EQUIVARIANT SPLITTING: SELF-SUPERVISED LEARN-ING FROM INCOMPLETE DATA

Anonymous authors

Paper under double-blind review

ABSTRACT

Self-supervised learning for inverse problems allows to train a reconstruction network from noise and/or incomplete data alone. These methods have the potential of enabling learning-based solutions when obtaining ground-truth references for training is expensive or even impossible. In this paper, we propose a new self-supervised learning strategy devised for the challenging setting where measurements are observed via a single incomplete observation model. We introduce a new definition of equivariance in the context of reconstruction networks, and show that the combination of self-supervised splitting losses and equivariant reconstruction networks results in unbiased estimates of the supervised loss. Through a series of experiments on image inpainting, accelerated magnetic resonance imaging, and compressive sensing, we demonstrate that the proposed loss achieves state-of-the-art performance in settings with highly rank-deficient forward models.

1 Introduction

Inverse problems are ubiquitous in many sensing and imaging applications. They are written as

$$y = Ax + \varepsilon \tag{1}$$

where $A \in \mathbb{R}^{m \times n}$ is the known forward matrix, $x \in \mathbb{R}^n$ is the ground truth image to be estimated, $y \in \mathbb{R}^m$ is the observed measurement vector, and $\varepsilon \in \mathbb{R}^m$ is the unknown noise, generally assumed to follow a Gaussian distribution. This model is suitable for many imaging modalities, including magnetic resonance imaging (MRI) for medical imaging (Zbontar et al., 2019), computed tomography (CT) (Withers et al., 2021), microscopy (Ragone et al., 2023), remote sensing (Fassnacht et al., 2024), and astronomical imaging (Vojtekova et al., 2021).

The number of linearly independent measurements is often smaller than the number of pixels in the target images due to physical and practical constraints. In this case, the forward matrix is rank-deficient, the forward process discards some of the information present in the target image. Further information about the target images is needed to solve the problem and different methods make different assumptions about the signal distribution.

Modern learning-based solvers generally obtain state-of-the-art performance by training on supervised pairs of ground-truth images and measurements. For certain applications, it is expensive or even impossible to obtain enough ground-truth data for supervised training (Belthangady & Royer, 2019). This is notably the case in astronomical imaging, microscopy and medical imaging.

Recent self-supervised methods overcome this limitation by learning to reconstruct without ground-truth data, relying only on a dataset of measurements, and they generally differ in the assumption they make on the forward model. For denoising problems where the forward matrix is the identity mapping, certain methods rely on knowing the exact noise distribution (Eldar, 2009; Pang et al., 2021; Monroy et al., 2025), others only assume it is entry-wise independent (Krull et al., 2019), while some make intermediate assumptions (Tachella et al., 2025a).

In settings where measurements are observed via multiple incomplete forward operators, such as accelerated MRI with masks varying across acquisitions or inpainting problems with missing pixels varying across images, the main approaches are splitting (Yaman et al., 2020; Millard & Chiew, 2023) and consistency across operators (Tachella et al., 2022). Splitting losses divide the measurements into input and target components, and are unbiased estimators of the supervised loss if there is enough diversity of operators in the dataset (Daras et al., 2023; Millard & Chiew, 2023).

In the more challenging case of measurement data obtained via a single incomplete forward operator, the main self-supervised approach is equivariant imaging (EI) (Chen et al., 2021; 2022) which makes the assumption that the target distribution is invariant to a certain group of transformation including geometric transformations (Wang & Davies, 2024; Scanvic et al., 2025) and range transformations such as intensity scalings (Sechaud et al., 2024). Experimental results show that it can obtain competitive performances to supervised methods even though it does not require ground-truth references for training (Chen et al., 2022). However, training with EI is typically slower than supervised learning, as it requires two to three evaluations of the model for every iteration, and can obtain subpar performances if the operator is highly incomplete.

In this work, we propose equivariant splitting (ES), a new self-supervised method for learning from measurements obtained via a single forward operator that combines the invariance to transformations assumption of equivariant imaging and the simplicity and computational efficiency of splitting methods. The key idea behind our method is the use of recent developments in equivariant architectures (Chaman & Dokmanić, 2021a; Puny et al., 2022) to design a training loss that performs implicit ground truth data augmentation without any transformation overhead. Our theoretical results show that our method yields (in expectation) the minimum mean squared error (MMSE) estimator as long as the model is expressive enough, which is not guaranteed for equivariant imaging and previous splitting methods. Additionally, our experiments demonstrate state-of-the-art performance on a wide range of self-supervised imaging problems including compressed sensing, image inpainting and accelerated MRI. This work is additional evidence that architectural constraints built upon equivariance are a powerful tool to solve ill-posed imaging inverse problems.

Our contributions are the following:

- 1. We propose a new definition for equivariance in inverse problems, and propose architectures that satisfy this property, including unrolled architectures.
- 2. We propose a new self-supervised loss that leverages equivariant networks (according to our definition above) whose global minimizer is the gold standard MMSE estimator under the assumption of an invariant signal distribution.
- 3. We demonstrate the performance of our method on a wide range of inverse problems including compressed sensing, inpainting and accelerated MRI.

2 RELATED WORK

Measurement splitting Various self-supervised losses consist of dividing measurement vectors into two components, one used as input and the other as target. It has been used to solve oversampled single-operator inverse problems including full-view CT (Hendriksen et al., 2020), and also undersampled multi-operator inverse problems with theoretical guarantees of producing estimates equivalent to supervised estimates in expectation, notably accelerated MRI and image inpainting with varying masks (Daras et al., 2023; Millard & Chiew, 2023). To the best of our knowledge, this work is the first to extend these methods to the more challenging single-operator undersampled setting, which notably includes sparse-view CT and accelerated MRI (with fixed mask).

Equivariant imaging It is possible to learn from incomplete data associated with a single degradation operator, as long as the underlying signal distribution is invariant to a group of transformations (Tachella et al., 2023). Equivariant imaging (Chen et al., 2021; 2022) assumes that the distribution of clean images remains unchanged under certain transformations, including translations, rotations and flips, and introduces a training loss that enforces the equivariance to these transformations of the entire measurement-reconstruction process, thereby constraining the set of learnable models. It has been used effectively to solve various inverse problems using adequate groups of transformations (Wang & Davies, 2024; Sechaud et al., 2024), but it is computationally expensive due to the two to three network evaluations. Moreover, the EI loss is not necessarily an unbiased estimator of the supervised loss, and it is unclear whether this approach recovers the optimal MMSE estimator in expectation.

Equivariant neural networks The design of equivariant networks is an active research topic and many different approaches exist. Some rely on data augmentation to make neural networks more equivariant to rotations and flips, while other rely on averaging on the group of transformations at test time (Rivera et al., 2021; Puny et al., 2022; Kaba et al., 2023; Sannai et al., 2024). A third line

of work focuses on architectures that are equivariant by design: Cohen & Welling (2016; 2017) propose convolutional layers that are equivariant to rotations and flips, and other works also rely on the design of more equivariant layers to improve translation-equivariance. Zhang (2019); Chaman & Dokmanić (2021b) propose equivariant pooling and downsampling/upsampling layers, and Karras et al. (2021); Michaeli et al. (2023) equivariant non-linearities and activation layers. Closer to our work, in the specific setting of inverse problems, Celledoni et al. (2021) propose the use of equivariant denoising blocks within unrolled architectures, and Terris et al. (2024) similarly propose to render plug-and-play denoisers more equivariant by averaging over transformations at test time. We take these analyses further by providing a clear definition of equivariance in inverse problems and showing which popular posterior estimators and related architectures verify these properties.

3 BACKGROUND

Solving the inverse problem in eq. (1) amounts to designing a reconstruction function $f(\boldsymbol{y}, \boldsymbol{A}) \approx \boldsymbol{x}$ estimating a ground truth signal $\boldsymbol{x} \in \mathbb{R}^n$ from its measurement vector $\boldsymbol{y} \in \mathbb{R}^m$ and forward matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$. In practice, it is often implemented as a neural network parametrized by a set of weights. Supervised methods assume the existence of a finite dataset containing pairs of ground truth signals and measurements $\{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i \in \mathcal{I}}$ that are used to learn the reconstruction function $f(\boldsymbol{y}, \boldsymbol{A})$. The main approach is to minimize a training loss equal to the mean squared error (Ongie et al., 2019)

$$\min_{f} \left\{ \frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} \mathcal{L}_{SUP}(\boldsymbol{x}_i, \boldsymbol{y}_i, \boldsymbol{A}, f) \right\}, \quad \mathcal{L}_{SUP}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{A}, f) = \|f(\boldsymbol{y}, \boldsymbol{A}) - \boldsymbol{x}\|^2.$$
 (2)

While this approach obtains state-of-the-art performance, it cannot be used in the absence of ground-truth data. Self-supervised methods overcome this limitation using a finite dataset containing only measurements $\{y_i\}_{i\in\mathcal{I}}$, and a training loss $\mathcal{L}(y, A, f)$ that need no ground truth data to be evaluated, and which is designed to approximate well the supervised objective in eq. (2).

In denoising problems (A = I) with Gaussian noise of known variance, Recorrupted2Recorrupted (R2R) (Pang et al., 2021) and SURE (Metzler et al., 2020) provide unbiased estimators of the supervised loss. The R2R loss is computed by adding synthetic Gaussian noise $\omega \sim \mathcal{N}(\mathbf{0}, \sigma^2 I)$ to the measurements, creating input-target pairs as $(y + \alpha \omega, y - \frac{\omega}{\alpha})$ for some $\alpha \in (0, +\infty)$. However, if measurements are observed by an incomplete operator A with a non-trivial nullspace, these losses fail to learn in the nullspace, approximating $f(y, A) = A^{\dagger} A \mathbb{E}_{x|y,A} \{x\} + v(y, A)$, with v being any function taking values in the nullspace of A (Chen et al., 2021).

In the rest of this section, we present the two main self-supervised losses that can learn beyond the nullspace, measurement splitting in Section 3.1 and equivariant imaging in Section 3.2. By combining their main ideas, we obtain our new self-supervised loss introduced in Section 4.

3.1 MEASUREMENT SPLITTING

One strategy to address the limitations imposed by the nullspace of a single operator is to employ multiple operators (Millard & Chiew, 2023). The key idea is that different operators generally do not share the same nullspace; thus, observing measurements through the image spaces of multiple operators allows access to the whole space \mathbb{R}^n . Formally, they assume that measurements are obtained according to $\mathbf{y} \sim p(\mathbf{y}|\mathbf{A}\mathbf{x})$ where the measurement operator \mathbf{A} is itself drawn from a distribution $p(\mathbf{A})$, and differ for each acquisition.

Splitting losses divide the measurements into two components $\boldsymbol{y} = [\boldsymbol{y}_1^\top, \boldsymbol{y}_2^\top]^\top$ with corresponding operators $\boldsymbol{A} = [\boldsymbol{A}_1^\top, \boldsymbol{A}_2^\top]^\top$. The network is then trained to predict the entire measurements vector \boldsymbol{y} from only one of its two components \boldsymbol{y}_1 , using the training loss

$$\mathcal{L}_{\text{SPLIT}}(\boldsymbol{y}, \boldsymbol{A}, f) = \mathbb{E}_{\boldsymbol{y}_1, \boldsymbol{A}_1 | \boldsymbol{y}, \boldsymbol{A}} \left\{ \| \boldsymbol{A} f(\boldsymbol{y}_1, \boldsymbol{A}_1) - \boldsymbol{y} \|^2 \right\}, \tag{3}$$

where $p(y_1, A_1 \mid y, A)$ is a random splitting distribution, which is chosen on a per-problem basis. The loss encourages the model to learn in the nullspace of each operator by predicting the unobserved part. In practice, the expectation in eq. (3) is estimated using a single split y_1 for each training batch.

