

UNIVERSITÀ DEGLI STUDI DI GENOVA

Ph.D. in Mathematics and Applications

XXXVIII cycle



**Università
di Genova**

Generative models for inverse problems

Supervisors

Prof. Matteo SANTACESARIA

Prof. Vito Paolo PASTORE

Candidate

Paolo ANGELLA

July 2026

Generative models for inverse problems

Paolo Angella

Abstract

Inverse problems are ubiquitous in scientific and engineering applications, where the goal is to recover an unknown signal from indirect and often corrupted observations. These problems are typically ill posed, requiring the incorporation of prior information to ensure stable and meaningful solutions. In recent years, deep generative models have emerged as powerful tools for encoding complex data distributions, offering a promising alternative to classical regularization techniques.

This thesis investigates the use of generative models for solving inverse problems, with a particular focus on medical imaging applications and theoretical advancements in generative sensing. The work is structured into two main parts.

In the first part, we address the problem of motion artifact correction in brain Magnetic Resonance Imaging (MRI). We evaluate two primary applications of diffusion models: deploying them for direct motion correction, and utilizing them for data augmentation to generate synthetic pairs for supervised training. Through extensive numerical experiments on real MRI datasets, we quantitatively and qualitatively outline the strengths and weaknesses of these methods, while highlighting important considerations regarding model reliability and clinical applicability.

The second part focuses on a theoretical extension of generative approaches to inverse problems. We introduce an infinite dimensional formulation of generative compressed sensing, where signals are assumed to lie in the range of a generative operator rather than being sparse in some predefined basis. Within this framework, we establish recovery guarantees and analyze the role of the sampling strategy, showing how adaptive measurements informed by the generative model can significantly improve reconstruction performance. Finally we perform numerical experiments, including applications to function space, to support the theoretical findings.

Overall, this thesis demonstrates that generative models provide a flexible and effective framework for tackling inverse problems, bridging the gap between practical applications and theoretical understanding, and opening new directions for data driven reconstruction methods.

Table of Contents

1	Introduction	1
1.1	Inverse Problems	2
1.1.1	Examples of inverse problems	3
1.1.2	Variational techniques.	4
1.1.3	The Bayesian approach.	6
1.2	Deep Learning	7
1.2.1	Application to inverse problems	8
1.3	Generative Modeling for Inverse Problems	9
1.3.1	Generative models examples	9
1.3.2	Solving Inverse Problems with Generative Models.	11
1.4	Overview and Contributions	13
2	Diffusion models for motion correction in magnetic resonance imaging	15
2.1	Introduction	15
2.1.1	Motion artifacts	15
2.1.2	Classical techniques	16
2.1.3	Deep learning	16
2.1.4	Literature review	17
2.1.5	Denoising diffusion probabilistic model	18
2.1.6	U-net	19
2.2	Assessing the use of diffusion models for direct motion artifact correction in brain MRI	20
2.2.1	Methods	20
2.2.2	Experiments	21
2.2.2.1	Dataset	21
2.2.2.2	Experiment Details	21
2.2.3	Results	21
2.2.3.1	Parameter Tuning	22
2.2.3.2	Transverse view	23
2.2.3.3	Sagittal view	23
2.2.3.4	Coronal view	23
2.2.4	Summary and discussion	24
2.3	Using diffusion models for data pair generation in MRI motion correction	24
2.3.1	Dataset	24

2.3.2	Implementation details	25
2.3.2.1	Diffusion model	25
2.3.2.2	U-net	25
2.3.2.3	Data and parameter selection	26
2.3.2.4	Computational requirements	28
2.3.3	Results	28
2.3.3.1	Challenging settings	29
2.3.4	Summary and discussion	30
3	Infinite dimensional generative sensing	35
3.1	Introduction	35
3.1.1	Related work	36
3.1.2	Structure of the chapter	37
3.2	Setup and Main Results	38
3.2.1	Local Coherence and Sampling Strategies	40
3.2.2	Recovery Guarantees	41
3.2.2.1	The Balancing Property	42
3.3	Numerical Validation	43
3.3.1	Experimental Setup and Methodology	44
3.3.1.1	Dataset and Architecture	44
3.3.1.2	Training	45
3.3.1.3	Fitting the Latent Space	45
3.3.1.4	Reconstruction	46
3.3.2	Results	46
3.3.2.1	Local coherence and adaptive sampling	46
3.3.2.2	Adaptive sampling	47
3.3.2.3	Discretization invariance of the reconstruction error	47
3.3.2.4	GMM as regularizer	50
3.4	Proofs	51
3.4.1	Setup and local coherence	51
3.4.2	Generalized Restricted Isometry Property	54
3.4.3	Recovery bounds	58
3.5	Summary and discussion	59
	Bibliography	63

Chapter 1

Introduction

Thesis outline

This work is divided into three different chapters:

In the first one, we provide an overview of the fundamental concepts required for the subsequent chapters. We begin by presenting the general framework of inverse problems, followed by an introduction to deep learning methodologies. We then discuss how deep learning techniques can be employed to address inverse problems. Finally, we introduce generative models and examine their application, with particular emphasis on their role in solving inverse problems.

The second chapter is based on the following two articles, developed in the course of this thesis:

- P. Angella, V. P. Pastore, and M. Santacesaria, “**Assessing the Use of Diffusion Models for Motion Artifact Correction in Brain MRI,**” in *Proceedings of the 2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI)*.
- P. Angella *et al.*, “**DIMA: Diffusing Motion Artifacts for Unsupervised Correction in Brain MRI Images,**” *IEEE Access*, vol. 13, 2025.

These works constitute a direct application of the proposed framework, focusing on motion artifact correction in MRI through the use of diffusion models.

The last chapter is based on the preprint:

- P. Angella, V. P. Pastore, and M. Santacesaria, “**Infinite Dimensional Generative Sensing,**” arXiv preprint arXiv:2603.03196, 2026.

This work extends the proposed framework to the infinite-dimensional setting, providing a theoretical and methodological foundation for generative approaches to inverse problems in functional spaces.

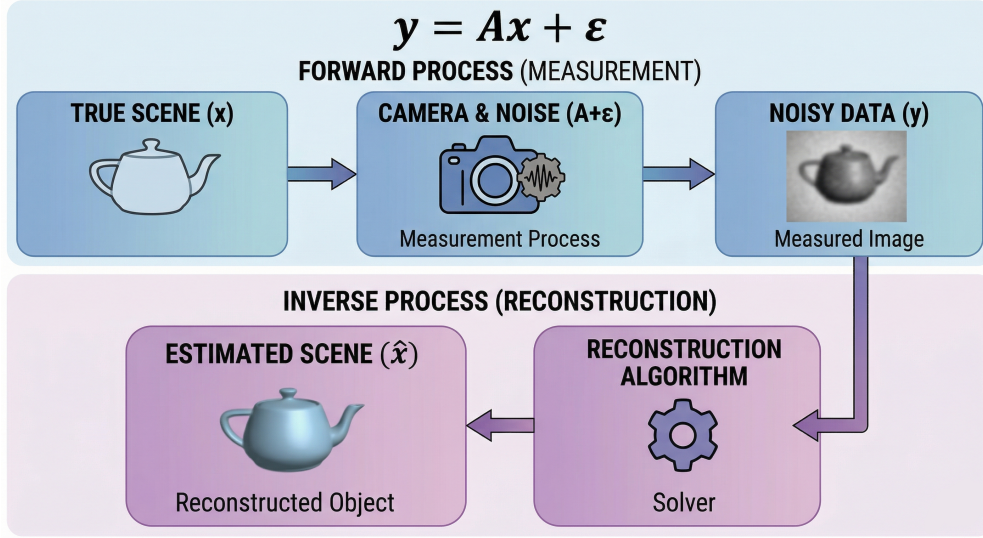


Figure 1.1: Inverse problem example.

1.1 Inverse Problems

An inverse problem is the task of recovering information of interest from indirect and noisy measurements. To model this, consider two Hilbert spaces X and Y , with $x \in X$ the quantity we want to recover, $y \in Y$ the observed measurements, $A : X \rightarrow Y$ a measurement operator, and $\varepsilon \in Y$ the noise that affects our measurement. We can then model our setup with

$$y = Ax + \varepsilon.$$

This framework can be made more general, for instance A may be nonlinear or the noise non additive, but this formulation is sufficient for the present discussion. We can call the problem of obtaining y from x given A the “direct” one that is generally easy, while the opposite or “inverse” one is usually much harder. Inverse problems are an area of fundamental importance since they have applications in many fields such as medical imaging, geophysics, signal processing, industrial testing, and others [1, 2].

To understand the challenge of solving inverse problems, it is helpful to first define what makes a mathematical problem “well posed”. According to Jacques Hadamard [3], a problem is **well posed** if it satisfies three properties:

1. **Existence:** There is at least one solution x for every observation y , meaning the range of the operator covers the observation space.
2. **Uniqueness:** There is at most one solution x for every observation y .
3. **Stability:** The solution depends continuously on the data, such that a bounded perturbation in y guarantees a bounded perturbation in x .

In practice, inverse problems are often **ill posed**, meaning they violate one or more of these properties. Some examples of inverse problems are as follows.

1.1.1 Examples of inverse problems

Deblurring. In many imaging systems, the measured image is a smoothed version of the object being imaged. Here A measure by acting as a convolution operator, $Ax = k * x$ for some blur kernel k (also known as a point spread function) on some domain Ω [4]. The continuous formulation is given by:

$$(Ax)(t) = \int_{\Omega} k(t - \tau)x(\tau)d\tau$$

and recovery of x amounts to inverting this smoothing integral.

Computed Tomography (CT). In CT imaging, A is the Radon transform, which maps a function x defined on a two dimensional domain to its line integrals over a collection of directions [5]. For a line $L_{s,\theta}$ parameterized by distance s and angle θ , the measurement is:

$$y(s, \theta) = \int_{L_{s,\theta}} x(\mathbf{r})d\ell.$$

In practice, this means we can determine the internal density of an object (like a human body) simply by measuring how X rays are absorbed from various angles, avoiding destructive procedures.

Fourier subsampling. In applications like magnetic resonance imaging (MRI), we ideally want to sample the complete Fourier transform of the signal being reconstructed. In practice, however, it is only feasible to acquire a subset of the required data [6]. Because we are missing frequencies, the number of acquired measurements is significantly smaller than the total number of unknown image pixels. We can express this linear inverse problem mathematically as

$$y = \mathcal{F}_{\Omega}x + \varepsilon$$

where y represents the subsampled data and $\mathcal{F}_{\Omega}$ is the undersampled Fourier operator. Since this system is strictly underdetermined, there are infinitely many potential images x that perfectly match the acquired measurements in the frequency space. This ambiguity directly violates the uniqueness property required for a well posed problem.

Denoising. Finally we have a task at the edge of inverse problems and imaging, it arises when $A = I$, the identity operator. Here the measurements are a direct observation of the signal but with some corruption given by the noise:

$$y = x + \varepsilon$$

and the task is to recover x from the noisy observation y .

Since the commonalities between all this different models, different strategies have been developed to tackle them.

¹Image credit: <https://compphys.go.ro/nuclear-magnetic-resonance-and-fourier-transform/>

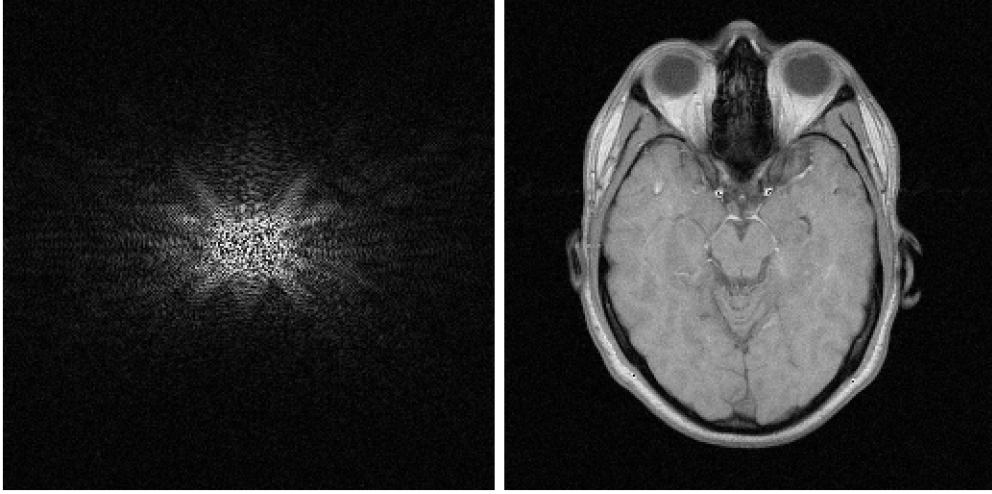


Figure 1.2: The measurement obtained by an MRI scan and the reconstructed image.¹

1.1.2 Variational techniques.

Unregularized Recovery. The most direct approach to solving the inverse problem $y = Ax + \varepsilon$ is to look for a signal x whose forward projection Ax best matches the observed measurements y . This leads to the classical least squares formulation, which minimizes the discrepancy between the predicted and observed data:

$$\min_x \|Ax - y\|_2^2.$$

This formulation assumes that the measurement noise ε is small enough and that the forward operator A is sufficiently well conditioned. When A has full column rank, the minimizer admits the closed form solution given by the Moore-Penrose pseudoinverse [7]:

$$x_{\text{est}} = (A^T A)^{-1} A^T y.$$

While this solution is optimal in the least squares sense, it becomes highly unstable when the inverse problem is ill posed. In particular, if $A^T A$ is poorly conditioned or nearly singular, even small perturbations in the measurements can be greatly amplified in the reconstructed signal. As a result, the recovered solution may be dominated by noise and therefore unusable in practice.

Regularization Techniques. A common strategy to mitigate the instability of inverse problems is to incorporate additional prior information about the solution. This general principle is referred to as regularization, where the reconstruction problem is augmented with a penalty term that promotes desirable properties in the recovered signal.

A broad class of methods can be written in the variational form:

$$\min_x \|Ax - y\|_2^2 + \lambda R(x),$$

where $R(x)$ is a regularization functional encoding prior assumptions, and $\lambda > 0$ controls

the trade-off between data fidelity and prior enforcement.

One of the most classical choices is Tikhonov regularization [8], corresponding to a quadratic penalty:

$$R(x) = \|\Gamma x\|_2^2,$$

where the operator Γ reflects structural assumptions on the solution. For instance, $\Gamma = I$ penalizes large amplitudes, while differential operators promote smoothness. In this case, the problem admits the closed-form solution

$$x_\lambda = (A^T A + \lambda \Gamma^T \Gamma)^{-1} A^T y.$$

Beyond quadratic penalties, more sophisticated regularizers have been introduced to better capture the underlying structure of signals. A prominent example is Total Variation (TV) regularization[9], which promotes piecewise constant solutions by penalizing the gradient magnitude, making it particularly effective for imaging applications with sharp edges. Similarly, sparsity-promoting regularizers such as ℓ_1 -norm[10] penalties encourage solutions with few nonzero coefficients in a suitable representation.

In general, the regularization term improves stability by controlling the effective conditioning of the inverse problem and suppressing the amplification of noise during inversion, at the cost of introducing bias. Different choices of $R(x)$ reflect different prior assumptions and lead to different reconstruction properties, highlighting the importance of selecting a regularizer consistent with the structure of the underlying signal.

Compressed Sensing. While quadratic regularization encourages smooth solutions, many real world signals possess a different form of structure: they are sparse (meaning most of their components are exactly zero) or compressible (meaning they can be accurately represented by just a few nonzero coefficients) after an appropriate transformation. This observation motivates the framework of compressed sensing [11, 12]. The key assumption is that the signal x can be represented using only a small number of significant coefficients in some basis or dictionary Ψ . Instead of penalizing energy, compressed sensing promotes sparsity through an ℓ_1 regularization term:

$$\min_x \frac{1}{2} \|Ax - y\|_2^2 + \lambda \|\Psi x\|_1.$$

The ℓ_1 norm acts as a convex surrogate for the nonconvex ℓ_0 sparsity measure and encourages solutions with many coefficients exactly equal to zero. Remarkably, under suitable conditions on the measurement operator A , such as the Restricted Isometry Property (RIP) [13], this convex optimization problem can recover the true sparse signal exactly from a surprisingly small number of measurements. Specifically, a matrix A satisfies the Restricted Isometry Property of order s if there exists a restricted isometry constant $\delta_s \in (0, 1)$ such that the following condition holds for all vectors x with at most s nonzero entries:

$$(1 - \delta_s) \|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta_s) \|x\|_2^2.$$

This property guarantees that the measurement matrix A nearly preserves the Euclidean

length of all sparse signals, ensuring that they are not mapped to the null space of the operator. This insight has had a profound impact on fields such as medical imaging, where it enables accelerated acquisition strategies. In particular, for matrices with i.i.d. subgaussian entries, the RIP of order s holds with high probability once the number of measurements satisfies

$$m = O\left(s \log\left(\frac{n}{s}\right)\right),$$

showing that the required sampling rate scales with the sparsity s rather than the ambient dimension n [14].

1.1.3 The Bayesian approach.

An alternative and conceptually powerful perspective is provided by Bayesian inference, where the inverse problem is formulated in terms of probability distributions [15, 16]. In this framework, both the unknown signal and the measurements are treated as random variables.

Bayes theorem combines the likelihood of observing the data given a candidate solution with prior knowledge about the signal:

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}.$$

Here, $p(y|x)$ represents the likelihood, which describes the measurement process and noise model, while $p(x)$ encodes prior beliefs about the structure of the signal. The denominator $p(y)$ represents the marginal likelihood or evidence. It acts purely as a normalizing constant to ensure the posterior is a valid probability distribution that integrates to exactly one. Because it does not depend on the unknown signal x , it is safely ignored during the optimization steps used to find specific point estimators.

A commonly used estimator is the Maximum A Posteriori (MAP) estimate, which selects the most probable solution given the observed data:

$$x_{MAP} = \arg \min_x (-\log p(y|x) - \log p(x))$$

This formulation reveals a deep connection between probabilistic inference and classical regularization. In particular, assuming Gaussian noise leads to a quadratic data fidelity term, while choosing a Gaussian prior on x results in a quadratic regularization penalty. Under these assumptions, the MAP estimator coincides exactly with the Tikhonov regularized solution [17]. This equivalence highlights how many deterministic reconstruction methods can be interpreted as special cases of Bayesian inference with specific prior assumptions.

Besides the MAP estimate, the Bayesian framework naturally accommodates other point estimators derived from the posterior distribution. A prominent example is the Conditional Mean estimator, which minimizes the mean square error. It is defined as the

expected value of the posterior:

$$x_{CM} = \int x p(x|y) dx.$$

While the MAP estimate identifies the peak of the probability density, the Conditional Mean provides the center of mass, which can be more robust for complex or highly skewed distributions.

Another foundational estimator is the Maximum Likelihood (ML) estimate. Unlike MAP, the ML approach does not incorporate any prior knowledge about the signal, implicitly assuming a uniform prior. It seeks the solution that maximizes the probability of the observed data:

$$x_{ML} = \arg \max_x p(y|x) = \arg \min_x (-\log p(y|x))$$

The ML estimate is particularly useful when prior information is unavailable or when the data is informative enough to overwhelm the prior. However, for ill posed inverse problems, ML estimation often leads to unstable or noisy solutions, underscoring the necessity of the prior term in the MAP formulation.

Perhaps the most significant advantage of this probabilistic formulation, compared to the variational approach, is the ability to quantify uncertainty. Variational methods typically yield a single deterministic solution without any measure of confidence. In contrast, the Bayesian method yields the entire posterior distribution $p(x|y)$. This comprehensive statistical description allows us to evaluate exactly how certain a solution is by computing covariance matrices or credible regions. Consequently, we can rigorously assess the reliability of the reconstructed signal and understand which features are strictly determined by the data versus those driven by prior beliefs.

1.2 Deep Learning

Deep learning has emerged as a powerful paradigm for solving complex computational tasks, including inverse problems [18, 19]. At its core, deep learning relies on artificial neural networks, which are composed of interconnected layers of artificial neurons [20]. A standard feedforward network applies a series of linear transformations followed by nonlinear activation functions, allowing it to approximate highly complex mapping functions between inputs and outputs. Mathematically, a neural network f_θ with L layers and learnable parameters $\theta = \{W_l, b_l\}_{l=1}^L$ can be expressed as a composition of functions:

$$f_\theta(y) = \sigma(W_L \sigma(\dots \sigma(W_1 y + b_1) \dots) + b_L)$$

where W_l are the weight matrices, b_l are the bias vectors, and σ represent a pointwise nonlinear activation function such as the ReLU, which is defined as $ReLU(x) = \max(0, x)$.

For spatial data such as images, Convolutional Neural Networks (CNNs) have become the standard architecture [21]. Instead of fully connected layers, CNNs utilize convolutional filters that slide across the input, capturing local patterns like edges and

textures. This translation invariant property makes them exceptionally well suited for imaging inverse problems like deblurring and medical image reconstruction.

1.2.1 Application to inverse problems

The most straightforward application of deep learning to inverse problems is the end to end supervised approach [22]. Given a large dataset of N paired noisy measurements and ground truth signals $\{(y_i, x_i)\}_{i=1}^N$, a neural network f_θ , belonging to a high dimensional class of parameterized functions, is trained to directly map the observation to the clean signal. This is achieved by finding the optimal parameters θ^* that minimize a chosen empirical loss function $\mathcal{L}$, such as the mean squared error:

$$\theta^* = \arg \min_{\theta} \frac{1}{N} \sum_{i=1}^N \mathcal{L}(f_\theta(y_i), x_i).$$

The minimum is usually obtained by using stochastic gradient descent. While the data-driven mapping f_θ can be extremely fast during inference, it often lacks guarantees of physical consistency with the forward model A and struggles to generalize outside the training distribution. Furthermore, standard supervised networks often yield overly smoothed reconstructions when faced with highly ill posed problems, as they tend to output the conditional mean of all possible solutions rather than a sharp image.

A common alternative is the **postprocessing** approach [23]. Here, an initial reconstruction $\tilde{x}$ is obtained using a simple analytical method, such as the pseudoinverse $A^\dagger$, and a neural network f_θ is subsequently trained to remove artifacts and restore high frequency details. The final estimate is formulated as:

$$x_{est} = f_\theta(A^\dagger y)$$

To address these limitations and bridge the gap between pure data-driven methods and classical model based approaches, researchers developed physics-informed deep learning strategies.

A prominent example is **deep unrolling**, also known as algorithm unrolling [24]. This technique takes a classic iterative optimization algorithm, such as proximal gradient descent, and truncates it to a fixed number of iterations. The standard proximal gradient step to minimize a data fidelity term with a regularization penalty $R(x)$ is given by:

$$x_{k+1} = \text{prox}_{\gamma R}(x_k - \gamma A^T(Ax_k - y))$$

where γ is the step size and A^T is the adjoint operator. In deep unrolling, the handcrafted proximal operator $\text{prox}_{\gamma R}$ is replaced by a learnable convolutional neural network $\mathcal{H}_{\theta_k}$ at each iteration k . The unrolled update rule becomes:

$$x_{k+1} = \mathcal{H}_{\theta_k}(x_k - \gamma A^T(Ax_k - y))$$

By mimicking these iterations, these hybrid models combine the mathematical reliability

and strict fidelity to the measurement physics with the expressive power of deep learning.

Another highly successful framework is the **plug and play** prior methodology [25]. Instead of explicitly defining a mathematical penalty like the Tikhonov matrix, plug and play methods replace the regularization step in an iterative algorithm, such as Half Quadratic Splitting, with an advanced pretrained neural network denoiser $\mathcal{D}_\sigma$. The optimization splits into two alternating steps. First, a data consistency update is computed:

$$z_k = \arg \min_x \frac{1}{2} \|Ax - y\|_2^2 + \frac{\rho}{2} \|x - v_{k-1}\|_2^2,$$

followed by a denoising update that acts as the learned prior:

$$v_k = \mathcal{D}_\sigma(z_k).$$

With this approach, these methods combine the interpretability of physical models with the unparalleled representational power of deep networks.

Despite these advances, deterministic networks still struggle to fully quantify uncertainty or reconstruct fine high frequency details in strictly underdetermined scenarios. To capture the full posterior distribution and generate highly detailed and realistic reconstructions, the field is increasingly shifting towards probabilistic approaches, specifically generative modeling [26].

1.3 Generative Modeling for Inverse Problems

While classical methods rely on handcrafted priors, generative models aim to learn the true underlying probability distribution of the data, denoted as $p(x)$, directly from a large dataset. By capturing the complex statistical dependencies within the data, these models serve as exceptionally powerful priors for solving ill posed inverse problems [27]. The most prominent families of generative models include Normalizing Flows, Variational Autoencoders, Generative Adversarial Networks, and Diffusion Models.

