
Equivariant Priors for Compressed Sensing with Unknown Orientation

Anna Kuzina ^{†1} Kumar Pratik ^{*2} Fabio Valerio Massoli ^{*2} Arash Behboodi ²

Abstract

In compressed sensing, the goal is to reconstruct the signal from an underdetermined system of linear measurements. Thus, prior knowledge about the signal of interest and its structure is required. Additionally, in many scenarios, the signal has an unknown orientation prior to measurements. To address such recovery problems, we propose using equivariant generative models as a prior, which encapsulate orientation information in their latent space. Thereby, we show that signals with unknown orientations can be recovered with iterative gradient descent on the latent space of these models and provide additional theoretical recovery guarantees. We construct an equivariant variational autoencoder and use the decoder as generative prior for compressed sensing. We discuss additional potential gains of the proposed approach in terms of convergence and latency.

1. Introduction

Compressed sensing (CS) deals with the problem of reconstructing an unknown underlying signal $\mathbf{x} \in \mathbb{R}^m$ from m linear measurements generated as

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \varepsilon \quad (1)$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$ is a task-specific measurement matrix and ε is additive noise. A wide range of practical problems can be formulated as a recovery of an unknown signal from the noisy linear measurements. Medical applications include computed tomography (Chen et al., 2008), where one seeks to recover an unknown image from a sinogram (set of linear measurements) and rapid MRI (Lustig et al., 2007), where

^{*}Equal contribution ¹Vrije Universiteit Amsterdam, Netherlands. [†]Work done during the internship in Qualcomm AI research. ²Qualcomm AI Research, Qualcomm Technologies Netherlands B.V. Qualcomm AI Research is an initiative of Qualcomm Technologies, Inc.. Correspondence to: Anna Kuzina <av.kuzina@yandex.ru>.

an MR image is reconstructed from an undersampled k-space. It is also used in wireless communication for channel estimation problems (Paredes et al., 2007). Single-pixel imaging (Duarte et al., 2008) is another example of linear inverse problems.

Since the number of measurements is assumed to be smaller than the signal space dimension, this system of linear equations is under-determined, and extra assumptions are needed to recover the solution. One common approach would be to use the sparsity assumption about the signal of interest in a given basis (Tibshirani, 1996; Candès et al., 2006; Donoho, 2006). There are other notions of structure beyond sparsity such as low rankness. More complicated structures can be captured using generative modelling. This field is usually referred to as generative compressed sensing. Generative CS uses a pre-trained generative model to capture the prior over the space of signals. In this case, the signal recovery is done by finding the latent code of the generative model, for example, using gradient descent (Bora et al., 2017). There are, however, some challenges in using generative priors for compressed sensing, such as convergence and latency issues. Gradient descent methods sometimes need many iterations and restarts for convergence (see (Whang et al., 2021b) and references and discussions therein).

In this work, we consider a more general setup, where along with the known linear forward operator, the input is also transformed by an unknown group operation. A typical example is a rotation by an angle α . Such examples occur in many imaging applications where the signal pose is unknown before the imaging. Another example is cryo-electron microscopy (Singer, 2018). We are interested in this problem:

$$\mathbf{y} = \mathbf{A}\mathcal{T}_g\mathbf{x} + \varepsilon, \quad (2)$$

where $\mathcal{T}_g : \mathcal{G} \times \mathcal{X} \rightarrow \mathcal{X}$ is an action of a group element g in the group \mathcal{G} on the signal space \mathcal{X} . For example, it can be a rotation by some angle which is, in many cases, captured by a linear transformation. The group transformation \mathcal{T}_g is assumed to be unknown.

To use standard generative priors for this task, the signal pose should be estimated jointly with the latent code. It is therefore natural to include group action information in the latent space. In this paper, we propose equivariant generative priors for this task. A generative model $G(\cdot)$ is equiv-

ariant if for group representations \mathcal{T}_g^x on the signal space and \mathcal{T}_g^z on the latent space, we have $\mathcal{T}_g^x G(z) = G(\mathcal{T}_g^z z)$ for all $g \in \mathcal{G}$ and z . In this way, unknown transformations are coded in the latent space and the task boils down to the estimation of the transformed (e.g. rotated) latent code.

In this work we focus on the group of rotations, however, our approach can be also applied to other groups. We show how the current generative compressed sensing approaches can be adapted to the setup of the unknown signal orientation. We compare different baseline generative priors in this setup and propose a rotation equivariant prior based on equivariant variational autoencoder model, which can equally be used for a conventional compressed sensing problem and, at the same time, is well equipped to deal with the unknown orientation of the concerned signal.

The contributions of this paper are as follows. We propose using equivariant generative priors for solving compressed sensing problems with unknown rotations. We show how standard recovery guarantees for generative priors are transferred to the equivariant models and compressed sensing with unknown orientation. We build an equivariant variational autoencoder. We experimentally show that the proposed prior is useful for both the conventional and the rotation-aware compressed sensing. In particular, these models provide promising gains in terms of latency and convergence. In addition, integrating the inductive bias about rotation equivariance into the model leads to a more parameter efficient generative priors.

2. Related Work

Generative Compressed Sensing It is common to address a compressed sensing problem using the maximum likelihood approach. Assuming that the additive noise ε in the linear inverse problem (1) is standard Gaussian, the likelihood would also follow Gaussian distribution $p(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{y}|\mathbf{A}\mathbf{x}, \mathbf{I})$. Maximum-Likelihood estimation of the unknown signal is, therefore, a solution of the following optimization problem:

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2. \quad (3)$$

(Bora et al., 2017) proposed to use a generative model as a prior in this setup. It is assumed that the generative model is equipped with the generator $G(\cdot)$, which maps latent variables z (often assumed to be standard Gaussian) to the space of signals. This way, generative model serves as a regularizer and the optimization problem (3) can be reformulated as

$$\mathbf{x}^* = G(z^*), \quad (4)$$

$$z^* = \arg \min_z \|\mathbf{y} - \mathbf{A}G(z)\|_2^2. \quad (5)$$

Using MLP-VAE and DCGAN models as a prior was shown to perform better than the sparsity prior (Bora et al., 2017). There are many follow-up works on better optimization over latent space (Daras et al., 2021; Lei et al., 2019). The authors in (Hussein et al., 2020) introduce the Image-Adaptive GAN model, where the generator is updated together with the latent code. Normalizing Flow model is used as prior in (Asim et al., 2019) and shown that it outperforms previous GAN-based approaches for a variety of inverse problems, including compressed sensing.

It was observed that the maximum likelihood estimation (equation 5) usually requires a proper regularization for faster convergence. Natural way to impose such regularization would be maximum a posteriori estimate of the unknown signal:

$$\mathbf{x}^{\text{MAP}} = \arg \max_{\mathbf{x}} \log p(\mathbf{y}|\mathbf{x}) + \log p(\mathbf{x}). \quad (6)$$

MAP formulation was introduced in (Whang et al., 2021a) and can be as well written in terms of the generative prior and Gaussian noise model:

$$\mathbf{x}^{\text{MAP}} = G(z^*), \quad (7)$$

$$z^* = \arg \min_z \|\mathbf{y} - \mathbf{A}G(z)\|_2^2 + \log p_G(G(z)). \quad (8)$$

It is worth mentioning that the latter term of equation 8 limits the applicability of the formulations to the generative priors with the explicit density functions, e.g. normalizing flows.

In some applications it is also beneficial to recover the whole distribution of the unknown signal \mathbf{x} given the observation \mathbf{y} . Since the true posterior distribution is either too hard or even impossible to recover, one can use stochastic variational inference (SVI) to get the approximate posterior:

$$q_{\mathbf{x}}^* = \arg \min_{q \in \mathcal{Q}} D_{\text{KL}} [q(\mathbf{x}) \| p(\mathbf{x}|\mathbf{y})]. \quad (9)$$

This approach was recently applied to compressed sensing (Whang et al., 2021c). It was proposed to construct $q_{\mathbf{x}}$ in such a way that it comprises of two normalizing flow models. The first, pre-generator q_z , is attached to a prior generative model and is trained to match the posterior distribution of the unknown signal.

$$q_z^* = \arg \min_{q_z} D_{\text{KL}} [q_z \| p_z] + \mathbb{E}_{q_z} [\|\mathbf{y} - \mathbf{A}G(z)\|_2^2]. \quad (10)$$

The problem can be further amortized (Kingma & Welling, 2013; Rezende et al., 2014), so that the variational posterior is conditioned on the observation $q_{\mathbf{x}}(\mathbf{x}|\mathbf{y})$. This significantly reduced the computational complexity of the method. The point estimate of the signal can also be obtained in this case. One can sample from the variational posterior and average the results, which gives a Monte-Carlo estimate of the conditional expectation $\mathbb{E}[\mathbf{x}|\mathbf{y}]$.

Equivariant Generative Models Group equivariant neural networks (Cohen & Welling, 2016a) incorporate symmetries of the data in the architecture. The authors in (Weiler & Cesa, 2019) provide a framework to construct neural networks equivariant under all the isometries of \mathbb{R}^2 plane. Reincorporating symmetries in the architecture was shown to improve generalization and sample efficiency for discriminative models. In (Karras et al., 2021) authors show that rotation and translation equivariance can be beneficial for GANs as well. In this work, we introduce VAE with fully equivariant latent space. Benefits of the equivariant decoder were also shown in (Bepler et al., 2019), where authors use a spacial generator network, which allows representing rotation and translation as a coordinate space transformation. Homeomorphic-VAE (Falorsi et al., 2018) focuses on generative models with latent space that has $\text{SO}(3)$, in general, a Lie group, values. In this work, we consider a conventional Euclidean latent space with real-valued entries. On this space, we define the action of a discrete group of 2D rotations (C_n).

Another promising direction in this area is employing equivariant normalizing flow models. Recent works have introduced equivariant generative models based on continuous normalizing flows (Köhler et al., 2020; Satorras et al., 2021), which are, however difficult to scale.

Compressed Sensing and Unknown Orientation In this work, we focus on the reconstruction of the signal with an unknown orientation. The authors in (Davenport et al., 2007) consider reconstructing the rotation angle of the input and focuses primarily on compressive classification and not signal reconstruction. We aim at restoring the signal regardless of the fact that it was rotated or not. Multi-reference alignment (Bendory et al., 2017) aims at solving inverse problems from multiple observations of a single source in different unknown orientations. The main challenge, and focus of research, in these works, is the alignment of these observations. For us, there is only a single observation, and the alignment step does not apply. To the best of our knowledge, this is the first work considering compressed sensing problems with unknown orientation.