3.2 EQUIVARIANT IMAGING

Equivariant imaging (Chen et al., 2021; 2022) relies on the assumption that the distribution of images is invariant under a group of transformations T_g , $g \in \mathcal{G}$, to learn beyond the nullspace of measurements obtained via a *single* operator. In this setting, the reconstruction model is expected to be able to estimate ground truth images x as well as their transformations $T_g x$ in a coherent manner, i.e., such that the entire measurement-reconstruction pipeline f(Ax) is equivariant with respect to the transformations $f(AT_g x, A) = T_g f(Ax, A)$. In order to achieve this, they propose a self-supervised loss which consists in a traditional measurement consistency term (replaced by SURE in the presence of noise), along with an equivariance-promoting term

$$\mathcal{L}_{EI}(\boldsymbol{y}, \boldsymbol{A}, f) = \|\boldsymbol{A}f(\boldsymbol{y}, \boldsymbol{A}) - \boldsymbol{y}\|^2 + \lambda \mathbb{E}_q \{ \|\boldsymbol{T}_q f(\boldsymbol{y}, \boldsymbol{A}) - f(\boldsymbol{A}\boldsymbol{T}_q f(\boldsymbol{y}, \boldsymbol{A}), \boldsymbol{A})\|^2 \}, \tag{4}$$

where $\lambda>0$ is a trade-off coefficient. Even though it has been shown to be particularly effective on a wide variety of problems (Wang & Davies, 2024; Sechaud et al., 2024), it typically requires from two to three evaluations of the neural network which makes it very computationally expensive (Xu et al., 2025). Moreover, the equivariant loss is only effective at enforcing equivariance when the learned estimator achieves almost perfect reconstructions, $f(y, A) \approx x$ (Chen et al., 2021), which is not the case for very ill-conditioned problems and which leads to the method having multiple possible solutions in general.

4 Method

In this section, we present our method that combines measurement splitting and EI introduced in Section 3. In order to learn from incomplete measurements obtained via a single operator, we rely on the same assumption of EI:

Assumption 1. The distribution of ground truth images p(x) is invariant to the transformations $\{T_g\}_{g\in\mathcal{G}}$

$$p(T_g x) = p(x), \forall g \in \mathcal{G}, \forall x \in \mathbb{R}^n.$$
 (5)

The set of transformations is a design choice of the method to be chosen on a per-problem basis. Corollary 1 helps to choose them for specific problems, and we use it in the experiments.

In Section 4.1, we present our proposed loss and state its optimality under mild assumptions in Theorem 1 and Proposition 1. In Section 4.2, we present a new definition of equivariance for reconstruction functions and state sufficient conditions for common architectures to be equivariant in Theorem 2. In Section 4.3, we finally present a computational synergy between our loss and these equivariant architectures stated in Theorem 3. See Appendix C for the detailed proofs.

4.1 Proposed loss

Under Assumption 1, the measurements can be understood in a different way than they are traditionally. Indeed, measurements y are generally thought as being associated to the ground truth image x and the forward matrix A, but they can equally be understood as being associated to the virtual ground truth image $x_g = T_g^{-1}x$ and the virtual forward matrix $A_g = AT_g$, with

$$y = Ax + \varepsilon = AT_gT_g^{-1}x + \varepsilon = A_gx_g + \varepsilon.$$
 (6)

The implicit multi-operator structure (with operators $\{AT_g\}_{g\in\mathcal{G}}$) of the problem hints that we should be able to leverage the splitting approaches presented in Section 3. Moreover, since we use the same invariance assumption as EI, we can combine the two approaches to obtain equivariant splitting (ES), a new self-supervised loss \mathcal{L}_{ES} that has the advantages of both methods.

Noiseless measurements The ES self-supervised loss is expressed as

$$\mathcal{L}_{ES}(\boldsymbol{y}, \boldsymbol{A}, f) \triangleq \mathbb{E}_{g} \left\{ \mathcal{L}_{SPLIT}(\boldsymbol{y}, \boldsymbol{A}\boldsymbol{T}_{g}, f) \right\}$$
(7)

$$= \mathbb{E}_g \left\{ \mathbb{E}_{\boldsymbol{y}_1, \boldsymbol{A}_1 | \boldsymbol{y}, \boldsymbol{A} \boldsymbol{T}_g} \left\{ \| \boldsymbol{A} \boldsymbol{T}_g f(\boldsymbol{y}_1, \boldsymbol{A}_1) - \boldsymbol{y} \|^2 \right\} \right\}$$
(8)

where $A_1 \sim p(A_1|AT_g) \triangleq p(A_1|g)$ is a random splitting of AT_g .

Theorem 1. In the case of noiseless measurements with p(x) \mathcal{G} -invariant Assumption 1, if the matrix $Q_{A_1} \triangleq \mathbb{E}_{g|A_1} \left\{ (AT_g)^\top AT_g \right\}$ has full rank for some split A_1 , then the splitting method yields the same MMSE-optimal reconstructions as the supervised method, i.e.,

$$f^*(\boldsymbol{y}_1, \boldsymbol{A}_1) = \mathbb{E}_{\boldsymbol{x}|\boldsymbol{y}_1, \boldsymbol{A}_1} \left\{ \boldsymbol{x} \right\}. \tag{9}$$

Having a matrix Q_{A_1} invertible is a sufficient condition for sharing the minimizer of the supervised loss, but not a necessary one. At test time, one may employ several splits A_1 and average the corresponding reconstructions.

Proposition 1. If the matrix $\bar{Q}_{A} \triangleq \mathbb{E}_{A_1|A} \{Q_{A_1}\}$ is invertible and f minimizes $\mathbb{E}_{y} \{\mathcal{L}_{ES}(y, A, f)\}$. Then the reconstruction function

$$\overline{f}(\boldsymbol{y}, \boldsymbol{A}) \triangleq \mathbb{E}_{\boldsymbol{y}_1, \boldsymbol{A}_1 | \boldsymbol{y}, \boldsymbol{A}} \left\{ \bar{\boldsymbol{Q}}_{\boldsymbol{A}}^{-1} \boldsymbol{Q}_{\boldsymbol{A}_1} f(\boldsymbol{y}_1, \boldsymbol{A}_1) \right\}$$
(10)

satisfies

$$\overline{f}(\boldsymbol{y}, \boldsymbol{A}) = \mathbb{E}_{\boldsymbol{y}_1, \boldsymbol{A}_1 | \boldsymbol{y}, \boldsymbol{A}} \left\{ \bar{\boldsymbol{Q}}_{\boldsymbol{A}}^{-1} \boldsymbol{Q}_{\boldsymbol{A}_1} \mathbb{E}_{\boldsymbol{x} | \boldsymbol{y}_1, \boldsymbol{A}_1} \left\{ \boldsymbol{x} \right\} \right\}. \tag{11}$$

where eq. (11) is a convex combination of MMSE estimators for different splittings.

In practice, often neither \bar{Q}_A nor Q_{A_1} can be computed in closed-form, and we use a non-weighted average over random splittings

$$\overline{f}(\boldsymbol{y}, \boldsymbol{A}) := \frac{1}{J} \sum_{j=1}^{J} f(\boldsymbol{y}_{1}^{(j)}, \boldsymbol{A}_{1}^{(j)}) \text{ with } (\boldsymbol{y}_{1}^{(j)} \boldsymbol{A}_{1}^{(j)}) \sim p(\boldsymbol{y}_{1} \boldsymbol{A}_{1} | \boldsymbol{y}, \boldsymbol{A} \boldsymbol{T}_{g})$$
(12)

where g is chosen randomly over the group of transformations for each split.

As with EI, the forward operator should not be equivariant with respect to the choice of transformations, in order to learn beyond the nullspace of the operator:

Corollary 1. In order for the matrices Q_{A_1} or \bar{Q}_A to have full rank, it is necessary that A is not equivariant:

$$\exists g \in \mathcal{G}, AT_q \neq T_q A. \tag{13}$$

Noisy measurements The ES loss can be split into two separate terms, one enforcing measurement consistency, and the other prediction accuracy:

$$\mathcal{L}_{\mathrm{ES}}(\boldsymbol{y},\boldsymbol{A},f) = \mathbb{E}_g \left\{ \mathbb{E}_{\boldsymbol{y}_1,\boldsymbol{A}_1|\boldsymbol{y},\boldsymbol{A}\boldsymbol{T}_q} \left\{ \|\boldsymbol{A}_1 f(\boldsymbol{y}_1,\boldsymbol{A}_1) - \boldsymbol{y}_1\|^2 + \|\boldsymbol{A}_2 f(\boldsymbol{y}_1,\boldsymbol{A}_1) - \boldsymbol{y}_2\|^2 \right\} \right\},$$

where A_1 and A_2 are a splitting of AT_g . If the measurements are noisy, the first term can be replaced by a self-supervised denoising loss. In particular, if measurements are corrupted by Gaussian noise of standard deviation σ , we replace the first term by the R2R loss, yielding:

$$\mathcal{L}_{\text{G-ES}}(\boldsymbol{y}, \boldsymbol{A}, f) = \mathbb{E}_{g, \boldsymbol{y}_1, \boldsymbol{A}_1, \boldsymbol{\omega} | \boldsymbol{y}, \boldsymbol{A} \boldsymbol{T}_g} \left\{ \left\| \boldsymbol{A}_1 f(\boldsymbol{y}_1 + \alpha \boldsymbol{\omega}, \boldsymbol{A}_1) - \left(\boldsymbol{y}_1 - \frac{\boldsymbol{\omega}}{\alpha} \right) \right\|^2 + \left\| \boldsymbol{A}_2 f(\boldsymbol{y}_1 + \alpha \boldsymbol{\omega}, \boldsymbol{A}_1) - \boldsymbol{y}_2 \right\|^2 \right\}$$

with $\omega \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ and a hyper-parameter $\alpha \in (0, +\infty)$. Since R2R provides an unbiased estimate of the clean measurement consistency term (Pang et al., 2021), we can apply Theorem 1 to show that minimizing this loss (in expectation) also results in MMSE estimators (if the conditions on Q_{A_1} or \bar{Q}_A are verified). In the case of non-Gaussian noise, the R2R loss can be replaced by its non-Gaussian extension (Monroy et al., 2025). As with random splits, the expectation over ω is computed using a random realization per batch. At test time, we modify eq. (12) to average over both splits and synthetic noise additions.

4.2 EQUIVARIANT RECONSTRUCTORS

The ES loss requires a model evaluation for every mask and transformation. We show that, instead of sampling a random transformation each evaluation, imposing architectural equivariance constraints removes the need to explicitly compute the transforms.

Image-to-image functions $\phi(x)$ are equivariant if they satisfy (Cohen & Welling, 2016)

 $\phi(T_q \mathbf{x}) = T_q \phi(\mathbf{x}), \ \forall \mathbf{x} \in \mathbb{R}^n, \ \forall g \in \mathcal{G}.$ (14)

In this work, we introduce an extension of this definition to reconstruction functions f(y, A). To the best of our knowledge, this is the first work that introduces this definition.

Definition 1. We say that the reconstruction function f(y, A) is an equivariant reconstructor if

$$f(\boldsymbol{y}, \boldsymbol{A}\boldsymbol{T}_g) = \boldsymbol{T}_g^{-1} f(\boldsymbol{y}, \boldsymbol{A}), \ \forall \boldsymbol{y} \in \mathbb{R}^m, \forall g \in \mathcal{G}, \forall \boldsymbol{A} \in \mathbb{R}^{m \times n}.$$
(15)

This property is very general and the class of classical reconstruction functions that satisfy it is large. **Theorem 2.** *The reconstruction functions defined in points below are all equivariant as in eq.* (15).