1.3.1 Generative models examples

Normalizing Flows. Normalizing flows construct complex distributions by applying a sequence of invertible and differentiable transformations to a simple base distribution, such as a standard Gaussian [28]. Let z be a latent variable sampled from a simple prior $p_Z(z)$, and let f be an invertible mapping such that $x = f(z)$. Using the change of variables formula, the probability density of the target data distribution $p_X(x)$ can be computed exactly:

$$p_X(x) = p_Z(f^{-1}(x)) \left| \det \left(\frac{\partial f^{-1}(x)}{\partial x} \right) \right|.$$

The primary advantage of normalizing flows is this exact likelihood computation, which allows for direct maximum likelihood training. After training, new data can be generated straightforwardly by sampling a latent variable $z \sim p_Z(z)$ from the simple base distribution and passing it through the learned transformation $x = f(z)$. However, designing

expressive yet strictly invertible transformations can be architecturally restrictive, which limits the flexibility of this particular approach.

Variational Autoencoders (VAEs). VAEs are probabilistic models that map data into a lower dimensional latent space [29]. They consist of a probabilistic encoder $q_\phi(z|x)$ that approximates the intractable true posterior, and a generative decoder $p_\theta(x|z)$ that reconstructs the data from the latent representation z . Since direct likelihood maximization is intractable, VAEs are trained by maximizing the Evidence Lower Bound:

$$\mathcal{L}(\theta, \phi) = \mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x|z)] - D_{KL}(q_\phi(z|x) \| p(z)).$$

Here, the first term ensures reconstruction fidelity, while the Kullback-Leibler divergence D_{KL} penalizes deviations of the approximate posterior from the prior $p(z)$. After training, new samples are generated by drawing a latent variable $z \sim p(z)$ from the prior distribution and passing it through the decoder network to produce $x \sim p_\theta(x|z)$. VAEs are computationally efficient and provide a well structured latent representation, but the probabilistic decoding process often leads to blurry samples compared to other generative models.

Generative Adversarial Networks (GANs). GANs take a completely different approach by framing generation as a two player game between a Generator G and a Discriminator D [30]. The generator attempts to produce realistic data $x = G(z)$ from noise $z \sim p_z$, while the discriminator tries to distinguish between real data from the dataset p_{data} and fake data produced by the generator. The optimization is a minimax objective:

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}}[\log D(x)] + \mathbb{E}_{z \sim p_z}[\log(1 - D(G(z)))]$$

Through this adversarial training process, the generator learns a mapping from the latent space to the data distribution. Once training is complete, generating new data is straightforward: a random latent vector z is sampled from the prior distribution and passed through the generator to produce a synthetic sample $x = G(z)$. GANs are renowned for producing exceptionally sharp and realistic images. However, they suffer from training instability and mode collapse, where the generator only learns a small subset of the data distribution, making them sometimes unreliable for inverse problems in critical applications.

Diffusion Models. Currently representing the state of the art in generative modeling, diffusion models operate by gradually corrupting data with Gaussian noise over a series of steps until the signal is completely destroyed, and then learning to reverse this process [31, 32]. In the continuous time formulation, the forward corruption process is described by a stochastic differential equation:

$$dx = f(x, t)dt + g(t)dw,$$

where w is the standard Wiener process, $f(x, t)$ is the drift coefficient, and $g(t)$ is the diffusion coefficient [26]. The remarkable property of this process is that its reversal is

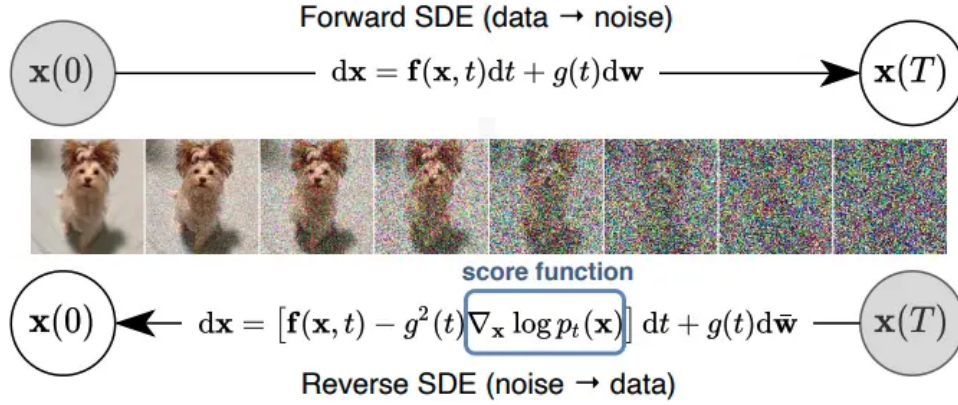


Figure 1.3: Diffusion model scheme.²

also a stochastic differential equation, given by:

$$dx = [f(x, t) - g(t)^2 \nabla_x \log p_t(x)]dt + g(t)d\bar{w}.$$

Here, $\bar{w}$ is a reverse time Wiener process, and $\nabla_x \log p_t(x)$ is the score function of the perturbed data distribution. A neural network is trained to estimate this score function via score matching. After training, new data can be generated by starting from pure Gaussian noise and numerically simulating the reverse time stochastic differential equation, progressively removing noise until a structured sample from the learned data distribution is obtained. Diffusion models offer stable training and unprecedented sample quality, though iterative sampling can be computationally expensive due to the large number of denoising steps required.

1.3.2 Solving Inverse Problems with Generative Models.

Once trained, a generative model implicitly or explicitly represents the prior distribution $p(x)$ of the signals of interest. This learned distribution can be incorporated into either the Bayesian or the variational framework for inverse problems, where the goal is to recover an unknown signal x from indirect and noisy measurements y . Recollecting the Bayesian formulation

$$p(x|y) \propto p(y|x)p(x)$$

where , the likelihood term $p(y|x)$ enforces consistency with the measurements, while the prior $p(x)$ captures the natural structure of the signals, generative models provide a powerful way to represent the prior distribution. Instead of designing handcrafted regularizers such as sparsity-promoting ones or total variation, the model learns complex structures directly from data, including textures, anatomical constraints, and long range correlations. Once the model is trained, the inverse problem becomes one of sampling

²Score based Generative Modeling through Stochastic Differential Equations [26]

from or optimizing over the posterior distribution $p(x|y)$.

Different classes of generative models lead to different reconstruction strategies. For example, when using generative adversarial networks or variational autoencoders, the signal is typically parameterized through a latent variable z , such that

$$x = G(z).$$

The inverse problem, in the variational formulation, can then be solved by searching for a latent code that best explains the measurements [33]:

$$z^* = \arg \min_z \|AG(z) - y\|^2.$$

The reconstruction is subsequently obtained as $x^* = G(z^*)$. This formulation restricts the solution to lie on the range of the generator, which can significantly reduce the dimensionality of the problem and enforce strong structural priors. Such approaches have been successfully applied to compressed sensing, image super resolution, and inpainting, where the generator fills in missing or corrupted information in a realistic manner.

Normalizing flows offer another attractive framework because they provide an explicit and tractable likelihood [34]. Since the mapping between latent variables and data is invertible, it is possible to directly evaluate the density $p(x)$ and compute gradients of the log likelihood. This enables optimization based methods that explicitly combine the likelihood and prior terms:

$$x^* = \arg \max_x \log p(y|x) + \log p(x).$$

The exact likelihood available in flow based models makes them particularly appealing for Bayesian inference and uncertainty quantification in inverse problems.

Diffusion models provide yet another paradigm, which has recently emerged as one of the most powerful approaches for solving inverse problems [35]. In this framework, the generative model learns the score function $\nabla_x \log p(x)$ of the data distribution. This quantity can be directly incorporated into posterior sampling procedures. By applying Bayes theorem at the level of the score, we obtain

$$\nabla_x \log p(x|y) = \nabla_x \log p(x) + \nabla_x \log p(y|x).$$

The first term in the right hand side corresponds to the unconditional score learned from data, while the second term enforces measurement consistency with the forward model.

This decomposition enables a wide range of reconstruction algorithms based on stochastic differential equations or Langevin dynamics. Starting from pure noise, the algorithm iteratively refines the sample by following the posterior score, gradually producing solutions that both resemble realistic data and satisfy the measurement constraints. One of the key advantages of this approach is its flexibility: the same pretrained generative model can be used to solve many different inverse problems simply by modifying the likelihood term [36].

This framework has been successfully applied to a wide variety of imaging tasks [37]. In magnetic resonance imaging (MRI), diffusion-based priors can reconstruct high quality images from highly undersampled Fourier measurements, significantly accelerating acquisition times [38]. In computed tomography (CT), generative priors can mitigate artifacts arising from sparse view or low dose acquisitions [39]. In image inpainting, the model can plausibly reconstruct missing regions by conditioning on the known pixels [40]. Similarly, in super resolution problems the model generates high resolution details that are consistent with the observed low resolution measurements [41].

Beyond imaging, generative priors have also been applied to problems such as deblurring, phase retrieval, and seismic imaging [37, 42]. In each of these cases, the learned data distribution provides a powerful constraint that helps resolve ambiguities inherent in the forward model. By combining statistical learning with physical measurement models, generative approaches offer a principled and flexible framework for tackling high dimensional inverse problems that would otherwise be extremely challenging to solve.

1.4 Overview and Contributions

This thesis is organized into two main parts, each focusing on different aspects of the proposed framework for solving inverse problems using generative models.

The first part, based on the two papers [43, 44], focuses on practical applications, with particular emphasis on medical imaging. In Chapter 2, we investigate the use of diffusion models for motion artifact correction in brain MRI, one of the most common sources of image degradation, typically caused by involuntary patient movement during acquisition. Such artifacts can significantly impair image quality and, consequently, the reliability of downstream clinical analysis. Within this context, we show how generative models, specifically diffusion models, can be effectively employed to perform image restoration in a data-driven yet principled manner. We obtain reconstruction methods that leverage learned data distributions while maintaining a clear connection to the underlying physical acquisition process. This combination allows for improved reconstruction quality compared to traditional approaches, particularly in challenging scenarios. At the same time, we critically analyze the limitations of these methods, emphasizing that their application may raise significant trustworthiness and reliability concerns if not properly handled. In particular, the use of generative priors can lead to the introduction of hallucinated features or the suppression of clinically relevant details, which is especially problematic in sensitive domains such as medical imaging. The proposed approaches are validated on realistic scenarios using real-world data, demonstrating both their robustness and their effectiveness in handling complex, real world degradations. These experiments highlight not only the potential of diffusion-based methods for motion correction, but also the importance of careful methodological design and evaluation when deploying generative models in practice.

The second part, in Chapter 3 which is based on the preprint [45], we introduce an infinite dimensional formulation of generative sensing, a methodology inspired by classical compressed sensing. In contrast to standard sparsity-based assumptions, we

model the unknown signal as belonging to the range of a generative model, thereby replacing handcrafted priors with learned ones. This perspective enables a principled extension of generative approaches to infinite-dimensional (functional) settings. We provide a rigorous mathematical framework for this formulation, establishing conditions under which stable and accurate recovery is possible. This infinite-dimensional viewpoint allows us to generalize classical finite-dimensional results and sheds new light on the role of learned priors in inverse problems. Finally, we support the theoretical findings with numerical experiments, focusing on the reconstruction of solutions to Darcy flow problems.

Chapter 2

Diffusion models for motion correction in magnetic resonance imaging

2.1 Introduction

Magnetic Resonance Imaging (MRI) [46, 47] is a powerful medical imaging technique. However, the technique suffers from a critical limitation: patients must remain extraordinarily still during imaging. This immobility requirement creates significant challenges across different patient groups. Young children, claustrophobic individuals, and elderly patients often struggle to remain motionless for extended periods. Additionally, some movements are simply unavoidable, such as those caused by breathing and heart function. The motion-induced artifacts in MRI are uniquely complex, stemming directly from the image formation process. Unlike traditional photography, where movement creates localized blurring, MRI’s image construction involves sampling points in Fourier space (known as k-space). In this domain, even minor local movements can dramatically distort the entire image [48]. This occurs because MRI images are created by sampling points from the Fourier transform of the object and then computing an inverse Fourier transform with incomplete data. As an inverse problem, these motion-induced errors can be substantially amplified. To mitigate these challenges, researchers have developed two primary intervention strategies: prospective and retrospective techniques. Prospective methods are implemented during the scanning process, such as collecting redundant data points at different times to compensate for patient movement. However, these approaches are often machine-specific, technique-dependent, and can increase overall scan duration. Retrospective techniques, in contrast, focus on correcting artifacts after the image has been captured. Our proposed method belongs to this second category.

2.1.1 Motion artifacts

While MRI offers superior soft tissue contrast and non-invasive imaging capabilities without ionizing radiation, its relatively long acquisition times make it susceptible

to patient movement, which can result in various image artifacts including ghosting, blurring, geometric distortion, and reduced peak signal-to-noise ratio (PSNR). These artifacts can compromise diagnostic accuracy and automated analysis, potentially leading to false positives or negatives, and may necessitate scan repetition, which increases costs and delays treatment. Motion correction approaches can be classified by several factors: motion type (rigid translation/rotation vs. deformable), timing (inter-image, inter-scan, or intra-scan), pattern (periodic, quasi-periodic, continuous, or sporadic), and in 2D imaging, whether the motion occurs in-plane or through-plane. While faster imaging techniques have reduced acquisition times from dozens of minutes to seconds, these improvements often involve trade-offs in resolution, contrast, or SNR. The ongoing development of motion correction techniques reflects the critical importance of addressing this fundamental challenge in MRI technology, as improving motion correction capabilities directly impacts clinical outcomes and research validity [49].

2.1.2 Classical techniques

Classical methods to mitigate motion artifacts in MRI primarily focus on reducing patient movement and correcting the acquired data. Physical restraints and breath-holding techniques are employed to minimize motion during image acquisition. Prospective motion correction [50] actively adjusts scan parameters in real-time to compensate for detected movement, while retrospective methods apply post-processing algorithms to realign and correct motion-affected image segments. Additionally, navigator echoes, temporal filtering, and multi-shot acquisition strategies with motion-robust reconstruction algorithms [48] can help reduce the impact of motion on image quality. These techniques, while effective, often require careful optimization and may not fully address complex motion patterns, particularly in challenging clinical scenarios.

2.1.3 Deep learning

Recent advances in deep learning have enabled significant progress in retrospective motion artifact correction for MRI [46, 47]. These methods rely on training machine learning models, predominantly deep neural networks, through supervised approaches where data quality is paramount. Two primary strategies exist for obtaining suitable training data. The first involves acquiring paired scans of the same subject—one motion-free and one with motion artifacts. However, this approach presents substantial limitations: such paired acquisitions rarely occur in clinical practice, requiring expensive dedicated experiments that may not accurately represent involuntary patient movements encountered in real-world scenarios. The second approach involves simulating motion artifacts through k-space manipulation, which while eliminating the need for additional scans, requires detailed knowledge of the MRI acquisition parameters and may produce artifacts that inadequately represent clinical reality. Our contribution was divided in two different efforts. In [43] we showed that directly using generative models, in particular diffusion models, can produce high-quality reconstructions but they are also susceptible to hallucination. This poses risks in diagnostic applications. Nonetheless,

we critically evaluated their use for correcting motion artifacts in 2D brain MRI scans. In [44] we proposed a method to address the limitation of requiring paired data for training by utilizing unpaired motion-affected and motion-free images from routine clinical examinations. This eliminates the requirement for k-space modification or sequence-specific technical information, thereby enabling broader application across various scanning protocols and hardware configurations.

2.1.4 Literature review

Prospective methods either use sampling patterns that sample the same points in the k-space at different times to estimate and correct the patient motion [51, 52] or use external tracking devices such as cameras to achieve the same objective [53, 54]. Among the retrospective methods we cite the ones using compressed sensing techniques [55, 56] that leverage a priori knowledge on MR images to perform the motion correction task. Numerous deep learning approaches have been proposed for MRI motion artifact correction. In [57], CNNs were used for motion correction in 3D MRI data, while [58] employed a conditional generative adversarial network (cGAN) for artifact reduction. [59] introduced a deep residual network with densely connected multi-resolution blocks, and [60] developed stacked U-Nets with self-assisted priors, such as contiguous slices, for improved corrections. In [61], a neural network-based regularization term was integrated with Autofocusing, a classic optimization-based method, and [62] employed a diffusion model for motion correction. Recently, [63] proposed an annealed score-based diffusion model, which consists in iteratively applying forward and reverse diffusion processes. All of these methods rely on supervised learning and require paired motion-corrupted and motion-free images, which are often unavailable and necessitate the simulation of motion artifacts.

Different approaches [64, 65, 66, 67, 68] faced the data availability problem by using an image-to-image translation framework, exploiting general techniques such as [69], and with a framework inspired by [70]. Exploiting synthetic data, these methods enabled artifact removal without paired training data through innovative approaches like adversarial autoencoders, k-space subsampling, and novel reconstruction methods, thereby enhancing image quality across respiratory, brain, and dynamic imaging scenarios. Specifically, in [64], the authors employ an adversarial network to learn the translation mappings between the motion-affected and motion-free images; in [65] instead they trained an adversarial autoencoder network in a way that forces the encoder to remove motion artifacts; in [66] they developed an unsupervised deep learning scheme through outlier rejecting bootstrap sub-sampling and aggregation, using data in the k-space; [67] introduced an adversarial framework with a new generator architecture and loss function to perform unsupervised correction; finally in [68] they improved the vanilla CycleGAN by adding loss regularization with a module that estimates the motion severity of the input image.

All these works were originally evaluated on synthetic or proprietary datasets. However, a subsequent study [71] provided a comparative evaluation of the aforementioned techniques using the MR-ART dataset, which we also use in this work. The work [71] also

introduced the first application of diffusion models for motion artifact correction, where motion-free images are generated directly from their motion-affected counterparts. Their diffusion-based method significantly outperformed state-of-the-art approaches based on image-to-image translation.

Other works have used diffusion models in relevant tasks. For instance [72] proposed a method that performs MRI reconstruction jointly with motion-free estimation to improve image quality. In [73] they developed conditional diffusion models for various tasks, applying them to medical imaging problems. Finally [74] also addressed the problem of motion correction using a multi-phase approach that includes a diffusion model. Similarly to previous works, we exploited diffusion models for motion artifact correction like [71].

2.1.5 Denoising diffusion probabilistic model

Denoising Diffusion Probabilistic Models (DDPMs), commonly referred to as Diffusion Models for brevity, are generative models that create data by reversing a diffusion process. In the forward diffusion process, Gaussian noise is progressively added to an initial data point $\mathbf{x}_0$ (e.g., an MRI image) over T timesteps, resulting in a sequence of latent variables $\mathbf{x}_1, \dots, \mathbf{x}_T$. Each transition is governed by the following rule:

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{\alpha_t}\mathbf{x}_{t-1}, (1 - \alpha_t)\mathbf{I}),$$

where α_t are hyperparameters of the variance schedule parameters. The reverse process, that is learned with a neural network, learns to denoise each $\mathbf{x}_t$ to recover $\mathbf{x}_{t-1}$:

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \mu_\theta(\mathbf{x}_t, t), (1 - \alpha_t)\mathbf{I}).$$

This architecture enables DDPMs to generate samples by progressive denoising starting from pure Gaussian noise.

The values that regulate the amount of noise added, and consequently the amount that needs to be removed, are the coefficients α_t , which are fixed a priori. From these, we can define

$$\bar{\alpha}_t = \prod_{s=1}^t \alpha_s$$

which represents the amount of the original signal preserved after t noising steps, and

$$\sigma_t^2 = 1 - \alpha_t$$

that denotes the amount of noise added at each step. For a more in depth theoretical discussion, we refer the reader to [75]. In our two works we leveraged the DDPM to either introduce or remove motion artifacts. A DDPM typically generates images by starting from random noise and performing T denoising steps. In our approach, described in Algorithm 1, we adapt the DDPM to begin with an initial image, add Gaussian noise progressively up to a certain step n (where $n < T$), and then allow the model to denoise for the remaining n steps. This method provides bidirectional capabilities depending on

the training domain. If the model is trained on motion-free images and is fed motion-affected ones during inference, it effectively removes motion artifacts. Alternatively, it can introduce realistic motion artifacts if trained on motion-affected images, by starting with a motion-free image. In either scenarios, simulating the forward diffusion process up to step n followed by partial denoising retains the original structural characteristics while applying the desired transformation. The choice of n is critical to achieving a balance; if n is too large, the resulting images may appear exaggerated or unrealistic, diverging from the characteristics of real world data. Conversely, if n is too low, the modifications may not be prominent enough, making them ineffective for generating meaningful paired training data or properly restoring the image. By tuning n , we ensure that the generated images are both realistic and diverse, enabling the creation of high quality paired datasets for supervised training.

Algorithm 1 Diffusion-based motion simulation. The algorithm is used to add simulated motion artifacts to motion-free images; this is done after training the network ϵ_θ on the motion-affected images

```

1: input  $\mathbf{Y}, n, \{\alpha_t, \bar{\alpha}_t, \sigma_t\}_{t=1}^n, \epsilon_\theta$ 
2:  $\mathbf{z} \sim \mathcal{N}(0, \mathbf{I})$ 
3:  $\mathbf{x}_n = \sqrt{\bar{\alpha}_n} \mathbf{Y} + (1 - \bar{\alpha}_n) \mathbf{z}$ 
4: for  $t = n, \dots, 1$  do
5:   if  $t > 1$   $\mathbf{z} \sim \mathcal{N}(0, \mathbf{I})$ , else  $\mathbf{z} = 0$ 
6:    $\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left( \mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\alpha_t}} \epsilon_\theta(\mathbf{x}_t, t) \right) + \sigma_t \mathbf{z}$ 
7: end for
8: return  $\mathbf{x}_0$ 
```

2.1.6 U-net

In [44] following the generation of simulated data, we employ a U-Net architecture [76] to learn a mapping between motion-corrupted images and their artifact-free counterparts. The U-Net’s design is particularly well-suited for biomedical image analysis due to its encoder-decoder structure with skip connections, which preserves fine spatial details while capturing hierarchical feature representations. To train the network, we utilize paired datasets where motion-affected images (simulated via our framework) serve as inputs, and their corresponding ground-truth, artifact-free images act as targets. During training, we optimize using the SSIM loss function. U-Nets also appear in [43] but only as a baseline, supervised method. Once trained, the U-Net operates as a standalone artifact correction model, processing unseen motion-degraded images in inference. To validate performance, we quantify restoration accuracy using as metrics structural similarity index (SSIM), mean squared error (MSE) and peak signal-to-noise ratio (PSNR).

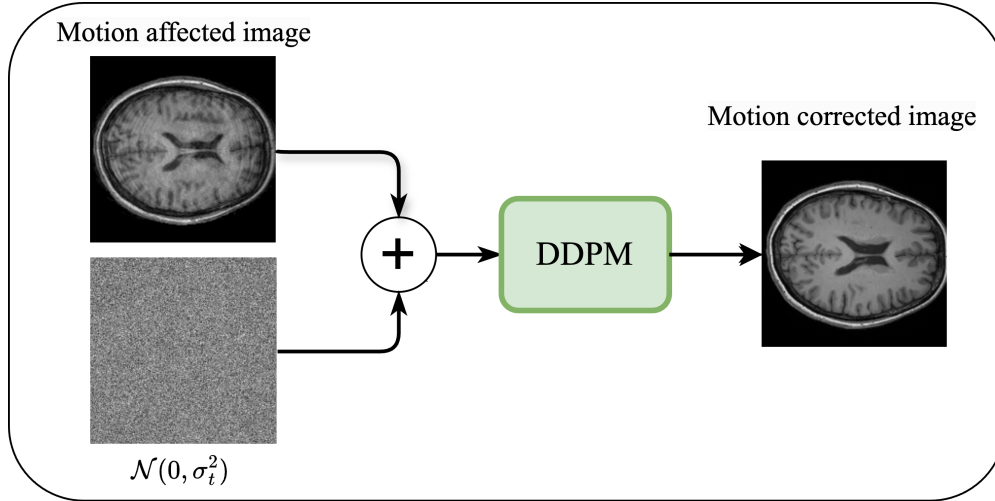


Figure 2.1: Our implementation follows a two-phase approach. In the first phase, a diffusion model (DDPM) is trained on a dataset of motion artifact-affected images. In the second phase, the trained model is used to introduce motion artifacts into clean images, generating paired datasets. These pairs enable supervised training in the subsequent step. See Algorithm 1 for more details.

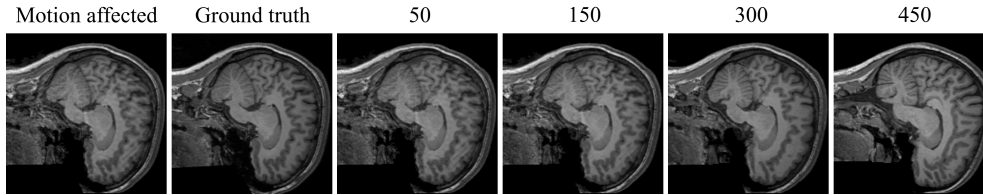


Figure 2.2: DDPM motion artifact correction with different time steps n .