3. Methodology

3.1. Background: Variational Autoencoder

Variational Autoencoder (VAE) (Kingma & Welling, 2013; Rezende et al., 2014) is a deep generative model which models a joint distribution of observed random variables $\mathbf{x} \in \mathcal{X}$ (e.g. $\mathcal{X} = \mathbb{R}^n$) and latent variables $\mathbf{z} \in \mathbb{R}^k$ as $p_\theta(\mathbf{x}, \mathbf{z}) = p_\theta(\mathbf{x}|\mathbf{z})p(\mathbf{z})$. The model is trained to maximize the marginal likelihood $p_\theta(\mathbf{x})$ for a given set of points $\mathbf{x}_1, \dots, \mathbf{x}_N$. Amortized version of variational inference (Jordan et al., 1999) is used to obtain a tractable objective: by

introducing a variational posterior $q_\phi(\mathbf{z}|\mathbf{x})$, also referred to as encoder, one can obtain a lower bound on the intractable marginal likelihood (ELBO):

$$\mathcal{L}(\theta, \phi) = \mathbb{E}_{p_e(\mathbf{x})} [\mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})} \ln p_\theta(\mathbf{x}|\mathbf{z}) - D_{\text{KL}}(q_\phi(\mathbf{z}|\mathbf{x})||p(\mathbf{z}))]. \quad (11)$$

3.2. Equivariant VAE

Requirements for the Equivariant Latent Space We start by discussing the requirements, which encoder and decoder of the VAE should satisfy. Forward pass through the VAE model consists of three steps. First, we apply a neural network parameterized by ϕ to get parameters of the variational posterior $q_\phi(\mathbf{z}|\mathbf{x})$. Secondly, we sample from this distribution $\tilde{\mathbf{z}} \sim q_\phi(\mathbf{z}|\mathbf{x})$. And, on the third step, we push $\tilde{\mathbf{z}}$ through the neural network parameterized by θ which returns parameters of the generative distribution $p_\theta(\mathbf{x}|\tilde{\mathbf{z}})$.

Consider a group \mathcal{G} . Our goal is to construct VAE, in which latent space is equivariant under the action of the group element $g \in \mathcal{G}$. Let $\mathcal{T}_g^x : \mathcal{G} \times \mathcal{X} \rightarrow \mathcal{X}$ be a transformation of a data point under the group action g and $\mathcal{T}_g^z : \mathcal{G} \times \mathbb{R}^k \rightarrow \mathbb{R}^k$ be a transformation of a latent code under the same action g . Then, for the encoder to be equivariant, we aim for the following property to hold:

$$\mathcal{T}_g^z \mathbf{z} \stackrel{d}{=} \mathbf{z}_g, \text{ where} \quad (12)$$

$$\mathbf{z}_g \sim q_\phi(\mathbf{z}|\mathcal{T}_g^x \mathbf{x}), \quad (13)$$

$$\mathbf{z} \sim q_\phi(\mathbf{z}|\mathbf{x}). \quad (14)$$

where $\stackrel{d}{=}$ stands for equality in distribution. In other words, sampling latent code for variational posterior conditioned on the transformed input should be the same as transforming a latent code which is sampled from the variational distribution conditioned on the non-transformed input.

Furthermore, symmetric property should hold for the decoder network. Namely,

$$\mathcal{T}_g^x \mathbf{x} \stackrel{d}{=} \mathbf{x}_g, \text{ where} \quad (15)$$

$$\mathbf{x}_g \sim p_\theta(\mathbf{x}|\mathcal{T}_g^z \mathbf{z}), \quad (16)$$

$$\mathbf{x} \sim p_\theta(\mathbf{x}|\mathbf{z}). \quad (17)$$

We now discuss how to construct the Encoder and Decoder networks to ensure that both requirements (12) and (15) hold.

Equivariant Encoder The common choice of the variational posterior distribution $q_\phi(\mathbf{z}|\mathbf{x})$ is Gaussian

$$\begin{aligned} q_\phi(\mathbf{z}|\mathbf{x}) &= \mathcal{N}(\mathbf{z}|\mu, \Sigma), \\ \mu &= \mu_\phi(\mathbf{x}), \\ \Sigma &= \mathbf{L}\mathbf{L}^T, \mathbf{L} = L_\phi(\mathbf{x}). \end{aligned} \quad (18)$$

In general case \mathbf{L} is a lower triangular matrix, which can be obtained by Cholesky decomposition of Σ . In practice, the covariance matrix is often chosen to be diagonal. That is \mathbf{L} is also diagonal: $L_\phi(\mathbf{x}) = \text{diag } \sigma_\phi(\mathbf{x})$.

Such parametrization is practical for several reasons. Firstly, Gaussian distribution is reparametrizable:

$$\mathbf{z} = \mu + \mathbf{L}\varepsilon, \quad (19)$$

This is crucial for backpropagation through the ELBO (11). Secondly, the KL-term between variational posterior and prior is to be computed at each iteration. The prior is usually chosen to be standard Gaussian, which gives a closed form for the KL-divergence:

$$D_{\text{KL}}[q_\phi(\mathbf{z}|\mathbf{x})||p(\mathbf{z})] = \frac{1}{2}[-\log \det \Sigma - d + \text{tr}\Sigma + \mu^T \mu]. \quad (20)$$

With the diagonal covariance matrix computing the log-determinant in equation 20 reduces to a summation of d scalars.

In this work, we also focus on the Gaussian q_ϕ . Then it is enough to match the first and second moments of $\mathcal{T}_g^z \mathbf{z}$ and \mathbf{z}_g to satisfy the desired properties of the equivariant latent space. When it comes to \mathbf{z}_g , it follows a Gaussian distribution, which gives us the following moments:

$$\mathbb{E}\mathbf{z}_g = \mu_\phi(\mathcal{T}_g^x \mathbf{x}), \quad (21)$$

$$\mathbb{E}(\mathbf{z}_g - \mathbb{E}\mathbf{z}_g)(\mathbf{z}_g - \mathbb{E}\mathbf{z}_g)^T = L_\phi(\mathcal{T}_g^x \mathbf{x})L_\phi(\mathcal{T}_g^x \mathbf{x})^T. \quad (22)$$

Next, we would like to find the moments of $\mathcal{T}_g^z \mathbf{z}$, knowing that \mathbf{z} is a Gaussian random variable (18). The mean value is given by

$$\mathbb{E}\mathcal{T}_g^z \mathbf{z} = \mathbb{E}\mathcal{T}_g^z[\mu + \mathbf{L}\varepsilon] = \mathcal{T}_g^z[\mu]. \quad (23)$$

And for the second moment, we have:

$$\mathbb{E}(\mathcal{T}_g^z \mathbf{z} - \mathbb{E}\mathcal{T}_g^z \mathbf{z})(\mathcal{T}_g^z \mathbf{z} - \mathbb{E}\mathcal{T}_g^z \mathbf{z})^T \quad (24)$$

$$= \mathbb{E}(\mathcal{T}_g^z[\mu + \mathbf{L}\varepsilon] + \mathcal{T}_g^z \mu)(\mathcal{T}_g^z[\mu + \mathbf{L}\varepsilon] + \mathcal{T}_g^z \mu)^T \quad (25)$$

$$= \mathbb{E}\mathcal{T}_g^z[\mathbf{L}\varepsilon]\mathcal{T}_g^z[\mathbf{L}\varepsilon]^T \quad (26)$$

$$= \mathcal{T}_g^z[\mathbf{L}]\mathcal{T}_g^z[\mathbf{L}]^T. \quad (27)$$

This gives us the following condition on the mean function:

$$\mu_\phi(\mathcal{T}_g^x \mathbf{x}) = \mathcal{T}_g^z[\mu_\phi(\mathbf{x})]. \quad (28)$$

This property can be satisfied when using an equivariant neural network to model the mean function. The covariance, on the other hand, should satisfy:

$$L_\phi(\mathcal{T}_g^x \mathbf{x})L_\phi^T(\mathcal{T}_g^x \mathbf{x}) = \mathcal{T}_g^z[L_\phi(\mathbf{x})] (\mathcal{T}_g^z[L_\phi(\mathbf{x})])^T. \quad (29)$$

The main restriction that results from the (29) is that the diagonal covariance matrix can violate the desired equivariance property. As mentioned earlier, one usually chooses $L_\phi(\mathbf{x}) = \text{diag } \sigma(\mathbf{x})$. In this case $L_\phi(\mathcal{T}_g^x \mathbf{x})$ will always be diagonal. However, $\mathcal{T}_g^z[L_\phi(\mathbf{x})]$ is not necessarily staying diagonal. Therefore, we have to use a non-diagonal covariance matrix. We still need to be able to sample \mathbf{z} , compute log-determinant on the forward pass and ensure that the matrix is positive definite. To this end, we propose to use a full-rank covariance matrix in our model:

$$\Sigma = V_\phi(\mathbf{x})V_\phi(\mathbf{x})^T + \eta\mathbf{I}, \quad (30)$$

where $V_\phi(\cdot)$ is an equivariant function, which outputs a full-rank matrix. We ensure that the resulting covariance matrix is positive definite by adding a small positive perm to the diagonal elements.

Equivariant Decoder For the decoder to be equivariant, we should ensure a similar property as we did for the encoder. We consider two distributions to model the conditional likelihood $p_\theta(\mathbf{x}|\mathbf{z})$. For RGB images, we can use Gaussian distribution. The same assumptions about the mean function and covariance should be satisfied in this case.

For grey-scale images, Bernoulli distribution is used: $p_\theta(\mathbf{x}|\mathbf{z}) = \mathcal{B}e(\mathbf{x}|f_\theta(\mathbf{z}))$. In this case, it is straightforward to respect the property (15):

$$f_\theta(\mathcal{T}_g^z \mathbf{z}) = \mathcal{T}_g^x[f_\theta(\mathbf{z})], \quad (31)$$

which means that we should just ensure that f_θ is an equivariant function. We summarize the forward pass through the resulting model in the Algorithm 1.

3.3. Generative Compressed Sensing with Unknown Orientation

We consider the following forward process:

$$\mathbf{y} = \mathbf{A}\mathcal{T}_g^x \mathbf{x} + \varepsilon, \quad (32)$$

where the sensing matrix \mathbf{A} is known and g stands for the unknown rotation angle (element of the group of rotations). Our goal is to reconstruct the rotated signal $\mathcal{T}_g^x \mathbf{x}$. The way to address this problem depends on the type of prior generative model that we are using. Below we consider three different scenarios.