1 Artifact ron

1. Artifact removal network. For a denoiser $\phi(\mathbf{x})$ equivariant in the sense of eq. (14),

$$f(\boldsymbol{y}, \boldsymbol{A}) = \phi(\boldsymbol{A}^{\top} \boldsymbol{y}), or \ f(\boldsymbol{y}, \boldsymbol{A}) = \phi(\boldsymbol{A}^{\dagger} \boldsymbol{y}).$$
 (16)

2. Unrolled network. For $\phi(x)$ equivariant, any $\gamma \in \mathbb{R}$ and data fidelity $d(\mathbf{A}x, y)$, with

$$m{x}_0=m{0},\quad m{x}_{k+1}=\phiig(m{x}_k-\gamma
abla_{m{x}_k}d(m{A}m{x}_k,m{y})ig)$$
 for $k=0,\ldots,L-1$ and $f(m{y},m{A})=m{x}_L.$

3. **Reynolds averaging.** For a possibly non-equivariant reconstructor r(y, A), with

$$f(\boldsymbol{y}, \boldsymbol{A}) = \frac{1}{|\mathcal{G}|} \sum_{g \in \mathcal{G}} T_g r(\boldsymbol{y}, \boldsymbol{A} T_g).$$
(18)

(17)

4. Maximum a posteriori (MAP). For a distribution p(x) invariant as in eq. (5), with

$$f(\boldsymbol{y}, \boldsymbol{A}) = \underset{\boldsymbol{x} \in \mathbb{R}^n}{\operatorname{argmax}} \left\{ p(\boldsymbol{x} \mid \boldsymbol{y}, \boldsymbol{A}) \right\}.$$
 (19)

5. Minimum mean squared error (MMSE). For a distribution p(x) invariant as in eq. (5),

$$f(\mathbf{y}, \mathbf{A}) = \mathbb{E}_{\mathbf{x}|\mathbf{y}, \mathbf{A}} \left\{ \mathbf{x} \right\}. \tag{20}$$

For additional motivation and details about these reconstructor architectures, see Appendix A.

4.3 Efficient loss evaluation with equivariant reconstructors

For equivariant reconstructors, the ES loss in eq. (7) reduces to the splitting loss in eq. (3).

 Theorem 3. If f(A, x) is an equivariant reconstructor, then ES is equivalent to the splitting loss

$$\mathcal{L}_{ES}(\boldsymbol{y}, \boldsymbol{A}, f) = \mathcal{L}_{SPLIT}(\boldsymbol{y}, \boldsymbol{A}, f). \tag{21}$$

We emphasize that the condition for a reconstructor to be equivariant is different from the condition enforced by the EI loss, i.e., $f(AT_gx, A) = T_gf(Ax, A)$. They are only equivalent if the reconstruction function is a perfect one-to-one mapping over all possible images.

In our experiments, we build equivariant reconstructors using i) artifact removal networks with a translation equivariant UNet denoiser Chaman & Dokmanić (2021b) and ii) unrolled networks with a denoiser architecture equivariant to rotations and flips via averaging (Sannai et al., 2021).

5 EXPERIMENTS

We assess the effectiveness of the proposed self-supervised loss using experiments conducted on different inverse problems. For each experiment, we train a model corresponding to our method as well as baseline methods and compare their performance. The inverse problems we consider are 1) inpainting, 2) compressive sensing and 3) accelerated MRI. We also validate our theoretical

predictions by testing the effect of using an equivariant architecture. In Section 5.1 we detail our experiments on compressive sensing, in Section 5.2 on image inpainting and in Section 5.3 on accelerated MRI, and in Section 5.4 we present an ablation study on the effect of equivariant architectures. For additional details about the experiments, see Appendix B.

In each experiment, we compare against multiple baselines: a supervised baseline using the supervised loss described in eq. (2), the EI baseline described in eq. (4), a measurement consistency baseline (MC and SURE) Eldar (2009) and a learning-free baseline which are either the measurements directly, or their image under the adjoint or the pseudo-inverse of the forward operator. We use the same architecture and the same training procedure for every method to ensure a fair comparison. We report the peak signal-to-noise ratio (PSNR) and the structural similarity index measure (SSIM) (Wang et al., 2004) of the final reconstructions for the different methods. They are distortion metrics indicating how close the reconstructions are to the ground truth images. We do not include perception metrics which are known to be at odds with them (Blau & Michaeli, 2018). In each case, we also compute equivariance metrics (EQUIV) for translations or rotations and flips.

We design a model architecture from the principles introduced in Section 4, with a variant equivariant to shifts, one equivariant to rotations and flips, and one without equivariance. It uses an existing unrolled architecture (Aggarwal et al., 2019) with a prior step implemented as a standard UNet (Ronneberger et al., 2015), or that of Chaman & Dokmanić (2021b) to enforce the equivariance to shifts. It also optionally uses Reynolds' averaging to enforce the equivariance to rotations and flips. We use a single equivariant variant dictated by Corollary 1 for each problem, the shift-equivariant one for compressive sensing and inpainting, and that equivariant to rotations and flips for MRI.

5.1 Compressive sensing

We use images from MNIST to serve as ground truth. Measurements are obtained by multiplying the ground truth images viewed as vectors of size 784 instead of 28×28 pixel images by the same (known) rectangular measurement matrix with fewer columns than rows, without adding further noise. Figure 1 shows the performance of the different methods for the different compression rates. Our method performs almost as well as the supervised baseline, while the equivariant imaging baseline performs close to the supervised baseline only for higher compression rates.

5.2 IMAGE INPAINTING

The dataset consists of 128×128 images from DIV2K (Agustsson & Timofte, 2017) corrupted with a single binary mask which keeps about 30% of the pixels, without additional noise. Among the 900 images in the dataset, 800 are used for training while the remaining 100 are used for testing. For the supervised method, we use different crops at each evaluation. Table 2 and Figure 2 show that ES performs almost as well as the supervised baseline, and better than EI.

Table 1: MRI performance results. ES (ours) performs better than EI and SURE (baselines), while performing almost as well as the supervised baseline in terms of reconstruction quality (PSNR, SSIM) and measured equivariance (EQUIV). In **bold**, the best self-supervised metrics (avg \pm st.d.).

	MRI (×8 Accel., 40 dB SNR)			
Method	PSNR ↑	SSIM ↑	EQUIV ↑	
Supervised	28.74 ± 2.81	0.6445 ± 0.1094	31.71 ± 2.83	
ES (Ours) EI SURE	28.54 ± 2.75 27.88 ± 2.64 24.45 ± 1.86	0.6195 ± 0.1188 0.5731 ± 0.1299 0.5479 ± 0.0740	31.53 ± 2.74 30.79 ± 2.64 27.35 ± 1.90	
IDFT	23.62 ± 1.90	0.5052 ± 0.0900	25.99 ± 1.94	

5.3 MAGNETIC RESONANCE IMAGING

The dataset contains 320×320 images from FastMRI (Zbontar et al., 2019) subsampled in the Fourier domain by a single binary mask corresponding to an acceleration of 8, as well as Gaussian noise with a standard deviation of 0.005 corresponding to a signal-to-noise ratio (SNR) of 40

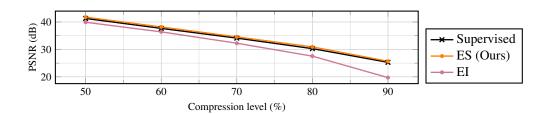


Figure 1: **Compressive sensing results.** ES (ours) performs similarly as the supervised baseline, unlike EI (baseline) whose performance gap widens with higher compression levels.

Table 2: **Inpainting results.** ES (ours) performs better than EI (baseline), both in terms of reconstruction quality (PSNR, SSIM) and measured equivariance (EQUIV), while performing competitively against the supervised baseline. In **bold**, the best self-supervised metrics (avg \pm st.d.).

			, e ,
Method	PSNR ↑	SSIM ↑	EQUIV ↑
Supervised	28.46 ± 2.97	0.8982 ± 0.0411	28.46 ± 2.97
ES (Ours)	27.45 ± 2.86	0.8737 ± 0.0461	27.46 ± 2.85
EI	25.89 ± 2.65	0.8332 ± 0.0521	25.89 ± 2.65
MC	8.22 ± 2.47	0.0983 ± 0.0551	8.22 ± 2.47
Incomplete image	8.22 ± 2.47	0.0973 ± 0.0542	N/A

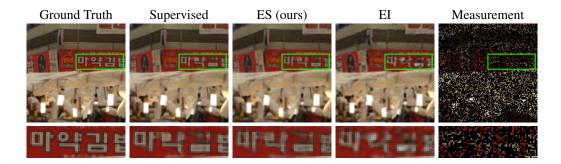


Figure 2: **Sample reconstructions for image inpainting.** ES (ours) produces images perceptually closer to the supervised baseline than EI (baseline) which appears blurry.

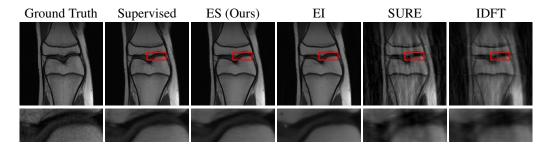


Figure 3: **Sample reconstructions for MRI** (**x8 Accel., 40 dB SNR**). Unlike EI (baseline) which suffers from dot-shaped artifacts, ES (ours) is perceptually closer to the supervised baseline. In line with the theoretical predictions, SURE and IDFT (baselines) fail to recover information beyond the observed frequencies.

dB. Out of the 973 images in the full dataset, 900 are selected for training and the remaining 73 are used for testing. We also test our method on a noise dominated setting using a different mask corresponding to an acceleration of 6, with a higher noise level of 0.1 corresponding to an SNR of only 10 dB, see these results in Appendix B.4. The variant of our proposed loss we use in this experiment is the R2R one introduced in Section 4 with $\alpha=0.5$. Table 1 shows the performance of the different methods on the test set. Table 3 shows the synergy of our method with the equivariant architecture. Figure 3 shows sample reconstructions from the trained models.

5.4 ABLATION STUDY

Table 3 shows that architectures designed to be equivariant are measurably more equivariant across imaging modalities and training losses. Moreover, it shows that networks trained using the splitting loss perform better for equivariant architectures than for non-equivariant architectures, with an even greater gap than for supervised inpainting baselines, which confirms the theoretical analysis made in Section 4. Finally, we observe that non-equivariant architectures still lead to fairly measurably equivariant models. While surprising, this phenomenon has already been witnessed and is usually referred to as learned equivariance, whereby the training data and inductive biases lead to fairly equivariant learned models (Gruver et al., 2024). We believe that this learned equivariance is responsible for the high performance of splitting methods even when using non-equivariant architectures. For additional results on the impact of equivariant architectures, see Appendix B.4.

Table 3: **Impact of using equivariant architectures.** In accordance with the theoretical results described in Section 4, there is a synergy between the splitting loss and equivariant architectures resulting in higher performance. Non-equivariant models have surprisingly high equivariance measures (EQUIV) which might explain their high performance when using the splitting loss. Eq. arch. denotes whether the architecture is equivariant. In **bold**, the best self-supervised metrics (avg \pm st.d.).

,.				
		Image inpainting		
Training loss	Eq. arch.	PSNR ↑	SSIM ↑	EQUIV ↑
Supervised	✓	28.46 ± 2.97	0.8982 ± 0.0411	28.46 ± 2.97
•	×	28.62 ± 3.03	0.9002 ± 0.0414	27.85 ± 2.71
Splitting (Ours)	✓	27.45 ± 2.86	0.8737 ± 0.0461	27.46 ± 2.85
	×	27.20 ± 2.83	0.8652 ± 0.0461	26.52 ± 2.60
		MRI (×8 Accel., 40 dB SNR)		
Training loss	Eq. arch.	PSNR ↑	SSIM ↑	EQUIV ↑
Supervised	✓	28.74 ± 2.81	0.6445 ± 0.1094	31.71 ± 2.83
	×	28.48 ± 2.68	0.6381 ± 0.1082	28.78 ± 1.95
Splitting (Ours)	✓	28.54 ± 2.75	0.6195 ± 0.1188	31.53 ± 2.74
	×	28.18 ± 2.58	0.6104 ± 0.1176	27.28 ± 2.10

6 Conclusion

In this work, we propose a new self-supervised loss for solving inverse problems which bridges the gap between existing equivariance and splitting-based self-supervised losses. We motivate the design of our loss by showing that minimizing the expected loss results in MMSE estimators. We further validate our method using numerical simulations on different image distributions and imaging modalities, including inpainting of natural images and MRI. These results suggest that the proposed method compares favorably to the equivariant imaging baseline and is close to supervised methods. To the best of our knowledge, this work is the first to leverage equivariant networks to learn from incomplete data alone, going beyond the usual goal of improving the generalization of the networks to unseen transformations at test time. Our method provides a new way to evaluate the benefits of using different equivariant architectures, and can benefit from future advances made in this field. More broadly, our work is further evidence that invariance is a promising prior for learning from incomplete data.