2.2 Assessing the use of diffusion models for direct motion artifact correction in brain MRI

In this section, we focused on comparing two different methods for motion correction, both of which operate under the assumption that paired data of motion-affected and motion-free images are unavailable. This assumption holds because such paired scans can only be obtained through dedicated experiments, which are not feasible in standard clinical practice.

2.2.1 Methods

The first of the two methods compared involves a U-Net trained in a supervised manner using synthetically generated motion-affected images. The second method employs an unconditional diffusion model, specifically a Denoising Diffusion Probabilistic Model (DDPM) [75], trained only on motion-free images to accomplish the same task (see Fig. 2.1 for a schematic overview).

The key challenge lies in adapting the DDPM for motion correction, as it typically generates images by starting from random noise and performing T denoising steps. In

our approach, described in Algorithm 1, we instead begin with the motion-affected image, add Gaussian noise up to a certain step n (where $n < T$), and then allow the model to denoise for the remaining n steps. The choice of n is crucial: if too high, the model may hallucinate details; if too low, the correction will be insufficient (see Fig. 2.2 for a qualitative example).

To evaluate the performance of the various approaches, we employ standard metrics for motion artifact correction assessment: the Structural Similarity Index Measure (SSIM), Normalized Mean Squared Error (NMSE), and Peak Signal-to-Noise Ratio (PSNR).

2.2.2 Experiments

2.2.2.1 Dataset

All experiments are conducted using the MR-ART dataset [77], which contains brain MRI scans of 144 patients, captured both while they remained still and while moving their heads to introduce motion artifacts. As mentioned earlier, having paired data like this is uncommon. Therefore, we employ data from Olsson et al. [78], which includes simulated motion artifacts. The dataset consists of 3D volumes, from which sagittal, coronal, and transverse views were extracted for use in the experiments. Our preprocessing step includes normalization and registration between the motion-free and motion-affected images from MR-ART, as the volumes are often misaligned.

2.2.2.2 Experiment Details

All experiments are conducted on an NVIDIA A30 GPU using the PyTorch framework. The architecture of the U-Net follows the original of [76], while the diffusion model is based on the architecture of [79]. Specifically, 500 timesteps were used. For DDPM training, the learning rate was set to 5×10^{-5} , and for the U-Net at 1×10^{-3} , in both cases the batch size was 6. Early stopping was applied to all models to prevent overfitting on the training set. A subset of 30 patients was used for training, 10 for validation, and 25 for testing. This approach was adopted to ensure efficient use of computational resources while maintaining robust performance.

2.2.3 Results

We first describe our parameter-tuning procedure, and then report the obtained results on the three available anatomical planes. Our analysis compares three approaches: (1) our diffusion model approach, (2) a U-Net trained on synthetic images from [78], and (3) a benchmark U-Net trained directly on paired motion-affected and motion-free images from the ground truth dataset. This last approach serves as an upper bound for performance, as it uses real paired data that would typically be unavailable in clinical settings.

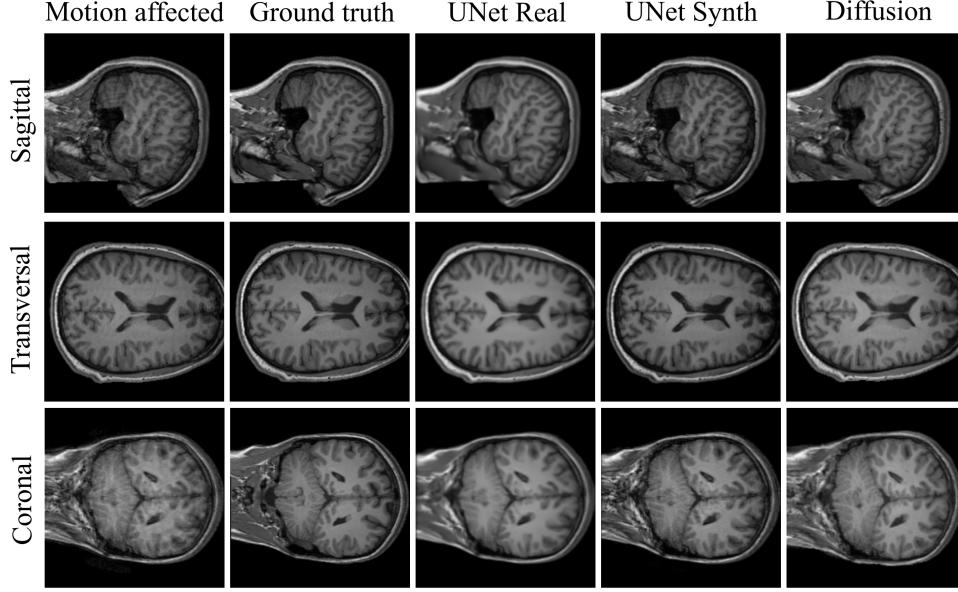


Figure 2.3: Comparison of the different approaches on all three views.

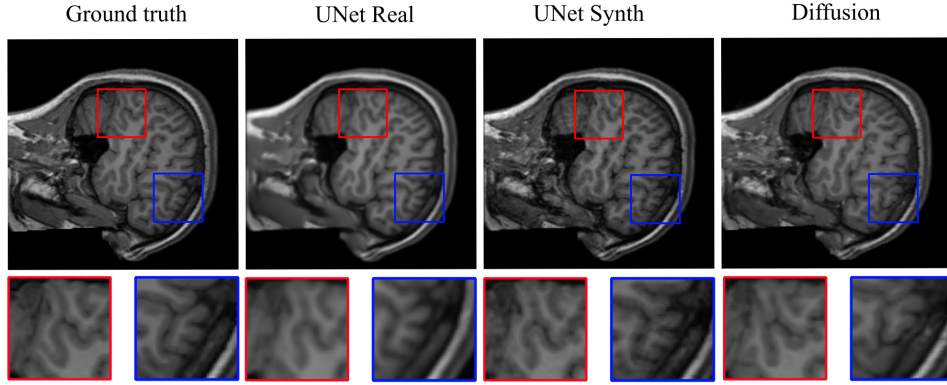


Figure 2.4: Example of hallucination of the DDPM.

2.2.3.1 Parameter Tuning

In the original MR-ART dataset, each clean image is paired with two corresponding images featuring induced motion. Therefore, for each U-Net training session, we utilize two synthetic motion-affected images for every motion-free image. In [78], four sets of synthetic images were produced for each reference image. For a fair comparison, we select the best two synthetic images by training a U-Net on each possible combination (referred to as A, B, C, and D) and selecting the one yielding the best results on the validation set. All sets produce very similar outcomes, with slightly better performance for the BC set. Thus, we select this set to compare with our diffusion model-based approach. An important parameter is the timestep at which to start the denoising process. Here, there is a trade-off between artifact removal and hallucination by the model, as shown in Fig. 2.2. We selected 150 timesteps as it provided an optimal balance between reconstruction quality and the extent of hallucinations.

Label	SSIM	NMSE	PSNR
UNet Real	0.858 $\pm$ 0.079	0.059 $\pm$ 0.054	22.9 $\pm$ 2.7
UNet Synth	0.844 $\pm$ 0.088	0.055 $\pm$ 0.062	24.5 $\pm$ 2.8
Diffusion	0.839 $\pm$ 0.073	0.047 $\pm$ 0.053	24.9 $\pm$ 2.4

Table 2.1: Transverse view metrics.

Label	SSIM	NMSE	PSNR
UNet Real	0.782 $\pm$ 0.055	0.084 $\pm$ 0.119	21.9 $\pm$ 2.1
UNet Synth	0.741 $\pm$ 0.060	0.098 $\pm$ 0.124	22.2 $\pm$ 1.9
Diffusion	0.666 $\pm$ 0.180	0.132 $\pm$ 0.182	22.0 $\pm$ 3.4

Table 2.2: Sagittal view metrics.

2.2.3.2 Transverse view

In the transverse view, the Diffusion model performs best for NMSE and PSNR, while the UNet achieves the highest SSIM. The UNet Synth model performs decently but slightly trails behind the Diffusion model in NMSE and PSNR, despite being close to the UNet in SSIM. See Fig. 2.3 (middle) for a qualitative comparison of the motion-free images predicted with the investigated approaches. In this view, DDPM is not prone to hallucination. To better frame the obtained results in the state-of-the-art, we report the performance on the transverse view available from [71], exploiting popular generative models including CycleGAN [80], Pix2Pix [69], and a conditional diffusion probabilistic model. Our investigated methods outperform the available state-of-the-art results for all the evaluation metrics. Notably, the transverse is the only view employed in the literature existing on motion artifact correction with diffusion models for the MR-ART dataset [71].

2.2.3.3 Sagittal view

For the sagittal view, the UNet performs the best in terms of SSIM and NMSE, though the UNet Synth slightly surpasses it in PSNR. The Diffusion model performs poorly in this view, with significantly higher NMSE and a much lower SSIM (See Fig. 2.3 (top) for a qualitative comparison) Furthermore, DDPM is prone to hallucination, affecting predicted motion-free images (see Fig. 2.4 for an example).

2.2.3.4 Coronal view

In the coronal view, the UNet Synth performs the best overall, with the highest SSIM and PSNR, though the UNet has a lower NMSE. The Diffusion model does well in NMSE

Label	SSIM	NMSE	PSNR
UNet Real	0.806 $\pm$ 0.103	0.064 $\pm$ 0.054	22.6 $\pm$ 2.9
UNet Synth	0.810 $\pm$ 0.121	0.074 $\pm$ 0.065	23.5 $\pm$ 2.8
Diffusion	0.765 $\pm$ 0.108	0.068 $\pm$ 0.051	23.5 $\pm$ 2.5

Table 2.3: Coronal view metrics.

and PSNR but falls behind in SSIM compared to the UNet models, with predicted motion-free images affected by hallucination. See Fig. 2.3 (bottom) for a qualitative comparison.

2.2.4 Summary and discussion

This study critically evaluates the use of diffusion models for correcting motion artifacts in 2D brain MRI scans. Our findings reveal that while diffusion models can achieve competitive performance with supervised methods in certain scenarios, their effectiveness varies significantly with data heterogeneity and anatomical planes. The primary challenge lies in balancing reconstruction quality against the risk of hallucinations, which could be particularly problematic in diagnostic settings. Our results suggest that diffusion models hold promise for MRI motion artifact correction, but their clinical application requires careful consideration of specific use cases and potential risks. Future work should focus on developing methods to detect and mitigate hallucinations, exploring hybrid approaches that combine the strengths of both diffusion and supervised methods, and conducting clinical validation studies to assess the impact on diagnostic accuracy.

2.3 Using diffusion models for data pair generation in MRI motion correction

In this section, we explore the use of diffusion models to generate motion artifact-corrupted MRI images from clean scans. These synthetic data enable supervised learning by providing paired clean and corrupted images for training a network to remove motion artifacts.

2.3.1 Dataset

Our experiments were conducted using two datasets. The first dataset is the MR-ART dataset [77], which includes brain MRI scans from 144 patients, where we extracted around 100 images from each scan for each possible view (sagittal, transversal and coronal). For each patient, there are three scans: one motion-free scan acquired while the patient remained still and two motion-affected scans acquired while the patient purposefully moved their head to induce artifacts. Additionally, we utilized simulated motion artifact data provided in [78], which generated four motion-affected simulations for each motion-free scan in the MR-ART dataset. This is done through k-space alteration techniques that mathematically model the complex physical processes underlying image degradation during magnetic resonance scanning, we point out that generating this simulations requires knowing the details of the particular MRI sequences used to acquire the scans. The second dataset [81], which we refer to as IN-SCANNER, consists of MRI scans from 25 patients, each with one motion-free and one motion-affected scan. It is important to note that paired motion-free and motion-affected MRI scans are exceptionally rare, particularly for non-brain MRI. This scarcity arises because accidental patient movement during scans is challenging to capture in a controlled manner. The

motion artifacts in the MR-ART dataset were generated through purposeful movement, which differs from the unintentional and often irregular motion seen in real-world clinical settings. Both datasets comprise 3D volumes, from which we extracted slices from all possible acquisition planes (sagittal, transversal and coronal) for our experiments. As part of the preprocessing pipeline, we applied normalization and performed registration to align the motion-free and motion-affected images in the MR-ART dataset, as the volumes were often misaligned.

2.3.2 Implementation details

2.3.2.1 Diffusion model

We trained a Denoising Diffusion Probabilistic Model (DDPM) based on the implementation from [79]. The model consists of a U-net architecture that accepts both an image and an integer representing the denoising time-step as inputs. We trained this DDPM on motion-affected images from 30 patients in the dataset (for reference, there are around 120 images for each patient), using an additional 15 for validation. 500 time-steps were used and we implemented early stopping to prevent overfitting.

We use a linear noise scheduler for σ_t^2 , ranging from 0.02 to 0.03. The U-Net used as the denoiser features an encoder that progressively reduces the spatial resolution through four stages with channel sizes of $[1, 2, 4, 8]$. Each stage consists of four residual blocks. Self-attention is applied at the 32×32 resolution using multi-head attention with four heads. The decoder mirrors the encoder, incorporating upsampling operations and skip connections to fuse encoder features at corresponding resolutions. These patients' data were excluded from all subsequent experiments to maintain data independence. To generate synthetic motion-affected images starting from the available motion-free ones, and employing the trained diffusion model, we select an integer n smaller than the total number of time-steps, add noise to each image slice according to the DDPM scheduler's specifications, and feed the noisy image to be denoised at the n -th timestep (see Algorithm 1). This process can be repeated multiple times to improve the result by simply taking the image generated and feeding it back to the DDPM, repeating the process. This approach is used because running multiple iterations allows for the introduction of more artifacts while minimizing hallucinations, compared to a single, deeper denoising process with more steps. There is a crucial balance to strike when choosing the timestep value. A timestep that's too large results in heavily hallucinated images, while one that's too small produces images too similar to the originals. Either extreme yields pairs that are less useful for training. See Figure 2.6 for a qualitative comparison of our diffusion-based motion artifact simulation with real motion artifacts from the MR-ART dataset and synthetic motion artifact from [78].

2.3.2.2 U-net

For rigorous comparative analysis, we simultaneously trained identical U-net neural network architectures using three alternative image sets. The first alternative training set is comprised of ground truth images sourced directly from the MR-ART dataset, providing

a baseline of what can be obtained with paired data. The second alternative training set consisted of images with simulated motion artifacts generated by [78]. Finally we have our own proposed method that uses synthetic images created by the diffusion model trained on real motion-affected images. In all cases we use the original motion-free images paired with the motion-affected real/simulated ones for training. By employing diverse image sets to train these networks, we developed a solid experimental framework that allows us to thoroughly assess the effectiveness and adaptability of our generation method. This methodical comparison allows us to assess not only the efficacy of our proposed method, but also its potential advantages over traditional motion artifact correction techniques. Our experimental design intentionally maintained consistent network architectures across all training scenarios, ensuring that any observed performance differences could be attributed directly to variations in the training data rather than network structure. This approach provides a motion-free, controlled methodology for understanding the nuanced challenges of motion artifact correction in medical imaging.

2.3.2.3 Data and parameter selection

All U-net networks were trained with a learning rate of 0.001, a batch size of 6, and the Structural Similarity Index Measure (SSIM) [82] as the loss function. We implemented early stopping to prevent overfitting. The test set consisted of ground truth pairs from the MR-ART dataset, and performance was evaluated using four metrics: Structural Similarity Index Measure (SSIM), Normalized Mean Squared Error (NMSE), Peak Signal-to-Noise Ratio (PSNR) and Learned Perceptual Image Patch Similarity (LPIPS)[83]. The MR-ART dataset pairs each motion-free reference image with two motion-affected variants. To maintain consistency in our U-Net training, we adopted the same approach by using two simulated motion-affected images per motion-free image in both simulated methods. Since the dataset provided by [78] contains four distinct sets of motion-affected images (A, B, C, and D) for each reference image, we conducted separate U-Net training experiments for every possible combination of these sets (results in Table 2.4). This systematic approach enabled us to identify the optimal image set for training. In addition to image sets, we investigated the number of patients required to achieve optimal performance, as motion simulations are computationally expensive. We trained all U-Net configurations using 10, 30, and 50 patients while maintaining a fixed validation set of 15 patients. These tests were performed exclusively on sagittal slices, with the selected parameters subsequently applied to the other two image views. During training, validation was performed using the same input images as those in the training set. However, to more accurately represent real-world performance, the data selection results (SSIM metrics shown in Figure 2.7) were calculated using original paired images from the MR-ART dataset. The results show that 30 patients represents a sufficient sample size, showing significant improvement over 10 patients while performing comparably to 50 patients. Detailed metrics for U-Nets trained with 30 patients across different training sets are presented in Table 2.4.

Notably, all combinations yielded highly similar results. We ultimately selected

Sets	SSIM	NMSE	PSNR	LPIPS
AB	0.704 ± 0.067	0.117 ± 0.06	21.3 ± 2.0	0.126 ± 0.044
AC	0.705 ± 0.069	0.113 ± 0.071	21.3 ± 2.2	0.126 ± 0.045
AD	0.703 ± 0.070	0.114 ± 0.064	21.3 ± 2.2	0.129 ± 0.046
BC	0.709 ± 0.068	0.112 ± 0.063	21.4 ± 2.1	0.125 ± 0.044
BD	0.704 ± 0.068	0.112 ± 0.059	21.3 ± 2.1	0.129 ± 0.045
CD	0.707 ± 0.069	0.117 ± 0.063	21.1 ± 2.1	0.126 ± 0.044

Table 2.4: Performance metrics for different dataset combinations of the data made available by [78].

Set	Denoising steps	Iterations
T	120	4
Z	170	5
H	280	2
J	330	1

Table 2.5: Parameters of motion simulation for the diffusion model.

combination "BC" due to its marginally higher SSIM and lower NMSE values, though we emphasize that no significant differences exist between combinations. For consistency, we also evaluated four potential data simulation parameters (H, J, T, and Z) against the control set and those from [78]. The parameters for each set, shown in Table 2.5, primarily involved the number of denoising steps, which determines how far deep in the DDPM architecture the process starts, and iterations, meaning how many times Algorithm 1 is repeated. These values were heuristically selected to balance motion artifacts and hallucinations. Following the same protocol, we tested all combinations with 10, 30, and 50 patients (results in Figure 2.8), again identifying 30 patients as sufficient. Detailed metrics for these U-Nets appear in Table 2.6. While results showed smaller variations between parameter choices, we selected combination TZ due to its slight performance advantage. We emphasize that the patients used for parameter selection were distinct from those in the final test set.

Label	SSIM	NMSE	PSNR	LPIPS
HJ	0.686 ± 0.068	0.122 ± 0.061	19.5 ± 2.0	0.208 ± 0.044
HT	0.709 ± 0.077	0.106 ± 0.058	20.8 ± 2.3	0.162 ± 0.041
HZ	0.706 ± 0.074	0.11 ± 0.058	20.1 ± 2.1	0.155 ± 0.045
JT	0.711 ± 0.071	0.109 ± 0.059	20.3 ± 2.1	0.159 ± 0.048
JZ	0.714 ± 0.068	0.111 ± 0.057	20.0 ± 2.0	0.163 ± 0.045
TZ	0.713 ± 0.068	0.106 ± 0.057	20.5 ± 2.1	0.128 ± 0.047

Table 2.6: Performance metrics for different dataset combinations of our simulated images.

Label	SSIM	NMSE	PSNR	LPIPS
SIMUNET [78]	0.786 ± 0.035	0.064 ± 0.016	23.6 ± 1.0	0.125 ± 0.044
DIMA	0.788 ± 0.036	0.058 ± 0.017	23.0 ± 1.2	0.128 ± 0.047
SUPUNET	0.820 ± 0.030	0.052 ± 0.015	23.5 ± 1.3	0.147 ± 0.039
CycleGAN [71]	0.670	0.225	21.7	
Pix2Pix [71]	0.640	0.242	21.4	
Unet [71]	0.730	0.195	22.5	
Diff [71]	0.680	0.191	21.9	

Table 2.7: Experiments results for the sagittal acquisition plane. The difference of PSNR between our 3 implemented methods is not statistically significant. All the other pairs of results are.

2.3.2.4 Computational requirements

All experiments presented in this work were performed on an NVIDIA RTX A5000 GPU with 24 GB of VRAM. On this hardware, training the diffusion model with a dataset of 6,500 images takes approximately 25 seconds per epoch. During inference, generating a new image—starting from Gaussian noise and performing all denoising steps—requires around 30 seconds. Training the U-Nets is considerably faster, taking only a few seconds per epoch. The inference time of the U-Nets, which are the only components intended for deployment in clinical settings, is on the order of milliseconds per image and requires at most 3 GB of VRAM.

2.3.3 Results

We compare the performance of different U-Net models trained under varying conditions. To facilitate this analysis, we refer to the following methods:

1. **SIMUNET:** A U-Net trained with traditional motion simulation techniques, utilizing data from [78].
2. **DIMA:** A U-Net employing our proposed diffusion-based motion simulation technique.
3. **SUPUNET:** A U-Net trained in a supervised manner using pairs of real motion-affected and motion-free images, serving as an upper-bound benchmark for the performance of the two preceding methods.

A paired t-test was performed for each experiment. Results are statistically significant ($p < 0.01$) unless otherwise indicated in the caption.

The comparative performance across different anatomical views and datasets reveals several key insights. See Figure 2.10 for a qualitative comparison. For sagittal views (Table 2.7), our approach demonstrates comparable performance with the ground truth benchmark (SSIM: 0.788 ± 0.036 vs 0.820 ± 0.030), while substantially outperforming traditional architectures like CycleGAN and Pix2Pix. The view-specific analysis reveals interesting variations in performance. While our method shows results in line to the simulation-based approach in all the views (SSIM: 0.852 ± 0.037 vs 0.858 ± 0.034 , Table 2.8,

Label	SSIM	NMSE	PSNR	LPIPS
SIMUNET [78]	0.858 $\pm$ 0.034	0.046 $\pm$ 0.019	23.7 $\pm$ 1.5	0.085 $\pm$ 0.028
DIMA	0.852 $\pm$ 0.037	0.047 $\pm$ 0.019	23.1 $\pm$ 1.5	0.091 $\pm$ 0.030
SUPUNET	0.863 $\pm$ 0.029	0.054 $\pm$ 0.015	22.0 $\pm$ 1.2	0.120 $\pm$ 0.027

Table 2.8: Experiments results for the transversal acquisition plane. The difference of NMSE between SIMUNET and DIMA is not statistically significant. All the other pairs of results are.

Label	SSIM	NMSE	PSNR	LPIPS
SIMUNET [78]	0.806 $\pm$ 0.040	0.059 $\pm$ 0.020	23.6 $\pm$ 1.3	0.085 $\pm$ 0.028
DIMA	0.807 $\pm$ 0.057	0.053 $\pm$ 0.020	23.4 $\pm$ 1.4	0.096 $\pm$ 0.026
SUPUNET	0.832 $\pm$ 0.034	0.055 $\pm$ 0.017	22.8 $\pm$ 1.3	0.125 $\pm$ 0.028

Table 2.9: Experiments results for the coronal acquisition plane. The difference of SSIM between SIMUNET and DIMA is not statistically significant. All the other pairs of results are.

and SSIM: 0.807 ± 0.057 vs 0.806 ± 0.040 , Table 2.9), it performs better compared to ground truth benchmark in the transversal view. This anatomical dependency likely reflects differences in motion artifact complexity across imaging planes, with transversal views potentially containing more challenging rotational artifacts that our current simulation parameters capture less effectively. Crucially, when tested on the completely independent IN-SCANNER dataset (Table 2.10), for which we used 11 patients for the training of the diffusion model, 11 for the training of the U-nets and 4 in test, our approach demonstrates to generalize well, with SSIM (0.743 ± 0.076) and PSNR (23.2 ± 1.7 dB) metrics closely matching performance obtained when training with real pairs (SSIM: 0.761 ± 0.065 , PSNR: 22.8 ± 1.5 dB). This is particularly notable given that our method requires no k-space manipulation or sequence-specific knowledge - a significant practical advantage over traditional acquisition-dependent approaches.

2.3.3.1 Challenging settings

In cases where motion artifacts are particularly severe, our method may reconstruct structures inaccurately. This occurs when the original data are so strongly degraded that recovering the underlying image information becomes nearly impossible. In practical applications, it is therefore important to assess the level of degradation in the motion-affected images before applying motion correction, as excessively corrupted inputs may lead to unreliable reconstructions. We show an example of such a case in figure 2.9.

Label	SSIM	NMSE	PSNR	LPIPS
DIMA	0.743 $\pm$ 0.076	0.155 $\pm$ 0.054	23.2 $\pm$ 1.7	0.145 $\pm$ 0.009
SUPUNET	0.761 $\pm$ 0.065	0.156 $\pm$ 0.046	22.8 $\pm$ 1.5	0.125 $\pm$ 0.010

Table 2.10: Experiments results for the alternative IN-SCANNER dataset, for the sagittal acquisition plane.