Algorithm 1 Forward pass through equivariant VAE

Input: x
 $\mu = \mu_\phi(\mathbf{x})$ \triangleright Forward pass through the encoder
 $\mathbf{V} = V_\phi(\mathbf{x})$
 $\Sigma = \mathbf{V}\mathbf{V}^T + \eta\mathbf{I}$
 $L = \text{Chol}(\Sigma)$ \triangleright Cholesky decomposition of Σ
 $\tilde{\mathbf{z}} = \mu + \mathbf{L}\varepsilon, \varepsilon \sim \mathcal{N}(0, \mathbf{I})$
 $\text{KL} = \frac{1}{2} \sum_i (-\log L_{ii} - 1 + L_{ii}^2 + \mu_i^2)$
 $p_x = f_\theta(\tilde{\mathbf{z}})$ \triangleright Forward pass through the decoder
 $\text{Re} = -\log p_\theta(\mathbf{x}|p_x)$ \triangleright Reconstruction Loss
 $\mathcal{L}(\phi, \theta) = -\text{Re} - \text{KL}$ \triangleright Compute ELBO
Return: $-\mathcal{L}(\phi, \theta)$

Standard Prior Firstly, we may have a prior generative model, which is uninformed of the rotation. That is, it was trained on the signals in the non-rotated (canonical) orientation and it is only able to generate such. However, it is still possible to reformulate the optimization problem (5), so that we are able to reconstruct the rotated signal:

$$\mathbf{z}^*, g^* = \arg \min_{\mathbf{z}, g} \|\mathbf{y} - \mathbf{A}\mathcal{T}_g^x G(\mathbf{z})\|_2^2, \quad (33)$$

$$\mathbf{x}^* = \mathcal{T}_{g^*}^x G(\mathbf{z}^*). \quad (34)$$

However, we have observed that optimizing for the latent code and the rotation angle simultaneously using the gradient descent does not produce reasonable results. However, coordinate gradient descent can be applied instead. In the latter case, we alternate the gradient steps to update the latent code and the rotation angle. This allows us to use different learning rates and considerably improves the convergence.

Conditional Prior Secondly, we may have a conditional generative model as a prior. In this case, the model is trained on rotated images. As a result, it can generate a rotated image from a latent code and with the desired angle. The optimization task for an MLE solution will then be the following:

$$\mathbf{z}^*, g^* = \arg \min_{\mathbf{z}, g} \|\mathbf{y} - \mathbf{A}G(\mathbf{z}, g)\|_2^2, \quad (35)$$

$$\mathbf{x}^* = G(\mathbf{z}^*, g^*). \quad (36)$$

Prior with rotation-aware latent space Finally, we may train a prior with a latent code for all the rotated images. For example, we can augment the dataset with the rotated sample while training the prior. Another option would be to train an equivariant generative model. The latter approach does not require the rotated samples during training. However, the equivariance property still allows it to produce the rotated samples.

$$\mathbf{z}^* = \arg \min_{\mathbf{z}} \|\mathbf{y} - \mathbf{A}G(\mathbf{z})\|_2^2, \quad (37)$$

$$\mathbf{x}^* = G(\mathbf{z}^*). \quad (38)$$

The straightforward advantage of this approach is that we have the same objective as a conventional generative compressed sensing and there is no need to optimize with respect to the rotation angle. Furthermore, in case of equivariant prior, we do not need any additional (augmented) data while training the generative model.

3.4. Recovery guarantees for equivariant priors

With some conditions on the sensing matrix, generative compressed sensing has recovery guarantees in classical setup. The main idea is presented in (Bora et al., 2017) with follow-up works on more general sensing setups (Liu & Scarlett, 2020; Liu et al., 2020). We can provide the same recovery guarantees for equivariant priors. Before stating the result, we introduce some notations. A generative model $G(\cdot)$ is defined as a mapping from k -dimensional ball of radius r in \mathbb{R}^k to \mathbb{R}^n . We assume that the action of rotation in \mathbb{R}^n is given by a unitary representation \mathcal{T}_g^x .

Theorem 3.1. *Consider an equivariant generative model $G : B_k(r) \rightarrow \mathbb{R}^n$, which is assumed to be an L -Lipschitz function. Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ be a random Gaussian matrix with i.i.d. entries $\mathcal{N}(0, m^{-1})$. Assume that a signal $\mathbf{x}^* \in \mathbb{R}^n$ is observed in an unknown orientation according to $\mathbf{y} = \mathbf{A}\mathcal{T}_g^x \mathbf{x}^* + \varepsilon$. Suppose that $\hat{\mathbf{z}}$ minimizes $\|\mathbf{y} - \mathbf{A}G(\mathbf{z})\|$ to within error of δ_{approx} . For measurement number $m = \Omega(k \log(Lr/\delta))$, the following holds with probability $1 - e^{-\Omega(m)}$*

$$\|G(\hat{\mathbf{z}}) - \mathcal{T}_g^x \mathbf{x}^*\|_2 \leq 6 \min_{\mathbf{z} \in B_k(r)} \|G(\mathbf{z}) - \mathbf{x}^*\|_2 + 3\|\varepsilon\|_2 + 2\delta_{\text{approx}} + 2\delta.$$

Since equivariant models incorporate the rotation in the latent space, the proof remains the same. Nonetheless, we review it in the Appendix.

Note that the upper bound on the reconstruction error does not depend on the rotation operation \mathcal{T}_g^x . The term $\min_{\mathbf{z} \in B_k(r)} \|G(\mathbf{z}) - \mathbf{x}^*\|_2$ is evaluated only on the canonical pose. This is a built-in feature of the equivariant models, namely that,

$$\begin{aligned} \|G(\mathbf{z}) - \mathcal{T}_g^x \mathbf{x}^*\|_2 &= \|\mathcal{T}_g^{xH} G(\mathbf{z}) - \mathbf{x}^*\|_2 \\ &= \|G(\mathcal{T}_g^z \mathbf{z}) - \mathbf{x}^*\|_2. \end{aligned}$$

Since \mathcal{T}_g^z is a unitary matrix, it maps $B_k(r)$ to itself, and

$\mathcal{T}_g^{zH} z$ remains inside $B_k(r)$. This means that:

$$\begin{aligned} \min_{z \in B_k(r)} \|G(z) - \mathcal{T}_g^x \mathbf{x}_*\|_2 &= \min_{z \in B_k(r)} \|G(\mathcal{T}_g^{zH} z) - \mathbf{x}_*\|_2 \\ &= \min_{z \in B_k(r)} \|G(z) - \mathbf{x}_*\|_2. \end{aligned}$$

This upper bound remains the same for generative compressed sensing with no rotation. Although this is just an upper bound, it suggests that there should not be a difference in the performance of equivariant priors on rotated and non-rotated data. Our experimental results confirm that there is a small difference in performance for rotated and non-rotated signals.

4. Experiments

We consider two different setups. The conventional compressed sensing discussed in section 2 (no rotation) and compressed sensing with unknown orientation discussed in section 3.3 (unknown rotation).

Datasets We conduct experiments on two different datasets. We start with benchmarking experiments on MNIST. Subsequently, concerning a real-world application of the proposed approach, we conduct experiments on the Low Dose CT Image and Projection Data¹ (MAYO) dataset (Moen et al., 2021), which consist of three types of data: DICOM-CT-PD projection data, DICOM image data, and Excel clinical data reports. To our aim, we use the DICOM subset only. Images are divided into three sets labelled N for neuro, C for chest, and L for liver each of which comprises 512x512 images from 50 different patients. To train the generative priors, we consider the L subset which is made of $\sim 7K$ samples that we divide into train, validation and test sets comprising $\sim 80\%$, $\sim 10\%$, and $\sim 10\%$ of the images, respectively. Before feeding a model, we randomly crop the image and then rescale it to 128x128, and finally, we normalize the pixels value in $[0, 1]$.

Prior Generative Models We compare various generative models as prior on the signals that we aim to reconstruct. Broadly speaking, we consider two types of generative models in our experiments: VAE and Normalizing Flow. In the former case, we have MLP-VAE, fully convolutional VAE (Conv-VAE) and Equivariant VAE (Eq-VAE: proposed model). And in the latter, as a flow prior, we consider a multi-scale RealNVP model. To bring the non-equivariant models, i.e., flow and VAE, at par with the Eq-VAE in terms of their ability to generate rotated images so that they can be effectively used as a prior for CS with unknown rotation, we consider two routes, 1. Conditional model: we include a conditional flow (Cond. Flow) model, which has rotation

¹<https://www.aapm.org/grandchallenge/lowdosect/>

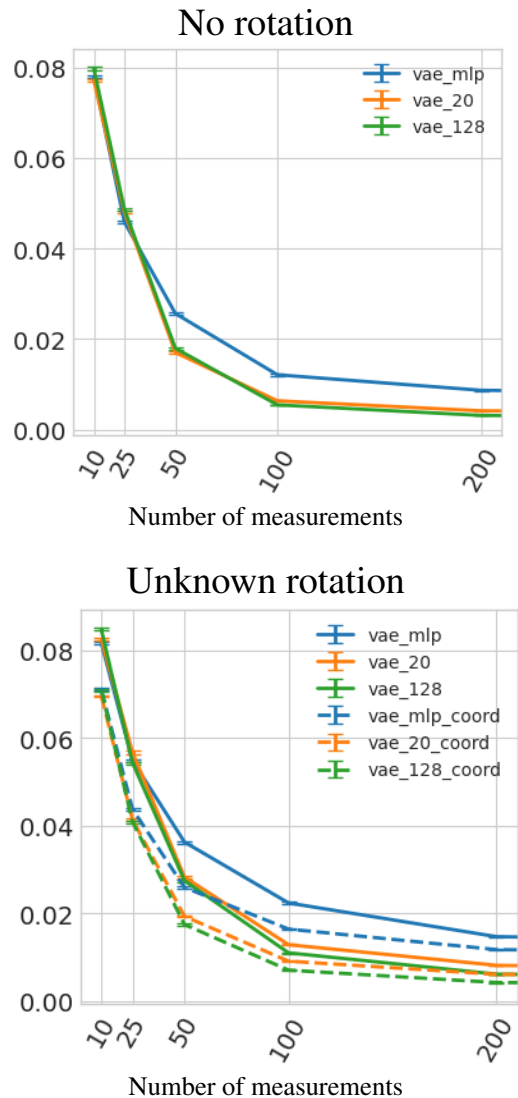


Figure 1: Comparison of MLP-VAE and Convolutional VAE with the latent dimensions 20 and 128 on MNIST dataset. Average MSE for 10000 test points.

angle as an additional input and is trained on the augmented (rotated) dataset, 2. Augmented models: another way to augment the capabilities of conventional generative models is to train them on augmented (rotated) datasets. On this line, we include augmented VAE (Aug-VAE), and augmented flow (Aug-Flow) as conventional VAE and Flow but explicitly trained on augmented (addition of random rotation as a pre-processing step) dataset. In our MNIST experiments, we noticed impractically high latency of Cond. Flow during both the generative prior training and the following CS experiment. The problem gets highly exacerbated for the MAYO dataset as it is more complex than MNIST, and hence, we decided to drop the Cond. Flow for MAYO due to its impractical latency.

