR ↑	SSIM ↑
Supervised	✓	28.46 ± 2.97	0.8982 ± 0.0411
	✗	28.62 ± 3.03	0.9002 ± 0.0414
Splitting (Ours)	✓	27.45 ± 2.86	0.8737 ± 0.0461
	✗	27.20 ± 2.83	0.8652 ± 0.0461

Training loss	Eq. arch.	MRI ($\times 8$ Accel.)	
		PSNR ↑	SSIM ↑
Supervised	✓	28.74 ± 2.81	0.6445 ± 0.1094
	✗	28.48 ± 2.68	0.6381 ± 0.1082
Splitting (Ours)	✓	28.54 ± 2.75	0.6195 ± 0.1188
	✗	28.18 ± 2.58	0.6104 ± 0.1176

What is Q_{A_1} for equivariant splitting?

- Q_{A_1} is a **weight** depending on all possible T_g , knowing A_1 .
- The more **probable** T_g is, the more it **impacts** Q_{A_1} .
- Q_{A_1} invertible \Rightarrow the entire space \mathbb{R}^n is covered by all $(AT_g)^\top AT_g$.