2.3.4 Summary and discussion

In this work, we proposed a novel technique that leverages diffusion models to address the challenge of requiring paired data for machine learning training in MRI motion artifact correction. By utilizing diffusion models, our approach generates high-quality synthetic paired data, effectively circumventing the need for real paired datasets, which are often difficult to obtain in medical imaging. Three key findings emerge:

1. Our simulation framework achieves performance comparable to state-of-the-art motion artifact simulation while eliminating the need for k-space manipulation
2. View-specific performance variations suggest opportunities for plane-adaptive parameter tuning
3. The strong cross-dataset generalization indicates clinical applicability beyond controlled simulation environments

The marginal performance differences between methods (typically less than 3% SSIM variation) suggest that the choice between approaches may ultimately depend on clinical constraints rather than pure metric superiority. Our method provides a viable alternative for scenarios requiring model-agnostic acquisition compatibility. Importantly, our method achieved these results without requiring k-space sampling information or detailed knowledge of the MR sequence parameters, representing a significant practical advantage over simulation-based approaches. The consistent performance across different anatomical views, combined with competitive metrics compared to both simulated and real reference data, demonstrates the robustness of our approach for clinical applications. However, challenges such as computational cost and the need to simulate more complex motion patterns remain areas for improvement. Furthermore, this framework could be extended to tackle more general blind inverse problems, such as image deblurring or super-resolution, where paired data is scarce and underlying degradation processes are unknown. These results encourage further exploration of diffusion-based methods not only in MRI but also across broader imaging domains, potentially revolutionizing unsupervised correction techniques.

In future work it would also be insightful to include clinical evaluation from experts to validate both the quality of the generated artifacts as well as the performance of the method as a whole.

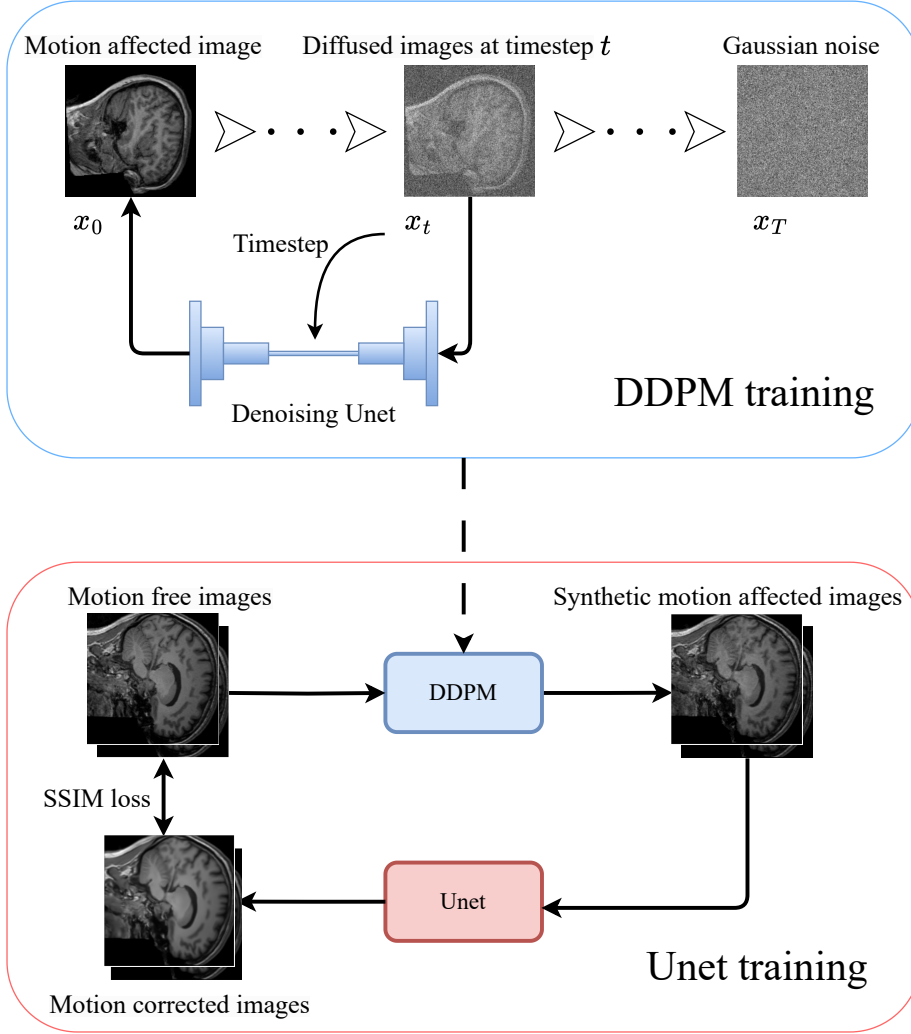


Figure 2.5: Our implementation follows a two-phase approach. In the first phase, a diffusion model (DDPM) is trained on a dataset of motion artifact-affected images. In the second phase, the trained model is used to introduce motion artifacts into clean images, generating paired datasets. These pairs enable supervised training in the subsequent step. See Algorithm 1 for more details.

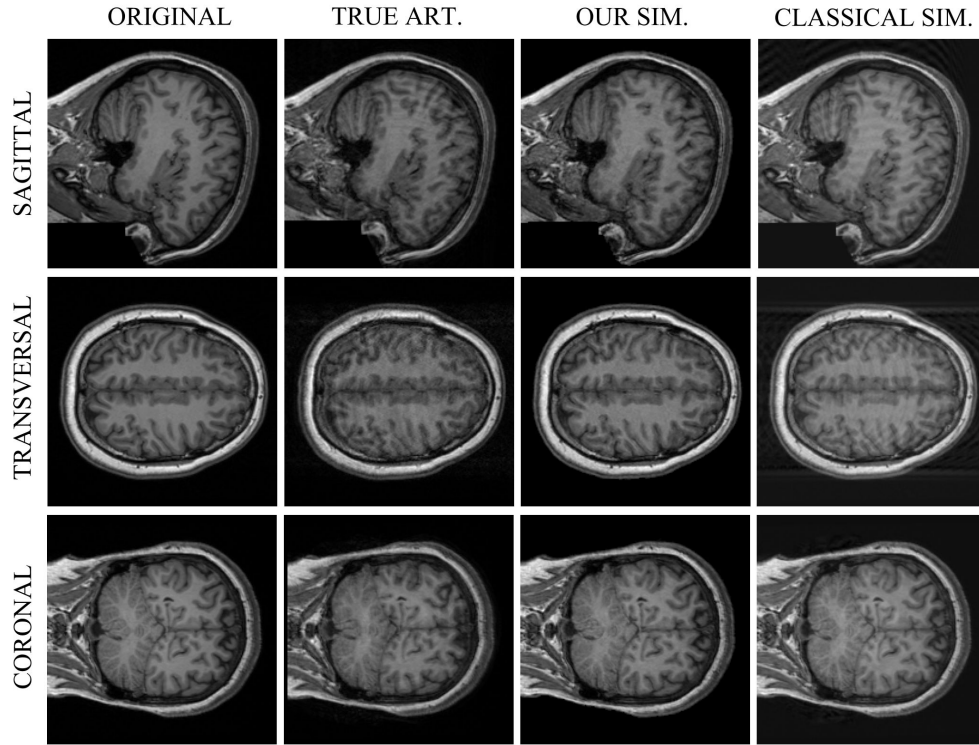


Figure 2.6: Example of the two motion simulation techniques used next to the original images.

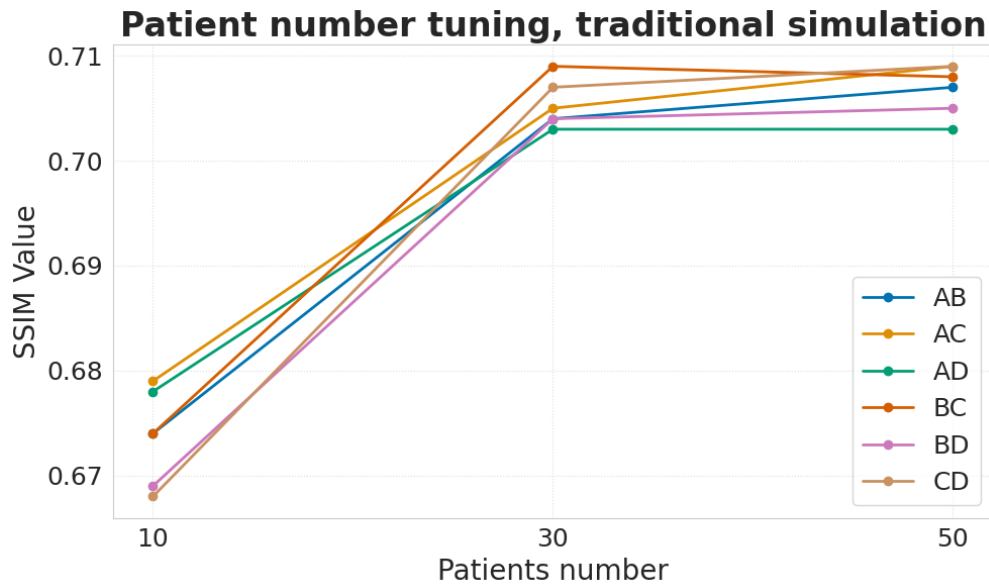


Figure 2.7: Graph comparing the performance obtained using different simulation parameters from [78] with varying numbers of patients.

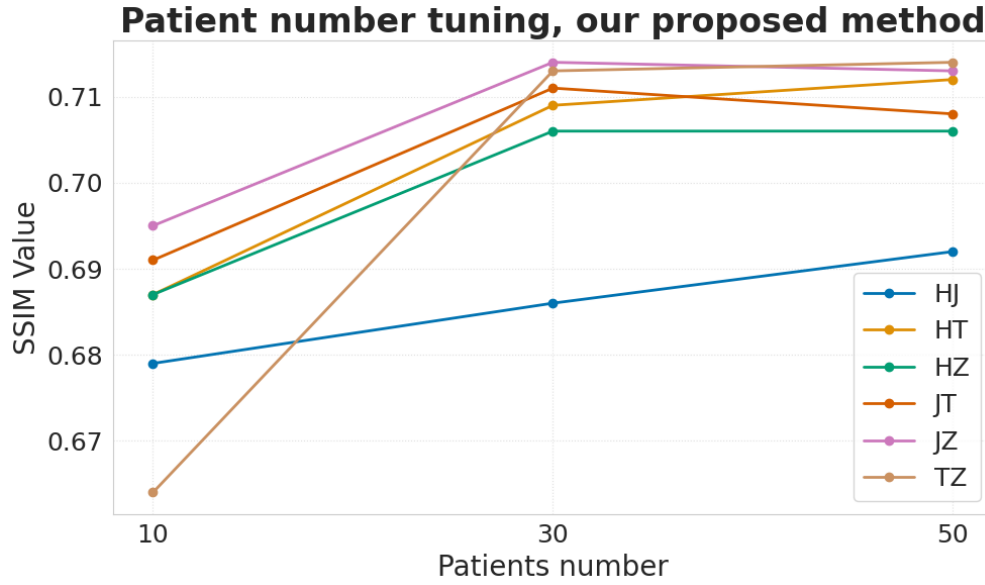


Figure 2.8: Graph comparing the performance obtained using different simulation parameters in our proposed methods with varying numbers of patients.

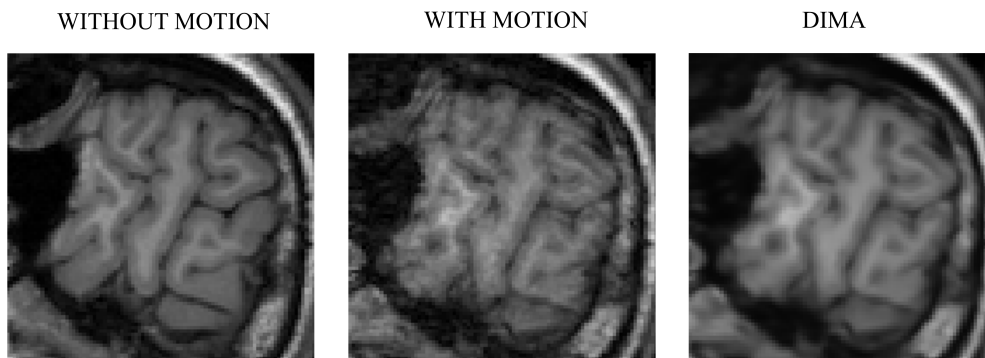


Figure 2.9: Example of the performance of DIMA on a particularly challenging motion-affected image.

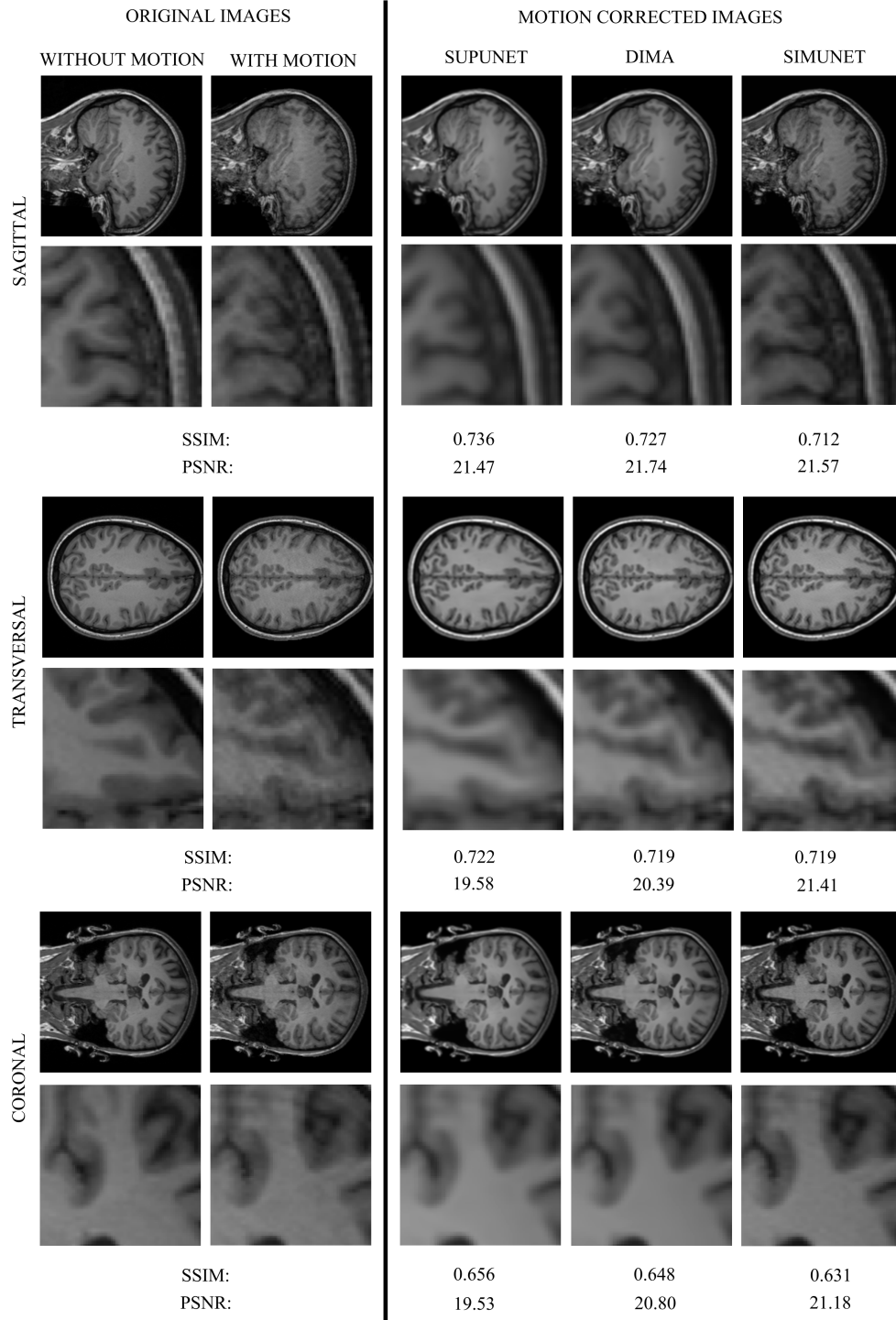


Figure 2.10: Example of the output of the various U-nets on the test set. From left to right: ground truth motion-free and motion-affected images, three different outputs of the three different networks.

Chapter 3

Infinite dimensional generative sensing

3.1 Introduction

The recovery of high dimensional signals from limited measurements is a fundamental challenge in signal processing, medical imaging, and scientific computing. Classical compressed sensing theory [12, 11] revolutionized this field by establishing that signals admitting a sparse representation in some basis can be recovered from a number of linear measurements proportional to their sparsity, rather than their ambient dimension, up to a logarithmic factor.

With the advent of machine learning, the rigidity of the sparsity assumption has been challenged by the emergence of deep generative models [29, 30, 84]. A substantial body of work in the literature has demonstrated that generative priors where the unknown signal is restricted to lie in the range of a low dimensional generator $G: \mathbb{C}^k \rightarrow \mathbb{C}^N$ can outperform sparsity-based classical methods, often requiring significantly fewer measurements [85, 86, 87, 88, 89]. This aligns with the manifold hypothesis [90]: signals concentrate on low-dimensional geometries that exhibit deformation stability and anisotropic regularity [91, 92].

In the finite dimensional setting, theoretical guarantees for these approaches rely on Restricted Isometry Properties [13, 14], which ensure that the measurement operator preserves in some way the geometry of the generator’s range. The problem lies in the fact that many inverse problems of practical interest such as those arising in Partial Differential Equations, Magnetic Resonance Imaging, or Computed Tomography are inherently infinite dimensional [2]. The underlying signals are functions in Hilbert spaces, not fixed-length vectors as the literature usually assumes.

Discretizing these problems *a priori* leads to distinct theoretical and practical issues, such as the “inverse crime” or a dependency on mesh resolution that obscures the true complexity of the signal [93]. This makes the generalization of the deep prior approach to compressed sensing dependent on a proper infinite-dimensional formulation. That is exactly the objective of this work.

We model the inverse problem as recovering an unknown function $x_0 \in \ell^2(\mathbb{N})$ from

noisy, weighted evaluations of a unitary operator $\mathcal{F}$ (e.g., the Fourier transform). We assume x_0 lies in the range of a Generalized Generative Network $G: \mathbb{C}^k \rightarrow \ell^2(\mathbb{N})$, a class of deep ReLU networks acting between finite dimensional spaces with a final linear layer that has $\ell^2(\mathbb{N})$ as co-domain. Taking the finite-dimensional results of [88] as our starting point, this work generalizes their framework to the infinite-dimensional setting.

In this chapter we give the following contributions:

- We introduce an infinite dimensional notion of local coherence, which quantifies the alignment between the generative model’s range and the measurement basis. We derive an optimal, coherence based sampling distribution p^* that minimizes the sample complexity.
- We establish a Gen-RIP for infinite dimensional operators. We prove that the number of samples m required for a stable embedding of the generator’s range scales logarithmically with the intrinsic dimension k and polynomially with the local coherence $\mu_{\mathcal{U}}$, independent of the ambient dimension.
- We extend our analysis to consider probability distributions that are not supported on the entire domain. We introduce a balancing property to control the approximation error in these practical scenarios.
- Through numerical experiments on Darcy flow, we demonstrate that coherence based sampling significantly outperforms uniform sampling. Furthermore, we uncover a counter intuitive phenomenon where lower resolution generators yield superior reconstruction accuracy in undersampled regimes, suggesting that limited resolution acts as a beneficial implicit regularizer.

Finally, we remark that while our framework is formulated in an infinite-dimensional ambient space, the generative priors considered here are inherently finite-dimensional, mapping a low-dimensional latent code to a finite-dimensional subspace of $\ell^2(\mathbb{N})$.

3.1.1 Related work

Infinite dimensional compressed sensing. The transition from finite to infinite dimensional compressed sensing was driven by the need to avoid the so called inverse crime and the discretization errors found in classical formulations. The theory of generalized sampling established that stable recovery of functions is possible using genuinely infinite dimensional measurement models, provided that discretization is handled properly [94] (see [93] for extensions to nonlinear inverse problems). Based on this, a theory of compressed sensing in infinite dimension was developed [95, 96, 97, 98, 99]. This framework has since been extended to a variety of bases and sampling strategies. In particular, rigorous recovery guarantees have been derived for sparse expansions in Legendre polynomial bases [100], as well as for sampling schemes exhibiting local or structured behavior, for which oracle type local recovery guarantees have been established [101]. More recently, a unified theory of infinite dimensional compressed sensing for inverse problems has been developed [102, 103]. This framework enables the recovery of signals from local measurements (such

as those arising from the Radon transform) without discretizing the forward operator, although the associated guarantees primarily focus on wavelet-based sparsity rather than learned generative priors.

Infinite dimensional generative models. The extension of deep generative models to function spaces has emerged as an active area of research, motivated by the need for resolution invariant representations. While early approaches largely focused on scaling discrete architectures, more recent work formulates generative modeling directly in Hilbert spaces. Representative examples include infinite dimensional Variational Autoencoders and score-based diffusion models defined on function spaces, which aim to learn probability measures over infinite dimensional domains [104, 105]. Particularly relevant to our setting is the class of Continuous Generative Neural Networks [106], and of Functional Autoencoders [107], which are explicitly designed to learn resolution independent mappings between finite dimensional latent spaces and function spaces (e.g., $L^2(\mathbb{R}^d)$). This makes them a natural candidate for the generator G in our theoretical framework.

Implicit Generative Priors. In the last few years, score-based generative models and diffusion probabilistic models have emerged as a dominant paradigm for solving inverse problems [26, 37]. Unlike the explicit low-dimensional generator $G : \mathbb{C}^k \rightarrow \mathcal{H}$ employed in our framework, these methods typically utilize an implicit prior given by score function, enabling high-quality reconstruction through stochastic sampling processes such as Langevin dynamics (see [108] for a recent review). While these approaches often achieve state-of-the-art perceptual quality, our work focuses on the rigorous recovery guarantees enabled by the deterministic geometry of explicit low-dimensional manifolds.

Operator Learning and Scientific Machine Learning. Our work is also relevant to the rapidly growing field of scientific machine learning, and in particular to operator learning [109], where neural networks are trained to approximate mappings between infinite dimensional function spaces. Architectures such as the Fourier Neural Operator (FNO) [110] and DeepONet [111] have demonstrated strong empirical performance in the solution of parametric partial differential equations. Generative approaches have also been explored in this context, for instance GenMod combines generative modeling with spectral representations to efficiently approximate the solution manifold of PDEs with random inputs [112]. See [113, 114] for error analysis and approximation theory results. These methods share our goal of resolution independence, but they typically rely on large supervised datasets to learn the forward or inverse operator directly. In contrast, our theory offers a complementary perspective, with the potential to inform optimal data acquisition strategies for training neural operators or for recovering inputs in data-scarce regimes.

3.1.2 Structure of the chapter

The rest of the chapter is organized as follows. Section 3.2 formulates the infinite-dimensional inverse problem and defines the class of Generalized Generative Networks used as priors. We introduce the concept of infinite-dimensional local coherence to derive

optimal sampling strategies and establish the balancing property to handle non-admissible distributions, and we present our main recovery guarantees. Section 3.3 validates this framework numerically by applying Functional Autoencoders [107] to the Darcy flow problem. Section 3.4 contains the proofs of the main results, including the derivation of the Generalized Restricted Isometry Property (Gen-RIP) and the error bounds for the recovered signal. Finally, Section 3.5 summarizes the contributions and discusses limitations and open directions.

3.2 Setup and Main Results

In this section, we establish the mathematical framework for solving inverse problems in infinite dimensional spaces using generative priors, this setup is a generalization, and closely follows [88]. Suppose we have a physical process modeled by an infinite dimensional unitary operator $F: \ell^2(\mathbb{N}) \mapsto \ell^2(\mathbb{N})$. Our objective is to recover an unknown signal $x_0 \in \ell^2(\mathbb{N})$ given a noisy measurement vector b , such that:

$$b = SDFx_0 + \eta,$$

where S is a subsampling operator, D a weighting operator (both defined below) and η represents measurement noise.

Classical Compressed Sensing relies on sparsity in a fixed basis. Here, we depart from sparsity and instead assume that the signal x_0 lies near the range of a deep generative model. Specifically, we constrain the solution to reside in the range of a generator:

$$G: \mathbb{C}^k \mapsto \ell^2(\mathbb{N}).$$

To formalize this prior, we introduce the definition of a generalized generative network suitable for our setting.