Metrics When it comes to measuring the success of the compressed sensing, we are interested in two criteria: reconstruction quality and convergence speed. We measure MSE to evaluate the quality of the reconstruction. For convergence, we assume that if the MSE (per pixel) is lower than 0.01, then the reconstruction is successful. We then report the proportion of converged points and the average number of iterations (or gradient descent steps) required for a method to converge.

4.1. Benchmark experiments on MNIST

Concerning compressed sensing experiments, we consider a different number of linear measurements to reconstruct the input signal. Specifically, in Table 1 we report CS results for the MNIST dataset considering 50, 100, and 200 measurements. For each of the mentioned values, and each prior model, we report the average value for the MSE and the number of iterations required to achieve a 100% success rate for reconstructing the input signals. As mentioned previously, we consider two different scenarios: no rotations and unknown rotations. In both cases, Eq-VAE prior report the best performance concerning MSE and the number of iterations required to reconstruct the input images. Indeed, although with all the priors we observe a 100% convergence rate for the CS experiments, Eq-VAE requires fewer iterations and reports a lower average MSE than all the other models.

Following (Bora et al., 2017) we train MLP-VAE with 20 dimensional latent space. However, as can be seen from the Figure 1, fully convolutional VAE with the same latent space dimensionality tends to perform better both with and without rotation. Presumably, the introduction of the fully convolutional architecture makes the model equivariant to the translation, which is beneficial for the compressed sensing. As results in Table 1 show, equivariance to rotation improves the performance even further.

Figure 1 also shows that larger latent space of the convolutional VAE gives a slight performance improvement. Thus we chose to train equivariant prior with the same latent dimension. When rotation is unknown, we also compare coordinate gradient descent (*dashed line*) with the joint optimization of the latent code and the rotation angle (*solid line*). In all the experiments, coordinate gradient descent shows better results.

4.2. CT scans reconstruction: MAYO dataset

We report in Table 2 results for the MAYO dataset. Differently from the MNIST result, in this case, we fix the number of measurements to 200 and evaluate the performance of the different generative priors on the CS task considering a different number of iterations. Specifically, we consider 50, 100, 150, and 200 iterations, and for each of them, we report

the average MSE and the percentage of points successfully reconstructed (i.e., with MSE per pixel below 0.01). Being the MAYO dataset more complex than the MNIST, in this case, we do not observe a clear dominance of one type of prior compared to the others. However, as a general remark, we can notice that using Eq-VAE the average percentage of converged points is higher than for other models for both no rotations, 85.0%, and unknown rotations, 85.5%. As a comparison, the second-best performing priors are the VAE with 75.3%, and the Flow with 85.2%, concerning no and unknown rotations, respectively. Instead, regarding the MSE, all the results agree within 1 standard deviation.

4.3. Discussion

As we discussed before, the upper bound on the reconstruction error in Theorem 3.1 suggests that there is no difference in the performance of equivariant models in rotated and non-rotated cases. The results for the MAYO dataset show a small performance drop from no rotation to unknown rotation cases. The discrepancy is a bit larger for MNIST experiments. The difference can be attributed to the fact that equivariant models are built for finite rotation groups while we test on continuous rotation groups. In Figure 2 we report a comparison between reconstructed rotated samples from Eq-VAE(left) and VAE (right). As expected, Eq-VAE successfully recovers the correct orientation for the reconstructed images while VAE struggles to achieve such a goal.

Group Choice We train the equivariant prior under the action of the cyclic group. This is a discrete group, size of which, is a hyperparameter to choose. However, in our compressed sensing experiments with unknown orientation, the rotation angle is sampled uniformly. Thus, a larger group size will potentially lead to a better-compressed sensing performance. We compared three different group sizes (see Appendix C.1) and conclude that group C_{16} results in the best reconstruction results in both the CS setups.

5. Conclusion

We consider a generative compressed sensing problem with unknown orientation of the measurement signal. This new setup motivates the usage of new generative prior models, which are capable of producing rotated images. We propose an equivariant variational autoencoder and extend theoretical convergence guarantees to the case of unknown rotation and equivariant generative prior. We experimentally show that the proposed equivariant prior is better or comparable to other benchmarks in terms of reconstruction quality and provide potential additional advantages in terms of convergence speed, as it usually requires fewer iterations (gradient descent steps).

Equivariant Priors for Compressed Sensing with Unknown Orientation

Table 1: CS results on MNIST. We report results averaged over 1000 test points. By Converged Points, we refer to points that, for a given number of iterations, resulted in MSE (per pixel) lower than $1.e^{-2}$. We emphasize in bold, results from the best model.

	Prior	50 measurements			100 measurements			200 measurements		
		MSE ($1.e^{-3}$)↓	Converged Points (%)↑	# iter↓	MSE ($1.e^{-3}$)↓	Converged Points (%)↑	# iter↓	MSE ($1.e^{-3}$)↓	Converged Points (%)↑	# iter ↓
No rotation	Flow	45.4	100	59	19.2	100	39	10.2	100	38
	Cond. Flow	22.6	100	35	10.1	100	22	5.8	100	22
	MLP-VAE	25.7	100	50	12.1	100	41	8.7	100	38
	Conv-VAE	18.0	100	48	5.5	100	39	3.2	100	35
	Eq-VAE	10.6	100	34	2.6	100	22	2.1	100	22
Unknown rotation	Flow	27.2	100	73	13.2	100	60	7.6	100	36
	Cond. Flow	17.9	100	42	7.2	100	32	6.0	100	30
	MLP-VAE	25.9	100	48	16.4	100	40	11.7	100	34
	Conv-VAE	17.5	100	51	7.0	100	39	4.2	100	32
	Eq-VAE	11.8	100	33	4.2	100	24	3.2	100	23

Table 2: CS results on MAYO. The reported results corresponding to 200 measurements. By Converged Points, we refer to points that, for a given number of iterations, resulted in MSE (per pixel) lower than $1.e^{-2}$. Concerning the MSE, we reported the mean value $\pm 1\sigma$. The MSE is computed across converged points only. We emphasize in bold, results from the best model.

	Prior	50 iterations		100 iterations		150 iterations		200 iterations	
		MSE ($1.e^{-3}$)↓	Converged Points (%)↑	MSE ($1.e^{-3}$)↓	Converged Points (%)↑	MSE ($1.e^{-3}$)↓	Converged Points (%)↑	MSE ($1.e^{-3}$)↓	Converged Points (%)↑
No rotation	Flow	8.9 ± 0.9	28.2	8.1 ± 1.2	65.9	7.3 ± 1.3	79.7	6.8 ± 1.4	86.8
	Aug-Flow	N/A	0.0	8.2 ± 1.2	40.1	7.4 ± 1.4	77.1	6.7 ± 1.4	83.4
	Conv-VAE	7.9 ± 1.3	39.9	6.8 ± 1.6	76.2	6.2 ± 1.8	89.6	5.9 ± 1.8	95.5
	Aug-VAE	7.5 ± 1.3	51.7	6.1 ± 1.6	75.7	5.7 ± 1.7	83.8	5.5 ± 1.6	88.1
	Eq-VAE	7.4 ± 1.4	73.5	6.4 ± 1.4	86.9	6.1 ± 1.5	89.2	6.0 ± 1.5	90.4
Unknown rotation	Flow	7.2 ± 1.4	76.4	5.8 ± 1.5	83.8	5.2 ± 1.6	87.5	5.0 ± 1.7	93.3
	Aug-Flow	N/A	0.0	8.2 ± 1.1	49.9	6.9 ± 1.3	80.9	6.0 ± 1.3	84.3
	Conv-VAE	7.1 ± 1.7	55.3	5.9 ± 2.1	80.5	5.4 ± 2.1	88.2	5.0 ± 2.0	90.2
	Aug-VAE	7.5 ± 1.5	59.3	5.9 ± 1.6	85.2	5.3 ± 1.5	89.1	5.1 ± 1.6	92.4
	Eq-VAE	7.6 ± 1.4	70.3	6.6 ± 1.5	87.5	6.1 ± 1.5	91.3	6.0 ± 1.6	92.7

References

Asim, M., Daniels, M., Leong, O., Ahmed, A., and Hand, P. Invertible generative models for inverse problems: mitigating representation error and dataset bias. *arXiv:1905.11672*, May 2019.

Bendory, T., Boumal, N., Ma, C., Zhao, Z., and Singer, A. Bispectrum inversion with application to multireference alignment. *IEEE Transactions on signal processing*, 66(4):1037–1050, 2017.

Bepler, T., Zhong, E. D., Kelley, K., Brignole, E., and Berger, B. Explicitly disentangling image content from translation and rotation with spatial-VAE. *arXiv:1909.11663*, September 2019.

Bora, A., Jalal, A., Price, E., and Dimakis, A. G. Compressed sensing using generative models. In Precup, D. and Teh, Y. W. (eds.), *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 537–546.

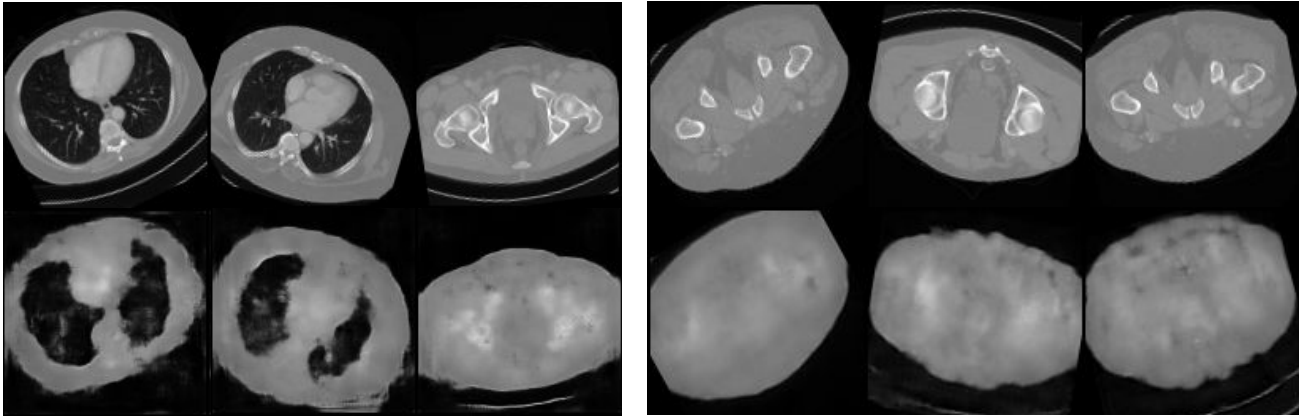


Figure 2: Example of reconstructed images using Eq-VAE (left-3) and VAE (right-3) as a prior. The top row shows ground truth rotated images prior to measurement, and the bottom row shows the corresponding CS reconstructed images. We notice in the third column (both left and right images), the VAE reconstructed image is 180° rotated with respect to the canonical configuration while the Eq. VAE is able to retrieve the configuration effectively.