REFERENCES

- Hemant Kumar Aggarwal, Merry P. Mani, and Mathews Jacob. MoDL: Model Based Deep Learning Architecture for Inverse Problems. *IEEE Trans. Med. Imaging*, 38(2):394–405, February 2019.
- Eirikur Agustsson and Radu Timofte. NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study. In *CVPRW*, pp. 1122–1131, July 2017.
 - Chinmay Belthangady and Loic A Royer. Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction. *Nature methods*, 16(12):1215–1225, 2019.
 - Yochai Blau and Tomer Michaeli. The Perception-Distortion Tradeoff. In *CVPR*, pp. 6228–6237, 2018.
 - Elena Celledoni, Matthias J. Ehrhardt, Christian Etmann, Brynjulf Owren, Carola-Bibiane Schönlieb, and Ferdia Sherry. Equivariant neural networks for inverse problems. *Inverse Problems*, 37(8):085006, July 2021.
 - Anadi Chaman and Ivan Dokmanić. Truly shift-invariant convolutional neural networks. In *CVPR*, pp. 3773–3783, 2021a.
 - Anadi Chaman and Ivan Dokmanić. Truly shift-equivariant convolutional neural networks with adaptive polyphase upsampling. In *Asilomar Conf Signals Syst Comput*, pp. 1113–1120, 2021b.
- Dongdong Chen, Julián Tachella, and Mike E. Davies. Equivariant Imaging: Learning Beyond the Range Space. In *ICCV*, pp. 4379–4388, 2021.
 - Dongdong Chen, Julián Tachella, and Mike E. Davies. Robust Equivariant Imaging: A fully unsupervised framework for learning to image from noisy and partial measurements. In *CVPR*, pp. 5647–5656, 2022.
 - Taco S. Cohen and Max Welling. Group Equivariant Convolutional Networks. In *ICML*, pp. 2990–2999, 2016.
 - Taco S. Cohen and Max Welling. Steerable CNNs. In *ICLR*, 2017.
 - Giannis Daras, Kulin Shah, Yuval Dagan, Aravind Gollakota, Alexandros G. Dimakis, and Adam Klivans. Ambient Diffusion: Learning Clean Distributions from Corrupted Data. In *NeurIPS*, pp. 288–313, 2023.
 - Leo Davy, Luis M. Briceno-Arias, and N. Pustelnik. Restarted contractive operators to learn at equilibrium, June 2025.
 - Yonina C. Eldar. Generalized SURE for Exponential Families: Applications to Regularization. *IEEE Trans Signal Process*, 57(2):471–481, February 2009.
 - Fabian Ewald Fassnacht, Joanne C White, Michael A Wulder, and Erik Næsset. Remote sensing in forestry: current challenges, considerations and directions. *Forestry: An International Journal of Forest Research*, 97(1):11–37, 2024.
 - Nate Gruver, Marc Finzi, Micah Goldblum, and Andrew Gordon Wilson. The Lie Derivative for Measuring Learned Equivariance, June 2024.
- Allard A. Hendriksen, Daniel M. Pelt, and K. Joost Batenburg. Noise2Inverse: Self-supervised deep convolutional denoising for tomography. *IEEE Trans. Comput. Imaging*, 6:1320–1335, 2020.
- Kyong Hwan Jin, Michael T McCann, Emmanuel Froustey, and Michael Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE Trans Image Process*, 26(9):4509–4522, 2017.
- Sékou-Oumar Kaba, Arnab Kumar Mondal, Yan Zhang, Yoshua Bengio, and Siamak Ravanbakhsh.
 Equivariance with Learned Canonicalization Functions. In *ICML*, pp. 15546–15566, 2023.
 - Tero Karras, Miika Aittala, Samuli Laine, Erik Härkönen, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Alias-free generative adversarial networks. In *NeurIPS*, volume 34, pp. 852–863, 2021.

- Achim Klenke. *Probability theory: a comprehensive course*. Springer, 2008.
- Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. Noise2Void Learning Denoising From Single Noisy Images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2129–2137, 2019.
 - Ilya Loshchilov and Frank Hutter. Decoupled Weight Decay Regularization, January 2019.
- Christopher A. Metzler, Ali Mousavi, Reinhard Heckel, and Richard G. Baraniuk. Unsupervised Learning with Stein's Unbiased Risk Estimator, July 2020.
 - Hagay Michaeli, Tomer Michaeli, and Daniel Soudry. Alias-Free Convnets: Fractional Shift Invariance via Polynomial Activations. In *CVPR*, pp. 16333–16342, 2023.
 - Charles Millard and Mark Chiew. A theoretical framework for self-supervised MR image reconstruction using sub-sampling via variable density Noisier2Noise. *IEEE Trans Comput Imaging*, 9:707–720, 2023.
 - Brayan Monroy, Jorge Bacca, and Julián Tachella. Generalized Recorrupted-to-Recorrupted: Self-Supervised Learning Beyond Gaussian Noise. In *CVPR*, pp. 28155–28164, 2025.
 - Greg Ongie, Rebecca Willett, Daniel Soudry, and Nathan Srebro. A Function Space View of Bounded Norm Infinite Width ReLU Nets: The Multivariate Case, October 2019.
 - Tongyao Pang, Huan Zheng, Yuhui Quan, and Hui Ji. Recorrupted-to-Recorrupted: Unsupervised Deep Learning for Image Denoising. In *CVPR*, pp. 2043–2052, June 2021.
 - Omri Puny, Matan Atzmon, Heli Ben-Hamu, Ishan Misra, Aditya Grover, Edward J. Smith, and Yaron Lipman. Frame Averaging for Invariant and Equivariant Network Design, March 2022.
 - Marco Ragone, Reza Shahabazian-Yassar, Farzad Mashayek, and Vitaliy Yurkiv. Deep learning modeling in microscopy imaging: A review of materials science applications. *Progress in Materials Science*, 138:101165, 2023.
 - Sebastián Rivera, Iván Ortíz, Tatiana Gelvez, Fernando Rojas, and Henry Arguello. Rotation invariant deep learning approach for image inpainting. In *STSIVA*, pp. 1–5, 2021.
 - Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, 2015.
 - Leonid I. Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, November 1992.
 - Akiyoshi Sannai, Makoto Kawano, and Wataru Kumagai. Equivariant and Invariant Reynolds Networks, October 2021.
 - Akiyoshi Sannai, Makoto Kawano, and Wataru Kumagai. Invariant and equivariant reynolds networks. *Journal of Machine Learning Research*, 25(42):1–36, 2024.
 - Jérémy Scanvic, Mike Davies, Patrice Abry, and Julián Tachella. Scale-Equivariant Imaging: Self-Supervised Learning for Image Super-Resolution and Deblurring, March 2025.
 - Victor Sechaud, Laurent Jacques, Patrice Abry, and Julián Tachella. Equivariance-based self-supervised learning for audio signal recovery from clipped measurements. In *EUSIPCO*, August 2024.
 - Julián Tachella, Dongdong Chen, and Mike Davies. Unsupervised learning from incomplete measurements for inverse problems. *Advances in Neural Information Processing Systems*, 35:4983–4995, 2022.
 - Julián Tachella, Dongdong Chen, and Mike Davies. Sensing Theorems for Unsupervised Learning in Linear Inverse Problems. *Journal of Machine Learning Research*, 24(39):1–45, 2023. ISSN 1533-7928.

- Julián Tachella, Mike Davies, and Laurent Jacques. UNSURE: Self-supervised learning with Unknown Noise level and Stein's Unbiased Risk Estimate. In *ICLR*, 2025a.
- Julián Tachella, Matthieu Terris, Samuel Hurault, Andrew Wang, Dongdong Chen, Minh-Hai Nguyen, Maxime Song, Thomas Davies, Leo Davy, Jonathan Dong, Paul Escande, Johannes Hertrich, Zhiyuan Hu, Tobías I. Liaudat, Nils Laurent, Brett Levac, Mathurin Massias, Thomas Moreau, Thibaut Modrzyk, Brayan Monroy, Sebastian Neumayer, Jérémy Scanvic, Florian Sarron, Victor Sechaud, Georg Schramm, Chao Tang, Romain Vo, and Pierre Weiss. DeepInverse: A Python package for solving imaging inverse problems with deep learning, May 2025b.
- Matthieu Terris, Thomas Moreau, Nelly Pustelnik, and Julian Tachella. Equivariant plug-and-play image reconstruction. In *CVPR*, pp. 25255–25264, 2024.
- Singanallur V Venkatakrishnan, Charles A Bouman, and Brendt Wohlberg. Plug-and-play priors for model based reconstruction. In *IEEE Glob Conf Signal Inf Process*, pp. 945–948, 2013.
- Antonia Vojtekova, Maggie Lieu, Ivan Valtchanov, Bruno Altieri, Lyndsay Old, Qifeng Chen, and Filip Hroch. Learning to denoise astronomical images with u-nets. *Monthly Notices of the Royal Astronomical Society*, 503(3):3204–3215, 2021.
- Andrew Wang and Mike Davies. Perspective-equivariance for unsupervised imaging with camera geometry. In Alessio Del Bue, Cristian Canton, Jordi Pont-Tuset, and Tatiana Tommasi (eds.), *ECCV Workshops*, pp. 116–134, 2024.
- Andrew Wang, Steven McDonagh, and Mike Davies. Benchmarking Self-Supervised Learning Methods for Accelerated MRI Reconstruction, May 2025.
- Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Trans Image Process*, 13(4):600–612, April 2004.
- Philip J Withers, Charles Bouman, Simone Carmignato, Veerle Cnudde, David Grimaldi, Charlotte K Hagen, Eric Maire, Marena Manley, Anton Du Plessis, and Stuart R Stock. X-ray computed tomography. *Nature Reviews Methods Primers*, 1(1):18, 2021.
- Guixian Xu, Jinglai Li, and Junqi Tang. Fast Equivariant Imaging: Acceleration for Unsupervised Learning via Augmented Lagrangian and Auxiliary PnP Denoisers, July 2025.
- Burhaneddin Yaman, Seyed Amir Hossein Hosseini, Steen Moeller, Jutta Ellermann, Kâmil Uğurbil, and Mehmet Akçakaya. Self-Supervised Learning of Physics-Guided Reconstruction Neural Networks without Fully-Sampled Reference Data. *Magnetic Resonance in Med*, 84(6):3172–3191, December 2020. ISSN 0740-3194, 1522-2594. doi: 10.1002/mrm.28378.
- Jure Zbontar, Florian Knoll, Anuroop Sriram, Tullie Murrell, Zhengnan Huang, Matthew J. Muckley, Aaron Defazio, Ruben Stern, Patricia Johnson, Mary Bruno, Marc Parente, Krzysztof J. Geras, Joe Katsnelson, Hersh Chandarana, Zizhao Zhang, Michal Drozdzal, Adriana Romero, Michael Rabbat, Pascal Vincent, Nafissa Yakubova, James Pinkerton, Duo Wang, Erich Owens, C. Lawrence Zitnick, Michael P. Recht, Daniel K. Sodickson, and Yvonne W. Lui. fastMRI: An Open Dataset and Benchmarks for Accelerated MRI, December 2019.
- Richard Zhang. Making Convolutional Networks Shift-Invariant Again. In *ICML*, pp. 7324–7334, 2019.