Definition 3.2.1 ((k, d) -Generalized Generative Network). Fix the integers $2 \leq k := k_0 \leq \dots \leq k_d$. We fix weight matrices $W^{(i)} \in \mathbb{C}^{k_i \times k_{i-1}}$, as well as a linear map $W: \mathbb{C}^{k_d} \mapsto \ell^2(\mathbb{N})$. We define a (k, d) -Generalized Generative Network as a function $G: \mathbb{C}^k \mapsto \ell^2(\mathbb{N})$ given by:

$$G(z) := W \sigma \left(W^{(d)} \sigma \left(\dots W^{(2)} \sigma \left(W^{(1)} z \right) \right) \right),$$

where σ is the ReLU activation function, defined as:

$$\sigma(x) = \max(0, x).$$

This is the natural extension of classical generative networks whose codomain is finite dimensional, the only difference is that we apply, as the final layer, a linear map W into an infinite dimensional space. Crucially, we restrict the intermediate linear maps $W^{(i)}$ to finite-dimensional codomains. Allowing them to map to infinite-dimensional spaces would render the generator's range infinite-dimensional, thereby invalidating the theoretical

argument presented in the sequel. In an infinite dimensional setting, uniform sampling is not achievable. Therefore, we must adopt a variable density sampling strategy. We introduce a probability distribution p on $\mathbb{N}$ which tells us how indices are selected. To analyze this process, we define the associated sampling, weighting, and projection operators.

Definition 3.2.2. Let p be a probability distribution on $\mathbb{N}$. Let $\{e_j\}_{j \in \mathbb{N}}$ denote the canonical basis of $\ell^2(\mathbb{N})$. We define the sampling operator $S : \ell^2(\mathbb{N}) \rightarrow \mathbb{C}^m$ as the linear map composed of m i.i.d. random vectors $s_1, \dots, s_m$ in $\ell^2(\mathbb{N})$, such that

$$\mathbb{P}(s_i = e_j) = p_j, \quad \forall j \in \mathbb{N}, \forall i \in [m],$$

where we have denoted $[m] = \{1, \dots, m\}$. The action of S is defined component-wise by the inner product:

$$(Sx)_i = \langle x, s_i \rangle, \quad \forall i \in [m].$$

To ensure that our estimator remains unbiased and to normalize the energy of the sampled components, we utilize a diagonal weighting operator D .

Definition 3.2.3. Given p a probability distribution of $\mathbb{N}$ we define a weight operator $D : \ell^2(\mathbb{N}) \mapsto \ell^2(\mathbb{N})$ so that

$$(Dx)_i = \begin{cases} \frac{x_i}{\sqrt{p_i}}, & \text{for } p_i > 0 \\ 0, & \text{for } p_i = 0 \end{cases}$$

Furthermore, we identify the support of our sampling distribution via the operator $\hat{I}$. This operator essentially acts as a projection onto the indices that have a non-zero probability of being sampled.

Definition 3.2.4. Given p a probability distribution of $\mathbb{N}$ we define the operator $\hat{I} : \ell^2(\mathbb{N}) \mapsto \ell^2(\mathbb{N})$

$$\hat{I} := \text{diag}(\mathbb{1}_{\text{supp } p}).$$

With these operators in place, our recovery strategy is formulated as an optimization problem. We seek a latent vector $\hat{z}$ that minimizes the data consistency error in the measurement domain:

$$\hat{z} \in \arg \min_{z \in \mathbb{C}^k} \|b - SDFG(z)\|.$$

We outline that D here is simply a weight operator needed to balance p since in general it will lead to a nonuniform sampling. The recovered signal is then given by:

$$\hat{x} = G(\hat{z}).$$

Our ultimate analytical goal is to derive an upper bound on the reconstruction error $\|\hat{x} - x_0\|$. The overall setup is similar to the finite dimensional case, with the key difference that we can handle both the infinite dimensional nature of the data and probability distributions that are not supported on the entire domain. In particular, the ability to

deal with distributions that fail to sample certain relevant coordinates is also important in the finite dimensional setting, but it has not been addressed in previous works in the literature.

3.2.1 Local Coherence and Sampling Strategies

To guarantee a better recovery rate, the sampling distribution p must be matched to the “geometry” of the operator F relative to the signal model. If F maps the signal model to specific coordinates in $\ell^2(\mathbb{N})$, we must sample those coordinates with higher probability. This alignment is captured by the concept of *local coherence*, borrowed from classical compressed sensing [115] and extended to generative models in [88].

Let $\mathcal{U} \subseteq \ell^2(\mathbb{N})$ be a union of N convex cones $\bigcup_{i=1}^N \mathcal{C}_i$. We define $\Delta(\mathcal{U}) := \bigcup_{i=1}^N \text{span}(\mathcal{C}_i)$. As for intuition, we should think about $\mathcal{U}$ as the range of our generator

Definition 3.2.5. Let $\{e_i\}_{i \in \mathbb{N}}$ denote the canonical basis of $\ell^2(\mathbb{N})$. Given a unitary operator F on $\ell^2(\mathbb{N})$, we define its i -th row as the vector $F_i := F^* e_i$. Given a finite union of convex cones $\mathcal{U} \subseteq \ell^2(\mathbb{N})$ contained in a finite-dimensional subspace, we define the local coherence $(\alpha_i)_{i \in \mathbb{N}}$ of $\mathcal{U}$ with respect to F as

$$\alpha_i = \sup_{x \in \Delta(\mathcal{U}), \|x\|=1} |\langle x, F_i \rangle| = \sup_{x \in \Delta(\mathcal{U}), \|x\|=1} |(Fx)_i|. \quad (3.1)$$

We can show that this is well defined in our framework of finite dimensional subspaces with the following lemma.

Lemma 3.2.6. Let $F: \ell^2(\mathbb{N}) \mapsto \ell^2(\mathbb{N})$ be a unitary operator and let $\mathcal{U} \subseteq \ell^2(\mathbb{N})$ be a finite union of convex cones contained in a subspace of dimension r , $\alpha = (\alpha_i)_{i \in \mathbb{N}}$ the local coherence of $\mathcal{U}$ as defined in (3.1). Then

$$\alpha \in \ell^2(\mathbb{N}) \quad \text{and} \quad \|\alpha\|^2 \leq r^2.$$

This is trivial in the finite case but not in the infinite dimensional one. The local coherence α_i quantifies the maximum correlation between the i -th measurement basis vector and any normalized difference vector in the model set. A sampling distribution is considered “good” if it sufficiently covers indices with high coherence.

Definition 3.2.7. Given F unitary in $\ell^2(\mathbb{N})$ and a finite union of convex cones $\mathcal{U} \subseteq \ell^2(\mathbb{N})$ contained in a finite-dimensional subspace, and p a probability distribution on $\mathbb{N}$. Let $(\alpha_i)_{i \in \mathbb{N}}$ be the local coherence of F with respect to $\mathcal{U}$. We will call $\mu_{\mathcal{U}}(F, p)$ the smallest constant for which

$$\alpha_j \leq \mu_{\mathcal{U}}(F, p) \sqrt{p_j} \quad \forall j: p_j > 0$$

holds. If $\mu_{\mathcal{U}}(F, p)$ is such that

$$\alpha_j \leq \mu_{\mathcal{U}}(F, p) \sqrt{p_j} \quad \forall j \in \mathbb{N},$$

we will say that p is **admissible**.

An admissible distribution ensures that no indices with non zero coherence are ignored. The following lemma identifies the optimal sampling distribution that minimizes the required sample complexity.

Lemma 3.2.8. *Given F unitary in $\ell^2(\mathbb{N})$ and a finite union of convex cones $\mathcal{U} \subseteq \ell^2(\mathbb{N})$ contained in a finite-dimensional subspace, the minimum of $\mu_{\mathcal{U}}(F, p)$ over all admissible p is achieved by p^* where*

$$p_j^* = \frac{\alpha_j^2}{\|\alpha\|^2}, \quad (3.2)$$

in which case $\mu_{\mathcal{U}}(F, p^*)^2 = \|\alpha\|^2$.

We observe that the optimal distribution p^* corresponds to sampling according to the normalized Christoffel function (often referred to as statistical leverage scores). This strategy is fundamental to optimal least squares approximation, as analyzed in [116] and recently reviewed in [117].

Our analysis often requires the assumption $\mu_{\mathcal{U}}(F, p) \geq 1$. As shown in Lemma 3.4.5 below, this condition is mild; it effectively rules out pathological cases (e.g., vanishing coherence) and provides a practical criterion for numerical verification. Moreover, Lemma 3.4.4 establishes that this condition holds automatically whenever p is admissible.

3.2.2 Recovery Guarantees

We state our main recovery result. This theorem provides an error bound for the recovered signal $\hat{x}$. It distinguishes between the general case (where the sampling might ignore some active indices) and the ideal case where the distribution is admissible.

We adopt standard Minkowski notation for set operations; specifically, for a given set A , we denote its difference set as $A - A := \{a - b \mid a, b \in A\}$.

Theorem 3.2.9. *Let F be a unitary operator on $\ell^2(\mathbb{N})$, and let p be a probability distribution on $\mathbb{N}$ with associated operators S , D , and $\hat{I}$. Let G be a (k, d) -generalized generative network, and let $\mathcal{T} = R(G) - R(G)$ be its difference set, such that $\mu_{\mathcal{T}}(F, p) \geq 1$. If the number of measurements $m \in \mathbb{N}$ satisfies*

$$m \geq 16\mu_{\mathcal{T}}(F, p)^2 \left[2kd \log \left(\frac{2ekd}{k} \right) + \log \left(\frac{4k}{\varepsilon} \right) \right],$$

then, with probability at least $1 - \varepsilon$, for any $x_0 \in \ell^2(\mathbb{N})$, $b := SDFx_0 + \eta$, $x^\perp := x_0 - \Pi_{R(G)}(x_0)$, and estimator $\hat{x} \in R(G)$ satisfying $\|SDF\hat{x} - b\| \leq \min_{x \in R(G)} \|SDFx - b\| + \hat{\varepsilon}$, the following holds:

$$\|\hat{x} - x_0\| \leq \|x^\perp\| + \sqrt{\frac{2}{m}} \left[2\|SDFx^\perp\| + 2\|\eta\| + \hat{\varepsilon} \right] + \left\| \hat{I}^\perp F(\hat{x} - \Pi_{R(G)}(x_0)) \right\|. \quad (3.3)$$

Furthermore, if p is an **admissible** distribution, the final tail term vanishes and the bound simplifies to:

$$\|\hat{x} - x_0\| \leq \|x^\perp\| + \sqrt{\frac{2}{m}} \left[2\|SDFx^\perp\| + 2\|\eta\| + \hat{\varepsilon} \right]. \quad (3.4)$$

We observe that the best choice for the sampling probability distribution is the one outlined in Lemma 3.2.8, so if α is the local coherence of F with respect to $\mathcal{T}$, then the bound becomes

$$m \geq 16\|\alpha\|^2 \left[2kd \log \left(\frac{2ek_d}{k} \right) + \log \left(\frac{4k}{\epsilon} \right) \right].$$

When p is admissible, the resulting bounds coincide with those known in the finite dimensional setting [87, 88]. In contrast, the case of non admissible p yields genuinely new bounds, representing a novel contribution in both the finite dimensional and infinite dimensional regimes.

Remark 3.2.10. Due to the weighting operator D in our measurement model $b := SDFx_0 + \eta$, the noise contribution in the final bound scales as $\frac{1}{\sqrt{m}}\|\eta\|$. This provides a favorable asymptotic behavior, as the error vanishes for finite-energy noise, or remains strictly bounded for noise models scaling with $\sqrt{m}$.

3.2.2.1 The Balancing Property

In many infinite dimensional scenarios, an admissible distribution (which requires $p_j > 0$ whenever $\alpha_j > 0$) might be impractical if the coherence has infinite support, or if we are constrained to sample only a finite subset of indices. In such cases, the term $\|\hat{I}^\perp F(\hat{x} - \Pi_{R(G)}(x_0))\|$ in Theorem 3.2.9 does not vanish. To control this error, we introduce the *balancing property*, a concept borrowed from infinite dimensional compressed sensing [95].

Definition 3.2.11. Given F unitary in $\ell^2(\mathbb{N})$, p a probability distribution on $\mathbb{N}$, with $\hat{I}$ its associated operators. G a (k, d) -generalized generative model. We say that the balancing property is verified for $\theta \in (0, 1)$, if

$$\|\hat{I}^\perp F \Pi_{R(G)}\| \leq \theta.$$

This property essentially requires that the energy of the signals in the model range is concentrated on the support of the sampling distribution p . If this holds, we can still derive a strong recovery bound, albeit with an amplification factor of $(1 - \theta)^{-1}$.

Remark 3.2.12. One might question whether the balancing property is a realistic assumption. We note that for any desired tolerance $\theta > 0$, any unitary operator F , and any finite-dimensional generative model G , there exists a distribution p such that the balancing property holds. To see this, consider a uniform distribution over the first ℓ indices, $p_\ell(n) = \frac{1}{\ell} \mathbb{1}_{[n \leq \ell]}$, and let $\hat{I}_\ell$ be the corresponding operator projecting onto these first ℓ coordinates. Its orthogonal complement, $\hat{I}_\ell^\perp$, projects onto the remaining tail indices.

Let $\mathcal{V}$ denote the finite-dimensional subspace containing the range $R(G)$, and let $R = F(\mathcal{V})$. We have:

$$\|\hat{I}_\ell^\perp F \Pi_{\mathcal{V}}\| \leq \sup_{x \in R, \|x\|=1} \|\hat{I}_\ell^\perp x\|.$$

To bound this tail energy, let $(u_1, \dots, u_r)$ be an orthonormal basis for R . Any unit vector $x \in R$ can be expressed as $x = \sum_{i=1}^r c_i u_i$ with $\sum_{i=1}^r |c_i|^2 = 1$. By the Cauchy-Schwarz

inequality, the n -th component satisfies $|x_n|^2 \leq \sum_{i=1}^r |(u_i)_n|^2$. Summing over the tail yields:

$$\sup_{x \in R, \|x\|=1} \|\hat{I}_\ell^\perp x\|^2 = \sup_{x \in R, \|x\|=1} \sum_{n > \ell} |x_n|^2 \leq \sum_{n > \ell} \sum_{i=1}^r |(u_i)_n|^2.$$

Since each basis vector u_i belongs to $\ell^2(\mathbb{N})$, the finite sum of their tail energies must strictly vanish as $\ell \rightarrow \infty$. Thus, $\|\hat{I}_\ell^\perp F \Pi_V\| \rightarrow 0$. Therefore, by choosing a sufficiently large support size ℓ , we can satisfy the balancing condition for any arbitrary θ .

Corollary 3.2.13. *Given F unitary in $\ell^2(\mathbb{N})$, p a probability distribution on $\mathbb{N}$, $m \in \mathbb{N}$ with S and D its associated operators. Let G be a (k, d) -generalized generative network and $\mathcal{T} = R(G) - R(G)$ such that $\mu_{\mathcal{T}}(F, p) \geq 1$. Then if the balancing property is verified for $\theta \in (0, 1)$ and*

$$m \geq 16\mu_{\mathcal{T}}(F, p)^2 \left[2kd \log \left(\frac{2ek_d}{k} \right) + \log \left(\frac{4k}{\epsilon} \right) \right],$$

the following holds with probability at least $1 - \epsilon$:

for any $x_0 \in \ell^2(\mathbb{N})$, let $b := SDFx_0 + \eta$ where $\eta \in \mathbb{C}^m$ and $x^\perp := x_0 - \Pi_{R(G)}x_0$. Let $\hat{x} \in R(G)$ satisfy $\|SDF\hat{x} - b\| \leq \min_{x \in R(G)} \|SDFx - b\| + \hat{\epsilon}$. Then

$$\|\hat{x} - x_0\| \leq \frac{1}{1 - \theta} \left(\|x^\perp\| + \sqrt{\frac{2}{m}} [2\|SDFx^\perp\| + 2\|\eta\| + \hat{\epsilon}] \right)$$

In other words, we retain the same form of stability guarantees as in the admissible case, at the cost of a worse constant, while using a sampling strategy that allows us to handle non admissible and thus more realistic probability distributions such as the one presented in the remark above.

3.3 Numerical Validation

The theoretical guarantees derived in the previous chapters are formulated within the infinite-dimensional Hilbert space $\ell^2(\mathbb{N})$. While this continuum formulation provides rigorous bounds, direct numerical verification in the exact continuum limit is computationally infeasible. Consequently, any numerical validation relies on finite dimensional discretizations. To ensure that our observations are not merely artifacts of specific grid choices, but rather accurate reflections of the underlying function-space theory, it is crucial to employ an architecture that operates consistently across varying resolutions. To this end, we leverage the *Functional Autoencoder* (FAE) framework introduced by [107]. Unlike standard discrete autoencoders, FAEs are designed to be natively mesh free, learning operators that map between function spaces rather than fixed Euclidean vectors. By adopting this architecture, we can empirically demonstrate the central claim of this work: that the proposed recovery guarantees and adaptive sampling strategies are robust to the discretization dimension. In the following sections, we apply this framework to the Darcy flow equation, a classic application for scientific machine learning. We perform experiments across varying spatial resolutions, ranging from 32×32 to 128×128 , to

verify that the reconstruction error and local coherence estimates exhibit asymptotic behavior consistent with the theoretical infinite dimensional limit, independent of the specific mesh granularity used during training.

3.3.1 Experimental Setup and Methodology

The following sections detail the numerical validation of our theoretical framework, specifically addressing the inverse problem associated with the steady state Darcy flow equation [118] [119]. This governing equation, defined as:

$$-\nabla \cdot (a(x)\nabla u(x)) = f(x), \quad x \in \Omega,$$

describes fluid flow through a porous medium on a bounded domain $\Omega \in \mathbb{R}^2$, where $a(x)$ represents the permeability coefficient of the medium, $u(x)$ denotes the resulting pressure field (the solution), and $f(x)$ is the source term.

Our objective is to recover the pressure field $u \in L^2(\Omega)$ from a finite set of noisy Fourier measurements. Formally, we consider the inverse problem

$$b = SDFu + \eta,$$

where $\mathcal{F}$ denotes the discrete Fourier transform at a high-fidelity resolution (fixed at 128×128 in our experiments), S is a sampling operator that selects a subset of m frequencies based on a specific probability distribution (uniform or coherence-based), D is the associated weighting operator to ensure unbiasedness, and η represents measurement noise. Since the noise dependence was not the main focus of this work, all the experiments are performed with noiseless measurements. We approach this ill-posed problem by restricting the solution to the range of a FAE, $G : \mathbb{R}^k \rightarrow L^2(\Omega)$, trained to approximate the solution manifold of the Darcy flow. The reconstruction is performed by searching for the optimal latent code $\hat{z}$ that minimizes the data fidelity loss:

$$\hat{z} = \underset{z \in \mathbb{R}^k}{\operatorname{argmin}} \|SDF(U_r G_r(z)) - b\|_2^2,$$

where G_r is the generator trained at resolution r , and U_r is a fixed upscaling operator mapping the generator’s output to the measurement resolution. This formulation allows us to decouple the generation resolution from the measurement resolution, testing the resolution-invariance of our recovery guarantees.

3.3.1.1 Dataset and Architecture

We validate our framework using the Darcy flow dataset provided in [110], whose specific implementation details are given in [110, Appendix A.3.2]. These solutions exhibit complex, multi scale features that challenge standard linear reconstruction methods. The original dataset contains 2048 images at 421×421 resolution, that for our experiments we resize to 32×32 , 64×64 and 128×128 .

We utilize the FAE introduced in [107]. However, a critical modification is made to the architecture: we replace the Gaussian Error Linear Unit (GeLU) activation functions with Rectified Linear Units (ReLU). The original FAE architecture consists of an encoder and a decoder, both parametrised as mesh-invariant neural networks. The encoder integrates a pointwise kernel network over the input domain, augmenting spatial coordinates with 16 random Fourier features before passing them through two hidden layers of width 64, producing a 64-dimensional latent code via a subsequent linear map. The decoder is a coordinate network that jointly takes the latent vector and a query coordinate as input, again augmented with 16 random Fourier features, and propagates them through five hidden layers of width 100 to produce the reconstructed function value at that coordinate. This coordinate-based, mesh-invariant design allows both networks to be evaluated on arbitrary discretizations independent of the resolution used during training. In the original formulation, GeLU activations are used throughout both the encoder and decoder; in our implementation, we substitute these with ReLU activations, this modification is necessary to ensure the network is within the theoretical guarantees developed in this work. Apart from this substitution, the rest of the network is kept identical to the original implementation and satisfies the assumptions required for our theoretical analysis.

3.3.1.2 Training

To demonstrate the resolution independent capabilities of the FAE, the training is carried out at three distinct spatial resolutions: 32×32 , 64×64 , and 128×128 . The FAE training is performed in an entirely unsupervised manner. The network, consisting of an encoder $\mathcal{E}$ and a decoder $\mathcal{D}$, is optimized to compress the input function into a low dimensional latent representation $z \in \mathbb{R}^k$ (in our experiments $k = 64$) and subsequently decompress it back to the original function domain. The objective is to minimize the reconstruction error while learning a discretization invariant representation of the physical operator. We adopt the same training protocol described in [107], and specifically, we train our FAE for 50,000 epochs, with a learning rate equal to 0.001 and a batch size of 32.

3.3.1.3 Fitting the Latent Space

Once the autoencoder training is completed, the decoder can be utilized independently as a generator $G(z)$ once we can sample z from the latent distribution. However the aggregated posterior distribution of the encoded training data often does not match a standard Gaussian as in variational autoencoders [29]. Consequently, simple Gaussian sampling would likely yield latent codes that map to out of distribution functions. To accurately capture the non-Gaussian structure of the learned latent manifold, we fit a Gaussian Mixture Model (GMM) with $K = 10$ components to the latent codes of the training data:

$$p(z) = \sum_{i=1}^K \pi_i \mathcal{N}(z | \mu_i, \Sigma_i). \quad (3.5)$$

This GMM serves two critical roles: it provides a valid generative mechanism for sampling, and it acts as a structured prior during the recovery phase, allowing us to penalize solutions that drift away from the support of the training distribution.

3.3.1.4 Reconstruction

The recovery problem is formulated as finding the latent space element z that best matches the observed sub-sampled Fourier measurements. A crucial aspect of our implementation is the handling of multi-resolution generators against a fixed high-resolution ground truth that we fix at 128×128 . Let u_{target} be the ground truth signal at that resolution. For a generator G_r trained at resolution r (where $r \in \{32, 64, 128\}$), we solve the following optimization problem:

$$\hat{z} = \arg \min_{z \in \mathbb{R}^k} \|SDF(U_r^{128}G_r(z)) - SDF(u_{\text{target}})\|^2$$

where U_r^{128} denotes a bicubic upscaling operator that maps the generator output to the target resolution, and $\mathcal{F}$, S , and D are the equivalent of the operators introduced in section 3.2, with the difference that now we have to operate in finite dimension, so $\mathcal{F}$ is the FFT in dimension 128×128 , S is the sampling operator that samples m Fourier coefficient based on the probability we have chosen (in our experiments either the one induced by the local coherence or uniform), and finally D weights the samples in the same manner. This formulation ensures that all models are evaluated against the same high-fidelity physical ground truth regardless of their internal resolution. We also ensure that the sampling operator is the same across different output dimensions since everything is up-scaled before measurement via the Fourier transform and sampling are performed. The optimization is performed using the Adam optimizer with a learning rate of 0.01 for 500 iterations. We employ an implementation that optimizes latent codes for multiple test instances simultaneously. We evaluate performance using the Mean Squared Error (MSE) on the spatial domain as the metric.

3.3.2 Results

We now present the results of our experiments. We used half of our datasets, so 1024 images, at the 3 different, for the training of the general model as well to fit the GMM model to the latent space. The other half of the images are used to compute their reconstruction from the undersampled Fourier transform. All the numerical results that follows are means across all the images of this test set.

3.3.2.1 Local coherence and adaptive sampling

To compute the optimal sampling distribution (3.2), one would need to evaluate the local coherence (Definition 3.2.5). However, this quantity is not explicitly computable, so we approximate it via a Monte Carlo procedure.

Specifically, we generate $N = 1024$ samples from the trained GMM, pass them through the generator, and upscale the outputs to the target resolution. We then compute the

maximum magnitude of the corresponding Fourier coefficients across these samples, obtaining

$$\alpha_j^r = \max_{i=1,\dots,N} \left| (\mathcal{F}U_r^{128}(G_r(z_i)))_j \right|,$$

where $r \in \{32, 64, 128\}$ and j indexes the Fourier frequencies. The sampling probabilities p_j are then obtained by normalizing α_j^r . The dependence on r arises because generators trained at different resolutions exhibit different local coherence structures. We compare this adaptive sampling strategy against a uniform sampling baseline.

From a theoretical perspective, local coherence should depend on normalized differences of signals in the model range rather than on individual samples. Motivated by this observation, we also tested an alternative Monte Carlo estimator based on self-differences. Concretely, given a batch of generated images $\{u_i\}_{i=1}^B$, we computed all pairwise differences $u_i - u_j$, normalized each difference, applied a channel-wise two-dimensional Fourier transform, and then maximized the modulus across the batch of normalized differences. The resulting coherence proxy was squared and normalized to obtain sampling probabilities.