PMLR, 2017.

- Candès, E. J., Romberg, J. K., and Tao, T. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, 2006.
- Cesa, G., Lang, L., and Weiler, M. A program to build e(n)-equivariant steerable CNNs. In *International Conference on Learning Representations*, 2022.
- Chen, G.-H., Tang, J., and Leng, S. Prior image constrained compressed sensing (piccs): a method to accurately reconstruct dynamic ct images from highly undersampled projection data sets. *Medical physics*, 35(2):660–663, 2008.
- Cohen, T. and Welling, M. Group equivariant convolutional networks. In *International conference on machine learning*, pp. 2990–2999. PMLR, 2016a.
- Cohen, T. S. and Welling, M. Steerable CNNs. In *ICLR 2017*, November 2016b.
- Daras, G., Dean, J., Jalal, A., and Dimakis, A. G. Intermediate Layer Optimization for Inverse Problems using Deep Generative Models. *arXiv: 2102.07364*, February 2021. URL <http://arxiv.org/abs/2102.07364>.
- Davenport, M. A., Duarte, M. F., Wakin, M. B., Laska, J. N., Takhar, D., Kelly, K. F., and Baraniuk, R. G. The smashed filter for compressive classification and target recognition. In *Computational Imaging V*, volume 6498, pp. 64980H. International Society for Optics and Photonics, 2007.
- Donoho, D. L. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, April 2006.
- Duarte, M. F., Davenport, M. A., Takhar, D., Laska, J. N., Sun, T., Kelly, K. F., and Baraniuk, R. G. Single-pixel imaging via compressive sampling. *IEEE signal processing magazine*, 25(2):83–91, 2008.
- Falorsi, L., de Haan, P., Davidson, T. R., De Cao, N., Weiler, M., Forré, P., and Cohen, T. S. Explorations in homeomorphic variational auto-encoding. *ICML workshop on Theoretical Foundations and Applications of Deep Generative Models*, 2018.
- Foucart, S. and Rauhut, H. *A Mathematical Introduction to Compressive Sensing*. Applied and Numerical Harmonic Analysis. Springer New York, New York, NY, 2013.
- Hussein, S. A., Tirer, T., and Giryas, R. Image-Adaptive GAN based reconstruction. *AAAI*, 34(04):3121–3129, April 2020.
- Jordan, M. I., Ghahramani, Z., Jaakkola, T. S., and Saul, L. K. An introduction to variational methods for graphical models. *Machine learning*, 37(2):183–233, 1999.
- Karras, T., Aittala, M., Laine, S., Härkönen, E., Hellsten, J., Lehtinen, J., and Aila, T. Alias-Free generative adversarial networks. *arXiv:2106.12423*, June 2021.
- Kingma, D. P. and Welling, M. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.
- Köhler, J., Klein, L., and Noé, F. Equivariant flows: exact likelihood generative learning for symmetric densities. In *International Conference on Machine Learning*, pp. 5361–5370. PMLR, 2020.
- Kondor, R. and Trivedi, S. On the generalization of equivariance and convolution in neural networks to the action of

- compact groups. In *International Conference on Machine Learning (ICML)*, 2018.
- Lei, Q., Jalal, A., Dhillon, I. S., and Dimakis, A. G. Inverting Deep Generative models, One layer at a time. In *Advances in Neural Information Processing Systems*, pp. 13910–13919, 2019.
- Liu, Z. and Scarlett, J. The Generalized Lasso with Nonlinear Observations and Generative Priors. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M. F., and Lin, H. (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 19125–19136. Curran Associates, Inc., 2020.
- Liu, Z., Gomes, S., Tiwari, A., and Scarlett, J. Sample Complexity Bounds for 1-bit Compressive Sensing and Binary Stable Embeddings with Generative Priors. In III, H. D. and Singh, A. (eds.), *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 6216–6225. PMLR, July 2020.
- Lustig, M., Donoho, D., and Pauly, J. M. Sparse mri: The application of compressed sensing for rapid mr imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007.
- Moen, T. R., Chen, B., Holmes III, D. R., Duan, X., Yu, Z., Yu, L., Leng, S., Fletcher, J. G., and McCollough, C. H. Low-dose ct image and projection dataset. *Medical physics*, 48(2):902–911, 2021.
- Paredes, J. L., Arce, G. R., and Wang, Z. Ultra-wideband compressed sensing: Channel estimation. *IEEE Journal of Selected Topics in Signal Processing*, 1(3):383–395, 2007.
- Rezende, D. J., Mohamed, S., and Wierstra, D. Stochastic backpropagation and approximate inference in deep generative models. In *International Conference on Machine Learning*, pp. 1278–1286. PMLR, 2014.
- Satorras, V. G., Hoogeboom, E., Fuchs, F. B., Posner, I., and Welling, M. E (n) equivariant normalizing flows. In *Advances in Neural Information Processing Systems*, 2021.
- Serre, J.-P. *Linear representations of finite groups*. Springer, 1977.
- Singer, A. Mathematics for cryo-electron microscopy. In *Proceedings of the International Congress of Mathematicians: Rio de Janeiro 2018*, pp. 3995–4014. World Scientific, 2018.
- Tibshirani, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.
- Weiler, M. and Cesa, G. General E(2)-Equivariant Steerable CNNs. In *Conference on Neural Information Processing Systems (NeurIPS)*, 2019.
- Whang, J., Lei, Q., and Dimakis, A. Solving inverse problems with a flow-based noise model. In Meila, M. and Zhang, T. (eds.), *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 11146–11157. PMLR, 2021a.
- Whang, J., Lindgren, E., and Dimakis, A. Composing normalizing flows for inverse problems. In Meila, M. and Zhang, T. (eds.), *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 11158–11169. PMLR, 2021b.
- Whang, J., Lindgren, E., and Dimakis, A. Composing normalizing flows for inverse problems. In Meila, M. and Zhang, T. (eds.), *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 11158–11169. PMLR, 2021c.

A. Group Theory and Equivariant Neural Networks

We review some basics of group theory and group convolutional neural networks. There are other ways of building equivariant networks beyond group convolutional neural networks, see for example (Weiler & Cesa, 2019; Cohen & Welling, 2016b; Kondor & Trivedi, 2018; Cesa et al., 2022). Our theoretical results and equivariant VAE constructions are oblivious to the choice of architecture.

For a set \mathcal{G} , a law of composition $\cdot : \mathcal{G} \times \mathcal{G} \rightarrow \mathcal{G}$ is a mapping that maps $h, g \in \mathcal{G}$ to $h \cdot g \in \mathcal{G}$. A *group* is a set \mathcal{G} with a law of composition $\cdot : \mathcal{G} \times \mathcal{G} \rightarrow \mathcal{G}$, called group law, which satisfies following properties:

- The law \cdot is associative: $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ for all $a, b, c \in \mathcal{G}$
- There is an identity element in \mathcal{G} , denoted by 1, such that $a \cdot 1 = 1 \cdot a = a$ for all $a \in \mathcal{G}$
- Every element of $a \in \mathcal{G}$ has an inverse, denoted by b such that $a \cdot b = 1$ and $b \cdot a = 1$.

A group \mathcal{G} is an abelian group if the group law is commutative, namely $a \cdot b = b \cdot a$. A group \mathcal{G} is finite if the set \mathcal{G} is finite. A group \mathcal{G} is a compact group if \mathcal{G} is a compact topological space with continuous group operation.

For an arbitrary space \mathcal{X} , we can define the action of the group \mathcal{G} by a mapping $\mathcal{T}_g : \mathcal{G} \times \mathcal{X} \rightarrow \mathcal{X}$, which maps $x \in \mathcal{X}$ and $g \in \mathcal{G}$ to $\mathcal{T}_g x$. For the identity element, x is mapped to itself. In this work, the action of the group is assumed to be linear on a vector space \mathcal{X} , and \mathcal{T}_g has a matrix representation. The action of the group \mathcal{G} satisfies group properties indicates above, and therefore, \mathcal{T}_g is a linear representation of the group \mathcal{G} on \mathcal{X} . Representation theory characterizes linear representations of groups on general vector spaces. We work with unitary representations, that is, \mathcal{T}_g is a unitary transformation.

A mapping $f : \mathcal{X} \rightarrow \mathcal{Y}$ is equivariant to the action of the group \mathcal{G} if

$$f(\mathcal{T}_g^{\mathcal{X}} x) = \mathcal{T}_g^{\mathcal{Y}} f(x),$$

where $\mathcal{T}_g^{\mathcal{X}}$ and $\mathcal{T}_g^{\mathcal{Y}}$ are linear representations of the group \mathcal{G} on \mathcal{X} and \mathcal{Y} .

Group Convolution One example of equivariant neural networks is group equivariant convolutional networks (Cohen & Welling, 2016a). The group convolution is defined on the space \mathcal{X} as:

$$\forall g \in \mathcal{G} \quad (\mathbf{w} \otimes_G \mathbf{x})(g) := \langle \mathcal{T}_g \mathbf{w}, \mathbf{x} \rangle, \quad (39)$$

where $\mathbf{w}, \mathbf{x} \in \mathcal{X}$. The group convolution is an equivariant transformation w.r.t. the actions of group \mathcal{G} . Note that $(\mathbf{w} \otimes_G \mathbf{x}) \in \mathbb{R}^{|\mathcal{G}|}$, and the actions of group \mathcal{G} on $\mathbf{x} \in \mathbb{R}^{|\mathcal{G}|}$ is permuting the entries of \mathbf{x} . This means that the regular representation \mathcal{G} is used as group representation (Serre, 1977).