A DETAILS ABOUT THE METHOD

- In Theorem 2, we consider reconstruction functions of 4 different structures. In this section, we give additional details about them.
- The artifact removal reconstructor architecture (Jin et al., 2017) in eq. (16) consists in a projection step that maps the measurements back into the image space using the adjoint or the pseudo-inverse of the forward matrix, which is immediately followed by a very general trainable network, often a UNet or another encoder-decoder type of network, or sometimes simply a fully convolutional

y

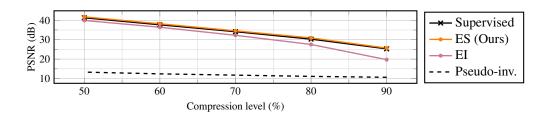


Figure 4: **Compressive sensing results.** Adds the pseudo-inverse reconstruction (pseudo-inv.) as a baseline to Figure 1

Equivariant reconstructor $\frac{1}{|\mathcal{G}|} \sum_{g \in \mathcal{G}} T_g r(y, AT_g) \longrightarrow f(y, A)$

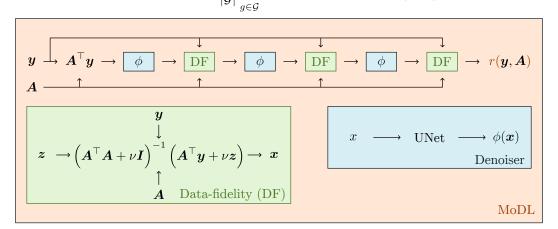


Figure 5: **Reconstructor equivariant to rotations and flips.** It is the Reynolds averaging of 90° rotations and horizontal and vertical flips as in eq. (18) of a non-equivariant reconstructor of the MAP type, as in eq. (19), implemented using a MoDL unrolled algorithm (Aggarwal et al., 2019) with 3 iterations, shared weights, and using a non-equivariant residual UNet (Ronneberger et al., 2015) as the denoiser architecture.

network. Theorem 2 states that this reconstructor architecture produces equivariant reconstructors as long as the denoiser architecture is, itself, equivariant in the sense of eq. (14).

Reynolds averaging for possibly non-equivariant reconstruction functions r(y, A) in eq. (18)

$$f(\mathbf{y}, \mathbf{A}) = \frac{1}{|\mathcal{G}|} \sum_{g \in \mathcal{G}} \mathbf{T}_g r(\mathbf{y}, \mathbf{A} \mathbf{T}_g). \tag{18}$$

has, as far as we know, not been defined in previous works. It is a natural extension of Reynolds averaging for possibly non-equivariant image-to-image functions or denoisers $\psi(x)$ (Sannai et al., 2024; Terris et al., 2024),

$$\phi(\mathbf{x}) = \frac{1}{|\mathcal{G}|} \sum_{g \in \mathcal{G}} T_g^{-1} \psi(T_g \mathbf{x})$$
(22)

which makes them equivariant in the sense of eq. (14), to reconstruction functions in order to make them equivariant in the sense of eq. (15). It is a fairly simple way to make a reconstructor equivariant and it is relatively inexpensive for small groups of transformations such as the group of 90° rotations and horizontal and vertical flips (Cohen & Welling, 2016). It is however too expensive to be used in practice when the group is relatively large, like the group of shifts, since it would require to evaluate the neural network for as many times as there are pixels in the input image, for each image.

The maximum a posteriori (MAP) reconstructor architecture defined in eq. (19)

$$f(\boldsymbol{y}, \boldsymbol{A}) = \underset{\boldsymbol{x} \in \mathbb{R}^n}{\operatorname{argmax}} \left\{ p(\boldsymbol{x} \mid \boldsymbol{y}, \boldsymbol{A}) \right\}.$$
 (19)

is a classical reconstructor architecture that has been used in different ways, including iterative algorithms with a hand-crafted prior (Rudin et al., 1992; Davy et al., 2025), plug-and-play architectures using an iterative approach but with the hand-crafted prior replaced with a pre-trained denoiser (Venkatakrishnan et al., 2013), and unrolled architectures where the optimization problem is done with a fixed number of steps and where parts of the algorithm are replaced with trainable modules (Aggarwal et al., 2019). It is often interpreted as a Bayesian maximum a posteriori estimator, but also commonly outside of a Bayesian framework as a variational approach with a data-fidelity term and a regularization term. It is one of the reconstructor architectures that we use for our experiments in Section 5. Theorem 2 states that these reconstructors are equivariant as long as the prior distribution is invariant in the sense of eq. (5), or equivalently, if the associated regularization function or negative log-prior is invariant.

The minimum mean squared error (MMSE) estimator in eq. (20)

$$f(\mathbf{y}, \mathbf{A}) = \mathbb{E}_{\mathbf{x}|\mathbf{y}, \mathbf{A}} \left\{ \mathbf{x} \right\} \tag{20}$$

is generally the target theoretical reconstructor as it achieves the highest theoretical PSNR (Tachella et al., 2023). It is generally estimated by reconstructors trained with a mean squared error loss (Chen et al., 2021), which is notably how we train the supervised baselines in the experiments in Section 5. Theorem 2 states that as long as the prior distribution is invariant in the sense of eq. (5), the MMSE reconstructor is equivariant in the sense of eq. (15).

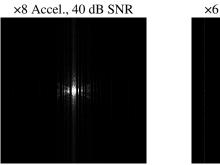
Table 4: **MRI results in another setting.** Supplementary results for a different MRI problem (×6 Accel., 10 dB SNR) than in Table 1 In **bold**, the best self-supervised metrics. Values: avg \pm st.d.

	MRI (×6 Accel., 10 dB SNR)			
Method	PSNR ↑	SSIM ↑	EQUIV ↑	
Supervised	27.39 ± 2.44	0.5243 ± 0.1373	30.38 ± 2.43	
ES (Ours) EI SURE	27.33 ± 2.45 27.23 ± 2.41 27.08 ± 2.29	0.5126 ± 0.1444 0.5110 ± 0.1421 0.5097 ± 0.1372	30.32 ± 2.44 30.21 ± 2.40 30.06 ± 2.28	
IDFT	23.85 ± 1.05	0.3878 ± 0.0272	25.14 ± 0.79	

B DETAILS ABOUT THE EXPERIMENTS

We use the optimizer AdamW (Loshchilov & Hutter, 2019) for every the training with different learning rates for the different inverse problems, a weight decay of 10^{-8} , beta coefficients equal to 0.9 and 0.999 and without the AMSGrad option. For longer trainings, we use step schedulers that divide the learning rate by a factor ranging from 2 to 10 at specific epochs, up to 3 or 4 times.

In our experiments, we make extensive use of the DeepInverse library (Tachella et al., 2025b) that provides an implementation of the various forward operators and training losses that we use. Every model is trained for up to 50 hours on a single GPU, either an NVIDIA H100, GH200 or RTX 4090.



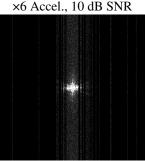


Figure 6: Examples of k-space measurements in MRI. (left) a less noisy problem with less measurements, (right) a noisier problem with more measurements.

B.1 DATASETS

Image inpainting The ground truth images are images obtained from the dataset DIV2K (Agustsson & Timofte, 2017) which contains pictures of natural scenes (landscapes, animals) by first resizing them to a resolution of 256×256 pixels before extracting a central 128×128 pixels crop. We synthesize the measurements by corrupting the ground truth images using a single binary mask sampled from a pixel-wise Bernoulli distribution with a 30% chance of keeping each pixel value. In this setting, we do not corrupt the measurements further with additional noise.

MRI The dataset consists in 973 pairs of ground truth images from the FastMRI dataset (Zbontar et al., 2019) and associated k-space measurements synthesized using a single coil sensitivity map. The ground truth images have a size of 320×320 pixels and each corresponds to the middle slice of a different 3d knee acquisition. The k-spaces are synthesized as discrete Fourier transforms subsampled on a single non-regular grid and further corrupted with additive white Gaussian noise with a standard deviation of 0.005 corresponding to an average SNR of about 40 dB. The subsampling grid models the coil sensitivity map, is sampled from a Gaussian distribution and corresponds to an acceleration of 8. The entire dataset is finally split into a train/val split containing 900 images and a test split containing the remaining 73 images. Our implementation uses the work from Wang et al. (2025).

We also consider an additional dataset obtained in the same way except using a different mask corresponding to an acceleration of 6 and with a higher noise level corresponding to a standard deviation of 0.1 or an average SNR of 10 dB in the k-space domain.

B.2 Network architectures

The unrolled architecture uses 3 iterations and the weights are shared across different iterations. Every UNet is residual, has 4 scales and has no normalization layer as we find them to be detrimental to the performance.

The transforms we use in the model architecture and in the metrics are grid-preserving: grid-aligned shifts, 90° rotations, and vertical and horizontal flips. The transforms we use in the EI are grid-aligned shifts and 1° rotations following the prior art Chen et al. (2021). Reynolds' averaging is

implemented using an unbiased Monte-Carlo estimator whereby a single random transform is sampled at every evaluation to save on computational cost.

B.3 MEASURED EQUIVARIANCE METRIC

We propose an equivariance metric similar to that used by Chaman & Dokmanić (2021b) adapted for our new definition of equivariant reconstructors. It is the average mean squared error associated with eq. (15) and expressed in dB for readability

$$EQUIV = -10\log_{10}\left(\mathbb{E}_{\boldsymbol{y},g}\left\{\left\|f(\boldsymbol{y},\boldsymbol{A}\boldsymbol{T}_g) - \boldsymbol{T}_g^{-1}f(\boldsymbol{y},\boldsymbol{A})\right\|^2\right\}\right). \tag{23}$$

It can be roughly understood as a PSNR for equivariance. In every experiment, we use the same group of transformations for EQUIV as we use for the equivariant reconstructor architectures.

B.4 RESULTS

Figure 4 shows additional compressive sensing results, Table 4 and Figure 7 show results on a MRI experiment with settings different from the main one, Table 5 shows extended results for the ablation study on equivariant architectures, and Figure 6 shows sample k-spaces from the MRI experiments. Figure 5 shows the equivariant reconstructor architecture that we adopt in the MRI experiments.

Table 5: Extended results on the impact of equivariant architectures. Adds to Table 3 the results for EI with a non-equivariant architecture for the inpainting task, results for the noise-dominated MRI task. In **bold**, the best self-supervised metrics. Values: $avg \pm st.d.$