While this estimator is more closely aligned with the theoretical definition of local coherence, it consistently underperformed the simpler maximum magnitude estimator described above. Empirically, the original strategy based on maximizing the Fourier magnitude of individual generated samples led to a sampling distributions with improved reconstruction performance.

Figure 3.1 compares the two approaches with the naive approach of uniform sampling. Despite its stronger theoretical motivation, the self-difference estimator does not translate into improved empirical performance in our setting.

3.3.2.2 Adaptive sampling

To evaluate the reconstruction performance, we vary the number of measurements m on a logarithmic scale, ranging from 1% to 32% of the total number of Fourier coefficients. This allows us to analyze the asymptotic behavior of the recovery error as the sampling density increases. As shown in Figure 3.2, adaptive sampling significantly outperforms the uniform baseline in the undersampled regime, exhibiting superior efficiency for a fixed error threshold. As the sampling rate increases, the two strategies asymptotically converge to comparable error levels. Moreover, this behavior is consistent across models trained on data from all three discretizations considered. In Figure 3.3, we provide a qualitative comparison of the reconstruction results at different sampling rates.

3.3.2.3 Discretization invariance of the reconstruction error

The main objective of this experiment is to show that the reconstruction performance is mostly invariant with respect to the output dimension. In Figure 3.4, we report the recovery error for three different models, each trained at one of the three discretization levels considered, and evaluated while generating outputs at three different resolutions (namely 32×32 , 64×64 , and 128×128).

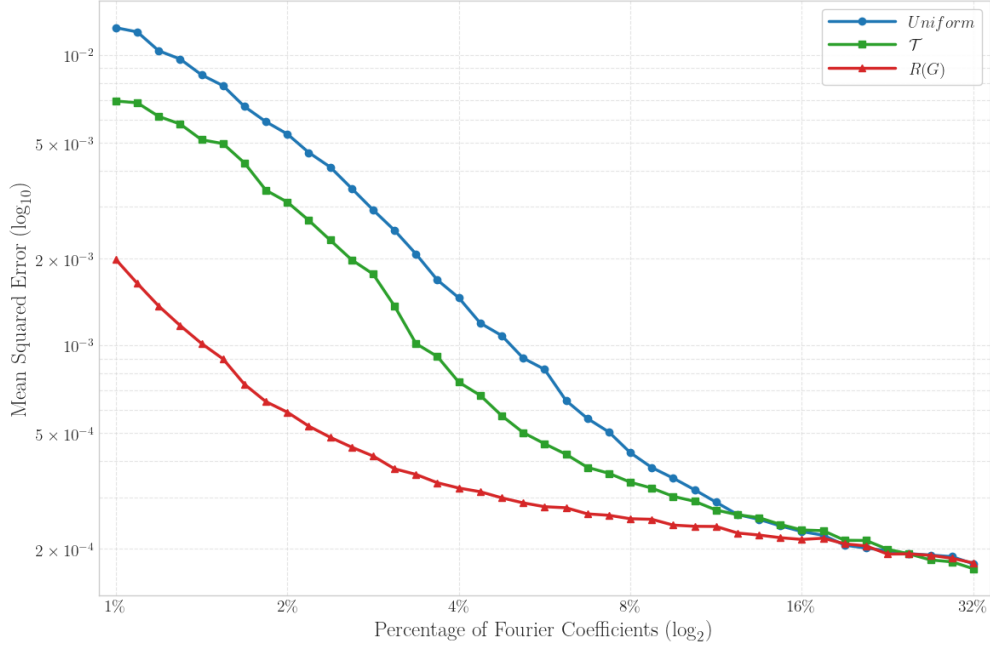


Figure 3.1: Comparison between the original maximum-magnitude based coherence estimator and the self-difference based estimator. Although the latter is theoretically better aligned with the definition of local coherence, the former yields superior empirical performance (model trained and with output in 64 x 64 resolution).

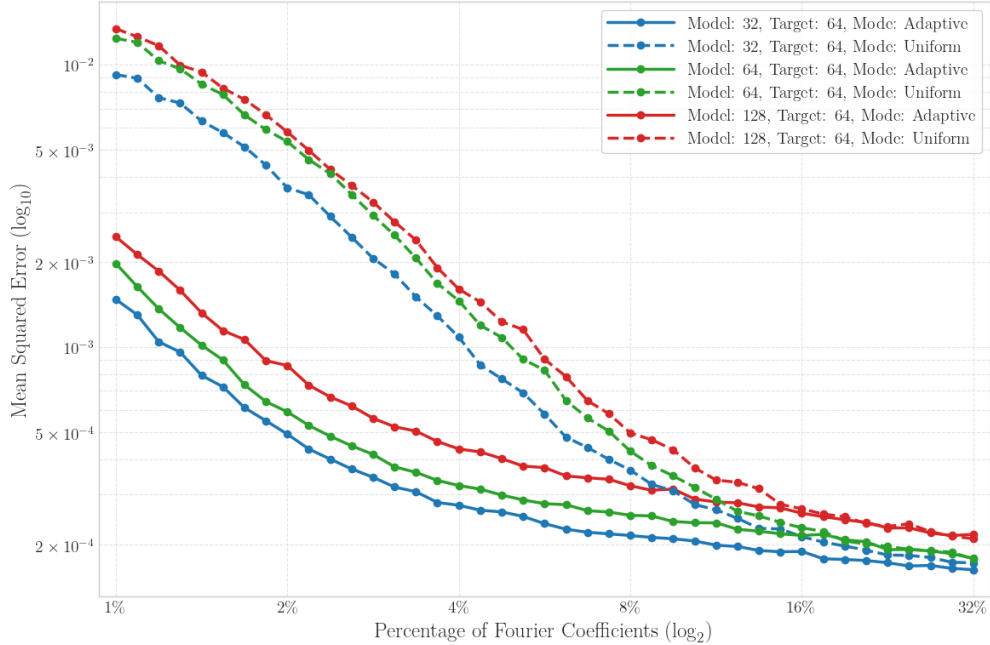


Figure 3.2: Reconstruction error versus sampling rate for uniform and adaptive Fourier sampling.

We observe that, in the asymptotic regime where a sufficient number of measurements is available, the recovery error is primarily determined by the training resolution of the model, and is essentially independent of the target output dimension. In contrast, in the severely undersampled regime, the model resolution has a mild impact on performance,

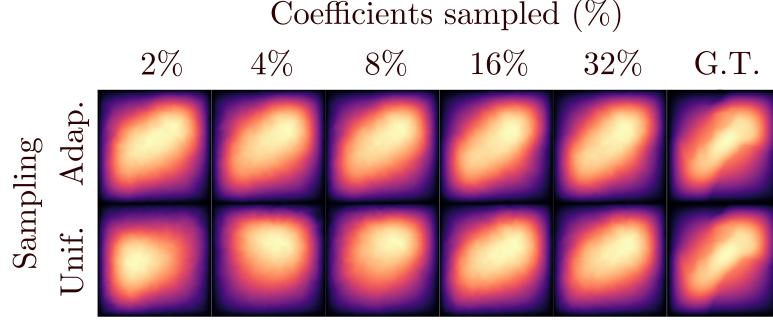


Figure 3.3: Reconstruction examples comparing uniform and adaptive sampling at different rates (model trained on 64×64 resolution).

with lower output dimensions leading to slightly better reconstructions.

Figure 3.5 qualitatively illustrates the reconstruction performance across varying output dimensions. We observe that in the sparse sampling regime, the recovery is dominated by low frequency features, which naturally favors models generating at lower resolutions. Furthermore, in the regime of abundant samples, models trained on lower resolutions exhibit superior performance. This behavior can be understood through the lens of the spectral bias of neural networks [120], which shows that deep networks prioritize learning low-frequency functions. In the severely undersampled regime, the available measurements primarily constrain the low-frequency components of the signal. A high-resolution generator, having a larger capacity, is prone to ‘hallucinating’ high-frequency artifacts that lie in the null space of the measurement operator. In contrast, a lower-resolution generator acts as an explicit low-pass filter, effectively regularizing the inverse problem by suppressing these spurious high-frequency components and ensuring the recovery aligns with the spectral bias inherent to neural networks.

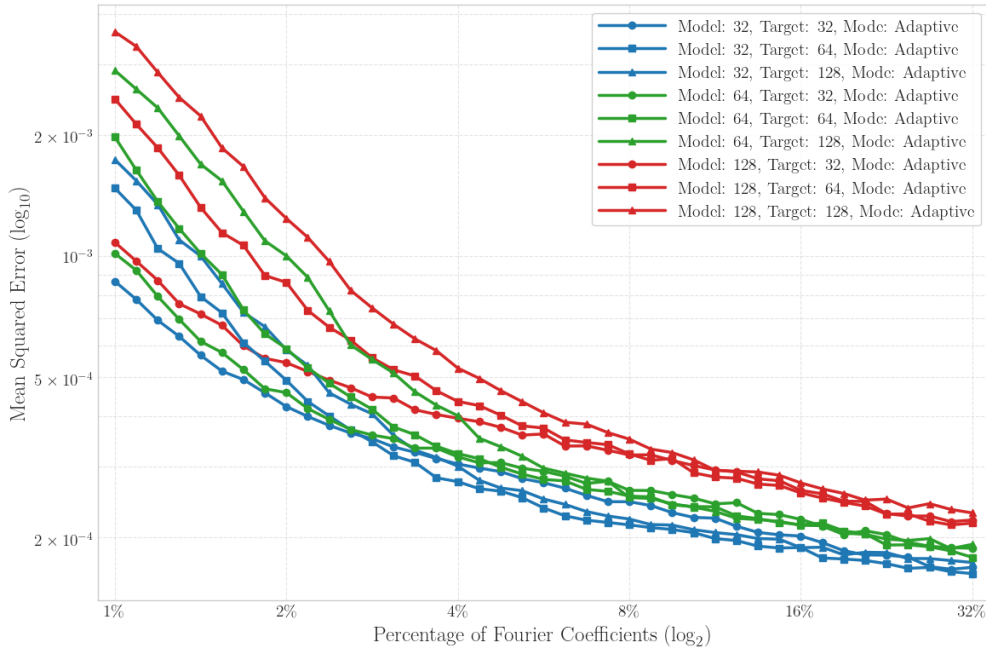


Figure 3.4: Recovery error across different training and output discretizations.

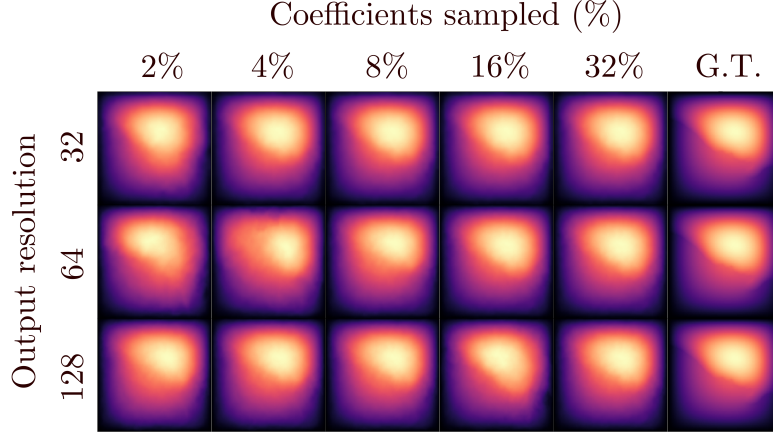


Figure 3.5: Example reconstructions for varying output resolutions (model trained on 64 x 64 resolution)

3.3.2.4 GMM as regularizer

In this experiment, we investigate the effect of incorporating the latent prior (3.5) as an explicit regularization term in the reconstruction optimization. Instead of solving a purely data fidelity problem, we augment the objective with a penalty that encourages latent variables to remain in regions of high probability under the fitted GMM. This promotes solutions that are more consistent with the latent distribution learned during training. Unlike the previous experiments, this study is not concerned with resolution invariance.

The resulting optimization problem can be written as

$$\hat{z} = \arg \min_{z \in \mathbb{R}^k} \left\| SDF(U_r^{128} G_r(z)) - SDF(u_{\text{target}}) \right\|^2 + \lambda (-\log p(z)),$$

where $p(z)$ is the GMM prior defined in (3.5) and $\lambda \geq 0$ controls the strength of the regularization.

Rather than fixing λ a priori, we perform a parameter-tuning procedure to identify the most effective regularization strength by using 10% of our dataset as validation. Specifically, for a discrete set of candidate values of λ , we compute the mean reconstruction error averaged across all sampling levels m . The value of λ that minimizes this aggregated error is then selected as the optimal regularization parameter, in our case we found $\lambda = 10^{-4}$. The performance obtained with this tuned regularizer is subsequently compared against the unregularized case ($\lambda = 0$) for the remaining data points.

Figure 3.6 reports the reconstruction performance using the optimally tuned regularization parameter as a function of the number of measurements. In the severely undersampled regime, the inclusion of the GMM prior leads to a clear improvement over the unregularized approach, as the prior effectively constrains the optimization to regions of the latent space corresponding to plausible generator outputs. In this context, plausible or realistic refers to samples that are likely under the generator induced data distribution, rather than to any external notion of physical realism.

As the number of measurements increases, the data fidelity term becomes increasingly

informative, and the advantage of regularization progressively diminishes. In the high sampling regime, the unregularized formulation performs better than the regularized one, indicating that the measurements alone are sufficient to accurately identify the latent code without introducing bias from the prior. Overall, these results highlight the complementary roles of the measurement operator and the latent prior, with the latter being most beneficial when observational information is limited.

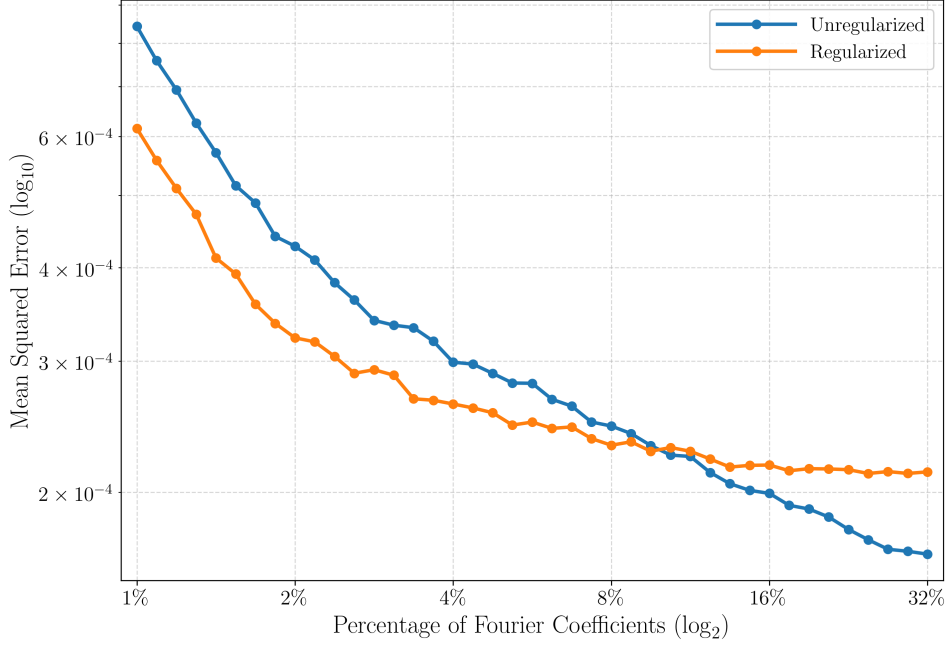


Figure 3.6: Comparison between our proposed method and classical compressed sensing (CS).

3.4 Proofs

We now provide all the results and proofs needed for the results of chapter 2

3.4.1 Setup and local coherence

Let us start by characterizing the range of a (k, d) -Generalized Generative Network, in particular obtaining that its range is contained in a finite dimensional space.

Lemma 3.4.1. *Let $2 \leq k := k_0 \leq \dots \leq k_d$ and G be a (k, d) -Generalized Generative Network. Then $R(G)$ is a union of no more than N at-most k dimensional polyhedral cones where*

$$N \leq \left(\frac{2ek_d}{k} \right)^{kd}.$$

Proof. The proof follows the counting argument of [87, Lemma S2.2]. We observe that the geometric mean of the layer widths is bounded by the maximum width k_d . Therefore, the range of the penultimate layer, $\tilde{G}(z) = \sigma(W^{(d)} \dots \sigma(W^{(1)}z))$, resides in $\mathbb{R}^{k_d}$ and is

a union of at most $N \leq \left(\frac{2ek_d}{k}\right)^{kd}$ polyhedral cones. The final layer of our generator is a linear operator $W : \mathbb{C}^{k_d} \rightarrow \ell^2(\mathbb{N})$. Since W is a linear map from a finite-dimensional space, it is necessarily bounded. A bounded linear operator maps a polyhedral cone in the domain to a polyhedral cone in the codomain (specifically, the image is the conic hull of the images of the generators). Thus, the geometric structure of the union of cones is preserved in $\ell^2(\mathbb{N})$ without increasing the number of components N . $\square$

Remark 3.4.2. Since $R(G)$ is a union of at most N k -dimensional cones we observe that the Minkowski sum $\mathcal{T} = R(G) - R(G)$ is a union of at most N^2 $2k$ -dimensional cones.

Let us now prove that local coherences are well defined.

Proof of Lemma 3.2.6. Let $(\varphi_i)_{i \in [r]}$ be a basis of $\Delta(\mathcal{U})$ and $(F_j := F^* e_j)_{j \in \mathbb{N}}$ the row of F which form an orthonormal basis of $\ell^2(\mathbb{N})$ since F is unitary. Then we can consider

$$\alpha_j = \sup_{c \in \mathbb{C}^r : \|c\|=1} |\langle F_j, \sum_{i \in [r]} c_i \varphi_i \rangle| = |\langle F_j, \sum_{i \in [r]} \bar{c}_{i,j} \varphi_i \rangle|,$$

where $\bar{c}_{i,j}$ are the coefficients attaining the supremum (which is achieved as a maximum by compactness). Then we can conclude with

$$\begin{aligned} \sum_{j \in \mathbb{N}} \alpha_j^2 &= \sum_{j \in \mathbb{N}} \langle F_j, \sum_{i \in [r]} \bar{c}_{i,j} \varphi_i \rangle^2 \\ &= \sum_{j \in \mathbb{N}} \left(\sum_{i \in [r]} \bar{c}_{i,j} \langle F_j, \varphi_i \rangle \right)^2 \\ &\leq r \sum_{j \in \mathbb{N}} \sum_{i \in [r]} \bar{c}_{i,j}^2 \langle F_j, \varphi_i \rangle^2 \\ &\leq r \sum_{j \in \mathbb{N}} \sum_{i \in [r]} \langle F_j, \varphi_i \rangle^2 \\ &= r \sum_{i \in [r]} \sum_{j \in \mathbb{N}} \langle F_j, \varphi_i \rangle^2 \\ &= r \sum_{i \in [r]} \|\varphi_i\|^2 = r^2 < +\infty. \end{aligned} \quad \square$$

The following two lemmas establish conditions under which $\mu_{\mathcal{U}}(F, p) \geq 1$, which are essential for our main results. In particular, the second lemma guarantees that this condition holds whenever p is admissible.

Lemma 3.4.3. *Let F be a unitary operator in $\ell^2(\mathbb{N})$, $\mathcal{U} \subseteq \ell^2(\mathbb{N})$ be a finite union of convex cones contained in a finite-dimensional subspace, and p be a probability distribution on $\mathbb{N}$. If there exists a unit vector $\hat{u} \in \Delta(\mathcal{U})$ such that $\|\hat{I}F\hat{u}\| = 1$, then $\mu_{\mathcal{U}}(F, p) \geq 1$.*

Proof. Consider $\beta_j = (F\hat{u})_j$. We have that $\mu_{\mathcal{U}}(F, p) \geq B$ for the smallest B that verifies

$$\beta_j \leq B\sqrt{p_j} \quad \forall j : p_j > 0,$$

we can take the squared norm on both sides to obtain

$$\|\hat{I}F\hat{u}\| \leq B,$$

since we look for the smallest constant that verifies the inequality and $\|\hat{I}F\hat{u}\| = 1$ we conclude that $B = 1$ and $\mu_{\mathcal{U}}(F, p) \geq B = 1$ $\square$

Lemma 3.4.4. *Let F be a unitary operator on $\ell^2(\mathbb{N})$, and let $\mathcal{U} \subseteq \ell^2(\mathbb{N})$ be a finite union of convex cones contained in a finite-dimensional subspace. If p is an admissible probability distribution on $\mathbb{N}$, then*

$$\mu_{\mathcal{U}}(F, p) \geq 1.$$

Proof. By Lemma 3.4.3, we only need to show there exists $\hat{u} \in \Delta(\mathcal{U})$ with $\|\hat{u}\| = 1$ such that $\|\hat{I}F\hat{u}\| = 1$, but this holds for all of them. If $(F\hat{u})_j \neq 0$, then by definition $\alpha_j > 0$. The admissibility of p guarantees that $p_j > 0$ for all such j , meaning the projection $\hat{I}$ acts as the identity on $F\hat{u}$. Since F is a unitary operator, we conclude $\|\hat{I}F\hat{u}\| = \|F\hat{u}\| = \|\hat{u}\| = 1$. $\square$

Intuitively, to guarantee $\mu_{\mathcal{U}}(F, p) \geq 1$, it suffices for the sampling distribution to place its support on a set of coordinates whose cumulative squared coherence is at least 1. This ensures the measurement operator provides sufficient coverage of the generator's range, as formalized in the following lemma.

Lemma 3.4.5. *Given F unitary in $\ell^2(\mathbb{N})$, $\mathcal{U} \subseteq \ell^2(\mathbb{N})$ a finite union of convex cones contained in a finite-dimensional subspace, and a probability distribution p with $\hat{I}$ the projection on the support of p . If we have*

$$\|\hat{I}\alpha\| \geq 1, \tag{3.6}$$

then $\mu_{\mathcal{U}}(F, p) \geq 1$.

Proof. From the definition of $\mu_{\mathcal{U}}(F, p)$, if we denote with Ω the support of p we have

$$\alpha_j^2 \leq \mu_{\mathcal{U}}(F, p)^2 p_j \quad \forall j \in \Omega.$$

Summing this inequality over all indices in Ω :

$$\sum_{j \in \Omega} \alpha_j^2 \leq \sum_{j \in \Omega} \mu_{\mathcal{U}}(F, p)^2 p_j = \mu_{\mathcal{U}}(F, p)^2 \sum_{j \in \Omega} p_j.$$

Therefore,

$$\sum_{j \in \Omega} \alpha_j^2 \leq \mu_{\mathcal{U}}(F, p)^2.$$

Taking the square root implies

$$\mu_{\mathcal{U}}(F, p) \geq \sqrt{\sum_{j \in \Omega} \alpha_j^2}.$$

Thus, if (3.6) holds, it follows directly that $\mu_{\mathcal{U}}(F, p) \geq 1$. $\square$

Finally we prove how we can characterize the best probability distribution for signal recovery using coherence.

Proof of Lemma 3.2.8. We simply have to observe that $\alpha_j^2 \leq \mu_{\mathcal{U}}(F, p)^2 p_j$ for all $j \in \mathbb{N}$ and then compute

$$\sum_{j \in \mathbb{N}} \alpha_j^2 \leq \sum_{j \in \mathbb{N}} \mu_{\mathcal{U}}(F, p)^2 p_j$$

which implies $\|\alpha\|^2 \leq \mu_{\mathcal{U}}(F, p)^2$. $\square$

3.4.2 Generalized Restricted Isometry Property

We now present two results establishing that our sensing operator preserves the geometry of the signal set. This behavior is analogous to the Restricted Isometry Property (RIP) in classical compressed sensing; accordingly, we refer to it as the Generalized Restricted Isometry Property (Gen-RIP).

Definition 3.4.6 (Generalized Restricted Isometry Property). Let $\mathcal{K} \subseteq \ell^2(\mathbb{N})$ be a signal set of interest and let $\mathcal{F}$ be a unitary operator and p a probability distribution on $\mathbb{N}$ with associated operators S , D , and $\hat{I}$. We say that $\mathcal{A}$ satisfies the Generalized Restricted Isometry Property (Gen-RIP) over $\mathcal{K}$ with constant $\delta \in (0, 1)$ if

$$\sup_{x \in \mathcal{K}, \|x\|=1} \left| \frac{1}{m} \|SDFx\|^2 - \|\hat{I}Fx\|^2 \right| \leq \delta$$

To establish this property for our sensing operator, we begin with an auxiliary result.

Lemma 3.4.7 (Matrix Bernstein [121, Theorem 5.4.1]). *Let $X_1, \dots, X_N$ be independent, mean zero, $n \times n$ symmetric random matrices, such that $\|X_i\| \leq K$ almost surely for all i . Then, for every $t \geq 0$, we have*

$$\mathbb{P} \left\{ \left\| \sum_{i=1}^N X_i \right\| \geq t \right\} \leq 2n \exp \left(-\frac{t^2/2}{\sigma^2 + Kt/3} \right),$$

where $\sigma^2 = \left\| \sum_{i=1}^N \mathbb{E} X_i^2 \right\|$.