Each group convolution can be seen as a function defined on the group \mathcal{G} , and therefore, after the first layer, we can focus on functions on \mathcal{G} determined by $|\mathcal{G}|$ -dimensional vectors. Given the group $\mathcal{G} = (g_0, g_1, \dots, g_i, \dots, g_{|\mathcal{G}|-1})$, and vectors $\mathbf{x}, \mathbf{w} \in \mathbb{R}^{|\mathcal{G}|}$, the group convolution can be defined by a matrix multiplication between the vector \mathbf{x} and a matrix \mathbf{W} :

$$(\mathbf{w} \otimes_G \mathbf{x})(g_i) = (\mathbf{W} \mathbf{x})[i] \quad (40)$$

where the matrix \mathbf{W} is given as follows:

$$\mathbf{W} = \begin{pmatrix} w(g_0) & \dots & w(g_{|\mathcal{G}|-1}) \\ w(g_1^{-1}g_0) & \dots & w(g_1^{-1}g_{|\mathcal{G}|-1}) \\ \vdots & \ddots & \vdots \\ w(g_{|\mathcal{G}|-1}^{-1}g_0) & \dots & w(g_{|\mathcal{G}|-1}^{-1}g_{|\mathcal{G}|-1}). \end{pmatrix} \quad (41)$$

A group convolutional neural network is defined by concatenating group convolution layers with pointwise non-linearities in the middle. At each layer, multiple kernel like \mathbf{w} can be chosen, each one corresponding to a single channel. It is possible to use other representations of the group \mathcal{G} to build equivariant networks (see (Cohen & Welling, 2016b; Weiler & Cesa, 2019) for some examples).

B. Recovery Guarantees and Theoretical Analysis

In this section, we present more details on theoretical guarantees for unknown rotation. We consider the following inverse problem with unknown rotation g :

$$\mathbf{y} = \mathbf{A}\mathcal{T}_g^x \mathbf{x}_* + \varepsilon, \quad (42)$$

where \mathcal{T}_g^x is the group representation in the input space. We work with unitary representations, which means that $(\mathcal{T}_g^x)^{-1} = (\mathcal{T}_g^x)^H$. We start by introducing some preliminary definitions.

B.1. Definitions

As we will see, we require the sensing matrix \mathbf{A} to satisfy certain constraints. The first one is set-restricted eigenvalue (S-REC) condition.

Definition B.1. A matrix \mathbf{A} is said to satisfy S-REC($\mathcal{S}, \gamma, \delta$) for some $\delta \geq 0$ and $\gamma > 0$ and a set \mathcal{S} if for all $\mathbf{x}, \mathbf{y} \in \mathcal{S}$:

$$\|\mathbf{A}(\mathbf{x} - \mathbf{y})\|_2 \geq \gamma \|\mathbf{x} - \mathbf{y}\| - \delta.$$

This is reminiscent of the null space property in compressed sensing. The idea is to guarantee that two points from the set of interest \mathcal{S} are *separated* enough after mapping \mathbf{A} , i.e., their difference lies away from the null space of \mathbf{A} .

The second condition on \mathbf{A} is that for every fixed $\mathbf{x} \in \mathbb{R}^n$, we have:

$$\|\mathbf{A}\mathbf{x}\|_2 \leq 2 \|\mathbf{x}\|_2. \quad (43)$$

Note that this bound needs to hold for a fixed \mathbf{x} and not all \mathbf{x} , and therefore, this is a weaker assumption than bounding the spectral norm. We will see how these conditions are satisfied for a random Gaussian matrix.

B.2. Proof

To train the generative prior, we assume that the data is given in a canonical orientation. The existence of a canonical orientation is not necessary. The following argument can be reworked by considering the quotient space and its proper parametrization instead. Without rotation, generative priors enjoy recovery guarantees of type provided in (Bora et al., 2017). For canonical orientations, the very same analysis holds for equivariant generative priors too. The difference appears for the scenario with unknown orientation g .

The proof for the equivariant priors follows standard steps exactly as (Bora et al., 2017). The reason is that unknown rotations are absorbed into the latent of the generative model so the main steps remain intact. We replicate the essence of the proof here to be self-contained.

The proof strategy for Theorem 3.1 is as follows. First, let $\hat{\mathbf{z}}$ be the estimated latent that minimizes $\|\mathbf{y} - \mathbf{A}G(\mathbf{z})\|$ to within error of δ_{approx} . This means that for all \mathbf{z} , we have

$$\|\mathbf{y} - \mathbf{A}G(\hat{\mathbf{z}})\|_2 \leq \|\mathbf{y} - \mathbf{A}G(\mathbf{z})\|_2 + \delta_{\text{approx}}.$$

Also, let \mathbf{z}_* be the latent code² minimizing $\|\mathbf{x}_* - G(\mathbf{z})\|_2$. Note that the equivariance condition implies that:

$$\mathcal{T}_g^x G(\mathbf{z}) = G(\mathcal{T}_g^z \mathbf{z}),$$

for group representation \mathcal{T}_g^x and \mathcal{T}_g^z . Therefore, the unitary property of \mathcal{T}_g^x implies that

$$\|\mathbf{x}_* - G(\mathbf{z})\|_2 = \|\mathcal{T}_g^x \mathbf{x}_* - \mathcal{T}_g^x G(\mathbf{z})\|_2 = \|\mathcal{T}_g^x \mathbf{x}_* - G(\mathcal{T}_g^z \mathbf{z})\|_2.$$

This means that $\mathcal{T}_g^z \mathbf{z}_*$ is the minimizer of $\|\mathcal{T}_g^x \mathbf{x}_* - G(\mathbf{z})\|_2$. According to the assumption of the theorem, and as seen above, $\hat{\mathbf{z}}$ and \mathbf{z}_* satisfy:

$$\|\mathbf{y} - \mathbf{A}G(\hat{\mathbf{z}})\|_2 \leq \|\mathbf{y} - \mathbf{A}\mathcal{T}_g^x G(\mathbf{z}_*)\|_2 + \delta_{\text{approx}}.$$

²Existence of $\hat{\mathbf{z}}$ and \mathbf{z}_* follows from compactness of domain and continuity of the objective function. For a more general case, the proof can be replicated by replacing $\hat{\mathbf{z}}$ and \mathbf{z}_* with their definition.

The goal is to convert this inequality to another one in terms of $\|\mathcal{T}_g^x \mathbf{x}_* - G(\hat{\mathbf{z}})\|_2$ and $\|\mathbf{x}_* - G(\mathbf{z}_*)\|_2$. The main steps are contained in the following derivation:

$$\begin{aligned}
 \|\mathcal{T}_g^x G(\mathbf{z}_*) - G(\hat{\mathbf{z}})\|_2 &= \|G(\mathcal{T}_g^z \mathbf{z}_*) - G(\hat{\mathbf{z}})\|_2 && \text{(equivariance)} \\
 &\leq \frac{\|\mathbf{A}G(\mathcal{T}_g^z \mathbf{z}_*) - \mathbf{A}G(\hat{\mathbf{z}})\|_2 + \delta}{\gamma} && \text{(S-REC)} \\
 &\leq \frac{\|\mathbf{y} - \mathbf{A}G(\hat{\mathbf{z}})\|_2 + \|\mathbf{y} - \mathbf{A}\mathcal{T}_g^x G(\mathbf{z}_*)\|_2 + \delta}{\gamma} && \text{(triangle ineq.)} \\
 &\leq \frac{2\|\mathbf{y} - \mathbf{A}\mathcal{T}_g^x G(\mathbf{z}_*)\|_2 + \delta_{approx} + \delta}{\gamma} && \text{(theorem assumption)} \\
 &\leq \frac{2\|\mathbf{A}\mathcal{T}_g^x \mathbf{x}_* - \mathbf{A}\mathcal{T}_g^x G(\mathbf{z}_*)\|_2 + 2\|\boldsymbol{\varepsilon}\|_2 + \delta_{approx} + \delta}{\gamma} && \text{(triangle ineq.)} \\
 &\leq \frac{4\|\mathcal{T}_g^x \mathbf{x}_* - \mathcal{T}_g^x G(\mathbf{z}_*)\|_2 + 2\|\boldsymbol{\varepsilon}\|_2 + \delta_{approx} + \delta}{\gamma} && \text{(condition (43))}
 \end{aligned}$$

These inequalities, Lemma 4.3 in (Bora et al., 2017), only require triangle inequalities, S-REC property and bound on the condition (43), where the last two conditions are satisfied by random Gaussian sensing matrices with i.i.d. entries and appropriately chosen variance. Using the following triangle inequality,

$$\|\mathcal{T}_g^x G(\mathbf{z}_*) - G(\hat{\mathbf{z}})\|_2 \geq \|\mathcal{T}_g^x \mathbf{x}_* - G(\hat{\mathbf{z}})\|_2 - \|\mathcal{T}_g^x \mathbf{x}_* - \mathcal{T}_g^x G(\mathbf{z}_*)\|_2, \quad (44)$$

and unitary property of \mathcal{T}_g^x

$$\|\mathcal{T}_g^x \mathbf{x}_* - \mathcal{T}_g^x G(\mathbf{z}_*)\|_2 = \|\mathbf{x}_* - G(\mathbf{z}_*)\|_2 \quad (45)$$

the theorem follows.

Let's consider the S-REC property and the condition (43). The later is satisfied by a random Gaussian matrix properly normalized. We use Gaussian concentration inequality to prove this.