Training loss Eq. arch. PSNR ↑ SSIM ↑ EQUIV ↑ Supervised ✓ 28.46 ± 2.97 0.8982 ± 0.0411 28.46 ± 2.97 × 28.62 ± 3.03 0.9001 ± 0.0415 27.85 ± 2.71 Splitting (Ours) ✓ 27.45 ± 2.86 0.8737 ± 0.0461 27.46 ± 2.85 × 27.20 ± 2.83 0.8651 ± 0.0463 26.52 ± 2.60 EI loss ✓ 25.89 ± 2.65 0.8332 ± 0.0521 25.89 ± 2.65 MC loss ✓ 26.33 ± 2.81 0.8451 ± 0.0536 25.58 ± 2.52 MC loss ✓ 8.22 ± 2.47 0.098 ± 0.055 8.22 ± 2.47 × 8.24 ± 2.48 0.100 ± 0.056 8.24 ± 2.48 Training loss Eq. arch. PSNR ↑ SSIM ↑ EQUIV ↑ Supervised ✓ 28.74 ± 2.81 0.6445 ± 0.1094 31.71 ± 2.83 × 28.48 ± 2.68 0.6381 ± 0.1082 28.78 ± 1.95 Splitting (Ours) ✓ 28.54 ± 2.75 0.6195 ± 0.1188 31.53 ± 2.74 Training loss			Image inpainting		
x 28.62 ± 3.03 0.9001 ± 0.0415 27.85 ± 2.71 Splitting (Ours) ✓ 27.45 ± 2.86 0.8737 ± 0.0461 27.46 ± 2.85 EI loss × 27.20 ± 2.83 0.8651 ± 0.0463 26.52 ± 2.60 EI loss ✓ 25.89 ± 2.65 0.8332 ± 0.0521 25.89 ± 2.65 × 26.33 ± 2.81 0.8451 ± 0.0536 25.58 ± 2.52 MC loss ✓ 8.22 ± 2.47 0.098 ± 0.055 8.22 ± 2.47 × 8.24 ± 2.48 0.100 ± 0.056 8.24 ± 2.48 MRI (×8 Accel., 40 dB SNR) Training loss Eq. arch. PSNR ↑ SSIM ↑ EQUIV ↑ Supervised ✓ 28.74 ± 2.81 0.6445 ± 0.1094 31.71 ± 2.83 × 28.48 ± 2.68 0.6381 ± 0.1082 28.78 ± 1.95 Splitting (Ours) ✓ 28.54 ± 2.75 0.6195 ± 0.1188 31.53 ± 2.74 × 28.18 ± 2.58 0.6104 ± 0.1176 27.28 ± 2.10 Training loss Eq. arch. PSNR ↑ SSIM ↑ EQUIV ↑	Training loss	Eq. arch.	PSNR ↑	SSIM ↑	EQUIV ↑
Splitting (Ours) ✓ 27.45 ± 2.86 0.8737 ± 0.0461 27.46 ± 2.85 EI loss × 27.20 ± 2.83 0.8651 ± 0.0463 26.52 ± 2.60 EI loss ✓ 25.89 ± 2.65 0.8332 ± 0.0521 25.89 ± 2.65 MC loss × 26.33 ± 2.81 0.8451 ± 0.0536 25.58 ± 2.52 MC loss ✓ 8.22 ± 2.47 0.098 ± 0.055 8.22 ± 2.47 × 8.24 ± 2.48 0.100 ± 0.056 8.24 ± 2.48 Training loss Eq. arch. PSNR ↑ SSIM ↑ EQUIV ↑ Supervised ✓ 28.74 ± 2.81 0.6445 ± 0.1094 31.71 ± 2.83 × 28.48 ± 2.68 0.6381 ± 0.1082 28.78 ± 1.95 Splitting (Ours) ✓ 28.54 ± 2.75 0.6195 ± 0.1188 31.53 ± 2.74 × 28.18 ± 2.58 0.6104 ± 0.1176 27.28 ± 2.10 MRI (×6 Accel., 10 dB SNR) Training loss Eq. arch. PSNR ↑ SSIM ↑ EQUIV ↑ Supervised ✓ 27.39 ± 2.44 0.5243 ± 0.1373 30.38 ± 2.4	Supervised	√	28.46 ± 2.97	0.8982 ± 0.0411	28.46 ± 2.97
EI loss	•	×	28.62 ± 3.03	0.9001 ± 0.0415	27.85 ± 2.71
EI loss	Splitting (Ours)	✓	27.45 ± 2.86	0.8737 ± 0.0461	27.46 ± 2.85
MC loss		×	27.20 ± 2.83	0.8651 ± 0.0463	26.52 ± 2.60
MC loss ✓ 8.22 ± 2.47 0.098 ± 0.055 8.22 ± 2.47 X 8.24 ± 2.48 0.100 ± 0.056 8.24 ± 2.48 MRI (×8 Accel., 40 dB SNR) Training loss Eq. arch. PSNR ↑ SSIM ↑ EQUIV ↑ Supervised ✓ 28.74 ± 2.81 0.6445 ± 0.1094 31.71 ± 2.83 × 28.48 ± 2.68 0.6381 ± 0.1082 28.78 ± 1.95 Splitting (Ours) ✓ 28.54 ± 2.75 0.6195 ± 0.1188 31.53 ± 2.74 × 28.18 ± 2.58 0.6104 ± 0.1176 27.28 ± 2.10 MRI (×6 Accel., 10 dB SNR) Training loss Eq. arch. PSNR ↑ SSIM ↑ EQUIV ↑ Supervised ✓ 27.39 ± 2.44 0.5243 ± 0.1373 30.38 ± 2.43 × 27.33 ± 2.42 0.5174 ± 0.1410 29.73 ± 2.20 Splitting (Ours) ✓ 27.33 ± 2.45 0.5126 ± 0.1444 30.32 ± 2.44	EI loss	\checkmark	25.89 ± 2.65	0.8332 ± 0.0521	25.89 ± 2.65
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$		×	26.33 ± 2.81	0.8451 ± 0.0536	25.58 ± 2.52
MRI (×8 Accel., 40 dB SNR) Training loss Eq. arch. PSNR ↑ SSIM ↑ EQUIV ↑ Supervised ✓ 28.74 ± 2.81 0.6445 ± 0.1094 31.71 ± 2.83 28.48 ± 2.68 0.6381 ± 0.1082 28.78 ± 1.95 Splitting (Ours) ✓ 28.54 ± 2.75 0.6195 ± 0.1188 31.53 ± 2.74 27.28 ± 2.10 MRI (×6 Accel., 10 dB SNR) × Training loss Eq. arch. PSNR ↑ SSIM ↑ EQUIV ↑ Supervised ✓ 27.39 ± 2.44 0.5243 ± 0.1373 30.38 ± 2.43 27.33 ± 2.42 0.5174 ± 0.1410 29.73 ± 2.20 Splitting (Ours) ✓ 27.33 ± 2.45 0.5126 ± 0.1444 30.32 ± 2.44	MC loss	\checkmark	8.22 ± 2.47	0.098 ± 0.055	8.22 ± 2.47
Training loss Eq. arch. PSNR ↑ SSIM ↑ EQUIV ↑ Supervised ✓ 28.74 ± 2.81 0.6445 ± 0.1094 31.71 ± 2.83 × 28.48 ± 2.68 0.6381 ± 0.1082 28.78 ± 1.95 Splitting (Ours) ✓ 28.54 ± 2.75 0.6195 ± 0.1188 31.53 ± 2.74 × 28.18 ± 2.58 0.6104 ± 0.1176 27.28 ± 2.10 MRI (×6 Accel., 10 dB SNR) Training loss Eq. arch. PSNR ↑ SSIM ↑ EQUIV ↑ Supervised ✓ 27.39 ± 2.44 0.5243 ± 0.1373 30.38 ± 2.43 × 27.33 ± 2.42 0.5174 ± 0.1410 29.73 ± 2.20 Splitting (Ours) ✓ 27.33 ± 2.45 0.5126 ± 0.1444 30.32 ± 2.44		×	8.24 ± 2.48	0.100 ± 0.056	8.24 ± 2.48
Supervised \checkmark 28.74 ± 2.81 0.6445 ± 0.1094 31.71 ± 2.83 X 28.48 ± 2.68 0.6381 ± 0.1082 28.78 ± 1.95 Splitting (Ours) \checkmark 28.54 ± 2.75 0.6195 ± 0.1188 31.53 ± 2.74 X 28.18 ± 2.58 0.6104 ± 0.1176 27.28 ± 2.10 MRI (×6 Accel., 10 dB SNR) Training loss Eq. arch. PSNR ↑ SSIM ↑ EQUIV ↑ Supervised \checkmark 27.39 ± 2.44 0.5243 ± 0.1373 30.38 ± 2.43 X 27.33 ± 2.42 0.5174 ± 0.1410 29.73 ± 2.20 Splitting (Ours) \checkmark 27.33 ± 2.45 0.5126 ± 0.1444 30.32 ± 2.44			MRI (×8 Accel., 40 dB SNR)		
× 28.48 ± 2.68 0.6381 ± 0.1082 28.78 ± 1.95 Splitting (Ours) ✓ 28.54 ± 2.75 0.6195 ± 0.1188 31.53 ± 2.74 × 28.18 ± 2.58 0.6104 ± 0.1176 27.28 ± 2.10 MRI (×6 Accel., 10 dB SNR) Training loss Eq. arch. PSNR ↑ SSIM ↑ EQUIV ↑ Supervised ✓ 27.39 ± 2.44 0.5243 ± 0.1373 30.38 ± 2.43 × 27.33 ± 2.42 0.5174 ± 0.1410 29.73 ± 2.20 Splitting (Ours) ✓ 27.33 ± 2.45 0.5126 ± 0.1444 30.32 ± 2.44	Training loss	Eq. arch.	PSNR ↑	SSIM ↑	EQUIV ↑
Splitting (Ours) ✓ 28.54 ± 2.75 0.6195 ± 0.1188 31.53 ± 2.74 × 28.18 ± 2.58 0.6104 ± 0.1176 27.28 ± 2.10 MRI (×6 Accel., 10 dB SNR) Training loss Eq. arch. PSNR ↑ SSIM ↑ EQUIV ↑ Supervised ✓ 27.39 ± 2.44 0.5243 ± 0.1373 30.38 ± 2.43 × 27.33 ± 2.42 0.5174 ± 0.1410 29.73 ± 2.20 Splitting (Ours) ✓ 27.33 ± 2.45 0.5126 ± 0.1444 30.32 ± 2.44	Supervised	✓	28.74 ± 2.81	0.6445 ± 0.1094	31.71 ± 2.83
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$		×	28.48 ± 2.68	0.6381 ± 0.1082	28.78 ± 1.95
Training loss Eq. arch. PSNR↑ SSIM↑ EQUIV↑ Supervised ✓ 27.39 ± 2.44 0.5243 ± 0.1373 30.38 ± 2.43 × 27.33 ± 2.42 0.5174 ± 0.1410 29.73 ± 2.20 Splitting (Ours) ✓ 27.33 ± 2.45 0.5126 ± 0.1444 30.32 ± 2.44	Splitting (Ours)	✓	28.54 ± 2.75	0.6195 ± 0.1188	31.53 ± 2.74
Training loss Eq. arch. PSNR↑ SSIM↑ EQUIV↑ Supervised ✓ 27.39 ± 2.44 0.5243 ± 0.1373 30.38 ± 2.43 × 27.33 ± 2.42 0.5174 ± 0.1410 29.73 ± 2.20 Splitting (Ours) ✓ 27.33 ± 2.45 0.5126 ± 0.1444 30.32 ± 2.44		×	28.18 ± 2.58	0.6104 ± 0.1176	27.28 ± 2.10
Supervised \checkmark 27.39 ± 2.44 0.5243 ± 0.1373 30.38 ± 2.43 \checkmark 27.33 ± 2.42 0.5174 ± 0.1410 29.73 ± 2.20 Splitting (Ours) \checkmark 27.33 ± 2.45 0.5126 ± 0.1444 30.32 ± 2.44			MRI (×6 Accel., 10 dB SNR)		
\times 27.33 ± 2.42 0.5174 ± 0.1410 29.73 ± 2.20 Splitting (Ours) \checkmark 27.33 ± 2.45 0.5126 ± 0.1444 30.32 ± 2.44	Training loss	Eq. arch.	PSNR ↑	SSIM ↑	EQUIV ↑
Splitting (Ours) $\sqrt{27.33 \pm 2.45}$ 0.5126 ± 0.1444 30.32 ± 2.44	Supervised	✓	27.39 ± 2.44	0.5243 ± 0.1373	30.38 ± 2.43
	-	×	27.33 ± 2.42	0.5174 ± 0.1410	29.73 ± 2.20
\times 27.20 ± 2.38 0.5095 ± 0.1430 28.66 ± 1.76	Splitting (Ours)	✓	27.33 ± 2.45	0.5126 ± 0.1444	30.32 ± 2.44
		×	27.20 ± 2.38	0.5095 ± 0.1430	28.66 ± 1.76

C PROOFS

For the sake of clarity, we state the propositions and theorems a second time before their proofs. We also state and prove the additional Lemma 1 which helps prove Theorem 1.

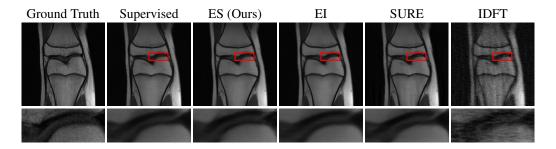


Figure 7: Sample reconstructions for noise dominated MRI (×6 Accel., 10 dB SNR). In the noise dominated setting, the different models perform more similarly than in the less noisy setting shown in Figure 3.

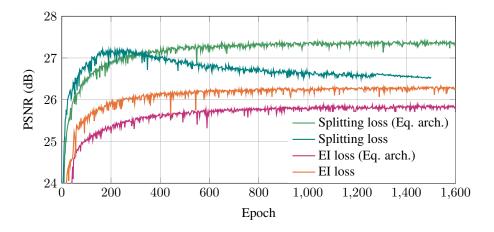


Figure 8: **Inpainting performance evolution during training.** Splitting methods perform better than EI independent of the network architecture.