Now we give the first results that applies to general sets $\mathcal{U}$.

Lemma 3.4.8. *Let F be a unitary operator on $\ell^2(\mathbb{N})$, and let $\mathcal{U} \subseteq \ell^2(\mathbb{N})$ be a r -dimensional subspace. Let p be a probability distribution on $\mathbb{N}$ with associated operators S , D , and $\hat{I}$, such that $\mu_{\mathcal{U}}(F, p) \geq 1$. Then, for $\varepsilon \in (0, 1)$ and $\delta \in (0, 1]$, we have that*

$$\mathbb{P} \left[\sup_{x \in \mathcal{U}, \|x\|=1} \left| \frac{1}{m} \|SDFx\|^2 - \|\hat{I}Fx\|^2 \right| \leq \delta \right] \geq 1 - \varepsilon$$

provided

$$m \geq \frac{4\mu_{\mathcal{U}}(F, p)^2}{\delta^2} \log \left(\frac{2r}{\varepsilon} \right).$$

Proof. Since $\mathcal{U} \subset \ell^2(\mathbb{N})$ is an r -dimensional subspace, we can fix an orthonormal basis $(\psi_i)_{i \in [r]}$ for it. We then define the operator $P_{\mathcal{U}}: \ell^2(\mathbb{N}) \rightarrow \mathbb{C}^r$ by

$$(P_{\mathcal{U}}(x))_i = \langle x, \psi_i \rangle$$

We observe that $P_{\mathcal{U}}^* P_{\mathcal{U}} = \Pi_{\mathcal{U}}$, i.e. $x = P_{\mathcal{U}}^* P_{\mathcal{U}} x$ for all $x \in \mathcal{U}$. Let us now compute

$$\begin{aligned} & \sup_{x \in \mathcal{U}, \|x\|=1} \left| \frac{1}{m} \|SDFx\|^2 - \|\hat{I}Fx\|^2 \right| \\ &= \sup_{x \in \mathcal{U}, \|x\|=1} \left| \frac{1}{m} \|SDFP_{\mathcal{U}}^* P_{\mathcal{U}}x\|^2 - \|\hat{I}FP_{\mathcal{U}}^* P_{\mathcal{U}}x\|^2 \right| \\ &= \sup_{u \in \mathbb{C}^r \cap \mathbb{S}^{r-1}} \left| \frac{1}{m} \|SDFP_{\mathcal{U}}^* u\|^2 - \|\hat{I}FP_{\mathcal{U}}^* u\|^2 \right| \\ &= \frac{1}{m} \sup_{u \in \mathbb{C}^r \cap \mathbb{S}^{r-1}} \left| u^* \left[(SDFP_{\mathcal{U}}^*)^* (SDFP_{\mathcal{U}}^*) - m (\hat{I}FP_{\mathcal{U}}^*)^* (\hat{I}FP_{\mathcal{U}}^*) \right] u \right|. \end{aligned}$$

We can write this as the operator norm

$$\frac{1}{m} \left\| P_{\mathcal{U}} F^* D S^* SDFP_{\mathcal{U}}^* - m P_{\mathcal{U}} F^* \hat{I} F P_{\mathcal{U}}^* \right\| = \frac{1}{m} \left\| \sum_{i=1}^m \left[P_{\mathcal{U}} F^* [D s_i s_i^* D - \hat{I}] F P_{\mathcal{U}}^* \right] \right\|.$$

First we observe that

$$\mathbb{E}[D s_i s_i^* D] = \hat{I}.$$

To simplify the notation, for each $i \in [m]$, we define the random vector $v_i \in \mathbb{C}^r$ as

$$v_i := P_{\mathcal{U}} F^* D s_i.$$

Since D is self-adjoint, the outer product is exactly $v_i v_i^* = P_{\mathcal{U}} F^* D s_i s_i^* D F P_{\mathcal{U}}^*$. This allows us to express the operator sum as

$$P_{\mathcal{U}} F^* D S^* SDFP_{\mathcal{U}}^* = \sum_{i=1}^m v_i v_i^*.$$

This allows us to write the expectation of each term in the sum as:

$$\mathbb{E}[v_i v_i^* - P_{\mathcal{U}} F^* \hat{I} F P_{\mathcal{U}}^*] = P_{\mathcal{U}} F^* \mathbb{E}[D s_i s_i^* D - \hat{I}] F P_{\mathcal{U}}^* = 0.$$

Consequently, the random matrices comprising the sum are mean-zero. Next, we bound their operator norms almost surely. Let us evaluate the Euclidean norm of v_i . Suppose the random vector s_i realizes as the canonical basis vector e_j for some $j \in \mathbb{N}$. We then have:

$$\begin{aligned} \|v_i\| &= \|P_{\mathcal{U}} F^* D e_j\| \\ &= \frac{1}{\sqrt{p_j}} \|P_{\mathcal{U}} F^* e_j\| \\ &= \frac{1}{\sqrt{p_j}} \sup_{x \in \mathcal{U} \cap B} |\langle x, F^* e_j \rangle| \end{aligned}$$

$$\leq \mu_{\mathcal{U}}(F, p) =: \mu,$$

where the inequality follows directly from the definition of the local coherence α_j and the sampling parameter $\mu_{\mathcal{U}}(F, p)$. Since this bound holds for any realization e_j , we have $\|v_i\| \leq \mu$ almost surely. We can now bound the operator norm of each mean-zero matrix:

$$\|v_i v_i^* - P_{\mathcal{U}} F^* \hat{I} F P_{\mathcal{U}}^*\| \leq \|v_i v_i^*\| + \|P_{\mathcal{U}} F^* \hat{I} F P_{\mathcal{U}}^*\| \leq \mu^2 + 1 \leq 2\mu^2,$$

where we used the rank-one identity $\|v_i v_i^*\| = \|v_i\|^2 \leq \mu^2$, the operator norm bound $\|P_{\mathcal{U}} F^* \hat{I} F P_{\mathcal{U}}^*\| \leq 1$, and our assumption that $\mu \geq 1$.

Then we compute

$$\begin{aligned} \sigma^2 &= \left\| \sum_{i=1}^m \mathbb{E} \left[\left(v_i v_i^* - P_{\mathcal{U}} F^* \hat{I} F P_{\mathcal{U}}^* \right)^2 \right] \right\| \\ &= \sup_{u \in \mathbb{C}^r \cap \mathbb{S}^{r-1}} \left\langle u, \sum_{i=1}^m \left(\mathbb{E} [v_i v_i^* v_i v_i^*] - \left(P_{\mathcal{U}} F^* \hat{I} F P_{\mathcal{U}}^* \right)^2 \right) u \right\rangle \\ &\leq \sup_{u \in \mathbb{C}^r \cap \mathbb{S}^{r-1}} \left\langle u, \sum_{i=1}^m \|v_i\|^2 \mathbb{E} [v_i v_i^*] u \right\rangle - \sum_{i=1}^m \|P_{\mathcal{U}} F^* \hat{I} F P_{\mathcal{U}}^* u\|^2 \\ &\leq \sup_{u \in \mathbb{C}^r \cap \mathbb{S}^{r-1}} \sum_{i=1}^m \mu^2 \langle u, \mathbb{E} [v_i v_i^*] u \rangle \\ &\leq \mu^2 m. \end{aligned}$$

The second equality holds because the operator inside the norm is positive semi-definite, meaning its operator norm coincides with the supremum of its quadratic form over the unit sphere. The subsequent inequalities follow by dropping the non-positive subtracted term and applying the bound $\|\mathbb{E}[v_i v_i^*]\| = \|P_{\mathcal{U}} F^* \hat{I} F P_{\mathcal{U}}^*\| \leq 1$.

From this, we apply the Matrix Bernstein inequality (Lemma 3.4.7) with dimension r , variance parameter $\sigma^2 \leq m\mu^2$, and spectral bound $L \leq 2\mu^2$, obtaining:

$$\mathbb{P} \left\{ \left\| \sum_{i=1}^m \left[P_{\mathcal{U}} F^* \left[D s_i s_i^* D - \hat{I} \right] F P_{\mathcal{U}}^* \right] \right\| \geq \gamma \right\} \leq 2r \exp \left(-\frac{\gamma^2/2}{m\mu^2 + \frac{2}{3}\mu^2\gamma} \right).$$

This bound translates to the operator norm over the subspace. Since $P_{\mathcal{U}}^*$ is an isometry from $\mathbb{C}^r$ to $\mathcal{U}$, the condition above is equivalent to:

$$\mathbb{P} \left\{ \sup_{x \in \mathcal{U}, \|x\|=1} \left| \frac{1}{m} \|SDFx\|^2 - \|\hat{I}Fx\|^2 \right| \geq \frac{\gamma}{m} \right\} \leq 2r \exp \left(-\frac{\gamma^2/2}{m\mu^2 + \frac{2}{3}\mu^2\gamma} \right).$$

Substituting $\delta = \frac{\gamma}{m}$, the exponent becomes:

$$-\frac{(m\delta)^2/2}{m\mu^2 + \frac{2}{3}\mu^2 m\delta} = -\frac{m\delta^2}{2\mu^2(1 + \frac{2}{3}\delta)}.$$

Assuming $\delta \leq 1$, we have $(1 + \frac{2}{3}\delta) \leq 5/3 < 2$, which implies $\frac{1}{2(1 + \frac{2}{3}\delta)} > \frac{1}{4}$. This yields the

simplified bound:

$$\mathbb{P} \left\{ \sup_{x \in \mathcal{U}, \|x\|=1} \left| \frac{1}{m} \|SDFx\|^2 - \|\hat{I}Fx\|^2 \right| \geq \delta \right\} \leq 2r \exp \left(-\frac{m\delta^2}{4\mu^2} \right).$$

Finally, requiring the right-hand side to be at most ε ,

$$2r \exp \left(-\frac{m\delta^2}{4\mu^2} \right) \leq \varepsilon,$$

we conclude that the condition holds provided that

$$m \geq \frac{4\mu_{\mathcal{U}}(F, p)^2}{\delta^2} \log \left(\frac{2r}{\varepsilon} \right). \quad \square$$

We now present our main Generalized Restricted Isometry Property (Gen-RIP) result, which establishes that the restricted isometry holds for signals within the range of a (k, d) -generalized generative network.

Theorem 3.4.9 (Gen-RIP). *Let F be a unitary operator on $\ell^2(\mathbb{N})$ and let p be a probability distribution on $\mathbb{N}$ with associated operators S , D , and $\hat{I}$. Let G be a (k, d) -generalized generative network and $\mathcal{T} = R(G) - R(G)$. For $\varepsilon > 0$, $\delta \in (0, 1)$, assume the number of measurements satisfies*

$$m \geq \frac{4\mu_{\mathcal{T}}(F, p)^2}{\delta^2} \left[2kd \log \left(\frac{2ek_d}{k} \right) + \log \left(\frac{4k}{\varepsilon} \right) \right],$$

where we assume $\mu_{\mathcal{T}}(F, p) \geq 1$. Then, with probability at least $1 - \varepsilon$, the following holds:

$$\sup_{x \in \mathcal{T}, \|x\|=1} \left| \frac{1}{m} \|SDFx\|^2 - \|\hat{I}Fx\|^2 \right| \leq \delta$$

Proof. By Lemma 3.4.2, $\mathcal{T}$ is a collection of at most N^2 $2k$ -dimensional sub-spaces. Therefore we can write $\mathcal{T} = \bigcup_{j=1}^N \mathcal{L}_j$, and apply on each $\mathcal{L}_j$ Lemma 3.4.8 with $\varepsilon' > 0$ and obtain

$$\begin{aligned} & \mathbb{P} \left\{ \sup_{x \in \mathcal{T}, \|x\|=1} \left| \frac{1}{m} \|SDFx\|^2 - \|\hat{I}Fx\|^2 \right| \geq \delta \right\} \\ & \leq \sum_{j=1}^N \mathbb{P} \left\{ \sup_{x \in \mathcal{L}_j, \|x\|=1} \left| \frac{1}{m} \|SDFx\|^2 - \|\hat{I}Fx\|^2 \right| \geq \delta \right\} \\ & \leq N\varepsilon'. \end{aligned}$$

To find the requested rate for m , we set $\varepsilon' = \varepsilon/N^2$ in Lemma 3.4.8, which requires:

$$m \geq \frac{4\mu_{\mathcal{U}}(F, p)^2}{\delta^2} \log \left(\frac{2kN}{\varepsilon} \right) = \frac{4\mu_{\mathcal{U}}(F, p)^2}{\delta^2} \left[\log N^2 + \log \left(\frac{4k}{\varepsilon} \right) \right].$$

Substituting the bound $\log N \leq kd \log \left(\frac{2ek_d}{k} \right)$ from Lemma 3.4.1 yields the desired sample complexity. $\square$

3.4.3 Recovery bounds

Now we have everything we need to prove the main theorem and corollary.

Proof of Theorem 3.2.9. Recall $x^\perp := x_0 - \Pi_{R(G)}(x_0)$. By the definition of the estimator and the triangle inequality, we have

$$\begin{aligned} \|SDF\hat{x} - b\| &\leq \min_{x \in R(G)} \|SDFx - b\| + \hat{\varepsilon} \\ &\leq \|SDF\Pi_{R(G)}(x_0) - b\| + \hat{\varepsilon} \\ &= \|SDFx^\perp + \eta\| + \hat{\varepsilon} \\ &\leq \|SDFx^\perp\| + \|\eta\| + \hat{\varepsilon}. \end{aligned}$$

By assumption, the number of measurements m is sufficiently large such that, by Corollary 3.4.9, the measurement operator SDF satisfies the restricted isometry property over the difference set $\mathcal{T} = \Delta(R(G))$ with constant $\delta = 1/2$, with probability at least $1 - \varepsilon$. Therefore, since the difference vector $\hat{x} - \Pi_{R(G)}(x_0) \in \Delta(R(G))$, we obtain

$$\begin{aligned} \|SDF\hat{x} - b\| &= \|SDF(\hat{x} - \Pi_{R(G)}(x_0)) - SDFx^\perp - \eta\| \\ &\geq \|SDF(\hat{x} - \Pi_{R(G)}(x_0))\| - \|SDFx^\perp\| - \|\eta\| \\ &\geq \sqrt{\frac{m}{2}} \|\hat{I}F(\hat{x} - \Pi_{R(G)}(x_0))\| - \|SDFx^\perp\| - \|\eta\|. \end{aligned}$$

Assembling the two inequalities gives

$$\|\hat{I}F(\hat{x} - \Pi_{R(G)}(x_0))\| \leq \sqrt{\frac{2}{m}} [2\|SDFx^\perp\| + 2\|\eta\| + \hat{\varepsilon}].$$

Applying the triangle inequality and utilizing the unitarity of F , we derive the general bound:

$$\begin{aligned} \|\hat{x} - x_0\| &\leq \|x_0 - \Pi_{R(G)}(x_0)\| + \|\hat{x} - \Pi_{R(G)}(x_0)\| \\ &\leq \|x^\perp\| + \|\hat{I}F(\hat{x} - \Pi_{R(G)}(x_0))\| + \|\hat{I}^\perp F(\hat{x} - \Pi_{R(G)}(x_0))\| \\ &\leq \|x^\perp\| + \frac{2\sqrt{2}}{\sqrt{m}} \|SDFx^\perp\| + \frac{2\sqrt{2}}{\sqrt{m}} \|\eta\| + \frac{\sqrt{2}}{\sqrt{m}} \hat{\varepsilon} + \|\hat{I}^\perp F(\hat{x} - \Pi_{R(G)}(x_0))\|. \end{aligned}$$

This proves (3.3).

For the second part, assume p is admissible. We observe that $\hat{I}^\perp Fu = 0$ for all $u \in \Delta(R(G))$. This holds because if the local coherence satisfies $\alpha_j > 0$, admissibility dictates that $p_j > 0$, implying the orthogonal projection component $\hat{I}_{j,j}^\perp = 0$. Conversely, if $\alpha_j = 0$, the j -th component is strictly zero by the definition of local coherence. Since $\hat{x} - \Pi_{R(G)}(x_0) \in \Delta(R(G))$, the tail term $\|\hat{I}^\perp F(\hat{x} - \Pi_{R(G)}(x_0))\|$ vanishes entirely, yielding (3.4). $\square$

Proof of Corollary 3.2.13. We simply need to observe that

$$\|\hat{I}^\perp F(\hat{x} - \Pi_{R(G)}(x_0))\| = \|\hat{I}^\perp F\Pi_{R(G)}(\hat{x} - x_0)\| \leq \theta \|\hat{x} - x_0\|,$$

simply substituting in (3.3) we obtain the thesis. $\square$

3.5 Summary and discussion

In this work we have established a theoretical foundation for infinite dimensional generative compressed sensing. By properly formulating the problem in $\ell^2(\mathbb{N})$ and accounting for the discretization gap we have shown that generative priors can be effectively used for functional recovery. We introduced the concept of infinite dimensional local coherence and utilized it to derive optimal sampling strategies. Our numerical experiments on Darcy flow validated these theoretical insights demonstrating that coherence based sampling significantly reduces the number of measurements required for accurate reconstruction.

Furthermore our empirical results revealed a valuable interaction between model resolution and sampling density. We found that in data scarce regimes employing a lower resolution generator with explicit upscaling provides a form of implicit regularization that stabilizes the inverse problem. Our findings suggest that the optimal generative prior is not necessarily the one with the highest fidelity, but rather the one that best matches the spectral content of the available measurements. By limiting resolution, we leverage the spectral bias of the network [120] to stabilize the inversion in data-scarce regimes.

While our theoretical framework is robust, it relies on the assumption that the generator employs piecewise linear activations (e.g., ReLU) to ensure the range is contained within a union of finite-dimensional polyhedral cones. Extending this theory to networks with smooth activations, such as GeLU or SiLU, remains an open theoretical challenge. Additionally, while our theory provides error bounds in the presence of measurement noise, our numerical validations were primarily focused on the noiseless regime. Future work could explore the empirical robustness of this framework under varying noise levels and distributions. Finally, translating these infinite-dimensional Gen-RIP guarantees to other practical measurement operators (especially non-unitary), such as the Radon transform for continuous computed tomography, is the natural next step.

Conclusions

Inverse problems remain a fundamental challenge across many scientific and engineering disciplines, often requiring sophisticated methodologies to recover meaningful information from indirect or corrupted measurements. This thesis explored the use of generative models as flexible and data driven tools for addressing such problems, combining both practical applications and theoretical analysis.

In the first part of the thesis, we addressed the problem of motion artifacts in magnetic resonance imaging, a crucial challenge given the widespread use of MRI scanners in clinical practice. Since paired datasets of motion corrupted and motion free scans are scarce, we chose to tackle the problem using generative modeling, which only requires unsupervised learning. In particular, this work investigated the use of diffusion models in two different settings. First, as a reconstruction tool that, once conditioned on a motion-corrupted image, generates the corresponding motion-free one. Here, numerical experiments showed that hallucinations pose a significant challenge for clinical applications. Second, we explored the use of diffusion models as generators of realistic synthetic paired data for supervised learning. This methodology proved capable of producing synthetic pairs that enable effective supervised training without relying on detailed knowledge of the MRI acquisition sequence, which is required by classical synthetic data generation approaches. Experiments on real MRI datasets showed competitive correction performance, highlighting the practical potential of generative approaches in clinically relevant scenarios. An interesting direction for future research would be to extend this framework to other blind inverse problems, as its agnostic nature with respect to the forward model makes it potentially applicable beyond MRI motion correction.

The second part of the thesis focused on generative compressed sensing, a framework inspired by compressed sensing in which, rather than assuming sparsity in a given basis, signals are modeled as belonging to the range of a generative operator. In particular, we studied the case of infinite dimensional signals, so where the generator also has an infinite dimensional range. Our goal was to derive theoretical recovery guarantees, specifically bounds on the number of measurements required to recover the signal up to a prescribed error. To achieve this, we introduced a generalized restricted isometry property, which enabled us to prove the main result of this section: a sampling rate that scales linearly with the intrinsic dimension of the generator, as expected. Moreover, we proved the existence of an optimal sampling strategy based on local coherence, which can be defined explicitly, although in practice it must be approximated. We validated our analysis through numerical experiments on the Darcy flow problem, confirming our theoretical findings. In particular, the optimal sampling strategy proved significantly more efficient than uniform sampling, while recovery rates remained equivalent across different resolutions. This behavior provides strong evidence that the proposed framework successfully captures the underlying infinite dimensional structure of the problem. A natural direction for future work would be to extend these theoretical guarantees to broader classes of generators, since the current results hold only for a relatively restricted

family of generative models.

Overall, this work shows that generative models provide a robust and flexible data-driven paradigm for solving inverse problems, contributing both to real-world applications and to a rigorous mathematical understanding of these tasks.

Bibliography

- [1] Richard C Aster, Brian Borchers, and Clifford H Thurber. *Parameter Estimation and Inverse Problems*. Elsevier, 2018 (cit. on p. 2).
- [2] Heinz Werner Engl, Martin Hanke, and Andreas Neubauer. *Regularization of Inverse Problems*. Vol. 375. Mathematics and Its Applications. Springer Science & Business Media, 1996 (cit. on pp. 2, 35).
- [3] Jacques Hadamard. *Lectures on Cauchy’s Problem in Linear Partial Differential Equations*. Yale University Press, 1923 (cit. on p. 2).
- [4] Per Christian Hansen, James G Nagy, and Dianne P O’leary. *Deblurring Images: Matrices, Spectra, and Filtering*. SIAM, 2006 (cit. on p. 3).
- [5] Frank Natterer. *The Mathematics of Computerized Tomography*. SIAM, 2001 (cit. on p. 3).
- [6] Michael Lustig, David Donoho, and John M Pauly. “Sparse MRI: The application of compressed sensing for rapid MR imaging”. In: *Magnetic Resonance in Medicine* 58.6 (2007), pp. 1182–1195 (cit. on p. 3).
- [7] Roger Penrose. “A generalized inverse for matrices”. In: *Mathematical Proceedings of the Cambridge Philosophical Society* 51.3 (1955), pp. 406–413 (cit. on p. 4).
- [8] Andreu Nikolaevich Tikhonov and Vasiliu Iakovlevich Arsenin. *Solutions of Ill-posed Problems*. Winston & Sons, 1977 (cit. on p. 5).
- [9] Leonid I Rudin, Stanley Osher, and Emad Fatemi. “Nonlinear total variation based noise removal algorithms”. In: *Physica D: nonlinear phenomena* 60.1-4 (1992), pp. 259–268 (cit. on p. 5).
- [10] Robert Tibshirani. “Regression shrinkage and selection via the lasso”. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 58.1 (1996), pp. 267–288 (cit. on p. 5).
- [11] David L Donoho. “Compressed sensing”. In: *IEEE Transactions on information theory* 52.4 (2006), pp. 1289–1306 (cit. on pp. 5, 35).
- [12] Emmanuel J Candès, Justin Romberg, and Terence Tao. “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information”. In: *IEEE Transactions on information theory* 52.2 (2006), pp. 489–509 (cit. on pp. 5, 35).

- [13] Emmanuel J Candes and Terence Tao. “Decoding by linear programming”. In: *IEEE transactions on information theory* 51.12 (2005), pp. 4203–4215 (cit. on pp. 5, 35).
- [14] Richard Baraniuk, Mark Davenport, Ronald DeVore, and Michael Wakin. “A simple proof of the restricted isometry property for random matrices”. In: *Constructive approximation* 28.3 (2008), pp. 253–263 (cit. on pp. 6, 35).
- [15] Jari Kaipio and Erkki Somersalo. *Statistical and Computational Inverse Problems*. Vol. 160. Springer Science & Business Media, 2006 (cit. on p. 6).
- [16] Andrew M Stuart. “Inverse problems: a Bayesian perspective”. In: *Acta numerica* 19 (2010), pp. 451–559 (cit. on p. 6).
- [17] Albert Tarantola. *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM, 2005 (cit. on p. 6).
- [18] Michael T McCann, Kyong Hwan Jin, and Michael Unser. “Convolutional Neural Networks for Inverse Problems in Imaging: A Review”. In: *IEEE Signal Processing Magazine* 34.6 (2017) (cit. on p. 7).
- [19] Simon Arridge, Peter Maass, Ozan Oktun, and Carola Bibiane Schonlieb. “Solving inverse problems using data driven models”. In: *Acta Numerica* 28 (2019) (cit. on p. 7).
- [20] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016 (cit. on p. 7).
- [21] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. “Deep learning”. In: *Nature* 521.7553 (2015) (cit. on p. 7).
- [22] Ge Wang. “A Perspective on Deep Imaging”. In: *IEEE Access* 4 (2016) (cit. on p. 8).
- [23] Kyong Hwan Jin, Michael T McCann, Emmanuel Froustey, and Michael Unser. “Deep Convolutional Neural Network for Inverse Problems in Imaging”. In: *IEEE Transactions on Image Processing* 26.9 (2017) (cit. on p. 8).
- [24] Vishal Monga, Yuelong Li, and Yonina C Eldar. “Algorithm Unrolling: Interpretable, Efficient Deep Learning for Signal and Image Processing”. In: *IEEE Signal Processing Magazine* 38.2 (2021) (cit. on p. 8).
- [25] Singanallur V Venkatakrishnan, Charles A Bouman, and Brendt Wohlberg. “Plug and Play priors for model based reconstruction”. In: *2013 IEEE Global Conference on Signal and Information Processing*. 2013 (cit. on p. 9).
- [26] Yang Song, Jascha Sohl Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. “Score Based Generative Modeling through Stochastic Differential Equations”. In: *International Conference on Learning Representations*. 2021 (cit. on pp. 9–11, 37).
- [27] Ajil Jalal, Sushrut Karmalkar Liu, Alexandros G Dimakis, and Eric Price. “Robust compressed sensing using generative models”. In: *Advances in Neural Information Processing Systems*. 2020 (cit. on p. 9).