Theorem B.2 (Theorem 8.34 - (Foucart & Rauhut, 2013)). *For a Lipschitz function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ with Lipschitz constant L , if \mathbf{g} is a standard Gaussian random vector, then for all $t > 0$, we have*

$$\mathbb{P}(f(\mathbf{g}) - \mathbb{E}[f(\mathbf{g})] \geq t) \leq \exp\left(-\frac{t^2}{4L^2}\right).$$

We choose f as a function of the matrix \mathbf{A} and defined by $f(\mathbf{A}) = \|\mathbf{A}\mathbf{x}\|_2$. First, see that $f(\cdot)$ is Lipschitz with $L = \|\mathbf{x}\|_2$:

$$\|\|\mathbf{A}\mathbf{x}\|_2 - \|\mathbf{B}\mathbf{x}\|_2\| \leq \|\mathbf{A}\mathbf{x} - \mathbf{B}\mathbf{x}\|_2 \leq \|\mathbf{A} - \mathbf{B}\|_2 \|\mathbf{x}\|_2 \leq \|\mathbf{A} - \mathbf{B}\|_F \|\mathbf{x}\|_2.$$

Therefore, we get, from the theorem, and for standard Gaussian random matrix \mathbf{A} :

$$\mathbb{P}(\|\mathbf{A}\mathbf{x}\|_2 - \mathbb{E}(\|\mathbf{A}\mathbf{x}\|_2) \geq t) \leq \exp\left(\frac{-t^2}{4\|\mathbf{x}\|_2^2}\right)$$

Note that $\mathbb{E}(f(\mathbf{A})) = \mathbb{E}(\|\mathbf{A}\mathbf{x}\|_2)$ can be upper bounded as:

$$\mathbb{E}(\|\mathbf{A}\mathbf{x}\|_2) \leq (\mathbb{E}(\|\mathbf{A}\mathbf{x}\|_2^2))^{1/2} = \left(\sum_{i=1}^m \mathbb{E}(\langle \mathbf{a}_i, \mathbf{x} \rangle^2)\right)^{1/2} = (m\|\mathbf{x}\|_2^2)^{1/2} = \sqrt{m}\|\mathbf{x}\|_2.$$

Finally, by choosing $t = \sqrt{m}\|\mathbf{x}\|_2$, we get with probability $1 - e^{-m^2/4}$ that:

$$\|\mathbf{A}\mathbf{x}\|_2 \leq t + \sqrt{m}\|\mathbf{x}\|_2 = 2\sqrt{m}\|\mathbf{x}\|_2$$

Normalizing entries of the standard Gaussian random matrix \mathbf{A} by $1/\sqrt{m}$ will provide the desired result by choosing $\mathbf{x} = \mathcal{T}_g^x \mathbf{x}_* - \mathcal{T}_g^x G(\mathbf{z}_*)$ which is a fixed vector independent of \mathbf{A} .

The next step is, therefore, S-REC property. Since we use an equivariant generative model, the ambiguity about the rotation operation is captured in the latent space. Therefore, the very same approach of (Bora et al., 2017) can be used to build ϵ -nets and use that to establish S-REC for the equivariant generative model. We re-state this theorem for completeness here.

Lemma B.3. *For L -Lipshcitz equivariant generative model $G : \mathbb{R}^k \rightarrow \mathbb{R}^n$ and $\alpha < 1$, the random Gaussian matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ with i.i.d. entries $\mathcal{N}(0, m^{-1})$ and $m = \Omega\left(\frac{k}{\alpha^2} \log(Lr/\delta)\right)$ satisfies S-REC($G(B_k(r)), 1 - \alpha, \delta$) with probability $1 - e^{-\Omega(\alpha^2 m)}$.*

Note that the proof of this lemma remains identical to (Bora et al., 2017), since the equivariance property does not play any role in the proof. It is only necessary to establish it for a generative model G and choosing $\mathcal{S} = G(B_k(r))$.

As a final remark, consider for the moment the case of non-equivariant priors where the rotation needs to be extracted jointly with the latent code. In this case S-REC would translate to:

$$\|\mathcal{T}_{g_*}^x G(\mathbf{z}_*) - \mathcal{T}_{\hat{g}}^x G(\hat{\mathbf{z}})\|_2 \leq \frac{\left\| \mathbf{A} \mathcal{T}_{g_*}^x G(\mathbf{z}_*) - \mathbf{A} \mathcal{T}_{\hat{g}}^x G(\hat{\mathbf{z}}) \right\|_2 + \delta}{\gamma}. \quad (46)$$

In this case, we would need to establish S-REC for the set $\cup_{g \in \mathcal{G}} \mathcal{T}_g^x G(B_k(r))$. If the sets $\mathcal{T}_g^x G(B_k(r))$ are disjoint for different g , the number of ϵ -nets would increase, proportional to the ϵ -net required to cover the group \mathcal{G} . This would contribute another factor inside the logarithm for sample complexity.

C. Additional results

C.1. Eq-VAE: group choice

Group choice can be rather important to the performance of compressed sensing. Thus, we compare results of three different prior: equivariant under the cyclic group of 4, 8 and 16 rotations. In Figure 3 we report MSE for CS on MNIST with 100 measurements. All three priors perform well, but the best results are achieved when we use the cyclic group of 16 rotations.

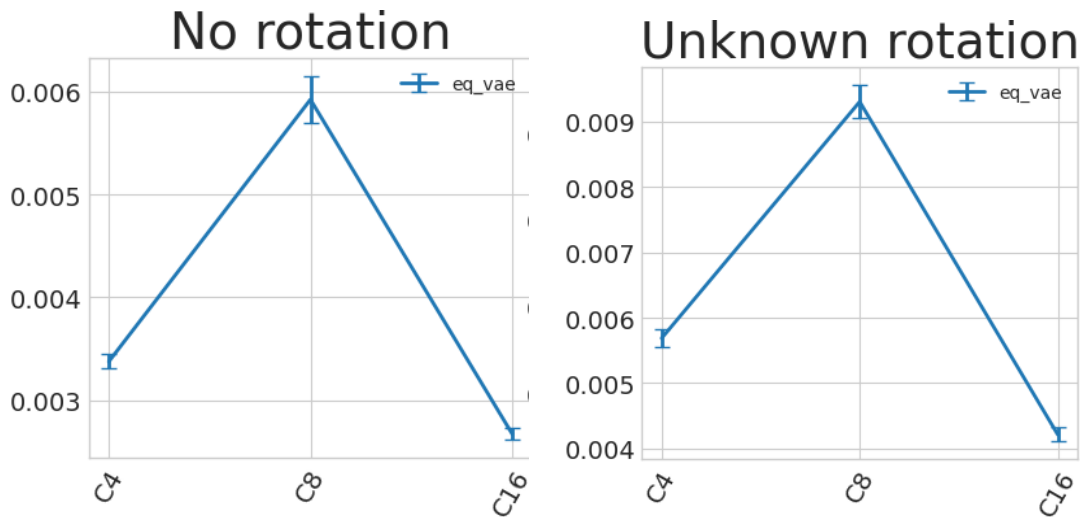


Figure 3: Compressed sensing with 100 measurements on MNIST. We compare VAE equivariant under the cyclic groups C_4 , C_8 and C_{16} .

C.2. Eq-VAE: Samples

We report the compressed sensing results for the C_{16} -equivariant VAE. Figure 4 shows, how the reconstructions change when we apply the group action to a latent code and push it through the decoder (for three different latent codes). This way we observe that even though the model was not trained on the rotated images, it can generate them perfectly well.

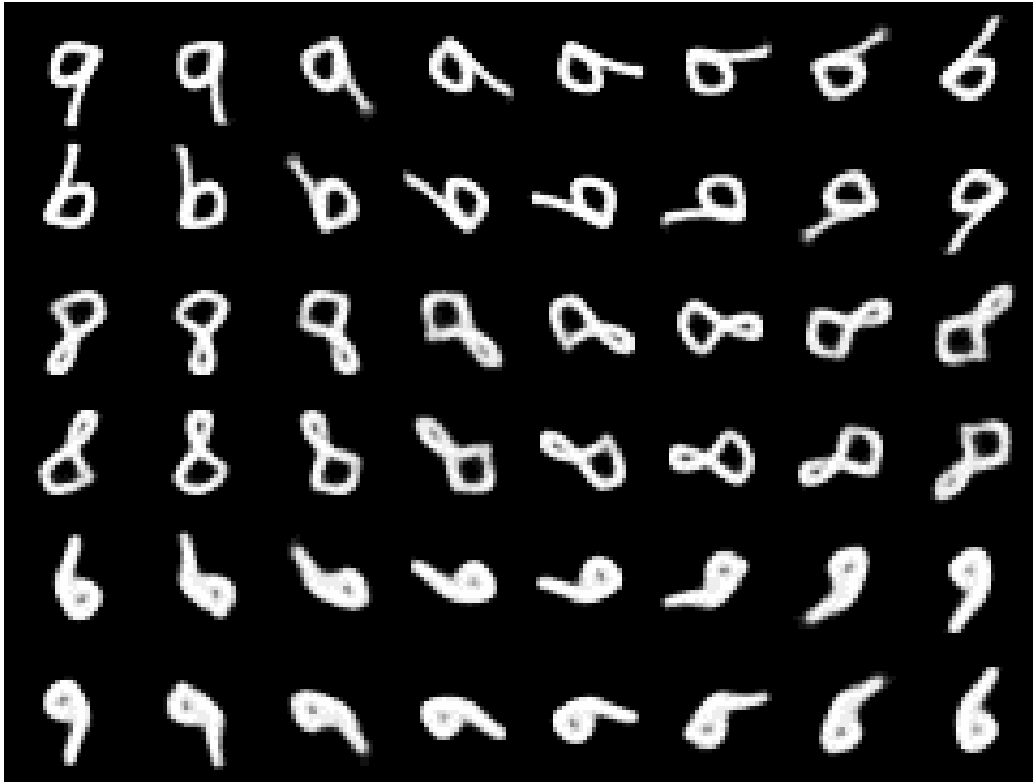


Figure 4: Samples from the equivariant VAE.

C.3. Eq-VAE Priors for MAYO

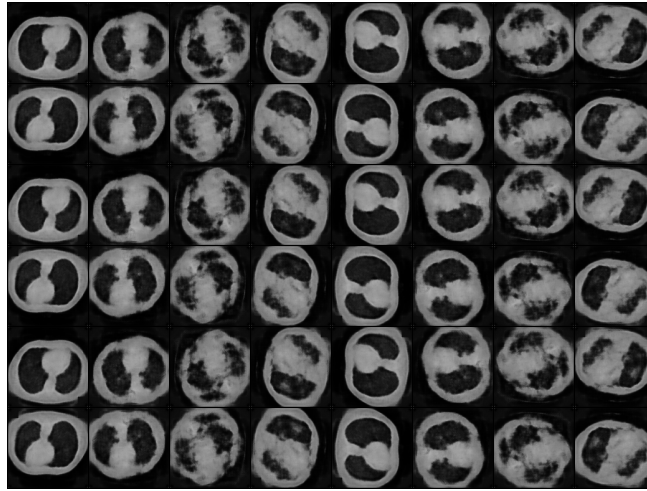


Figure 5: Rotated reconstructions from Eq-VAEs for MAYO dataset.