Lemma 1. The minimization problem

$$\min_{f} \mathbb{E}_{x,y} \left\{ \| \mathbf{A} f(y) - \mathbf{A} x \|^{2} \right\}$$
 (24)

admits as solutions the functions of the form

$$f(\mathbf{y}) = \mathbf{A}^{\dagger} \mathbf{A} \mathbb{E}_{\mathbf{x}|\mathbf{y}} \left\{ \mathbf{x} \right\} + (\mathbf{I} - \mathbf{A}^{\dagger} \mathbf{A}) v(\mathbf{y})$$
 (25)

where v(y) is any function.

Proof. Let's start by stating that $f(y) = \mathbb{E}_{x|y} \{x\}$ is the only solution of (Klenke, 2008)

$$\min_{f} \mathbb{E}_{\boldsymbol{x},\boldsymbol{y}} \left\{ \|f(\boldsymbol{y}) - \boldsymbol{x}\|^{2} \right\}. \tag{26}$$

For f any solution of eq. (24), applying it to $\tilde{f}(y) = Af(y)$ and $\tilde{x} = Ax$ gives

$$Af(y) = A\mathbb{E}_{x|y} \{x\}, \qquad (27)$$

and applying $f(y) = A^{\dagger}Af(y) + (I - A^{\dagger}A)f(y)$ with v(y) := f(y) yields eq. (25). Conversely, the objective in eq. (24) has the same value no matter the f satisfying eq. (25) and since at least one of them is solution of eq. (24), they all are.

Theorem 1. In the case of noiseless measurements with p(x) \mathcal{G} -invariant Assumption 1, if the matrix $Q_{A_1} \triangleq \mathbb{E}_{g|A_1} \left\{ (AT_g)^\top AT_g \right\}$ has full rank for some split A_1 , then the splitting method yields the same MMSE-optimal reconstructions as the supervised method, i.e.,

$$f^*(y_1, A_1) = \mathbb{E}_{x|y_1, A_1} \{x\}.$$
 (9)

Proof.

$$\begin{split} &\mathbb{E}_{\boldsymbol{y}}\{\mathcal{L}_{\mathrm{ES}}(\boldsymbol{y},\boldsymbol{A},f)\}\\ &=\mathbb{E}_{\boldsymbol{y}}\left\{\mathbb{E}_{\boldsymbol{g}}\{\mathbb{E}_{\boldsymbol{y}_{1},\boldsymbol{A}_{1}|\boldsymbol{y},\boldsymbol{A}\boldsymbol{T}_{g}}\left\{\|\boldsymbol{A}\boldsymbol{T}_{g}f(\boldsymbol{y}_{1},\boldsymbol{A}_{1})-\boldsymbol{y}\|^{2}\right\}\right\}\right\}\\ &=\mathbb{E}_{\boldsymbol{y},g}\{\mathbb{E}_{\boldsymbol{y}_{1},\boldsymbol{A}_{1}|\boldsymbol{y},g}\left\{\|\boldsymbol{A}\boldsymbol{T}_{g}f(\boldsymbol{y}_{1},\boldsymbol{A}_{1})-\boldsymbol{y}\|^{2}\right\}\right\}\\ &=\mathbb{E}_{\boldsymbol{y},\boldsymbol{y}_{1},\boldsymbol{A}_{1},g}\{\|\boldsymbol{A}\boldsymbol{T}_{g}f(\boldsymbol{y}_{1},\boldsymbol{A}_{1})-\boldsymbol{y}\|^{2}\}\\ &=\mathbb{E}_{\boldsymbol{x},\boldsymbol{y}_{1},\boldsymbol{A}_{1},g}\{\|\boldsymbol{A}\boldsymbol{T}_{g}f(\boldsymbol{y}_{1},\boldsymbol{A}_{1})-\boldsymbol{A}\boldsymbol{T}_{g}\boldsymbol{x}\|^{2}\}\\ &=\mathbb{E}_{\boldsymbol{y}_{1},\boldsymbol{A}_{1},\boldsymbol{x}}\mathbb{E}_{g|\boldsymbol{y}_{1},\boldsymbol{A}_{1}}\left\{\|\boldsymbol{A}\boldsymbol{T}_{g}(f(\boldsymbol{y}_{1},\boldsymbol{A}_{1})-\boldsymbol{x})\|^{2}\right\}\\ &=\mathbb{E}_{\boldsymbol{y}_{1},\boldsymbol{A}_{1},\boldsymbol{x}}\{(f(\boldsymbol{y}_{1},\boldsymbol{A}_{1})-\boldsymbol{x})^{\top}\boldsymbol{Q}_{\boldsymbol{A}_{1}}(f(\boldsymbol{y}_{1},\boldsymbol{A}_{1})-\boldsymbol{x})\}, \end{split}$$

where the third line use that $p(y_1, A_1 \mid y, AT_g) = p(y_1, A_1 \mid y, g)$ as A is fixed. The fifth line use the noiseless measurements assumption and the invariance of the distribution p(x). The last line uses definition of Q_{A_1} . By applying Lemma 1, the global minimizer of the expected loss is given by:

$$f^*(y_1, A_1) = Q_{A_1}^{\dagger} Q_{A_1} \mathbb{E}_{x|y_1, A_1} \{x\} + (I - Q_{A_1}^{\dagger} Q_{A_1}) v(y_1)$$
 (28)

where $v:\mathbb{R}^n o \mathbb{R}^n$ is any function. Moreover, since Q has full-rank, $Q_{A_1}^\dagger = Q_{A_1}^{-1}$ and then

$$f^*(y_1, A_1) = \mathbb{E}_{x|y_1, A_1} \{x\}.$$
 (29)

Proposition 1. If the matrix $\bar{Q}_{A} \triangleq \mathbb{E}_{A_1|A} \{Q_{A_1}\}$ is invertible and f minimizes $\mathbb{E}_{y} \{\mathcal{L}_{ES}(y, A, f)\}$. Then the reconstruction function

$$\overline{f}(\boldsymbol{y}, \boldsymbol{A}) \triangleq \mathbb{E}_{\boldsymbol{y}_1, \boldsymbol{A}_1 | \boldsymbol{y}, \boldsymbol{A}} \left\{ \overline{\boldsymbol{Q}}_{\boldsymbol{A}}^{-1} \boldsymbol{Q}_{\boldsymbol{A}_1} f(\boldsymbol{y}_1, \boldsymbol{A}_1) \right\}$$
(10)

satisfies

$$\overline{f}(\boldsymbol{y}, \boldsymbol{A}) = \mathbb{E}_{\boldsymbol{y}_1, \boldsymbol{A}_1 | \boldsymbol{y}, \boldsymbol{A}} \left\{ \bar{\boldsymbol{Q}}_{\boldsymbol{A}}^{-1} \boldsymbol{Q}_{\boldsymbol{A}_1} \mathbb{E}_{\boldsymbol{x} | \boldsymbol{y}_1, \boldsymbol{A}_1} \left\{ \boldsymbol{x} \right\} \right\}. \tag{11}$$

where eq. (11) is a convex combination of MMSE estimators for different splittings.

Proof. By applying eq. (28) to f in the definition of \overline{f} we obtain:

$$\begin{split} & \overline{f}(\boldsymbol{y}, \boldsymbol{A}) = \mathbb{E}_{\boldsymbol{y}_1, \boldsymbol{A}_1 | \boldsymbol{y}, \boldsymbol{A}} \left\{ \bar{\boldsymbol{Q}}_{\boldsymbol{A}}^{-1} \boldsymbol{Q}_{\boldsymbol{A}_1} \left(\boldsymbol{Q}_{\boldsymbol{A}_1}^{\dagger} \boldsymbol{Q}_{\boldsymbol{A}_1} \mathbb{E}_{\boldsymbol{x} | \boldsymbol{y}_1, \boldsymbol{A}_1} \left\{ \boldsymbol{x} \right\} + (\boldsymbol{I} - \boldsymbol{Q}_{\boldsymbol{A}_1}^{\dagger} \boldsymbol{Q}_{\boldsymbol{A}_1}) v(\boldsymbol{y}_1) \right) \right\} \\ & = \mathbb{E}_{\boldsymbol{y}_1, \boldsymbol{A}_1 | \boldsymbol{y}, \boldsymbol{A}} \left\{ \bar{\boldsymbol{Q}}_{\boldsymbol{A}}^{-1} \boldsymbol{Q}_{\boldsymbol{A}_1} \mathbb{E}_{\boldsymbol{x} | \boldsymbol{y}_1, \boldsymbol{A}_1} \left\{ \boldsymbol{x} \right\} \right\} + \mathbb{E}_{\boldsymbol{y}_1, \boldsymbol{A}_1 | \boldsymbol{y}, \boldsymbol{A}} \left\{ \bar{\boldsymbol{Q}}_{\boldsymbol{A}}^{-1} \boldsymbol{Q}_{\boldsymbol{A}_1} \left((\boldsymbol{I} - \boldsymbol{Q}_{\boldsymbol{A}_1}^{\dagger} \boldsymbol{Q}_{\boldsymbol{A}_1}) v(\boldsymbol{y}_1) \right) \right\} \\ & = \mathbb{E}_{\boldsymbol{y}_1, \boldsymbol{A}_1 | \boldsymbol{y}, \boldsymbol{A}} \left\{ \bar{\boldsymbol{Q}}_{\boldsymbol{A}}^{-1} \boldsymbol{Q}_{\boldsymbol{A}_1} \mathbb{E}_{\boldsymbol{x} | \boldsymbol{y}_1, \boldsymbol{A}_1} \left\{ \boldsymbol{x} \right\} \right\}. \end{split}$$

Corollary 1. In order for the matrices Q_{A_1} or \bar{Q}_A to have full rank, it is necessary that A is not equivariant:

$$\exists g \in \mathcal{G}, AT_g \neq T_g A. \tag{13}$$

Proof. Let's assume by contradiction that A is equivariant with respect to T_q

$$AT_a = T_a A, (30)$$

and let $x \in \ker(A)$.

$$egin{aligned} oldsymbol{Q}_{oldsymbol{A}_1} oldsymbol{x} &= \mathbb{E}_{g|oldsymbol{A}_1} \left\{ (oldsymbol{A}oldsymbol{T}_g)^ op oldsymbol{A}oldsymbol{T}_g oldsymbol{X}
ight\} \ &= \mathbb{E}_{g|oldsymbol{A}_1} \left\{ (oldsymbol{A}oldsymbol{T}_g)^ op oldsymbol{T}_g oldsymbol{A}oldsymbol{x}
ight\} \ &= oldsymbol{\mathbb{E}}_{g|oldsymbol{A}_1} \left\{ (oldsymbol{A}oldsymbol{T}_g)^ op oldsymbol{T}_g oldsymbol{0}
ight\} \ &= oldsymbol{0} \end{aligned}$$

Therefore,

$$\ker(\mathbf{Q}_{\mathbf{A}_1}) \supseteq \ker(\mathbf{A}) \supseteq \{\mathbf{0}\}. \tag{31}$$

The matrix Q_{A_1} has a non-trivial nullspace and thus cannot have full rank. Moreover, since this non-trivial nullspace is the same for all virtual operators AT_g , then \bar{Q}_A shares the same non-trivial nullspace.

Theorem 2. The reconstruction functions defined in points below are all equivariant as in eq. (15).

1. Artifact removal network. For a denoiser $\phi(x)$ equivariant in the sense of eq. (14),

$$f(\boldsymbol{y}, \boldsymbol{A}) = \phi(\boldsymbol{A}^{\top} \boldsymbol{y}), or \ f(\boldsymbol{y}, \boldsymbol{A}) = \phi(\boldsymbol{A}^{\dagger} \boldsymbol{y}).$$
 (16)

2. Unrolled network. For $\phi(x)$ equivariant, any $\gamma \in \mathbb{R}$ and data fidelity d(Ax, y), with

$$x_0 = 0, \quad x_{k+1} = \phi(x_k - \gamma \nabla_{x_k} d(\mathbf{A}x_k, y))$$
 (17)

for
$$k = 0, \dots, L - 1$$
 and $f(\boldsymbol{y}, \boldsymbol{A}) = \boldsymbol{x}_L$.