- [28] George Papamakarios, Eric Nalisnick, Danilo Jimenez Rezende, Shakir Mohamed, and Balaji Lakshminarayanan. “Normalizing flows for probabilistic modeling and inference”. In: *Journal of Machine Learning Research* 22.1 (2021) (cit. on p. 9).
- [29] Diederik P Kingma and Max Welling. “Auto-encoding variational bayes”. In: *International Conference on Learning Representations* (2013) (cit. on pp. 10, 35, 45).
- [30] Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. “Generative adversarial nets”. In: *Advances in neural information processing systems* 27 (2014) (cit. on pp. 10, 35).
- [31] Jonathan Ho, Ajay Jain, and Pieter Abbeel. “Denoising diffusion probabilistic models”. In: *Advances in Neural Information Processing Systems* 33 (2020) (cit. on p. 10).
- [32] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. “Deep Unsupervised Learning using Nonequilibrium Thermodynamics”. In: *Proceedings of the 32nd International Conference on Machine Learning*. Ed. by Francis Bach and David Blei. Vol. 37. Proceedings of Machine Learning Research. Lille, France: PMLR, July 2015, pp. 2256–2265 (cit. on p. 10).
- [33] Ashish Bora, Ajil Jalal, Eric Price, and Alexandros G Dimakis. “Compressed sensing using generative models”. In: *International Conference on Machine Learning*. 2017 (cit. on p. 12).
- [34] Muhammad Asim, Max Daniels, Oscar Leong, Ali Ahmed, and Paul Hand. “Invertible generative models for inverse problems: mitigating representation error and dataset bias”. In: *International Conference on Machine Learning* (2020) (cit. on p. 12).
- [35] Hyungjin Chung, Byeongsu Sim, Dohoon Ryu, and Jong Chul Ye. “Improving diffusion models for inverse problems using manifold constraints”. In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 25683–25696 (cit. on p. 12).
- [36] Giannis Daras, Hyungjin Chung, Chieh-Hsin Lai, Yuki Mitsufuji, Jong Chul Ye, Peyman Milanfar, Alexandros G. Dimakis, and Mauricio Delbracio. *A Survey on Diffusion Models for Inverse Problems*. 2024 (cit. on p. 12).
- [37] Hyungjin Chung, Jeongsol Kim, Michael Thompson Mccann, Marc Louis Klasky, and Jong Chul Ye. “Diffusion Posterior Sampling for General Noisy Inverse Problems”. In: *The Eleventh International Conference on Learning Representations*. 2023 (cit. on pp. 13, 37).
- [38] Hyungjin Chung and Jong Chul Ye. “Score-based diffusion models for accelerated MRI”. In: *Medical image analysis* 80 (2022), p. 102479 (cit. on p. 13).

- [39] Yang Song, Liyue Shen, Lei Xing, and Stefano Ermon. “Solving inverse problems in medical imaging with score-based generative models”. In: *arXiv preprint arXiv:2111.08005* (2021) (cit. on p. 13).
- [40] Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. “Repaint: Inpainting using denoising diffusion probabilistic models”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2022, pp. 11461–11471 (cit. on p. 13).
- [41] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. “Image super-resolution via iterative refinement”. In: *IEEE transactions on pattern analysis and machine intelligence* 45.4 (2022), pp. 4713–4726 (cit. on p. 13).
- [42] Daniele Trappolini, Laura Laurenti, Giulio Poggiali, Elisa Tinti, Fabio Galasso, Alberto Michelini, and Chris Marone. “Cold diffusion model for seismic denoising”. In: *Journal of Geophysical Research: Machine Learning and Computation* 1.2 (2024), e2024JH000179 (cit. on p. 13).
- [43] Paolo Angella, Vito Paolo Pastore, and Matteo Santacesaria. “Assessing the use of diffusion models for motion artifact correction in brain MRI”. In: *2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI)*. IEEE. 2025, pp. 1–4 (cit. on pp. 13, 16, 19).
- [44] Paolo Angella, Luca Balbi, Fabrizio Ferrando, Paolo Traverso, Rosario Varriale, Vito Paolo Pastore, and Matteo Santacesaria. “DIMA: DIffusing Motion Artifacts for unsupervised correction in brain MRI images”. In: *IEEE Access* 13 (2025), pp. 198649–198658 (cit. on pp. 13, 17, 19).
- [45] Paolo Angella, Vito Paolo Pastore, and Matteo Santacesaria. “Infinite dimensional generative sensing”. In: *arXiv preprint arXiv:2603.03196* (2026) (cit. on p. 13).
- [46] Yuchou Chang, Zhiqiang Li, Gulfam Saju, Hui Mao, and Tianming Liu. “Deep learning-based rigid motion correction for magnetic resonance imaging: a survey”. In: *Meta-Radiol* (2023), p. 100001 (cit. on pp. 15, 16).
- [47] Veronika Spieker, Hannah Eichhorn, Kerstin Hammernik, Daniel Rueckert, Christine Preibisch, Dimitrios C Karampinos, and Julia A Schnabel. “Deep learning for retrospective motion correction in MRI: a comprehensive review”. In: *IEEE Trans Med Imaging* (2023) (cit. on pp. 15, 16).
- [48] Maxim Zaitsev, Julian Maclaren, and Michael Herbst. “Motion artifacts in MRI: A complex problem with many partial solutions”. In: *J Magn Reson Imaging* 42.4 (2015), pp. 887–901 (cit. on pp. 15, 16).
- [49] Frank Godenschweger, Urte Kägebein, Daniel Stucht, Uten Yarach, Alessandro Sciarra, Renat Yakupov, Falk Lüsebrink, Peter Schulze, and Oliver Speck. “Motion correction in MRI of the brain”. In: *Physics in medicine & biology* 61.5 (2016), R32 (cit. on p. 16).

- [50] Julian Maclaren, Michael Herbst, Oliver Speck, and Maxim Zaitsev. “Prospective motion correction in brain imaging: a review”. In: *Magnetic resonance in medicine* 69.3 (2013), pp. 621–636 (cit. on p. 16).
- [51] James G Pipe. “Motion correction with PROPELLER MRI: application to head motion and free-breathing cardiac imaging”. In: *Magn Reson Med* 42.5 (1999), pp. 963–969 (cit. on p. 17).
- [52] Jan-Ray Liao, John M Pauly, Thomas J Brosnan, and Norbert J Pelc. “Reduction of motion artifacts in cine MRI using variable-density spiral trajectories”. In: *Magn Reson Med* 37.4 (1997), pp. 569–575 (cit. on p. 17).
- [53] Nick Todd, Oliver Josephs, Martina F Callaghan, Antoine Lutti, and Nikolaus Weiskopf. “Prospective motion correction of 3D echo-planar imaging data for functional MRI using optical tracking”. In: *Neuroimage* 113 (2015), pp. 1–12 (cit. on p. 17).
- [54] Nathan White, Cooper Roddey, Ajit Shankaranarayanan, Eric Han, Dan Rettmann, Juan Santos, Josh Kuperman, and Anders Dale. “PROMO: real-time prospective motion correction in MRI using image-based tracking”. In: *Magn Reson Med* 63.1 (2010), pp. 91–105 (cit. on p. 17).
- [55] Muhammad Usman, David Atkinson, Freddy Odille, Christoph Kolbitsch, Ghislain Vaillant, Tobias Schaeffter, Philip G Batchelor, and Claudia Prieto. “Motion corrected compressed sensing for free-breathing dynamic cardiac MRI”. In: *Magn Reson Med* 70.2 (2013), pp. 504–516 (cit. on p. 17).
- [56] Zai Yang, Cishen Zhang, and Lihua Xie. “Sparse MRI for motion correction”. In: *IEEE ISBI*. 2013, pp. 962–965 (cit. on p. 17).
- [57] Ben A Duffy, Wenlu Zhang, Haoteng Tang, Lu Zhao, Meng Law, Arthur W Toga, and Hosung Kim. “Retrospective correction of motion artifact affected structural MRI images using deep learning of simulated motion”. In: *MIDL*. 2018 (cit. on p. 17).
- [58] Patricia M Johnson and Maria Drangova. “Conditional generative adversarial network for 3D rigid-body motion correction in MRI”. In: *Magn Reson Med* 82.3 (2019), pp. 901–910 (cit. on p. 17).
- [59] Junchi Liu, Mehmet Kocak, Mark Supanich, and Jie Deng. “Motion artifacts reduction in brain MRI by means of a deep residual network with densely connected multi-resolution blocks (DRN-DCMB)”. In: *Magn Reson Imaging* 71 (2020), pp. 69–79 (cit. on p. 17).
- [60] Mohammed A Al-Masni, Seul Lee, Jaeuk Yi, Sewook Kim, Sung-Min Gho, Young Hun Choi, and Dong-Hyun Kim. “Stacked U-Nets with self-assisted priors towards robust correction of rigid motion artifact in brain MRI”. In: *Neuroimage* 259 (2022), p. 119411 (cit. on p. 17).

- [61] Ekaterina Kuzmina, Artem Razumov, Oleg Y Rogov, Elfar Adalsteinsson, Jacob White, and Dmitry V Dylov. “Autofocusing+: noise-resilient motion correction in magnetic resonance imaging”. In: *MICCAI*. Springer. 2022, pp. 365–375 (cit. on p. 17).
- [62] Brett Levac, Ajil Jalal, and Jonathan I Tamir. “Accelerated motion correction for MRI using score-based generative models”. In: *IEEE ISBI*. IEEE. 2023, pp. 1–5 (cit. on p. 17).
- [63] Gyutaek Oh, Sukyoung Jung, Jeong Eun Lee, and Jong Chul Ye. “Annealed score-based diffusion model for MR motion artifact reduction”. In: *IEEE Trans Comput Imaging* (2023) (cit. on p. 17).
- [64] Siyuan Liu, Kim-Han Thung, Liangqiong Qu, Weili Lin, Dinggang Shen, and Pew-Thian Yap. “Learning MRI artefact removal with unpaired data”. In: *Nat Mach Intell* 3.1 (2021), pp. 60–67 (cit. on p. 17).
- [65] Vahid Ghodrati, Mark Bydder, Fadil Ali, Chang Gao, Ashley Prosper, Kim-Lien Nguyen, and Peng Hu. “Retrospective respiratory motion correction in cardiac cine MRI reconstruction using adversarial autoencoder and unsupervised learning”. In: *NMR Biomed* 34.2 (2021), e4433 (cit. on p. 17).
- [66] Gyutaek Oh, Jeong Eun Lee, and Jong Chul Ye. “Unpaired MR motion artifact deep learning using outlier-rejecting bootstrap aggregation”. In: *IEEE Trans Med Imaging* 40.11 (2021), pp. 3125–3139 (cit. on p. 17).
- [67] Karim Armanious, Aastha Tanwar, Sherif Abdulatif, Thomas Küstner, Sergios Gatidis, and Bin Yang. “Unsupervised adversarial correction of rigid MR motion artifacts”. In: *IEEE ISBI*. IEEE. 2020, pp. 1494–1498 (cit. on p. 17).
- [68] Seuk Kim, Mohammed A Al-masni, Seul Lee, Sunyoung Jung, Kyu-Jin Jung, Chuanjiang Cui, Sung-Min Ghoh, Young Hun Choi, and Dong-Hyun Kim. “Unsupervised learning for motion correction and assessment in brain magnetic resonance imaging using severity-based regularized cycle consistency”. In: *Engineering Applications of Artificial Intelligence* 142 (2025), p. 109978 (cit. on p. 17).
- [69] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. “Image-to-image translation with conditional adversarial networks”. In: *CVPR*. 2017, pp. 1125–1134 (cit. on pp. 17, 23).
- [70] Karim Armanious, Chenming Jiang, Sherif Abdulatif, Thomas Küstner, Sergios Gatidis, and Bin Yang. “Unsupervised medical image translation using cycle-MedGAN”. In: *EUSIPCO*. IEEE. 2019, pp. 1–5 (cit. on p. 17).
- [71] Mojtaba Safari, Xiaofeng Yang, Ali Fatemi, and Louis Archambault. “MRI motion artifact reduction using a conditional diffusion probabilistic model (MAR-CDPM)”. In: *Med Phys* 51.4 (2024), pp. 2598–2610 (cit. on pp. 17, 18, 23, 28).

- [72] Brett Levac, Sidharth Kumar, Ajil Jalal, and Jonathan I Tamir. “Accelerated motion correction with deep generative diffusion models”. In: *Magnetic Resonance in Medicine* 92.2 (2024), pp. 853–868 (cit. on p. 18).
- [73] Alex Ling Yu Hung, Kai Zhao, Haoxin Zheng, Ran Yan, Steven S Raman, Demetri Terzopoulos, and Kyunghyun Sung. “Med-cdiff: Conditional medical image generation with diffusion models”. In: *Bioengineering* 10.11 (2023), p. 1258 (cit. on p. 18).
- [74] Feng Li, Zijian Zhou, Yu Fang, Jiangdong Cai, and Qian Wang. “MoCo-Diff: Adaptive Conditional Prior on Diffusion Network for MRI Motion Correction”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2024, pp. 411–421 (cit. on p. 18).
- [75] Jonathan Ho, Ajay Jain, and Pieter Abbeel. “Denoising Diffusion Probabilistic Models”. In: *Advances in Neural Information Processing Systems*. Ed. by H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin. Vol. 33. Curran Associates, Inc., 2020, pp. 6840–6851 (cit. on pp. 18, 20).
- [76] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-net: Convolutional networks for biomedical image segmentation”. In: *MICCAI*. Springer. 2015, pp. 234–241 (cit. on pp. 19, 21).
- [77] Ádám Nárai et al. “Movement-related artefacts (MR-art) dataset of matched motion-corrupted and clean structural MRI brain scans”. In: *Sci Data* 9.1 (2022) (cit. on pp. 21, 24).
- [78] Hampus Olsson et al. “Simulating rigid head motion artifacts on brain magnitude MRI data—Outcome on image quality and segmentation of the cerebral cortex”. In: *PLoS One* 19.4 (2024), e0301132 (cit. on pp. 21, 22, 24–29, 32).
- [79] Carolina Higuera, Byron Boots, and Mustafa Mukadam. “Learning to Read Braille: Bridging the Tactile Reality Gap with Diffusion Models”. In: (2023) (cit. on pp. 21, 25).
- [80] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. “Unpaired image-to-image translation using cycle-consistent adversarial networks”. In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2223–2232 (cit. on p. 23).
- [81] Heath R Pardoe and Samantha P Martin. *"In-scanner head motion and structural covariance networks"*. OpenNeuro, 2021 (cit. on p. 24).
- [82] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. “Image quality assessment: from error visibility to structural similarity”. In: *IEEE Trans Image Process* 13.4 (2004), pp. 600–612 (cit. on p. 26).
- [83] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. “The unreasonable effectiveness of deep features as a perceptual metric”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 586–595 (cit. on p. 26).

- [84] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. “Deep unsupervised learning using nonequilibrium thermodynamics”. In: *International conference on machine learning*. PMLR. 2015, pp. 2256–2265 (cit. on p. 35).
- [85] Ashish Bora, Ajil Jalal, Eric Price, and Alexandros G. Dimakis. “Compressed Sensing using Generative Models”. In: *Proceedings of the 34th International Conference on Machine Learning*. 2017, pp. 537–546 (cit. on p. 35).
- [86] Alireza Naderi and Yaniv Plan. “Sparsity-Free Compressed Sensing With Applications to Generative Priors”. In: *IEEE Journal on Selected Areas in Information Theory* 3.3 (2022), pp. 493–501 (cit. on p. 35).
- [87] Aaron Berk, Simone Brugiapaglia, Babhru Joshi, Yaniv Plan, Matthew Scott, and Ozgur Yilmaz. “A Coherence Parameter Characterizing Generative Compressed Sensing With Fourier Measurements”. In: *IEEE Journal on Selected Areas in Information Theory* 3.3 (2022), pp. 502–512 (cit. on pp. 35, 42, 51).
- [88] Aaron Berk, Simone Brugiapaglia, Yaniv Plan, Matthew Scott, Xia Sheng, and Ozgur Yilmaz. “Model-adapted Fourier sampling for generative compressed sensing”. In: *NeurIPS 2023 Workshop on Deep Learning and Inverse Problems*. 2023 (cit. on pp. 35, 36, 38, 40, 42).
- [89] Ben Adcock, Juan M. Cardenas, and Nick Dexter. “A Unified Framework for Learning with Nonlinear Model Classes from Arbitrary Linear Samples”. In: *Forty-first International Conference on Machine Learning (ICML)*. 2024 (cit. on p. 35).
- [90] Charles Fefferman, Sanjoy Mitter, and Hariharan Narayanan. “Testing the manifold hypothesis”. In: *Journal of the American Mathematical Society* 29.4 (2016), pp. 983–1049 (cit. on p. 35).
- [91] Stéphane Mallat. “Group invariant scattering”. In: *Communications on Pure and Applied Mathematics* 65.10 (2012), pp. 1331–1398 (cit. on p. 35).
- [92] Philipp Grohs and Gitta Kutyniok. “Parabolic molecules”. In: *Foundations of Computational Mathematics* 14.2 (2014), pp. 299–337 (cit. on p. 35).
- [93] Giovanni S Alberti and Matteo Santacesaria. “Infinite-dimensional inverse problems with finite measurements”. In: *Archive for Rational Mechanics and Analysis* 243.1 (2022), pp. 1–31 (cit. on pp. 35, 36).
- [94] Ben Adcock and Anders C Hansen. “A generalized sampling theorem for stable reconstructions in arbitrary bases”. In: *Journal of Fourier Analysis and Applications* 18.4 (2012), pp. 685–716 (cit. on p. 36).
- [95] Ben Adcock and Anders C Hansen. “Generalized sampling and infinite-dimensional compressed sensing”. In: *Foundations of Computational Mathematics* 16.5 (2016), pp. 1263–1329 (cit. on pp. 36, 42).

- [96] Ben Adcock, Anders C Hansen, Clarice Poon, and Bogdan Roman. “Breaking the coherence barrier: A new theory for compressed sensing”. In: *Forum of mathematics, sigma*. Vol. 5. Cambridge University Press. 2017, e4 (cit. on p. 36).
- [97] Ben Adcock. “Infinite-dimensional compressed sensing and function interpolation”. In: *Foundations of Computational Mathematics* 18.3 (2018), pp. 661–701 (cit. on p. 36).
- [98] Ben Adcock, Simone Brugiapaglia, and Clayton G Webster. *Sparse polynomial approximation of high-dimensional functions*. Vol. 25. SIAM, 2022 (cit. on p. 36).
- [99] Ben Adcock, Simone Brugiapaglia, Nick Dexter, and Sebastian Moraga. *On Efficient Algorithms for Computing Near-Best Polynomial Approximations to High-Dimensional, Hilbert-Valued Functions from Limited Samples*. European Mathematical Society - EMS - Publishing House GmbH, 2024 (cit. on p. 36).
- [100] Holger Rauhut and Rachel Ward. “Sparse Legendre expansions via ℓ_1 -minimization”. In: *Journal of Approximation Theory* 164.5 (2012), pp. 517–533 (cit. on p. 36).
- [101] Ben Adcock, Claire Boyer, and Simone Brugiapaglia. “On oracle-type local recovery guarantees in compressed sensing”. In: *Information and Inference: A Journal of the IMA* 10.1 (2021), pp. 1–49 (cit. on p. 36).
- [102] Giovanni S Alberti, Alessandro Felisi, Matteo Santacesaria, and S Ivan Trapasso. “Compressed sensing for inverse problems and the sample complexity of the sparse Radon transform”. In: *Journal of the European Mathematical Society* (2025) (cit. on p. 36).
- [103] Giovanni S Alberti, Alessandro Felisi, Matteo Santacesaria, and S Ivan Trapasso. “Compressed sensing for inverse problems II: applications to deconvolution, source recovery, and MRI”. In: *arXiv preprint arXiv:2501.01929* (2025) (cit. on p. 36).
- [104] Giulio Franzese and Pietro Michiardi. “Generative diffusion models in infinite dimensions: a survey”. In: *Philosophical Transactions of the Royal Society A* 383.2299 (2025) (cit. on p. 37).
- [105] Sam Bond-Taylor and Chris G. Willcocks. “ ∞ -Diff: Infinite Resolution Diffusion with Subsampled Mollified States”. In: *The Twelfth International Conference on Learning Representations*. 2024 (cit. on p. 37).
- [106] Giovanni S. Alberti, Matteo Santacesaria, and Silvia Sciutto. “Continuous Generative Neural Networks: A Wavelet-Based Architecture in Function Spaces”. In: *Numerical Functional Analysis and Optimization* 46.1 (2025), pp. 1–44 (cit. on p. 37).
- [107] Justin Bunker, Mark Girolami, Hefin Lambley, Andrew M Stuart, and TJ Sullivan. “Autoencoders in function space”. In: *Journal of Machine Learning Research* 26.165 (2025), pp. 1–54 (cit. on pp. 37, 38, 43, 45).

- [108] Giannis Daras, Hyungjin Chung, Chieh-Hsin Lai, Yuki Mitsufuji, Jong Chul Ye, Peyman Milanfar, Alexandros G Dimakis, and Mauricio Delbracio. “A survey on diffusion models for inverse problems”. In: *arXiv preprint arXiv:2410.00083* (2024) (cit. on p. 37).
- [109] Nicolas Boullé and Alex Townsend. “A mathematical guide to operator learning”. In: *Handbook of Numerical Analysis*. Vol. 25. Elsevier, 2024, pp. 83–125 (cit. on p. 37).
- [110] Zongyi Li, Nikola Borislavov Kovachki, Kamyar Azizzadenesheli, Burigede liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. “Fourier Neural Operator for Parametric Partial Differential Equations”. In: *International Conference on Learning Representations*. 2021 (cit. on pp. 37, 44).
- [111] Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. “Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators”. In: *Nature Machine Intelligence* 3.3 (2021), pp. 218–229 (cit. on p. 37).
- [112] Jacqueline Wentz and Alireza Doostan. “GenMod: A generative modeling approach for spectral representation of PDEs with random inputs”. In: *Journal of Computational Physics* 472 (2023), p. 111691 (cit. on p. 37).
- [113] Samuel Lanthaler, Siddhartha Mishra, and George E Karniadakis. “Error estimates for deepONets: A deep learning framework in infinite dimensions”. In: *Transactions of Mathematics and Its Applications* 6.1 (2022), tnac001 (cit. on p. 37).
- [114] Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. “Neural operator: Learning maps between function spaces with applications to pdes”. In: *Journal of Machine Learning Research* 24.89 (2023), pp. 1–97 (cit. on p. 37).
- [115] Emmanuel Candes and Justin Romberg. “Sparsity and incoherence in compressive sampling”. In: *Inverse problems* 23.3 (2007), p. 969 (cit. on p. 40).
- [116] Albert Cohen and Giovanni Migliorati. “Optimal weighted least-squares methods”. In: *SMAI Journal of Computational Mathematics* 3 (2017), pp. 181–203 (cit. on p. 41).
- [117] Ben Adcock. “Optimal sampling for least-squares approximation”. In: *Foundations of Computational Mathematics* 25.6 (2025), pp. 1975–2034 (cit. on p. 41).
- [118] John A Cherry and R Allan Freeze. *Groundwater*. Vol. 370. Englewood Cliffs, NJ: Prentice-Hall, 1979 (cit. on p. 44).
- [119] Joseph B Keller. “Darcy’s law for flow in porous media and the two-space method”. In: *Nonlinear partial differential equations in engineering and applied science*. Routledge, 2017, pp. 429–443 (cit. on p. 44).

- [120] Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred Hamprecht, Yoshua Bengio, and Aaron Courville. “On the spectral bias of neural networks”. In: *International conference on machine learning*. PMLR. 2019, pp. 5301–5310 (cit. on pp. 49, 59).
- [121] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*. Vol. 47. Cambridge university press, 2018 (cit. on p. 54).