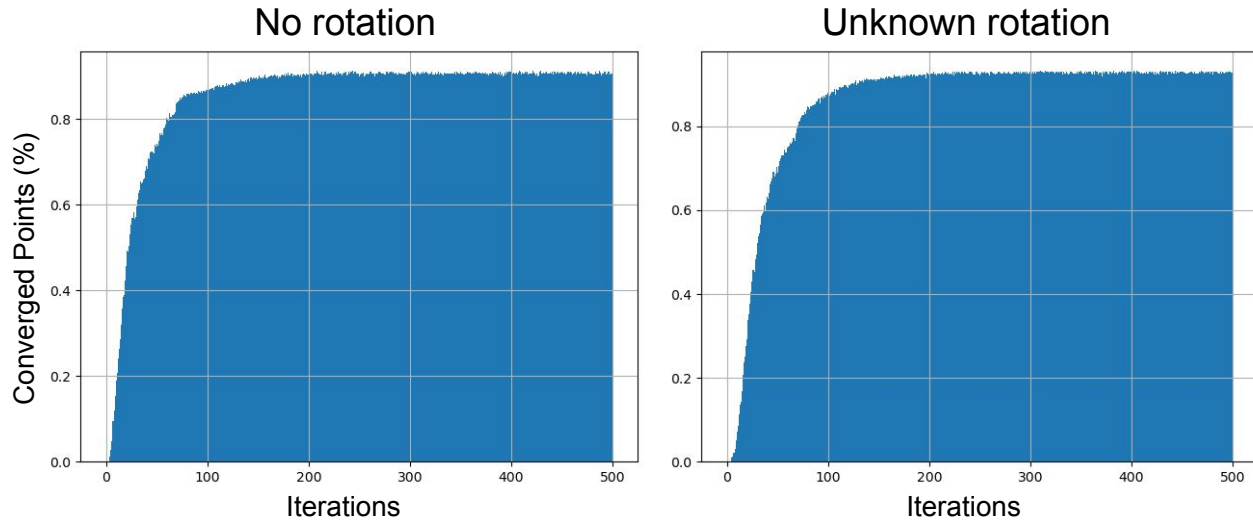


Figure 6: Number of iterations until convergence (i.e., $MSE \leq 1.e^{-2}$), for the MAYO dataset, using Eq-VAE as a generative prior. Left: No rotation angle. Right: Unknown rotations.

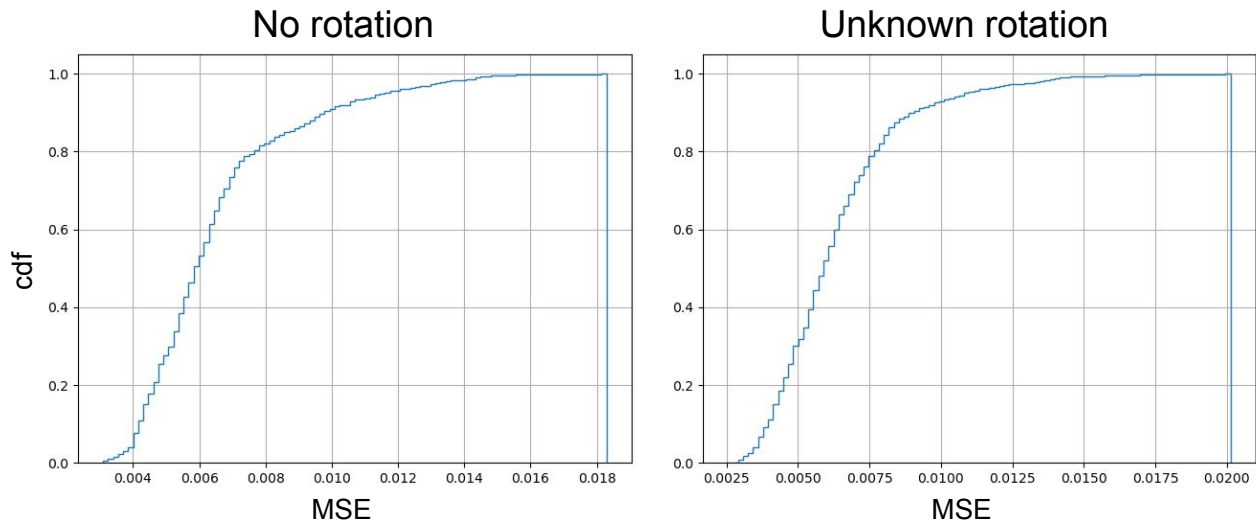


Figure 7: MSE from compressed sensing experiments, for the MAYO dataset, using Eq-VAE as a generative prior. Left: No rotation angle. Right: Unknown rotations.

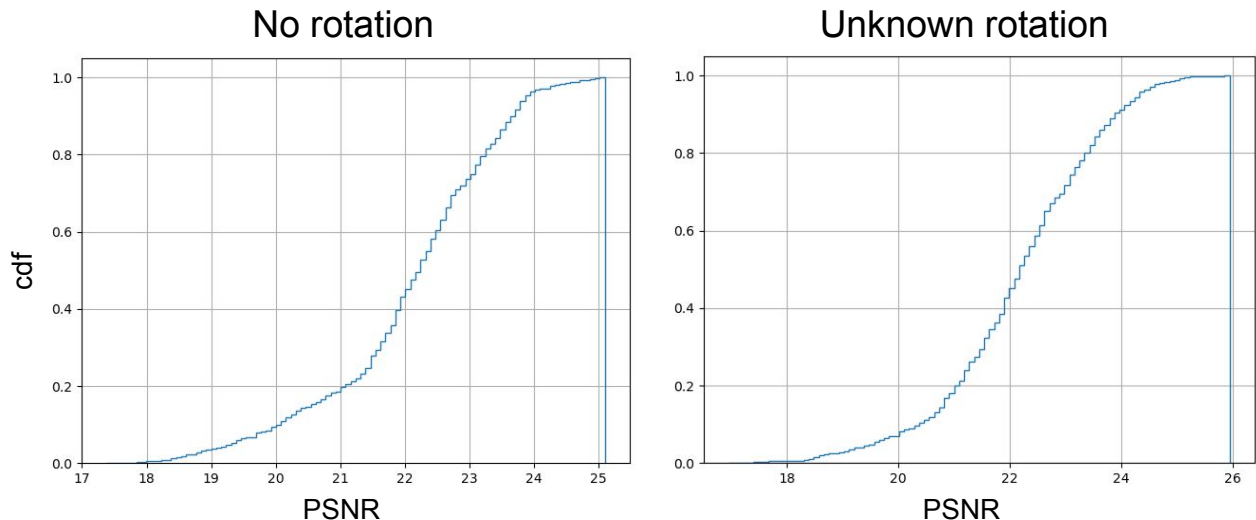


Figure 8: PSNR from compressed sensing experiments, for the MAYO dataset, using Eq-VAE as a generative prior. Left: No rotation angle. Right: Unknown rotations.

D. Experimental Setup

D.1. Training Generative Prior

D.1.1. CONVOLUTIONAL VAE AND EQUIVARIANT VAE

Architecture We train a VAE model, where both encoder and decoder are fully convolutional neural networks with the ReLU activations.

The representation space size is set to $z_{dim} = 128$.

For MNIST, the Conv-VAE encoder is fully convolutional architecture with kernel size, number of filters, stride, and padding as follows: Input signal (1) $\rightarrow [(3, 32, 2, 1) \rightarrow (3, 64, 2, 1) \rightarrow (3, 96, 2, 1) \rightarrow (3, 128, 2, 1) \rightarrow (3, 256, 2, 1) \rightarrow \text{flatten}(\cdot) \rightarrow (\mu_z, \log \sigma_z^2)]$. The decoder comprises of transpose convolutional layers. Following the preceding format, the architectural details of decoder is as follows: Input signal (128) $\rightarrow [(3, 128, 1, 0) \rightarrow (3, 96, 1, 0) \rightarrow (3, 64, 1, 1) \rightarrow$

$(4, 32, 2, 1) \rightarrow (4, 1, 2, 1) \rightarrow \text{flatten}(\cdot) \rightarrow (\mu_x)$.

For MAYO, the Conv-VAE encoder is fully convolutional architecture with kernel size, number of filters, stride, and padding as follows: Input signal (1) $\rightarrow [(5, 32, 1, 0) \rightarrow (3, 32, 1, 0) \rightarrow (3, 64, 2, 0) \rightarrow (3, 64, 1, 0) \rightarrow (3, 128, 2, 0) \rightarrow (3, 128, 2, 0) \rightarrow (3, 256, 2, 0) \rightarrow (3, 256, 2, 0) \rightarrow (2, 256, 2, 0) \rightarrow \text{flatten}(\cdot) \rightarrow (\mu_z, \log \sigma_z^2)]$. The decoder comprises of transpose convolutional layers. Following the preceding format, the architectural details of decoder is as follows: Input signal (128) $\rightarrow [(6, 256, 1, 0) \rightarrow (3, 256, 2, 0) \rightarrow (3, 128, 2, 0) \rightarrow (4, 128, 1, 0) \rightarrow (3, 64, 2, 0) \rightarrow (3, 32, 2, 0) \rightarrow (4, 32, 1, 0) \rightarrow (3, 1, 1, 0) \rightarrow \text{flatten}(\cdot) \rightarrow (\mu_x)]$.

For MAYO, the Equivariant VAE encoder comprises of group equivariant convolutional architecture with kernel size, number of filters, stride, and padding as follows: Input signal (1) $\rightarrow [(5, 16, 1, 0) \rightarrow (3, 16, 1, 0) \rightarrow (3, 16, 2, 0) \rightarrow (3, 16, 2, 0) \rightarrow (3, 32, 2, 0) \rightarrow (3, 32, 1, 0) \rightarrow (3, 32, 1, 0) \rightarrow (3, 64, 1, 0) \rightarrow (2, 256, 2, 0) \rightarrow \text{flatten}(\cdot) \rightarrow (\mu_z, \log \sigma_z^2)]$. The decoder comprises of equivariant transpose convolutional layers. Following the preceding format, the architectural details of decoder is as follows: Input signal (128) $\rightarrow [(6, 128, 1, 0) \rightarrow (3, 32, 2, 0) \rightarrow (3, 128, 2, 0) \rightarrow (4, 32, 1, 0) \rightarrow (3, 16, 2, 0) \rightarrow (3, 16, 2, 0) \rightarrow (4, 8, 1, 0) \rightarrow (3, 8, 1, 0) \rightarrow (3, 1, 1, 0) \rightarrow \text{flatten}(\cdot) \rightarrow (\mu_x)]$.

Note: Aug-VAE and Aug-Flow follow the same architecture as Conv-VAE and Flow respectively. The only difference between augmented and their non-augmented counterpart is that the augmented models are trained on augmented (rotated) dataset.

D.1.2. MULTI-SCALE REALNVP

Architecture The architecture of RealNVP comprises of 8 real-NVP blocks with scaling set to 3.