3. Reynolds averaging. For a possibly non-equivariant reconstructor r(y, A), with

$$f(\mathbf{y}, \mathbf{A}) = \frac{1}{|\mathcal{G}|} \sum_{g \in \mathcal{G}} \mathbf{T}_g r(\mathbf{y}, \mathbf{A} \mathbf{T}_g). \tag{18}$$

4. Maximum a posteriori (MAP). For a distribution p(x) invariant as in eq. (5), with

$$f(\boldsymbol{y}, \boldsymbol{A}) = \underset{\boldsymbol{x} \in \mathbb{R}^n}{\operatorname{argmax}} \left\{ p(\boldsymbol{x} \mid \boldsymbol{y}, \boldsymbol{A}) \right\}.$$
 (19)

5. Minimum mean squared error (MMSE). For a distribution p(x) invariant as in eq. (5),

$$f(\boldsymbol{y}, \boldsymbol{A}) = \mathbb{E}_{\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{A}} \left\{ \boldsymbol{x} \right\}. \tag{20}$$

Proof. We prove each case separately.

1067

1068 1069

1075

1077

1078

1079

1. Denoting $m{A}^ imes := m{A}^ op$ or $m{A}^ imes := m{A}^\dagger$, eq. (16) gives, as $(m{A}m{T}_g)^ imes = m{T}_q^{-1}m{A}^ imes$, 1026 1027 $f(\boldsymbol{y}, \boldsymbol{A}\boldsymbol{T}_a) = \phi\left(\boldsymbol{T}_a^{-1}\boldsymbol{A}^{\times}\boldsymbol{y}\right),$ 1028 (32)1029 and since $\phi(x)$ is equivariant, i.e., eq. (14) holds, it simplifies to eq. (15). 1030 1031 2. We start by making the notation show the explicit dependency on y and A: 1032 $x_0(y, A) = 0, \quad x_{k+1}(y, A) = \phi \Big(x_k(y, A) - \gamma \nabla_{x_k(y, A)} d(Ax_k(y, A), y) \Big).$ (33)1033 1034 and proceed to show that for $k=0,\ldots,L$ it holds that 1035 $\boldsymbol{x}_k(\boldsymbol{y}, \boldsymbol{A}\boldsymbol{T}_q) = \boldsymbol{T}_q^{-1}\boldsymbol{x}_k(\boldsymbol{y}, \boldsymbol{A}).$ (34)For k=0, it holds as $x_0(y, AT_q) = 0 = T_q^{-1}x_0(y, A).$ (35)1039 Let's assume that eq. (34) holds for k < L. Applying it and the chain rule in eq. (33) yields 1041 $x_{k+1}(y, AT_g) = \phi \Big(T_g^{-1} \left(x_k(y, A) - \gamma \nabla_{x_k(y, A)} d(Ax_k(y, A), y) \right) \Big).$ (36)1043 Finally, applying eq. (14) in this equation gives $x_{k+1}(y, AT_q) = T_q^{-1}x_{k+1}(y, A),$ (37)1045 and by induction, as $f(y, A) = x_L(y, A)$, eq. (15) holds. 1046 1047 3. From eq. (18), it holds that 1048 1049 $f(\boldsymbol{y}, \boldsymbol{A}\boldsymbol{T}_g) = \frac{1}{|\mathcal{G}|} \sum_{h \in \mathcal{G}} \boldsymbol{T}_h r(\boldsymbol{y}, \boldsymbol{A}\boldsymbol{T}_g \boldsymbol{T}_h),$ (38)1050 1051 which, as the group action property holds $T_gT_h=T_{gh}$, rewrites as 1052 $f(\boldsymbol{y}, \boldsymbol{A}\boldsymbol{T}_g) = \frac{1}{|\mathcal{G}|} \sum_{h \in \mathcal{G}} \boldsymbol{T}_h r(\boldsymbol{y}, \boldsymbol{A}\boldsymbol{T}_{gh}),$ (39)1056 Applying the change of variable h' = gh in this equation gives $f(\boldsymbol{y}, \boldsymbol{A}\boldsymbol{T}_g) = \frac{1}{|\mathcal{G}|} \sum_{h \in \mathcal{C}} \boldsymbol{T}_{g^{-1}h} r(\boldsymbol{y}, \boldsymbol{A}\boldsymbol{T}_h),$ 1058 (40)which finally, using group action property again $T_{g^{-1}h} = T_g^{-1}T_g$, gives eq. (15). 1062 4. Taking the negative natural logarithm in eq. (19) and using Bayes' theorem gives $f(\boldsymbol{y}, \boldsymbol{A}) = \underset{\boldsymbol{x} \in \mathbb{R}^n}{\operatorname{argmin}} \ \Big\{ d(\boldsymbol{A}\boldsymbol{x}, \boldsymbol{y}) + \rho(\boldsymbol{x}) \Big\},$ (41)

 $x \in \mathbb{R}^n$

where $d({m A}{m x},{m y}) = -\log p({m y}\mid {m A}{m x})$ and $ho({m x}) = -\log p({m x}).$ Applying $x' = {m T}_g\,{m x},$

$$f(\boldsymbol{y}, \boldsymbol{A}\boldsymbol{T}_g) = \boldsymbol{T}_g^{-1} \underset{\boldsymbol{x} \in \mathbb{R}^n}{\operatorname{argmin}} \left\{ d(\boldsymbol{A}\boldsymbol{x}, \boldsymbol{y}) + \rho \left(\boldsymbol{T}_g^{-1} \boldsymbol{x}\right) \right\}, \tag{42}$$

and eq. (5) makes $\rho(x)$ invariant as well $\rho(T_q^{-1}x) = \rho(x)$. Therefore, eq. (15) holds.

5. Let's assume that p(x) and p(A) are invariant in the sense of eq. (5). We first prove that

$$p(\mathbf{y}, \mathbf{A}T_q) = p(\mathbf{y}, \mathbf{A}). \tag{43}$$

Using eq. (5), the invariance of p(A) and the independence of x and A, we compute

$$p(\boldsymbol{y}, \boldsymbol{A}\boldsymbol{T}_g) = \mathbb{E}_{\boldsymbol{x}} \left\{ p(\boldsymbol{y}, \boldsymbol{A}\boldsymbol{T}_g \mid \boldsymbol{x}) \right\} = \mathbb{E}_{\boldsymbol{x}} \left\{ p(\boldsymbol{y}, \mid \boldsymbol{A}\boldsymbol{T}_g, \boldsymbol{x}) p(\boldsymbol{A}\boldsymbol{T}_g \mid \boldsymbol{x}) \right\}$$
$$= \mathbb{E}_{\boldsymbol{x}} \left\{ p(\boldsymbol{y} \mid \boldsymbol{A}\boldsymbol{T}_g, \boldsymbol{x}) p(\boldsymbol{A}\boldsymbol{T}_g) \right\} = \mathbb{E}_{\boldsymbol{x}} \left\{ p(\boldsymbol{y} \mid \boldsymbol{A}\boldsymbol{T}_g \boldsymbol{x}) p(\boldsymbol{A}) \right\}$$
$$= \mathbb{E}_{\boldsymbol{x}} \left\{ p(\boldsymbol{y} \mid \boldsymbol{A}\boldsymbol{x}) p(\boldsymbol{A}) \right\} = \mathbb{E}_{\boldsymbol{x}} \left\{ p(\boldsymbol{y} \mid \boldsymbol{A}, \boldsymbol{x}) p(\boldsymbol{A} \mid \boldsymbol{x}) \right\}$$
$$= \mathbb{E}_{\boldsymbol{x}} \left\{ p(\boldsymbol{y}, \boldsymbol{A} \mid \boldsymbol{x}) \right\} = p(\boldsymbol{y}, \boldsymbol{A}).$$

Next we start from $f(\boldsymbol{y}, \boldsymbol{A}\boldsymbol{T}_q) = \mathbb{E}_{\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{A}\boldsymbol{T}_q} \left\{ \boldsymbol{x} \right\}.$ (44)and applying the integral formula for expectations gives $f(\boldsymbol{y}, \boldsymbol{A}\boldsymbol{T}_g) = \int \boldsymbol{x} p(\boldsymbol{x} \mid \boldsymbol{y}, \boldsymbol{A}\boldsymbol{T}_g) dx.$ (45)Using Bayes' formula and $p(y \mid A, x) = p(y \mid Ax)$, it becomes $f(\boldsymbol{y}, \boldsymbol{A}\boldsymbol{T}_g) = \int \boldsymbol{x} \, \frac{p(\boldsymbol{y} \mid \boldsymbol{A}\boldsymbol{T}_g \boldsymbol{x})}{p(\boldsymbol{A}\boldsymbol{T}_g, \boldsymbol{y})} \, p(\boldsymbol{A}\boldsymbol{T}_g, \boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}.$ (46)By using eq. (43), the invariance of p(A) and the independence of A with x, we obtain $f(\boldsymbol{y}, \boldsymbol{A}\boldsymbol{T}_g) = \int \boldsymbol{x} \, \frac{p(\boldsymbol{y} \mid \boldsymbol{A}\boldsymbol{T}_g \boldsymbol{x})}{p(\boldsymbol{A}, \boldsymbol{y})} \, p(\boldsymbol{A}) p(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}.$ (47)With the change of variable $x' = T_q x$, and since T_q is unitary, we arrive at $f(\boldsymbol{y}, \boldsymbol{A}\boldsymbol{T}_g) = \int \boldsymbol{T}_g^{-1} \boldsymbol{x} \, \frac{p(\boldsymbol{y} \mid \boldsymbol{A}\boldsymbol{x})}{p(\boldsymbol{A}, \boldsymbol{y})} \, p(\boldsymbol{A}) p(\boldsymbol{T}_g^{-1} \boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}.$ (48)Finally, applying eq. (5) in this equation yields eq. (15). **Theorem 3.** If f(A, x) is an equivariant reconstructor, then ES is equivalent to the splitting loss $\mathcal{L}_{ES}(\boldsymbol{y}, \boldsymbol{A}, f) = \mathcal{L}_{SPLIT}(\boldsymbol{y}, \boldsymbol{A}, f).$ (21)*Proof.* We start from eq. (7) $\mathcal{L}_{ES}(\boldsymbol{y}, \boldsymbol{A}, f) \triangleq \mathbb{E}_q \left\{ \mathcal{L}_{SPLIT}(\boldsymbol{y}, \boldsymbol{AT}_q, f) \right\}$ (49) $= \mathbb{E}_{g} \left\{ \mathbb{E}_{\boldsymbol{y}_{1}, \boldsymbol{A}_{1} | \boldsymbol{y}, \boldsymbol{A} \boldsymbol{T}_{g}} \left\{ \|\boldsymbol{A} \boldsymbol{T}_{g} f(\boldsymbol{y}_{1}, \boldsymbol{A}_{1}) - \boldsymbol{y} \|^{2} \right\} \right\}$ (50)As A_1 is a splitting of AT_q , we can write $A_1 = MAT_q$ for M a splitting matrix. We obtain $\mathcal{L}_{ES}(\boldsymbol{y}, \boldsymbol{A}, f) = \mathbb{E}_q \left\{ \mathbb{E}_{\boldsymbol{M}|\boldsymbol{y},q} \left\{ \|\boldsymbol{A}\boldsymbol{T}_q f(\boldsymbol{M}\boldsymbol{y}, \boldsymbol{M}\boldsymbol{A}\boldsymbol{T}_q) - \boldsymbol{y}\|^2 \right\} \right\}$ (51)Applying eq. (15) and cancelling out T_g with T_g^{-1} yields $\mathcal{L}_{ES}(\boldsymbol{y}, \boldsymbol{A}, f) = \mathbb{E}_q \left\{ \mathbb{E}_{\boldsymbol{M}|\boldsymbol{y}, q} \left\{ \|\boldsymbol{A}f(\boldsymbol{M}\boldsymbol{y}, \boldsymbol{M}\boldsymbol{A}) - \boldsymbol{y}\|^2 \right\} \right\}.$ (52)By dropping the expectation in g and rewriting (My, MA) as (y_1, A_1) , this yields in eq. (21).